

Volume 11 Issue 5

May 2020



ISSN 2156-5570(Online)

ISSN 2158-107X(Print)



# Editorial Preface

## *From the Desk of Managing Editor...*

It may be difficult to imagine that almost half a century ago we used computers far less sophisticated than current home desktop computers to put a man on the moon. In that 50 year span, the field of computer science has exploded.

Computer science has opened new avenues for thought and experimentation. What began as a way to simplify the calculation process has given birth to technology once only imagined by the human mind. The ability to communicate and share ideas even though collaborators are half a world away and exploration of not just the stars above but the internal workings of the human genome are some of the ways that this field has moved at an exponential pace.

At the International Journal of Advanced Computer Science and Applications it is our mission to provide an outlet for quality research. We want to promote universal access and opportunities for the international scientific community to share and disseminate scientific and technical information.

We believe in spreading knowledge of computer science and its applications to all classes of audiences. That is why we deliver up-to-date, authoritative coverage and offer open access of all our articles. Our archives have served as a place to provoke philosophical, theoretical, and empirical ideas from some of the finest minds in the field.

We utilize the talents and experience of editor and reviewers working at Universities and Institutions from around the world. We would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations. Our high standards are maintained through a double blind review process.

We hope that this edition of IJACSA inspires and entices you to submit your own contributions in upcoming issues. Thank you for sharing wisdom.

**Thank you for Sharing Wisdom!**

**Managing Editor**

**IJACSA**

**Volume 11 Issue 5 May 2020**

**ISSN 2156-5570 (Online)**

**ISSN 2158-107X (Print)**

**©2013 The Science and Information (SAI) Organization**

# Editorial Board

## Editor-in-Chief

### **Dr. Kohei Arai - Saga University**

*Domains of Research: Technology Trends, Computer Vision, Decision Making, Information Retrieval, Networking, Simulation*

---

## Associate Editors

### **Chao-Tung Yang**

**Department of Computer Science, Tunghai University, Taiwan**

*Domain of Research: Software Engineering and Quality, High Performance Computing, Parallel and Distributed Computing, Parallel Computing*

### **Elena SCUTELNICU**

**"Dunarea de Jos" University of Galati, Romania**

*Domain of Research: e-Learning, e-Learning Tools, Simulation*

### **Krassen Stefanov**

**Professor at Sofia University St. Kliment Ohridski, Bulgaria**

*Domains of Research: e-Learning, Agents and Multi-agent Systems, Artificial Intelligence, Big Data, Cloud Computing, Data Retrieval and Data Mining, Distributed Systems, e-Learning Organisational Issues, e-Learning Tools, Educational Systems Design, Human Computer Interaction, Internet Security, Knowledge Engineering and Mining, Knowledge Representation, Ontology Engineering, Social Computing, Web-based Learning Communities, Wireless/ Mobile Applications*

### **Maria-Angeles Grado-Caffaro**

**Scientific Consultant, Italy**

*Domain of Research: Electronics, Sensing and Sensor Networks*

### **Mohd Helmy Abd Wahab**

**Universiti Tun Hussein Onn Malaysia**

*Domain of Research: Intelligent Systems, Data Mining, Databases*

### **T. V. Prasad**

**Lingaya's University, India**

*Domain of Research: Intelligent Systems, Bioinformatics, Image Processing, Knowledge Representation, Natural Language Processing, Robotics*

# CONTENTS

Paper 1: A Service Scheduling Security Model for a Cloud Environment

*Authors: Abdullah Sheikh, Malcolm Munro, David Budgen*

**PAGE 1-9**

Paper 2: Robust Speed Control for Networked DC Motor System

*Authors: Sokliep Pheng, Luo Xiaonan, Rachana Lav, Zhongshuai Wang, Zetao Jiang*

**PAGE 10-17**

Paper 3: Forest Fire Detection System using LoRa Technology

*Authors: Nicoleta Cristina GAITAN, Paula HOJBOTA*

**PAGE 18-21**

Paper 4: Detecting C&C Server in the APT Attack based on Network Traffic using Machine Learning

*Authors: Cho Do Xuan, Lai Van Duong, Tisenko Victor Nikolaevich*

**PAGE 22-27**

Paper 5: An Abnormal Behavior Detection Method using Optical Flow Model and OpenPose

*Authors: Zhu Bin, Xie Ying, Luo Guohu, Chen Lei*

**PAGE 28-34**

Paper 6: Clash between Segment-level MT Error Analysis and Selected Lexical Similarity Metrics

*Authors: Marija Brkic Bakaric, Kristina Tonkovic, Lucia Nacinovic Prskalo*

**PAGE 35-42**

Paper 7: Technologies for Making Reliable Decisions on a Variety of Effective Factors using Fuzzy Logic

*Authors: Yousef Ibrahim Daradkeh, Irina Tvoroshenko*

**PAGE 43-50**

Paper 8: Adaptive Hybrid Synchronization Primitives: A Reinforcement Learning Approach

*Authors: Fadai Ganjaliyev*

**PAGE 51-57**

Paper 9: Creating Knowledge Environment during Lean Product Development Process of Jet Engine

*Authors: Zehra C. Araci, Ahmed Al-Ashaab, Muhammad Usman Tariq, Jan H. Braasch, Emre M. C. Simsekler*

**PAGE 58-62**

Paper 10: The Prototype of Thai Blockchain-based Voting System

*Authors: Krittaphas Wisessing, Phattaradon Ekthammabordee, Thattapon Surasak, Scott C.-H. Huang, Chakkrit Preuksakarn*

**PAGE 63-68**

Paper 11: Satellite Image Database Search Engine which Allows Fuzzy Expression of Geophysical Parameters of Queries

*Authors: Kohei Arai*

**PAGE 69-73**

Paper 12: Difficulties in Teaching Online with Blackboard Learn Effects of the COVID-19 Pandemic in the Western Branch  
Colleges of Qassim University

*Authors: Fahad Alturise*

PAGE 74-81

Paper 13: Using Geographical Information System for Mapping Public Schools Distribution in Jeddah City

*Authors: Abdulkader A. Murad, Abdulmuakhir I. Dalhat, Ammar A Naji*

PAGE 82-90

Paper 14: An Efficient Approach for Storage of Big Data Streams in Distributed Stream Processing Systems

*Authors: Sultan Alshamrani, Quadri Waseem, Abdullah Alharbi, Wael Alosaimi, Hamza Turabieh, Hashem Alyami*

PAGE 91-98

Paper 15: Comparative Analysis of Methodologies for Domain Ontology Development: A Systematic Review

*Authors: Abdul Sattar, Ely Salwana Mat Surin, Mohammad Nazir Ahmad, Mazida Ahmad, Ahmad Kamil Mahmood*

PAGE 99-108

Paper 16: Performance Analysis of Transient Fault-Injection and Fault-Tolerant System for Digital Circuits on FPGA

*Authors: Sharath Kumar Y N, Dinesha P*

PAGE 109-115

Paper 17: Evaluation of the Diffusion Phenomenon using Information from Twitter

*Authors: Kohei Otake, Takashi Namatame*

PAGE 116-121

Paper 18: Machine Learning and Statistical Modelling for Prediction of Novel COVID-19 Patients Case Study: Jordan

*Authors: Ebaa Fayyumi, Sahar Idwan, Heba AboShindi*

PAGE 122-126

Paper 19: Analytical Study between Human Urban Planning and Geographic Information Systems: "The Case of the City of Casablanca"

*Authors: Mohamed Rtal, Mostafa Hanoun*

PAGE 127-132

Paper 20: Clustering-Based Trajectory Outlier Detection

*Authors: Eman O. Eldawy, Hoda M.O. Mokhtar*

PAGE 133-139

Paper 21: Measuring the Similarity between the Sanskrit Documents using the Context of the Corpus

*Authors: Jatinderkumar R. Saini, Prafulla B. Bafna*

PAGE 140-145

Paper 22: An Efficient Model for Mining Outlier Opinions

*Authors: Neama Hassan, Laila A. Abd-Elmegid, Yehia K. Helmy*

PAGE 146-153

Paper 23: Coronavirus Social Engineering Attacks: Issues and Recommendations

*Authors: Ahmed Alzahrani*

PAGE 154-161

**Paper 24: Hybrid based Energy Efficient Cluster Head Selection using Camel Series Elephant Herding Optimization Algorithm in WSN**

*Authors: N. Lavanya, T. Shankar*

**PAGE 162-169**

**Paper 25: Indexed Metrics for Link Prediction in Graph Analytics**

*Authors: Marcus Lim, Azween Abdullah, NZ Jhanjhi, Mahadevan Supramaniam*

**PAGE 170-178**

**Paper 26: Radar GPR Application to Explore and Study Archaeological Sites: Case Study**

*Authors: Ahmed Faize, Gamil Alsharahi, Mohammed Hamdaoui*

**PAGE 179-182**

**Paper 27: Generalized Approach to Analysis of Multifractal Properties from Short Time Series**

*Authors: Lyudmyla Kirichenko, Abed Saif Ahmed Alghawli, Tamara Radivilova*

**PAGE 183-198**

**Paper 28: Play-Centric Designing of a Serious Game Prototype for Low Vision Children**

*Authors: Nurul Izzah Othman, Nor Azan Mat Zin, Hazura Mohamed*

**PAGE 199-205**

**Paper 29: Principal Component Analysis on Morphological Variability of Critical Success Factors for Enterprise Resource Planning**

*Authors: Ayogeboh Epizitone, Oludayo. O. Olugbara*

**PAGE 206-217**

**Paper 30: A Comprehensive Science Mapping Analysis of Textual Emotion Mining in Online Social Networks**

*Authors: Shivangi Chawla, Monica Mehrotra*

**PAGE 218-229**

**Paper 31: Prediction of Heart Diseases (PHDs) based on Multi-Classifiers**

*Authors: Amirah Al Shammari, Haneen al Hadeaf, Hedia Zardi*

**PAGE 230-236**

**Paper 32: A New Approach to Predicting Learner Performance with Reduced Forgetting**

*Authors: Dagou Dangui Augustin Sylvain Legrand KOFFI, Tchimou N'TAKPE, Assouhoun ADJE, Souleymane OUMTANAGA*

**PAGE 237-244**

**Paper 33: Genetic Algorithm with Comprehensive Sequential Constructive Crossover for the Travelling Salesman Problem**

*Authors: Zakir Hussain Ahmed*

**PAGE 245-254**

**Paper 34: A Workflow Scheduling Algorithm for Reducing Data Transfers in Cloud IaaS**

*Authors: Jean Edgard GNIMASSOUN, Tchimou N'TAKPE, Gokou Hervé Fabrice DIEDIE, Souleymane OUMTANAGA*

**PAGE 255-263**

**Paper 35: Minimization of Spectrum Fragmentation for Improvement of the Quality of Service in Multifiber Elastic Optical Networks**

*Authors: Boris Stéphane ZOUNEME, Georges Nogbou ANOH, Souleymane OUMTANAGA*

**PAGE 264-271**

**Paper 36: The Effect of Firm's Size on Corporate Performance**

*Authors: Meiryani, Olivia, Jajat Sudrajat, Zaidi Mat Daud*

**PAGE 272-277**

**Paper 37: Image Captioning using Deep Learning: A Systematic Literature Review**

*Authors: Murk Chohan, Adil Khan, Muhammad Saleem Mahar, Saif Hassan, Abdul Ghafoor, Mehmood Khan*

**PAGE 278-286**

**Paper 38: A Trait-based Deep Learning Automated Essay Scoring System with Adaptive Feedback**

*Authors: Mohamed A. Hussein, Hesham A. Hassan, Mohammad Nassef*

**PAGE 287-293**

**Paper 39: Mutual Coexistence in WBANs: Impact of Modulation Schemes of the IEEE 802.15.6 Standard**

*Authors: Marwa BOUMAIZ, Mohammed EL GHAZI, Mohammed FATTAH, Anas BOUAYAD, Moulhime EL BEKKALI*

**PAGE 294-302**

**Paper 40: Ensemble Machine Learning Model for Higher Learning Scholarship Award Decisions**

*Authors: Wirawati Dewi Ahmad, Azuraliza Abu Bakar*

**PAGE 303-312**

**Paper 41: Where is the Highest Rate of Children with Anemia in Peru? An Answer using Grey Systems**

*Authors: Alexi Delgado, Nicolly Perez Ccancce*

**PAGE 313-317**

**Paper 42: Mapping UML Sequence Diagram into the Web Ontology Language OWL**

*Authors: Mo'men Elsayed, Nermeen Elkashef, Yasser F. Hassan*

**PAGE 318-326**

**Paper 43: Intelligent Fleet Management System for Open Pit Mine**

*Authors: Hajar BNOUACHIR, Meryiem CHERGUI, Nadia MACHKOUR, Mourad ZEGRARI, Aziza CHAKIR, Laurent DESHAYES, Atae SEMMAR, Hicham MEDROMI*

**PAGE 327-332**

**Paper 44: A Human Gait Recognition Against Information Theft in Smartphone using Residual Convolutional Neural Network**

*Authors: Gogineni Krishna Chaitanya, Krovi.Raja Sekhar*

**PAGE 333-340**

**Paper 45: The Cuckoo Feature Filtration Method for Intrusion Detection (Cuckoo-ID)**

*Authors: Wafa Alsharafat*

**PAGE 341-347**

**Paper 46: Segmentation of Fuzzy Enhanced Mammogram Mass Images by using K-Mean Clustering and Region Growing**

*Authors: Nidhi Singh, S. Veenadhari*

**PAGE 348-352**

**Paper 47: IoT Technology for Facilities Management: Understanding End user Perception of the Smart Toilet**

*Authors: Venushini Raendran, R. Kanesaraj Ramasamy, Intan Soraya Rosdi, Ruzanna Ab Razak, Nurazlin Mohd Fauzi*

**PAGE 353-359**

**Paper 48: New Learning Approach for Unsupervised Neural Networks Model with Application to Agriculture Field**

*Authors: Belattar Sara, Abdoun Otman, El khatir Haimoudi*

**PAGE 360-369**

**Paper 49: A Secured Large Heterogeneous HPC Cluster System using Massive Parallel Programming Model with Accelerated GPUs**

*Authors: Khalid Alsubhi*

**PAGE 370-376**

**Paper 50: Metaheuristic for the Capacitated Multiple Traveling Repairman Problem**

*Authors: Khanh-Phuong Nguyen, Ha-Bang Ban*

**PAGE 377-384**

**Paper 51: An Efficient Methodology for Water Supply Pipeline Risk Index Prediction for Avoiding Accidental Losses**

*Authors: Muhammad Shuaib Qureshi, Ayman Aljarbough, Muhammad Fayaz, Muhammad Bilal Qureshi, Wali Khan Mashwani, Junaid Khan*

**PAGE 385-393**

**Paper 52: Modeling of Coronavirus Behavior to Predict it's Spread**

*Authors: Shakir Khan, Amani Alfaifi*

**PAGE 394-399**

**Paper 53: Automatic Segmentation of Hindi Speech into Syllable-Like Units**

*Authors: Ruchika Kumari, Amita Dev, Ashwani Kumar*

**PAGE 400-406**

**Paper 54: Identification and Assesment of the Specific Absorption Rate (SAR) Generated by the most used Telephone in Peru in 2017**

*Authors: Natalia Indira Vargas-Cuentas, Mark Clemente-Arenas, Avid Roman-Gonzalez, Roxana Moran-Morales*

**PAGE 407-413**

**Paper 55: Multi Focus Image Fusion using Image Enhancement Techniques with Wavelet Transformation**

*Authors: Sarwar Shah Khan, Muzammil Khan, Yasser Alharbi*

**PAGE 414-420**

**Paper 56: Multiclass Pattern Recognition of Facial Images using Correlation Filters**

*Authors: Nisha Chandran S, Charu Negi, Poonam Verma*

**PAGE 421-428**

**Paper 57: Analytical Comparison Between the Information Gain and Gini Index using Historical Geographical Data**

*Authors: Majid Zaman, Sameer Kaul, Muheef Ahmed*

**PAGE 429-440**

**Paper 58: Disparity of Stereo Images by Self-Adaptive Algorithm**

*Authors: Md. Abdul Mannan Mondal, Mohammad Haider Ali*

**PAGE 441-454**

**Paper 59: Natural Language Processing based Anomalous System Call Sequences Detection with Virtual Memory Introspection**

*Authors: Suresh K. Peddoju, Himanshu Upadhyay, Jayesh Soni, Nagarajan Prabakar*

**PAGE 455-460**

**Paper 60: Parkinson's Disease Diagnosis using Spiral Test on Digital Tablets**

*Authors: Najd Al-Yousef, Raghad Al- Saikhan, Reema Al- Gowailly, Reem Al-Abdullatif, Felwa Al-Mutairi, Ouiem Bchir*

**PAGE 461-470**

**Paper 61: Still Image-based Human Activity Recognition with Deep Representations and Residual Learning**

*Authors: Ahsan Raza Siyal, Zuhaibuddin Bhutto, Syed Muhammad Shehram Shah, Azhar Iqbal, Faraz Mehmood, Ayaz Hussain, Saleem Ahmed*

**PAGE 471-477**

**Paper 62: Design and Experimental Analysis of Touchless Interactive Mirror using Raspberry Pi**

*Authors: Abdullah Memon, Syed Mudassir Kazmi, Atfiya Baqai, Fahim Aziz Umrani*

**PAGE 478-485**

**Paper 63: An Enhanced Distance Vector-Hop Algorithm using New Weighted Location Method for Wireless Sensor Networks**

*Authors: Fengrong Han, Izzeldin Ibrahim Mohamed Abdelaziz, Xinni Liu, Kamarul Hawari Ghazali*

**PAGE 486-495**

**Paper 64: Early Forest Fire Detection System using Wireless Sensor Network and Deep Learning**

*Authors: Wiame Benzekri, Ali El Moussati, Omar Moussaoui, Mohammed Berrajaa*

**PAGE 496-503**

**Paper 65: A Process Model Collection with Structural Variants and Evaluations**

*Authors: Rimsha Anam, Tahir Muhammad*

**PAGE 504-512**

**Paper 66: A Gamification Experience and Virtual Reality in Teaching Astronomy in Basic Education**

*Authors: Norka Bedregal-Alpaca, Olha Sharhorodska, Luis Jiménez-González, Robert Arce-Apaza*

**PAGE 513-521**

**Paper 67: Customer Churn Prediction Model and Identifying Features to Increase Customer Retention based on User Generated Content**

*Authors: Essam Abou el Kassem, Shereen Ali Hussein, Alaa Mostafa Abdelrahman, Fahad Kamal Alsheref*

**PAGE 522-531**

**Paper 68: Oscillation Preventing Closed-Loop Controllers via Genetic Algorithm for Biped Walking on Flat and Inclined Surfaces**

*Authors: Sabri Yilmaz, Metin Gokasan, Seta Bogosyan*

**PAGE 532-544**

**Paper 69: Using Fuzzy c-Means for Weighting Different Fuzzy Cognitive Maps**

*Authors: Mamoon Obiedat, Ali Al-yousef, Ahmad Khasawneh, Nabhan Hamadneh, Ashraf Aljammal*

**PAGE 545-551**

**Paper 70: Wind Power Integration with Smart Grid and Storage System: Prospects and Limitations**

*Authors: Saad Bin Abul Kashem, Muhammad E. H. Chowdhury, Amith Khandakar, Jubaer Ahmed, Azad Ashraf, Nushrat Shabrin*

**PAGE 552-569**

**Paper 71: Voice Scrambling Algorithm based on 3D Chaotic Map System (VSA3DCS) to Encrypt Audio Files**

*Authors: Osama M. Abu Zaid*

**PAGE 570-579**

**Paper 72: Air Quality Monitoring Device for Vehicular Ad Hoc Networks: EnvioDev**

*Authors: Josip Balen, Srdan Ljepic, Kristijan Lenac, Sadko Mandzuka*

**PAGE 580-590**

**Paper 73: Urbanization Change Analysis based on SVM and RF Machine Learning Algorithms**

*Authors: Farhad Hassan, Tauqeer Safdar, Ghulam Itaza, Aman Ullah Khan, Syed Muhammad Husnain Kazmi, Farah Murtaza*

**PAGE 591-601**

**Paper 74: Development of a Recurrent Neural Network Model for English to Yorùbá Machine Translation**

*Authors: Adebimpe Esan, John Oladosu, Christopher Oyeleye, Ibrahim Adeyanju, Olatayo Olaniyan, Nnamdi Okomba, Bolaji Omodunbi, Opeyemi Adanigbo*

**PAGE 602-609**

**Paper 75: Thinging-Oriented Modeling of Unmanned Aerial Vehicles**

*Authors: Sabah Al-Fedaghi, Jassim Al-Fadhli*

**PAGE 610-619**

**Paper 76: Deep Learning Model for Identifying the Arabic Language Learners based on Gated Recurrent Unit Network**

*Authors: Seifeddine Mechti, Roobaea Alroobaea, Moez Krichen, Saeed Rubaiee, Anas Ahmed*

**PAGE 620-627**

**Paper 77: Quantifying Feature Importance for Detecting Depression using Random Forest**

*Authors: Hatoon AlSaghi, Mourad Ykhlef*

**PAGE 628-635**

**Paper 78: Parallel QR Factorization using Givens Rotations in MPI-CUDA for Multi-GPU**

*Authors: Miguel Tapia-Romero, Amilcar Meneses-Viveros, Erika Hernández-Rubio*

**PAGE 636-645**

**Paper 79: VIPEye: Architecture and Prototype Implementation of Autonomous Mobility for Visually Impaired People**

*Authors: Waqas Nawaz, Khalid Bashir, Kifayat Ullah Khan, Muhammad Anas, Hafiz Obaid, Mughees Nadeem*

**PAGE 646-654**

**Paper 80: A High Performance System for the Diagnosis of Headache via Hybrid Machine Learning Model**

*Authors: Ahmad Qawasmeh, Noor Alhusan, Feras Hanandeh, Maram Al-Atiyat*

**PAGE 655-663**

**Paper 81: Efficient Cache Architecture for Table Lookups in an Internet Router**

*Authors: Hayato Yamaki*

**PAGE 664-672**

**Paper 82: A Robust Scheme to Improving Security of Data using Graph Theory**

*Authors: Khalid Bekkaoui, Soumia Ziti, Fouzia Omary*

**PAGE 673-680**

**Paper 83: 2D/3D Registration with Rigid Alignment of the Pelvic Bone for Assisting in Total Hip Arthroplasty Preoperative Planning**

*Authors: Christian A. Suca Velando, Eveling G. Castro Gutierrez*

**PAGE 681-688**

**Paper 84: Evaluating Contact Detection, Size Recognition and Grasping State of an Object using Soft Elastomer Gripper**

*Authors: Kazi Abaul Jamil*

**PAGE 689-694**

**Paper 85: Ensemble Methods to Detect XSS Attacks**

*Authors: PMD Nagarjun, Shaik Shakeel Ahamad*

**PAGE 695-700**

**Paper 86: Evaluation of a Model Maximizing the Quality Value of Selected Software Components in a Library**

*Authors: Koffi Kouakou Ive Arsene, Samassi Adama, Kouamé Appoh*

**PAGE 701-707**

**Paper 87: Feature Selection and Performance Improvement of Malware Detection System using Cuckoo Search Optimization and Rough Sets**

*Authors: Ravi Kiran Varma P, PLN Raju, K V Subba Raju, Akhila Kalidindi*

**PAGE 708-714**

# A Service Scheduling Security Model for a Cloud Environment

Abdullah Sheikh<sup>1</sup>, Malcolm Munro<sup>2</sup>, David Budgen<sup>3</sup>  
Department of Computer Science  
Durham University  
Durham, United Kingdom

**Abstract**—Scheduling tasks on a standalone system can be complex but applying it to a cloud environment can be even more complex because of the large amount of resources available. An added complexity in a Cloud environment is that of security. This paper addresses scheduling from a security point of view and presents a Scheduling Security Model and evaluates its effectiveness to meet user’s requirements with a number of worked examples with different scenarios.

**Keywords**—Scheduling; security; model; cost; cloud computing

## I. INTRODUCTION

Scheduling tasks on a standalone system to meet the needs of users can be complex but applying it to a cloud environment can be even more complex. This is because of the large amount of different resources available in a Cloud environment. An added complexity in a Cloud environment is that of security.

The National Institute of Standard and Technology (NIST) [1] gives a basic definition of Cloud Computing as a model for a customer to request an on-demand convenient services, that allow to use and access network resources such as servers, data storage, software and applications with indirect provider interactions. There are a various criteria to classify Scheduling Models in Cloud Computing to produce reliable, trusted and secure services, such as scheduling performance [2], cost scheduling [3], and scheduling based on security [4].

Also, NIST [1] specifies the deployment models to four deployment models, which are are: Public, Private, Community, Hybrid. These deployment models identify the accessibility to the cloud service [5]. Dillon et al. [6] indicate some security concerns about the deployment models that include data privacy and trust, service policies, and data transfer. So, these security concerns make the service provider aware to address these issues which have been focused by [7]. Also, [8] describe the main issues that can affect the cloud service models. Then security policies need to be applied to deal with data access and security.

As an example consider a service that consists of two tasks. First task is analysing data with basic security requirement. The second task is a high security task about saving private data. Thus this service requires two resources, one for each task, with different levels of security. In general, the higher the security the higher the cost of using a resource.

This requires that security levels need to be defined. In [9] the security levels for the SSM are driven from Watson [10]

as five security levels summarised on Fig. 1:

- 1) The first level 1: It is all about applying essential security level that can run tasks on trusted public resources.
- 2) The second level 2: It is all about applying more security setting that can run tasks on more trusted public resources.
- 3) The third level 3: It is all about applying highly security setting to run tasks on highly trusted public resources.
- 4) The fourth level 4: It is all about applying advance security setting to run tasks on trusted private resources.
- 5) The fifth level 5: It is all about applying more highly security setting to run tasks on highly trusted private resource.

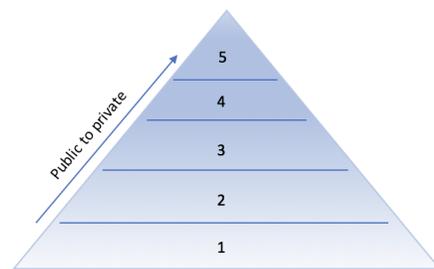


Fig. 1. The Five Security Levels Applied for Resources [9].

Thus for the previous example two resources are required. One of the tasks will be on a resource with Level 1 security and the other on a resource with Level 5 security.

A Scheduling Security Model (SSM) [11] [12] [13] is been devised to make security the primary factor for scheduling algorithm. Imbedded in the SSM is a cost calculation together with a scheduling algorithm. To resolve some scheduling issues within a resource each task is given an importance level. Other factors included in the model are times for each task, resource cost per hour, and a Quality of Service cost.

In the SSM a service consists of M tasks that can be executed on N resources. Each task is given a Security Level value of 1 to 5. Tasks with the same Security Level will be run on the same resource, thus N ranges from 1 to 5.

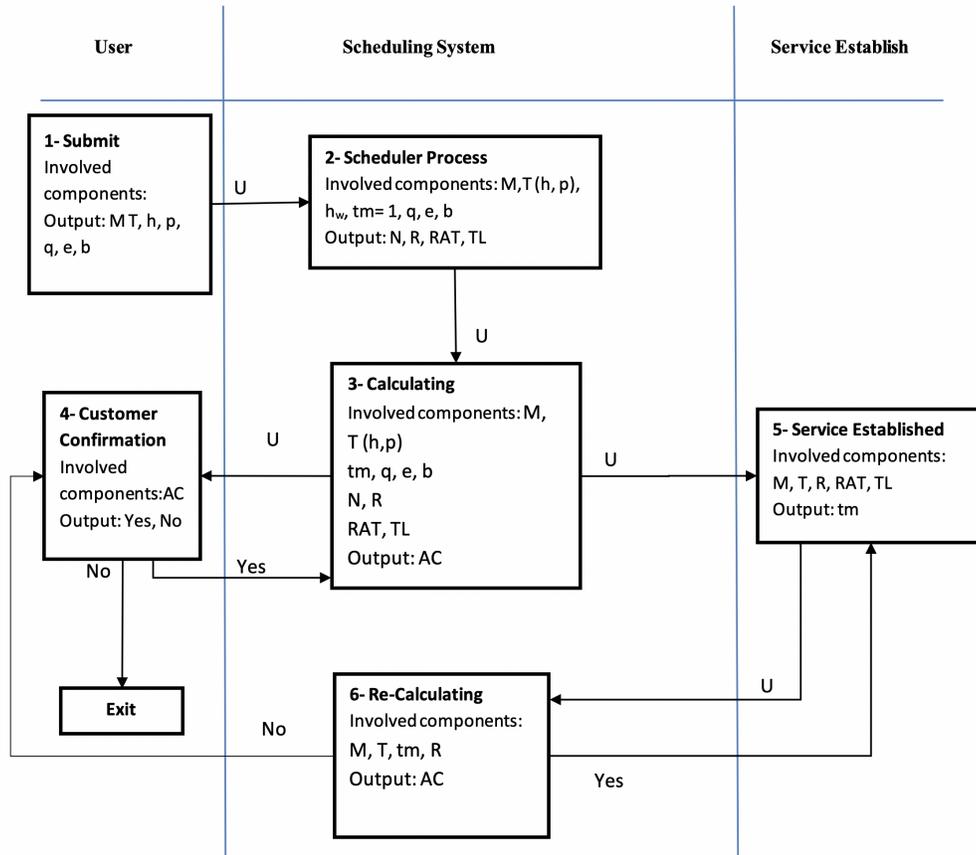


Fig. 2. Scheduling Security Model Function

The contribution of this paper is to address scheduling from a security point of view and present a Scheduling Security Model and evaluate its effectiveness to meet user's requirements with a number of worked examples with different scenarios.

## II. MODEL DEFINITION

This section serves as a definition of the SSM and its components. Then it summarises the SSM function.

The SSM [13] is defined formally as follows.

First, the SSM categorises all tasks based on the tasks security level. Then the SSM calculates initial cost for service. Then, the SSM requests confirmation from the customer to start the service. Next, establishing the service after the receiving the confirmations. Then the SSM recalculates the actual cost.

The suffixes used in the definitions are specified as follows:

$i$  : Tasks

$k$  : Resources

$j$  : Security Level

$l$  : Task Importance

The values in the definitions are:

- $tm$ : Time cost  
 $tm_i$ : time cost for task  $i$
- $q$ : Quality of Service for the service.
- $b$ : Customer budget for the service.
- $h$ : Security level for a task.  
 $h_j$ : Security level  $\in \{1, 2, 3, 4, 5\}$
- $hw$ : Security weight for the security level (for each task).  
 $hw_j$ : Security weight for security level  $j$ .
- Each task ( $t_i$ ) has a security level  $h_j$   
(and then a security weight  $hw_j$ )
- $R$ : Set of resources ( $R_k$ ,  $k \in \{1, 2, 3, 4, 5\}$ )  
(can use up to to 5 resources  $\{R_1, R_2, R_3, R_4, R_5\}$ )  
then, used resources will be numbered from 1 to  $N$   
 $N$ : is the number of resources used ( $N$  determined later)
- $R_k$ : is a set of tasks, where the  $hw_j = R w_k$
- $Rw$ : Resource security weight (for each resource)  
 $R w_k$ : Security weight for resource  $k$  ( $R_k$ )

$$Rw_k = hw_j$$

- $p$ : Tasks Importance,  $p_i$ : is the importance  $\{1, 2, 3\}$  task  $t_i$  has importance  $p_i$
- $M$ : Number of tasks.
- $N$ : Number of resources.
- $e$ : Maximum time.

The main components of the SSM:

- All tasks including the total number of tasks  $M$ , and Tasks Set  $T$ .
- The time duration for executing the task time cost : The elapsed time (EPT)  $tm_i$ .
- The required Quality of Service (QoS) level of the service  $q$ .
- The overall requested service budget  $b$ .
- The Security Level for each task  $h_j$ . The SSM will identify the security weight  $hw_j$  for each task  $t_i$ . Then The SSM allocates the tasks to the correspondents resources  $R$  with similar resource weight  $Rw_k$ . The SSM allocates and executes all tasks on the same Resource with same security weight.
- Task Importance Level  $p_i$  for each task, and the SSM defines these levels into three levels.

The reason for making the Task Importance in three levels is that the scheduling in the SSM is serving the security as a category, then there is a need to give each task within the same category an order to be executed. So, the order will be as identified with this three levels but if there are some tasks with the same Task Importance level then the scheduling will be for first come first served.

Table I shows a simple example of ordering tasks in same resource depends on task importance.

TABLE I. SIMPLE EXAMPLE OF ORDERING TASKS BASED ON TASK IMPORTANCE

Security Level(Weight)/Importance	1	2	3	$R_k$
1 (0.00)				
2 (0.25)				
3 (0.50)	$t_3$	$t_1$	$t_2$	$R_1$
4 (0.75)				
5 (1.00)				

Another simple example with tasks dependencies will be discussed in details later in the Scheduling Process. Tasks submitted  $t_1: h_1 = 2, p_1 = 1$ ,  $t_2: h_2 = 2, p_2 = 3$ , and  $t_3: h_3 = 4, p_3 = 2$ . Then the Task dependency  $t_3$  depends on  $t_1$ . The SSM will analyse this input and allocates Tasks to two Resources  $R_1$  and  $R_2$ . Tasks  $t_1$  and  $t_2$  will be allocated to  $R_1$  and  $t_3$  to  $R_2$ . But with the dependencies required  $t_1$  will assigned to the Fast-Track list. That means  $t_1$  will be executed first then  $t_3$  depends on  $t_1$  and it has the highest Security Level that lets the SSM to list first see Table II.

TABLE II. SIMPLE EXAMPLE OF ORDERING TASKS WITH DEPENDENCIES AND FAST-TRACK

Security Level(Weight)/Importance	1	2	3	$R_k$
1 (0.00)				
2 (0.25)	$t_1^{FT}$	$t_2$	$R_1$	
3 (0.50)				
4 (0.75)		$t_3$		$R_2$
5 (1.00)				

Table III shows a summary of the components that specify the customer requirement for requesting a service. Then SSM will analyse the requirements for the calculating step. Example of a service required is shown in Table IV.

- $AC$  Actual Cost of the service.
  - Resource Cost  $RC$  = Cost of resources for  $RT_k$  hours.

$$Resource\ Cost(RC) = \sum_{k=1}^N (RC_k * RT_k) \quad (1)$$

Where  $RC_k$  is Resource Cost for Resource  $k$  per hour and  $RT_k$  is the actual time used by Resource  $k$  in hours.

- QoS Cost =  $RC * Quality\ of\ Service\ (q)$  required

$$QoS\ Cost = RC * q \quad (2)$$

- Security Cost for each Resource

$$SC_k = RC_k * Rw_k * RT_k \quad (3)$$

Where  $Rw_k$  is Security weight for Resource  $k$ .

- Security Cost for all Resources

$$SC = \sum_{k=1}^N SC_k = \sum_{k=1}^N (RC_k * Rw_k * RT_k) \quad (4)$$

- Actual Cost ( $AC$ ) =  $RC + SC + QoS\ Cost$

Therefore

$$AC = \sum_{k=1}^N (RC_k * RT_k) + q * \sum_{k=1}^N (RC_k * RT_k) + \sum_{k=1}^N (RC_k * Rw_k * RT_k) \quad (5)$$

Then

$$AC = \sum_{k=1}^N ((RC_k * RT_k) * (1 + q + Rw_k)) \quad (6)$$

Fig. 2 shows the SSM function and summarised as the following:

TABLE III. SUMMARY OF THE CUSTOMER INPUTS

Component	Values	Range
Budget	$b$	$b > 0$
Maximum Time	$e$	$0 < e \leq 60$
QoS	$q$	0.0, 0.1, 0.2,...,1.0
Tasks	$t_1, t_2, t_3, \dots, t_i$ , indexed by $i, i \in \{1 - M\}$	M Tasks
Task Security Level	$h_1, h_2, h_3, \dots, h_j$ , indexed by $j$	$j \in \{1, 2, 3, 4, 5\}$
Task Importance	$p_1, p_2, p_3, \dots, p_l$ , indexed by $l$	$l \in \{1, 2, 3\}$

TABLE IV. EXAMPLE OF A SERVICE REQUIRED

Security Level(Weight)/Importance	1	2	3	$R_k$	$RC_k$
1 (0.00)					
2 (0.25)		$t_1$		$R_1$	20
3 (0.50)	$t_2$			$R_2$	20
4 (0.75)					
5 (1.00)					

1) **Submit:**

- Budget  $b$ .
- Maximum Time Required  $e$ : Minimum 60 minutes (one hour).
- The QoS  $q$ .
- The Number of Tasks  $M$ .
- $t_i: h_j \in \{1, 2, 3, 4, 5\}, p_l \in \{1, 2, 3\}$ .

2) **Scheduler Process:**

The SSM creates categories for all tasks. The number of categories is equal to the number of the security levels required. For each category, tasks will be ordered depending on the tasks importance. The task with higher task importance will be run first down to the task with lowest task importance. If there are more than one task with the same task importance the SSM will put them in the task number order.

3) **Calculations:**

Actual Cost  $AC$  calculated and compared to budget  $b$ .

4) **Customer Confirmation:**

The customer takes last decision to confirm the service.

5) **Establishing the Service:**

The SSM starts the service and all tasks allocated to resources.

6) **Re-Calculating:**

At this step, the SSM aims to Re-calculates the  $AC$  depends on the actual running time for all resources considering dependencies that might cause delays over other resources.

A. Example: 1

Table V shows the details of a service request:

TABLE V. SERVICE REQUIREMENT FOR EXAMPLE 1

Security Level(Weight)/Importance	1	2	3	Resource	$RC_k$
1 (0.00)					
2 (0.25)			$t_1$	$R_1$	20
3 (0.50)		$t_2$		$R_2$	20
4 (0.75)					
5 (1.00)					

1) *First Scenario: 1.1:* The running time for each resource identified as follows:

- $R_1 : t_1$  and  $tm_1 = 18$  minutes
- $R_2 : t_2$  and  $tm_2 = 13$  minutes

Therefore  $AC: tm_1 = 18, RT_1: \text{Time for } R_1 = 18$

$tm_2 = 13, RT_2: \text{Time for } R_2 = 13$

$$AC = ((20*1) * (1+0+0.25)) + ((20*1) * (1+0+0.50))$$

$$AC = 25 + 30 = 55$$

Fig. 3 shows tasks timeline over the allocated resources.

**Re-Calculation:**

The SSM calculates the actual running time as follows:  
 $AC = ((20*18/60)*(1+0+0.25)) + ((20*13/60)*(1+0+0.50))$

$$\text{Therefore, } AC = 7.5 + 6.5 = 14$$



Fig. 3. Example 1. Scenario 1.1. Tasks Timeline

III. EXAMPLES OF COSTS

The following examples discusses Re-calculation with different scenarios where QoS cost  $q = 0$ , and all times represented in minutes.

B. Example: 2

Table VI shows the details of a service request:

TABLE VI. SERVICE REQUIREMENT FOR EXAMPLE 2

Security Level(Weight)/ Importance	1	2	3	Resource	$RC_k$
1 (0.00)					
2 (0.25)	$t_1$	$t_2$		$R_1$	20
3 (0.50)					
4 (0.75)	$t_3$	$t_4$		$R_2$	20
5 (1.00)	$t_5$	$t_6$		$R_3$	20

1) *Second Scenario: 2.1:* This scenario shows the time for each resource and the tasks allocated with Fast-Track as follows:

- $R_1 : t_1^{FT}$  and  $tm_1 = 18, t_2$  and  $tm_2 = 15$
- $R_2 : t_3^{FT}$  and  $tm_3 = 13, t_4$  and  $tm_4 = 10$
- $R_3 : t_6$  and  $tm_6 = 10, t_5$  and  $tm_5 = 5$

The dependencies are:  $t_5$  depends on  $t_1$  and  $t_6$  depends on  $t_3$ . As a result of the dependencies, the SSM finds a delay of executing tasks  $t_6$  and  $t_5$ . So,  $AC$  calculated as follows:

$$tm_1 = 18, tm_2 = 15, RT_1 \text{ Time for } R_1 = 18+15 = 33$$

$$tm_3 = 13, tm_4 = 10, RT_2: \text{ Time for } R_2 = 13+10 = 23$$

$$tm_6 = 10, tm_5 = 5, RT_3: \text{ Time for } R_3 = 10+5 = 15$$

See Fig. 4 for tasks timeline. So, the SSM adds the total waiting time to  $RT_3$  because  $tm_3 < tm_1$ :  
 $R_3 = 13+10+5 = 28$ .

**Re-Calculation:**

So,  $AC = 13.75 + 13.42 + 18.67 = 45.84$   
If the SSM ignores the delay time as follows:

$$tm_1 = 18, tm_2 = 15, RT_1 \text{ Time for } R_1 = 18 + 15 = 33$$

$$tm_3 = 13, tm_4 = 10, RT_2: \text{ Time for } R_2 = 13 + 10 = 23$$

$$tm_6 = 10, tm_5 = 5, RT_3 \text{ Time for } R_3 = 10 + 5 = 15$$

**Re-Calculation:**

Then,  $AC = 13.75 + 13.42 + 10.00 = 37.17$   
As a result,  $AC$  of  $EPT < AC$  of the actual running time.

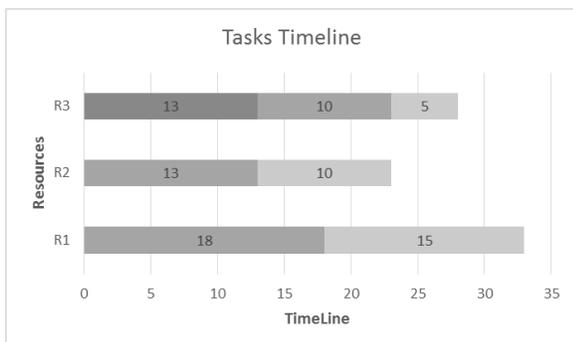


Fig. 4. Example 2. Scenario 2.1. Tasks Timeline

2) *Scenario: 2.2:* This scenario shows the time for each resource and the tasks allocated with Fast-Track as follows:

- $R_1 : t_1^{FT}$  and  $tm_1 = 8, t_2$  and  $tm_2 = 3$
- $R_2 : t_3^{FT}$  and  $tm_3 = 10, t_4$  and  $tm_4 = 4$
- $R_3 : t_6$  and  $tm_6 = 5, t_5$  and  $tm_5 = 7$

The dependencies are:  $t_5$  depends on  $t_1$  and  $t_6$  depends on  $t_3$ . As a result of the dependencies, the SSM finds a delay of executing tasks  $t_6$  and  $t_5$ . So,  $AC$  calculated as follows:

$$tm_1 = 8, tm_2 = 3, RT_1: \text{ Time for } R_1 = 8 + 3 = 11$$

$$tm_3 = 10, tm_4 = 4, RT_2: \text{ Time for } R_2 = 10 + 4 = 14$$

$$tm_6 = 5, tm_5 = 7, RT_3: \text{ Time for } R_3 = 5 + 7 = 12$$

See Fig. 5 for tasks timeline. So, the SSM adds the total waiting time to  $RT_3$  because  $tm_1 < tm_3$ :  
Time for  $R_3 = 8 + 7 + 5 = 20$ .

**Re-Calculation:**

So,  $AC = 4.58 + 8.17 + 13.33 = 26.08$   
If, the SSM ignores the delay time as follows:

$$tm_1 = 8, tm_2 = 3, RT_1: \text{ Time for } R_1 = 8 + 3 = 11$$

$$tm_3 = 10, tm_4 = 4, RT_2: \text{ Time for } R_2 = 10 + 4 = 14$$

$$tm_6 = 5, tm_5 = 7, RT_3: \text{ Time for } R_3 = 5 + 7 = 12$$

**Re-Calculation:**

Then,  $AC = 4.58 + 8.17 + 8.00 = 20.75$   
As a result,  $AC$  of  $EPT < AC$  of the actual running time.

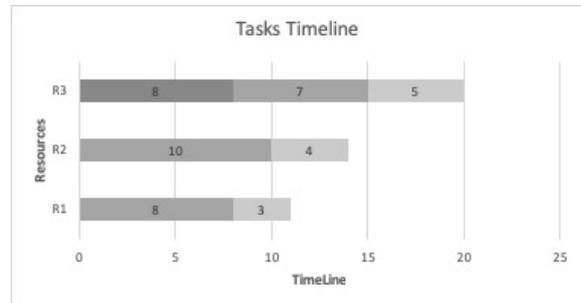


Fig. 5. Example 2. Scenario 2.2 Tasks Timeline

3) *Scenario: 2.3:* This scenario shows the time for each resource and the tasks allocated with Fast-Track as follows:

- $R_1 : t_1$  and  $tm_1 = 5, t_2$  and  $tm_2 = 5$
- $R_2 : t_3^{FT}$  and  $tm_3 = 5, t_4$  and  $tm_4 = 5$
- $R_3 : t_6$  and  $tm_6 = 5, t_5$  and  $tm_5 = 5$

The dependencies are:  $t_5$  depends on  $t_3$  and  $t_6$  depends on  $t_3$ . As a result of the dependencies, the SSM finds a delay of executing tasks  $t_6$  and  $t_5$ . So,  $AC$  calculated as follows:

$$tm_1 = 5, tm_2 = 5, RT_1: \text{ Time for } R_1 = 5+5 = 10$$

$$tm_3 = 5, tm_4 = 5, RT_2: \text{ Time for } R_2 = 5+5 = 10$$

$$tm_6 = 5, tm_5 = 5, RT_3: \text{ Time for } R_3 = 5+5 = 10$$

See Fig. 6 for tasks timeline. So, the SSM adds the total waiting time to  $RT_3$ . Therefore,  $RT_3 = 5+5+5 = 15$ .

**Re-Calculation:**

So,  $AC = 4.17 + 5.83 + 10.00 = 20$

If, the SSM ignores the delay time as follows:

$tm_1 = 5, tm_2 = 5, RT_1$ : Time for  $R_1 = 5+5 = 10$

$tm_3 = 5, tm_4 = 5, RT_2$ : Time for  $R_2 = 5+5 = 10$

$tm_6 = 5, tm_5 = 5, RT_3$ : Time for  $R_3 = 5+5 = 10$

**Re–Calculation:**

Then  $AC = 4.17 + 5.83 + 6.67 = 16.67$

As a result,  $AC$  of  $EPT < AC$  of the actual running time.

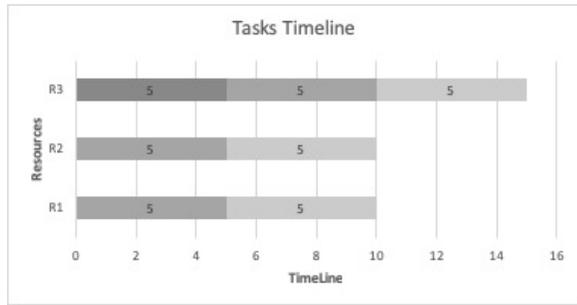


Fig. 6. Example 2. Scenario 2.3 Tasks Timeline

**C. Example: 3**

Table VII shows the details of a service request:

TABLE VII. SERVICE REQUIREMENT FOR EXAMPLE 3

Security Level(Weight)/Importance	1	2	3	Resource	$RC_k$
1 (0.00)					
2 (0.25)	$t_1$	$t_2$		$R_1$	20
3 (0.50)					
4 (0.75)	$t_3$	$t_4$		$R_2$	20
5 (1.00)	$t_5, t_7$	$t_6$		$R_3$	20

1) *Scenario: 3.1*: This scenario shows the time for each resource and the tasks allocated with Fast-Track as follows:

- $R_1 : t_1^{FT}$  and  $tm_1 = 8, t_2$  and  $tm_2 = 3$
- $R_2 : t_3^{FT}$  and  $tm_3 = 10, t_4$  and  $tm_4 = 4$
- $R_3 : t_6$  and  $tm_6 = 5, t_5$  and  $tm_5 = 7, t_7$  and  $tm_7 = 7$

The dependencies are:  $t_5$  depends on  $t_1$  and  $t_6$  depends on  $t_3$ . As a result of the dependencies, the SSM finds a delay of executing tasks  $t_6$  and  $t_5$ . So,  $AC$  calculated as follows:

$tm_1 = 8, tm_2 = 3, RT_1$ : Time for  $R_1 = 8+3 = 11$

$tm_3 = 10, tm_4 = 4, RT_2$ : Time for  $R_2 = 10+4 = 14$

$tm_6 = 5, tm_5 = 7, tm_7 = 7, RT_3$ : Time for  $R_3 = 5+7+7=19$

See Fig. 7 for tasks timeline. So, the SSM adds the delay time to  $RT_3$  because  $tm_1 < tm_3$ . Therefore,  $RT_3 = 8+7+5+7 = 27$ .

**Re–Calculation:**

So,  $AC = 4.58 + 8.17 + 18.00 = 30.75$

If, the SSM ignores the delay time as follows:

$tm_1 = 8, tm_2 = 3, RT_1$ : Time for  $R_1 = 8 + 3 = 11$

$tm_3 = 10, tm_4 = 4, RT_2$ : Time for  $R_2 = 10 + 4 = 14$

$tm_6 = 5, tm_5 = 7, tm_7 = 7, RT_3$ : Time for  $R_3 = 5+7+7= 19$ .

So,  $AC$  of  $EPT < AC$  of the actual running time. On other hand , if the SSM runs  $t_7$  first as follows:  $RT_3 = (8-7) + 7 + 5 + 7= 20$

**Re–Calculation:**

So,  $AC = 4.58 + 8.17 + 13.33 = 26.08$

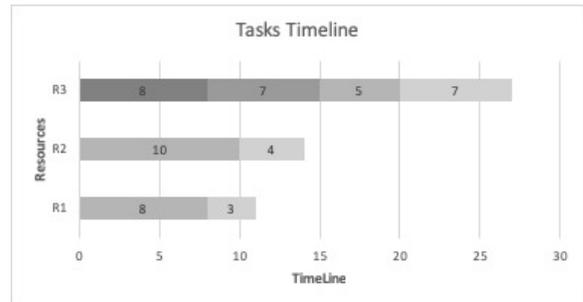


Fig. 7. Example 3. Scenario 3.1. Tasks Timeline

2) *Scenario: 3.2*: This scenario shows the time for each resource and the tasks allocated with Fast-Track as follows:

- $R_1 : t_1^{FT}$  and  $tm_1 = 8, t_2$  and  $tm_2 = 3$
- $R_2 : t_3^{FT}$  and  $tm_3 = 10, t_4$  and  $tm_4 = 4$
- $R_3 : t_6$  and  $tm_6 = 5, t_5$  and  $tm_5 = 7, t_7$  and  $tm_7 = 12$

The dependencies are:  $t_5$  depends on  $t_1$  and  $t_6$  depends on  $t_3$ . As a result of dependencies, the SSM finds a delay of executing tasks  $t_6$  and  $t_5$ . So,  $AC$  calculated as calculated follows:

$tm_1 = 8, tm_2 = 3, RT_1$ : Time for  $R_1 = 8+3= 11$

$tm_3 = 10, tm_4 = 4, RT_2$ : Time for  $R_2 = 10+4= 14$

$tm_6 = 5, tm_5 = 7, tm_7 = 12, RT_3$ : Time for  $R_3 = 5+7+12 = 24$

See Fig. 8 for tasks timeline. So, the SSM adds the delay time to  $RT_3$ , because  $tm_1 < tm_3$ . Therefore,  $RT_3 = 8+7+5+12 = 32$ .

**Re–Calculation:**

$AC = 4.58 + 8.17 + 21.33 = 34.08$

If, the SSM ignores the delay time as follows:

$tm_1 = 8, tm_2 = 3, RT_1$  Time for  $R_1 = 8 + 3 = 11$

$tm_3 = 10, tm_4 = 4, RT_2$ : Time for  $R_2 = 10 + 4 = 14$

$tm_6 = 5, tm_5 = 7, tm_7 = 12, RT_3$ : Time for  $R_3 = 5+7+12= 24$

On other hand, if the SSM runs  $t_7$  first as follows:

**Re–Calculation:**

So,  $AC = 4.58 + 8.17 + 16.00 = 28.75$

**D. Example: 4**

Table VIII shows the details of a service request:

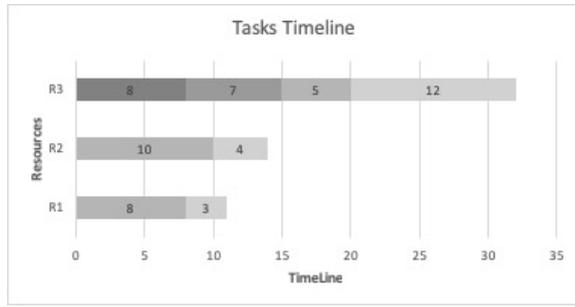


Fig. 8. Example 3. Scenario 3.2 Tasks Timeline

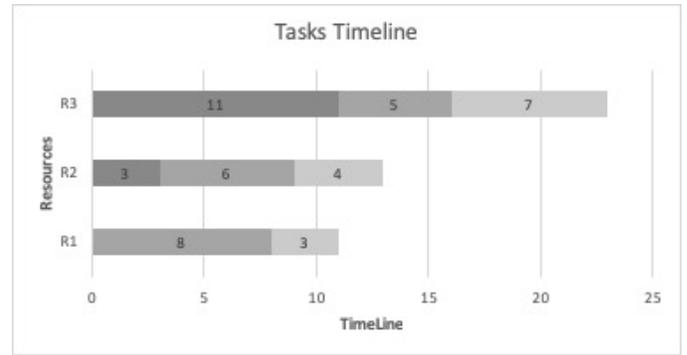


Fig. 9. Example 4. Scenario 4.1 Tasks Timeline

TABLE VIII. SERVICE REQUIREMENT FOR EXAMPLE 4

Security Level(Weight)/Importance	1	2	3	Resource	$RC_k$
1 (0.00)					
2 (0.25)	$t_1$	$t_2$		$R_1$	20
3 (0.50)					
4 (0.75)	$t_3$	$t_4$		$R_2$	20
5 (1.00)	$t_5$	$t_6$		$R_3$	20

1) *Scenario: 4.1:* This scenario shows the time for each resource and the tasks allocated with Fast-Track. The dependencies as follows:

- $t_6$  depends on  $t_1$
- $t_5$  depends on  $t_3$
- $t_3$  depends on  $t_2$

The resources allocated as follows:

- $R_1 : t_1^{FT}$  and  $tm_1 = 8$ ,  $t_2^{FT}$  and  $tm_2 = 3$
- $R_2 : t_3^{FT}$  and  $tm_3 = 6$ ,  $t_4$  and  $tm_4 = 4$
- $R_3 : t_6$  and  $tm_6 = 5$ ,  $t_5$  and  $tm_5 = 7$

See Fig. 9 for tasks timeline with tasks orders. So, the SSM finds a delay of executing tasks. As a results, the SSM adds  $tm_2$  to  $RT_2$ , then the SSM adds  $RT_1$  to  $RT_3$  to avoid the delays of running this service:

- $R_1 : t_1^{FT}$  and  $tm_1 = 8$ ,  $t_2^{FT}$  and  $tm_2 = 3$ ,  $RT_1 = 8+3 = 11$
- $R_2 : t_3^{FT}$  and  $tm_3 = 6$ ,  $t_4$  and  $tm_4 = 4$ ,  $RT_2 = 3+6+4 = 13$
- $R_3 : t_6$  and  $tm_6 = 5$ ,  $t_5$  and  $tm_5 = 7$ ,  $RT_3 = 11+5+7 = 23$

#### IV. COMPARISON AGAINST CURRENT MODELS

This section provide a comparison of the SSM with a current work.

The aim of Tripathy and Patra [14] approach is to execute the high priority tasks first over the allocated resources  $R_K$  needed. On the other hand, the SSM justified the priority just to order the tasks with five levels.

Tripathy and Patra approach allows a single task to run over a number of resources  $R_k$  at the same time. The Task set  $job_i(j,k,l)$  identified as the following:

- $i$  serves as tasks id
- $j$  identifies the number of resources required
- $k$  serves as task time
- $l$  to set task priority

For example, for a service request:

- $t_1:(N = 2, tm_1 = 5, pl_1 = 1)$
- $t_2:(N = 6, tm_2 = 10, pl_2 = 5)$
- $t_3:(N = 2, tm_3 = 5, pl_3 = 4)$
- $t_4:(N = 2, tm_4 = 5, pl_4 = 2)$

See Fig. 10 for tasks timeline and allocated resources for each task.

The SSM analyses this service request as follows:

- Set  $q = 0$ .
- $hw_j = 0.00$  for all task  $st_i$ , therefore,  $Rw_i = 0.00$  for all  $R_k$
- All  $RT_k$  for  $R_k = 15$ .
- $AC$  will be calculated from equation No. 6.  
 $R_k \text{ Cost} = (15*20)/60$   
 $AC = 5 * 6 = 30$

#### V. DISCUSSION

This section discusses previous examples and the comparison against current models.

In all examples, the SSM has applied the calculating steps produced by [9]. The calculating steps help to identify if the service includes tasks dependencies or not. Because of the dependencies can make the scheduling and the executing process very complicated.

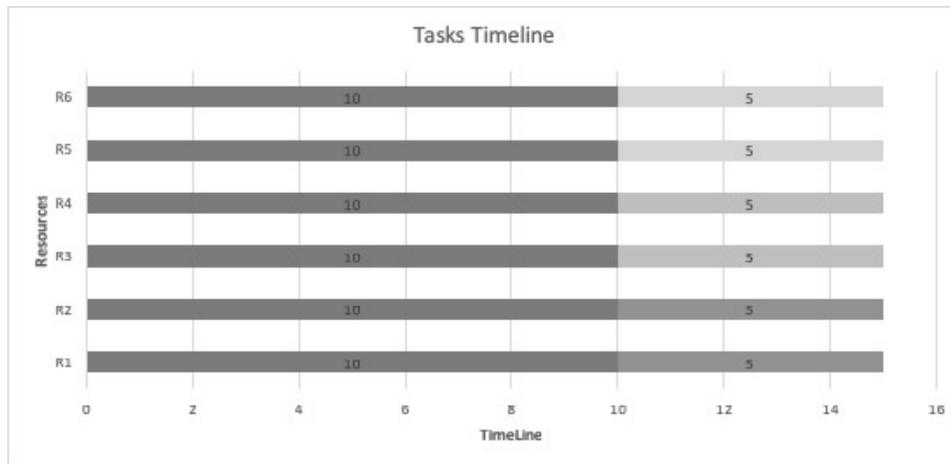


Fig. 10. Resources and Tasks Timeline

Each examples of different possible scenarios present different calculated service *AC* that can fulfil the customer expectations of service security, time, cost, and QoS.

serves these five features. However, there might be a slight delay of service time when applying the SSM on the same example included in Tripathy and Patra [14] approach.

TABLE IX. SSM FEATURES OVER ALL EXAMPLES

Feature/Examples	1	2	3	4	5
No. of Tasks (single task or set of tasks)	2	6,7	7	6	4
Dependencies	NA	√	√	√	NA
Security	√	√	√	√	basic
QoS q	0	0	0	0	0
No. of Task Importance/Priority levels	5	5	5	5	5,3
<i>AC</i> vs Initial Cost	less	less	less	less	less
<i>AC</i> of <i>EPT</i> vs <i>AC</i> of Running Time	equal	less	less	less	less

Table IX shows the differences in the services for all examples. The differences include the Number of tasks for each examples, Security, QoS, Number of Task Importance levels, *AC* against Initial Cost, and *AC* of *EPT* against *AC* of Running Time.

For all examples,  $AC < \text{the Initial cost}$ . However, it is possible to have a scenario that finds there is no big different of *AC* and the Initial cost.

Also, The number of tasks running in the service are different as the SSM has defined that a task can be a single task or set of tasks.

The SSM has more features presented in Fig. 11 that can be applied to the services than the current models. These Features are Security, Priority, QoS, Cost, and Time. In addition, these features can make the cloud services very flexible to meet customers requirements. For example, Security defined by the SSM to have five different levels, also there are a range of QoS levels to ensure that the service will be delivered in the required level. So, the comparison of the SSM with Tripathy and Patra [14] showed that the SSM can

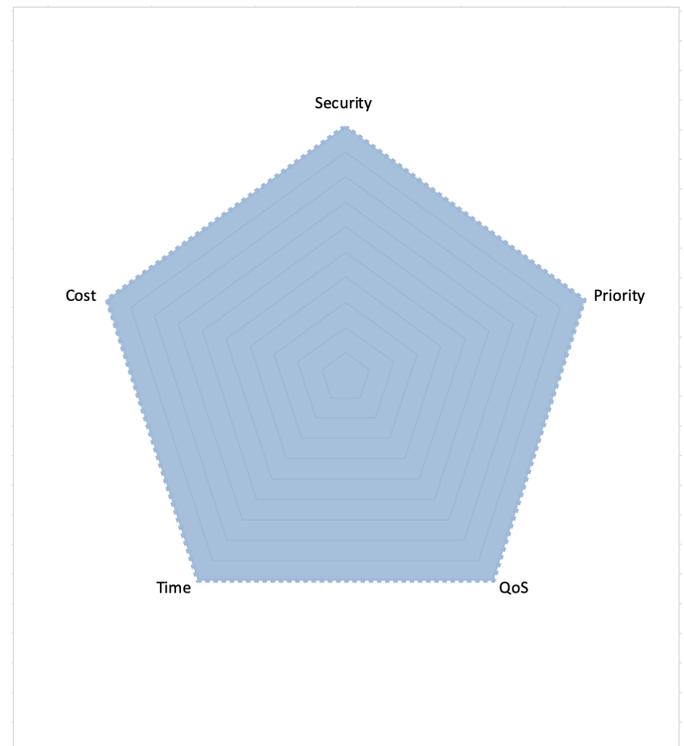


Fig. 11. SSM Features [13]

## VI. FUTURE WORK

This paper outcomes of the cost and effect suggested more investigating of the SSM. Also, further investigation required on QoS levels introduced by [15] that can be applied and considered for a service.

## VII. CONCLUSIONS

In this paper, the SSM has been examined through various worked examples. The Re-calculating step showed a major effect on the service cost  $AC$ . Furthermore, it discussed the SSM features over the examples and it showed the effect of the dependencies on the service scheduling and the service time. Finally, SSM works with more benefits than current models as it serves more features that can be applied to the services such as security, priority, QoS, Cost, and Time.

## ACKNOWLEDGMENT

Thanks to Taif University in Kingdom of Saudi Arabia for the funding.

## REFERENCES

- [1] P. Mell and T. Grance, *The NIST Definition of Cloud Computing*, Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology Gaithersburg, 2011.
- [2] L. Qiang, H. Qinfen, X. Limin and L. Zhoujun, *Adaptive management of virtualized resources in cloud computing using feedback control*, First International Conference on Information Science and Engineering, IEEE, pp. 99–102, 2009.
- [3] Y. Hu, HCH. Chen, PPC. Lee and Y. Tang, *NCCloud: applying network coding for the storage repair in a cloud-of-clouds*, FAST, pp.21, 2012.
- [4] V. Chang, D. Bacigalupo, G. Wills and D. De Roure, *A categorisation of cloud computing business models*, FAST, pp. 509–512, 2010.
- [5] D. Goutam, A. Verma, N. Agrawal, *The performance evaluation of proactive fault tolerant scheme over cloud using CloudSim simulator*, International Conference on the Applications of Digital Information and Web Technologies (ICADIWT), IEEE, 171–176, 2014.
- [6] T. Dillon, C. Wu, and E. Chang, *Cloud computing: issues and challenges*, *International Conference on Advanced Information Networking and Applications*, IEEE, 27–33, 2010.
- [7] S. Subashini and V. Kavitha, *A survey on security issues in service delivery models of cloud computing*, *Journal of network and computer applications*, Elsevier, 34, 1, 1–11, 2011.
- [8] A. Sheikh, M. Munro and D. Budgen, *Systematic Literature Review (SLR) of Resource Scheduling and Security in Cloud Computing*, *International Journal of Advanced Computer Science and Applications (IJCSA)*, IEEE, Vol. 10, No. 4, pp. 35–44, 2019.
- [9] A. Sheikh, M. Munro and D. Budgen, *Cost and Effect of Using Scheduling Security Model in a Cloud Environment*, *International Conference on Cyber Security and Cloud Computing and Edge Computing and Scalable Cloud (CSCloud/EdgeCom)*, IEEE, pp. 95–101, 2019.
- [10] P. Watson, *A multi-level security model for partitioning workflows over federated clouds*, *Journal of Cloud Computing*, Springer, 1, 1, pp 1-15, 2012.
- [11] A. Sheikh, M. Munro and D. Budgen, *SSM: Scheduling Security Model for a Cloud Environment*, *International Conference on Cloud and Big Data Computing (ICCBDC 2018)*, ACM, pp. 11–15, 2018.
- [12] A. Sheikh, M. Munro and D. Budgen, *Evaluating a Cloud Service using Scheduling Security Model (SSM)*, *International Journal of Advanced Computer Science and Applications (IJCSA)*, IEEE, Vol. 10, No. 10, pp. 525–532, 2019.
- [13] A. Sheikh, *Scheduling Security Model for a Cloud Environment*, PhD thesis, Durham University, 2020.
- [14] L. Tripathy and R.R. Patra, *Scheduling in cloud computing*, *International Journal on Cloud Computing: Services and Architecture (IJCCSA)*, 4, 5, pp.21-7, 2014.
- [15] H. Ramadan, and D. Kashyap, *Quality of Service (QoS) in Cloud Computing*, *International Journal of Computer Science and Information Technologies (IJCSIT)*, vol. 8, No. 3, pp. 318–320, 2017.

# Robust Speed Control for Networked DC Motor System

Sokliep Pheng<sup>1\*</sup>

School of Information and Communication Engineering  
Guilin University of Electronic Technology, Guilin, China

Luo Xiaonan<sup>2</sup>

Guangxi Key Laboratory of Intelligent Processing of  
Computer Image and Graphics  
School of Computer Science and Information Security  
Guilin University of Electronic Technology, Guilin, China

Rachana Lav<sup>3</sup>

School of Energy Science and Engineering  
University of Electronic Science and Technology  
Chengdu, China

Zhongshuai Wang<sup>4</sup>, Zetao Jiang<sup>5</sup>

School of Computer Science and Information Security  
Guilin University of Electronic Technology  
Guilin, China

**Abstract**—In this paper, we used observer-based  $H_\infty$  output feedback control problems for the communication from the controller to dc motor and considered data packet dropout characterized by the Bernoulli random binary distribution the disturbance. The uncertain parameters have also been considered in the network-based dc motor system. Firstly, we used the robust  $H_\infty$  output feedback control strategy to optimize controller gain and observer gain to guarantee the mean square stability. The observer-based  $H_\infty$  output feedback has been designed to achieve robust speed control in the mean-square sense and optimize the parameters of the control system while guaranteeing the robust  $H_\infty$  output feedback performance. Then, when data is transmitted in the control system, we illustrated that the system is stable and robust speed control can be achieved as well as the result realized.

**Keywords**—The Networked DC Motor System; Observer Design; Robust Speed Control; Data Dropout; LMI

## I. INTRODUCTION

In recent years, robust speed control of NCS is the most population has been widely utilized in many industrial applications, and has very interesting in the industrial electronics community for control applications. Compared to traditional dc motor control, we found that the system component is located in the same place and connected by point-to-point wiring. Although, many requirements, dc motor, and controller are so hard to be located in the same place, and thus, signals are required to be transmitted from one place to another place. For improved networked control dc motor system technologies to reduce the cost of installation, easy maintenance [1], there is steady state error to integrate communication networked apply to the designed controller in the system. The systems are the type of NCS in the control system. Recently, NCSs is still popular and enticed much attention from research teams. To guarantee the obtaining good results such as networked control system, packet dropout, observer-design  $H_\infty$  output feedback controller, and robust speed control have investigated with introduced in the literature [2].

The observer-designed  $H_\infty$  output feedback control strategy is performed in the controller designed to corresponding robust speed control and determined controller gain and observer gain in the system. The data packets have transmitted through the network to the plant and there are some packet dropouts during networked transmissions. In this paper, the robust speed control proposed with observer designed and random data dropout have been considered and model to issue that in NCS is the packet losses phenomenon have solved in form mean-square sense. In addition, the previous work that studied about Markovian-jumping to find parameter was utilized to model a discrete-time as a linear system with data packet losses in NCSs [3] other literature research about random packet loss, observer-design  $H_\infty$  feedback controller, robustly exponentially stabilize, prescribed disturbance rejection level in reference [4].

## II. RELATED WORK

According to our research literature, we have focused and studied on two main points.

1) *Observer designed to distribute*: Generally, designed observer-based  $H_\infty$  output feedback control strategy to optimize parameters performance, estimate state error, control the method utilized to achieve the exponentially stable in mean square sense for the network-based dc motor control system. This proposed is to archive robustly in the controller designed.

2) *Robust speed control*: After we got the observabilities, we also need robust speed control based on the exponentially stabilities mean square and robust  $H_\infty$  feedback control strategy simultaneously for the NCS system. Therefore, our research purpose is to combine observabilities and robust  $H_\infty$  feedback control strategy together which can be used for potential applications into the NCS system.

For controller designed in this paper we provide a scalar value  $\gamma > 0$ . For the system in:

$$\delta_{k+1} = \bar{A}\delta_k + \epsilon_k \tilde{A}\delta_k + \bar{B}\omega_k \quad (1)$$

\*Corresponding Author

is exponentially stabilities mean square and we apply  $H_\infty$  norm bound constraints [5-9]:

$$\sum_{k=0}^{\infty} E\{\|z_k\|^2\} < \gamma^2 \sum_{k=0}^{\infty} E\{\|\omega_k\|^2\} \quad (2)$$

is obtained for all of nonzero  $\omega_k$ , if only if that there is the existence of positive definite in matrices  $W_{11} \in R^{m \times m}$ ,  $W_{22} \in R^{(n-m) \times (n-m)}$ , and  $W_2 \in R^{n \times p}$ , and real matrices  $Z \in R^{m \times n}$  and  $N \in R^{n \times p}$ , such as we proved in the systems (3) are represented [10]:

$$\begin{bmatrix} -W_1 & * & * & * & * & * & * & * & * & * & * \\ 0 & -W_2 & * & * & * & * & * & * & * & * & * \\ 0 & 0 & -\gamma^2 I & * & * & * & * & * & * & * & * \\ W_1 A - \bar{\beta} B Z & \bar{\beta} B Z & W_1 B_1 & -W_1 & * & * & * & * & * & * & * \\ 0 & W_2 A - Y C_1 & W_2 B_1 - Y D_1 & 0 & -W_2 & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & -I & * & * & * & * & * \\ \beta B Z & \beta B Z & 0 & 0 & 0 & 0 & -W_1 & * & * & * & * \\ Y C_1 & 0 & 0 & 0 & 0 & 0 & 0 & -W_2 & * & * & * \\ 0 & 0 & 0 & H_1^T W_1 & H_2^T W_2 & 0 & 0 & 0 & -\epsilon I & * & * \\ \epsilon E & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -d \end{bmatrix} < 0 \quad (3)$$

$$W_1 := U_1^T W_{11} U_1 + U_2^T W_{22} U_2$$

$U_1$ , and  $U_2$  it takes from the system (5). Moreover, to determine for the parameters are provided by the following system [11]:

$$K = V \Sigma^{-1} W_{11}^{-1} \Sigma V^T Z, \quad L = W_2^{-1} Y \quad (4)$$

*Proofs:* Since there is the existing system  $W_{11} > 0$  and  $W_{22} > 0$ , such as:

$$W_1 = U_1^T W_{11} U_1 + U_2^T W_{22} U_2$$

Where

$U_1$  and  $U_2$  are defined in the system (2.20) below:

$$\tilde{B} = U B_2 V = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} B_2 V = \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} \quad (5)$$

The following from the Lemma 1, and there are the exists of a nonsingular matrices

$$W \in R^{m \times m}, \text{ such as } B_2 W = W_1 B_2$$

Now, let we computed such as a matrix  $W$  that is related  $B_2 W = W_1 B_2$  is following [12]:

$$W_1 U^T \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T = U^T \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T W \quad (6)$$

And

$$U^T \begin{bmatrix} W_{11} & 0 \\ 0 & W_{22} \end{bmatrix} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T = U^T \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T W \quad (7)$$

Which we are implied as:

$$W = (V^T)^{-1} \Sigma^{-1} W_{11} \Sigma V^T \quad (8)$$

For anyway, we can be integrated from (4) and (8) is that:

$$B_2 W = W_1 B_2, \quad Z = W K, \quad Y = W_2 L \quad (9)$$

Furthermore, it is so hard to see that the system (3) is that equivalent to the equation (10) below [13-15]:

$$\begin{bmatrix} -W_1 & * & * & * & * & * & * & * & * & * & * \\ 0 & -W_2 & * & * & * & * & * & * & * & * & * \\ 0 & 0 & -\gamma^2 I & * & * & * & * & * & * & * & * \\ W_1 A - \bar{\beta} W_1 B_2 K & \bar{\beta} W_1 B_2 K & W_1 B_1 & -W_1 & * & * & * & * & * & * & * \\ 0 & W_2 A - W_2 L C_1 & W_2 B_1 - W_2 L & 0 & -W_2 & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & -I & * & * & * & * & * \\ \beta W_1 B_2 K & \beta W_1 B_2 K & 0 & 0 & 0 & 0 & -W_1 & * & * & * & * \\ W_2 L C_1 & 0 & 0 & 0 & 0 & 0 & 0 & -W_2 & * & * & * \\ 0 & 0 & 0 & H_1^T W_1 & H_2^T W_2 & 0 & 0 & 0 & -\epsilon I & * & * \\ \epsilon E & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -d \end{bmatrix} < 0 \quad (10)$$

We considered the robust of  $H_\infty$  control problems are computed for the type of NCS system with random distribution data dropout in the communication. In the introducing of the random packet losses and observer-design  $H_\infty$  feedback control strategy to obtained robust speed control, exponentially stabilities mean square and we also obtained a prescribed  $H_\infty$  performant for disturbance rejection level in control system. It is introduced that the controller problems are designed under considerations is computable if only if the LMI of the system (3) is able implied.

By the production, it takes out that the optimization problems have formulated as following in the system, the  $H_\infty$  performant control strategy:

$$\min_{W_{11} > 0, W_{22} > 0, Z, Y} \gamma \quad \text{Subject into the system (3)} \quad (11)$$

Finally, after all of our proof that we are implemented the  $H_\infty$  performance is obtained the control parameters in the control system.

### III. EXPERIMENTS AND SIMULATION

#### A. Test Simulation Results of the Stabilities

This part of the simulation setup has considered on the network-based dc motor system, which the parameters are identifies from iNetCon-PC104 system. The simulation results of the network-based dc motor system under observer design to distributed into the system. The controller designed the system of closed-loop under observer based on  $H_\infty$  output feedback control problems to optimize the system parameters. The  $H_\infty$  performant control problems, which have been solved by LMI, by obtaining the strong of the robustness in the represented of the disturbance input for the closed-loop system. When the control signal has been transmitted, the packet losses rate is being considered in the NCS system and analysis the results are introduced of our method in this part. However, in the order of our studying for the random data

dropout, observer design to distributed based on the  $H_\infty$  output feedback control problems in the system to get the exponentially stabilities mean square and we can also obtain the prescribed of the  $H_\infty$  disturbance rejection level. In addition, the system parameters of the network-based dc motor system are demonstrated.

The iNetCon-PC104 systems of the parameters are identified to get the brushless dc motor parameters in a transfer function form and then we obtained the system matrices as below:

$$G(z^{-1}) = \frac{0 \quad 0.000859116877110923z^{-1} \quad -0.00171576967292345z^{-2} \quad 0.000859861408129589z^{-3}}{1 \quad -3.91599031549643z^{-1} \quad 5.75462836130592z^{-2} \quad -3.76106608588523z^{-3} \quad 0.922431267338640z^{-4}}$$

So it can be converted in the system matrices are obtained

$$A = \begin{bmatrix} 3.9160 & -5.7546 & 3.7611 & -0.9224 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \text{ and}$$

$$C_1 = [0 \quad 0.0009 \quad -0.0017 \quad 0.0009]$$

$$K = [4.6060 \quad -5.5768 \quad 3.7814 \quad -0.9215]$$

$$L = [-511860 \quad -528080 \quad -471780 \quad -357940]^T$$

In the simulation, the networked dc motor system we considered which described by matrices as discrete-time modelling for the sampling period  $t = 0.1$  s and then we designed [40].

$$H_1 = [0.1 \quad 0 \quad 0 \quad 0]^T, \quad E = [1 \quad 0 \quad 0 \quad 0], \quad B_1 = [0.5 \quad 0 \quad 0.2 \quad 0]^T$$

The disturbance input signal of this system is assumed that to be  $\omega = 1/k$ , and the time-varying of norm bound for the uncertain parameters  $F(k)$  is supposed as following:

$$F(k) = \begin{bmatrix} rand & 0 & 0 & 0 \\ 0 & rand & 0 & 0 \\ 0 & 0 & rand & 0 \\ 0 & 0 & 0 & rand \end{bmatrix}$$

In Fig. 1 as shown above, we used the observer designed to distribution in the network-based dc motor system for operation result and deal with random data dropout for the signals that it transmitted into the networked. In addition, we want to provide the estimate of the internal state output feedback of a given real system, and we have known that the state is necessary to realize in LMI method by using MATLAB toolbox. The observer designed to distribution based  $H_\infty$  feedback control strategy, and we can obtained the controller gain, observer gain and prescribed the  $H_\infty$

performant control strategy to achieve as in system (11). The structure designed in Fig. 1 as in the form for the system of closed-loop in the exponentially stabilities means square for the network-based dc motor system operation.

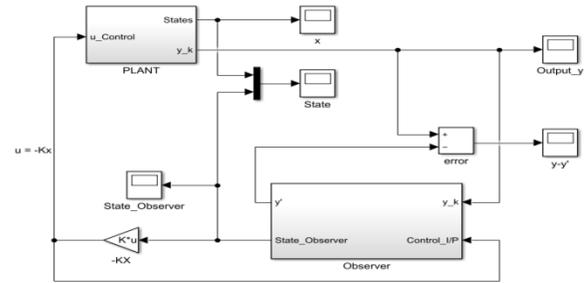


Fig. 1. Observer Design to Distribution.

### B. Packet Losses Rate

In this addition, we have changed three value of  $\bar{\beta}$  to performant in the simulation are given to demonstrate effectiveness for our studying method. We used to introduce the application for our method under observer designed to distribution based  $H_\infty$  performant control strategy deals with random data dropout. In the simulation, results have represented to describe the possibilities and stabilities by using performant control strategy. In addition, we make the comparison for the control performant result with difference data dropout value, and we have been chosen by the difference data dropout value such as  $\bar{\beta} = 0.55, 0.95$ . Because of this applying of the following value in the simulation results shown by the figure as state response, control output, state estimate error and the packet losses respectively.

*Step 1:* We designed the  $H_\infty$  performant control strategy with random for data dropout  $\bar{\beta} = 0.55$ , and the  $H_\infty$  performant is minimized as the index. Therefore, we greatly to an agreement with minimization of gamma problems. Solving of the  $H_\infty$  optimization control strategy (11) by using MATLAB of the LMI toolbox, and we obtain the minimization of gamma  $\gamma_{\min} = 0.8569$ .

The initial condition of the uncertainty linear system of NCS is  $x = [2 \quad 0 \quad 0 \quad 0]$  and the observer initial condition is  $\hat{x} = [0 \quad 0 \quad 0 \quad 0]^T$ , and the disturbance input we assumed that  $\omega = 1/k$ , the simulation results have shown below respectively.

*Step 2:* For the last step, we designed the  $H_\infty$  performant control problems with appropriated random data dropout  $\bar{\beta} = 0.95$ , and we are again considered with a suitable the  $H_\infty$  performant  $\gamma > 0$  is minimized as the index. Solving of the  $H_\infty$  optimization control problems (11) by using MATLAB of the LMI toolbox and we obtain a new of the minimization of gamma  $\gamma_{\min} = 0.4428$ .

For the comparison of the simulation results as shown in above figures we can also see that when the larger value of

$\bar{\beta} = 0.95$  we got probabilities distribution of the data dropout is smaller and the network-based dc motor system performant is faster become to zero which means faster to be stable. In the order of this system to analysis for the simulation results, we choose the difference value of  $\bar{\beta}$  to make a comparison in the system. The  $\bar{\beta}$  represent is the probabilities of data dropout for the NCS system is converged to zero and the observer gain and controller gain are realized by applying the LMI approach in MATLAB toolbox.

According to the numerical of simulation results for the network-based dc motor system, we can see that in these terms, the performant  $\gamma$  of observer design to distribution based on  $H_\infty$  output feedback control problems is close to probabilities of the related for NCS system data dropout. We desired to achieve a minimum of that is solved in the LMI toolbox system by optimization of the  $H_\infty$  performant control problem system in (11). So by using  $\bar{\beta} = 0.95$  we can conclude of the results that the probabilities of the NCS system data dropout are lower than other choose two value of  $\bar{\beta} = 0.55$ , and the dc motor of NCS system performant is obtain in better condition. The analysis for the simulation results of this work is represented from Fig. 2 to Fig. 7, respectively.

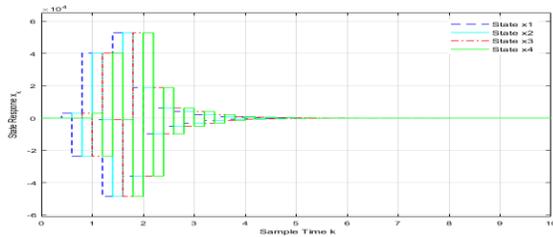


Fig. 2. Simulation of State Response  $x_k$  with  $\bar{\beta} = 0.55$

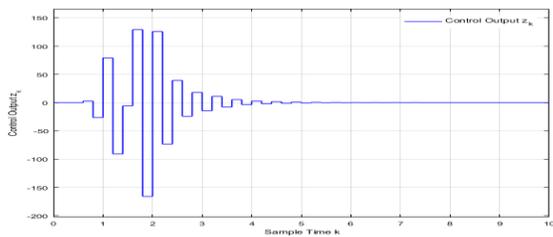


Fig. 3. Simulation of Control Output  $z_k$  with  $\bar{\beta} = 0.55$

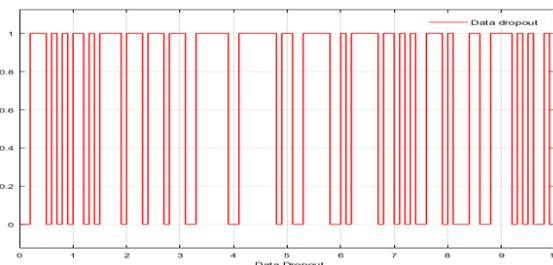


Fig. 4. Simulation of Data Dropout with  $\bar{\beta} = 0.55$

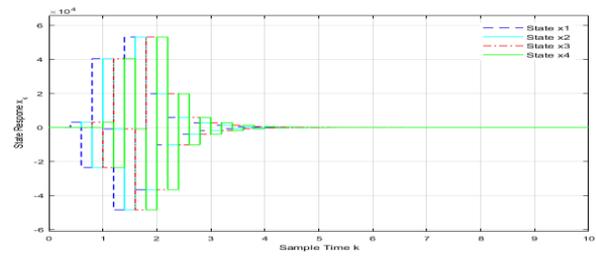


Fig. 5. Simulation of State Response  $x_k$  with  $\bar{\beta} = 0.95$

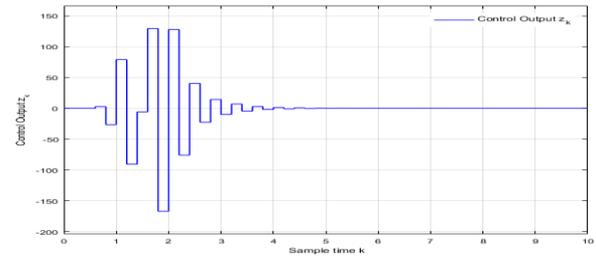


Fig. 6. Simulation of Control Output  $z_k$  with  $\bar{\beta} = 0.95$

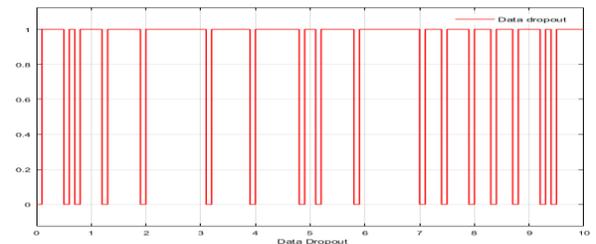


Fig. 7. Simulation of Data Dropout with  $\bar{\beta} = 0.95$

### C. Comparison Result with Previous Work

The NCS system is very interesting, become more investigated and widely used in such an application, and there are many studies in the past. For development technology and improve previous results, we choose this topic to study and compare it to the previous work. The network-based dc motor system, designed observer-based  $H_\infty$  performant control strategy deal with random data dropout is realized by the system of closed-loop is exponentially stabilities in the mean square sense that we can see in Fig. 8 become too stable faster than previous research are shown in Reference [11]. The simulation result has shown in Fig. 2 and Fig. 5 achieved by choosing three different values to make a comparison in this work. As above results, the value of  $\bar{\beta} = 0.95$  it became too stable faster than the other two values. The simulation results which have been given to introduce the state response, the state observer, data dropout and the control output  $z(k)$  with disturbance rejection level are obtained in the system of closed-loop is exponentially stabilities in mean square sense are illustrated in Fig. 2 and Fig. 5. It is facile to show that when we considered of difference value the probabilities of the data dropout, the system of closed-loop become stable under we designed observer-based  $H_\infty$  performant control strategy. When the packet losses rate is lower, it means that

the network-based dc motor system has become faster to stable. It shows that the designed observer is the reconstruction state can be tracked the original state, and the state in the system is very good. However, it still has some lack off in our proposed method. The advantage of this work, the state response and the observer state are faster to stable and the estimate error state is faster go to zero. In Fig. 8, as shown for the effectiveness of our studying for a new design observer to distribute based  $H_\infty$  performance strategy [11].

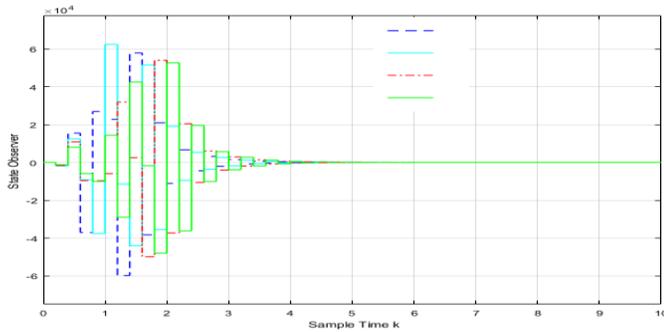


Fig. 8. Simulation of State Observer  $\hat{x}$  with  $\bar{\beta} = 0.95$

According to Fig. 8, the simulation results which have been demonstrated of the effectiveness of our proposed method by using a new designed observer-based  $H_\infty$  output feedback control strategy that the  $H_\infty$  performant have realized by MATLAB of the LMI toolbox. After we solved the LMI we obtained the controller gain and the observer gain to implement in the system of closed-loop is exponentially stabilities of the mean square. Moreover, we have analysed our proposed method of a new designed observer-based  $H_\infty$  performant control strategy by choosing the value of probabilities for random data dropout with  $\bar{\beta} = 0.95$ . The initial condition, for uncertainly linear system of the NCS, is assumed as  $x = [-4 \ 4 \ 0 \ 0]$  and the observer initial condition is also assumed as  $\hat{x} = [0 \ 0 \ 0 \ 0]^T$ , and the disturbance input we assumed that  $\omega = 1/k$ , for the simulation results that we have shown above respectively. The simulation results have become stable on 5s which mean that the networked dc motor system is faster than the previous research are shown in [11]. For the previous works are stable on the 20s. Consequently, our results in above mentioned, it concluded that the observer-based robust  $H_\infty$  feedback control strategy deal with random data dropout in the NCS system of closed-loop is robustly and exponentially stabilities in mean square sense to apply in the network-based dc motor system.

#### IV. RESULTS AND DISCUSSIONS

##### A. Simulation Result of Robust Speed Control

After we got the stabilities of the system, robust speed control has been considered to apply for the network-based dc motor system. It is demonstrated the effectiveness of a new observer design based on  $H_\infty$  performant control problems deal with random data dropout. We note that the network-based dc

motor system has been designed to drive for the dc motor to prescribe the robust speed control. However, the observer designed is to optimize for the parameters and eliminated the steady-state tracking error, and the random data dropout is analysed about the data transmitted through the NCS system performant applied for dc motor. The specified of  $H_\infty$  norm bound constraints in the system (3) have been solved in MATLAB of LMI toolbox.

The simulation by using iNetCon-PC104 system parameters with brushless dc motor, we are strongly focused about steady state speed tracking error under of our observer designed to distribute in the system that we have studied. The new system parameters of brushless dc motor that we identified from iNetCon-PC104 system to implement for the simulation results. The reference speed input we set 2000 rad/s in the simulation to illustrate the robust performance in the controller design. The effectiveness of our studied method by a new observer design to distribution based on  $H_\infty$  performant control problems deal with random data dropout we have obtained the robust speed in the system performant. Now let we consider the terms to drive the network-based dc motor system to set the reference speed input 2000 rad/s, we noted that under our proposed method is illuminated steady state speed error and the estimated state error and robust performance. To simulation of the network-based dc motor system, first we have considered on the system of closed-loop is exponentially stabilities mean square under using observer designed to distribution. The simulation results have done by implementing our method as well as shown:

In Fig. 9, the simulation result of brushless dc motor by communicated through networked with the iNetCon-PC104 system is robustly under using observer designed to distribute deal random data dropout into the system, the method want to achieve robust performance and stability. In the simulation result, we can succeed to control steady state speed tracking error and robustness of a network-based dc motor system under our proposed method, the robust speed performance is desirable with reference speed input 2000 rad/s. However, the modelling of the network-based dc motor system, first, we have been proofed is exponentially stabilities mean square with uncertainty parameters and disturbance under new observer design to a distribution deal with data dropout. Second, the robust performance of networked dc motor system obtained after the above completed.

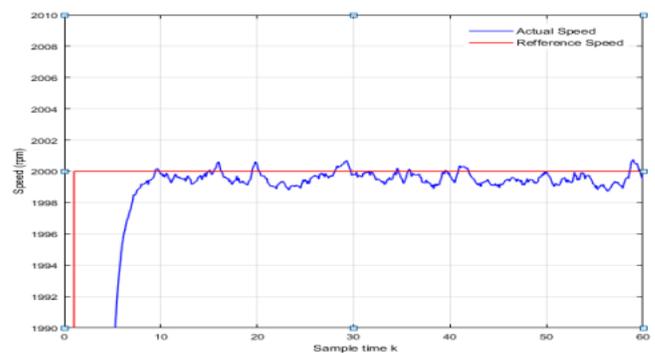


Fig. 9. Simulation Result of Robust Speed Control with Reference Speed 2000 Rpm.

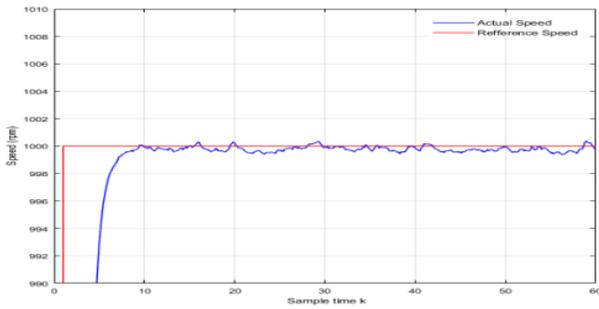


Fig. 10. Simulation Result of Robust Speed Control with Reference Speed 1000 Rpm.

After we settled reference speed 2000 rad/s and then we want to make simulation again with reference speed 1000 rad/s to mention how is robust speed of the networked dc motor system by using observer design based  $H_\infty$  output feedback controller. This simulation analysis is to demonstrate the robustness of the networked dc motor systems between the higher speed and the lower speed. The simulation result with reference speed 1000 rad/s have been done and evaluated in Fig. 10 is illustrated as following:

According to Fig. 10, the simulation result of the networked dc motor system demonstrated the speed error is eliminated and achieved the robust performance in the presence of bound modelling error. By using observer design based  $H_\infty$  output feedback controller, the actual speed of networked dc motor system achieved robustness in the controller design. The simulation result can compare between the higher speed and lower speed so the reference speed is lower and the speed error of networked dc motor system is smaller too. Fig. 9 and Fig. 10 to demonstrated the robustness of the networked dc motor system in our proposed method.

**B. Experiment Result of Robust Speed Control**

In this part of work, we are mainly analysed of the characteristic of the experiment result of robust speed control of an NCS system. In the experimental, we used the iNetCon-PC104 system to communicate from the controller through the network and the objective is brushless dc motor. For the above part of work, we have done simulation successfully for the designed observer to distribution based on  $H_\infty$  output feedback control problem. The experimental result of speed control for dc motor is robustly under our proposed method. Furthermore, we have experimented for the realization of brushless dc motor by the iNetCon-PC104 system to communicate through networked. The speed is robust which desirable under our proposed method.

For this part is very important to detail in the experiment which is the main part for our studied. It can be illustrated the speed that is robustly in the experimental result as well as figure:

In Fig. 11 above, we settled the reference speed input 2000, we saw that the speed was changed robustly under designed observer to distribution into the control system based on output feedback  $H_\infty$  performant control problems. In the experimental result, we have successfully controlled the speed

of a networked dc motor system for our proposed method. It has shown how it becomes robustly with speed reference input. The effectiveness of our studying we can use observer designed to apply for the iNetCon-PC104 system can drive the brushless dc motor. Finally, the observer designed to optimize the parameters and eliminated the steady-state error, the robust speed control and analysed about the data transmitted through the NCS system performant by iNetCon-PC104 system which applied for brushless dc motor.

In this part, we want to make the experimental of networked dc motor system that we have changed the reference speed between 2000rad/s and 1000 rad/s in the communicated through network between controller and the plant to illustrate how to achieve robust performance from high speed to low speed and from low speed to high speed. The experimental of networked dc motor system has been demonstrated in the following figure:

According to Fig. 12, the red line is the reference speed and the black line is the actual speed of the networked dc motor system. The experimental results of the networked dc motor system that we settled reference speed 2000 rad/s and 1000 rad/s to achieve robust performance. By proposed our control strategy can be drive the speed of networked dc motor system from the high speed to low speed and from the low speed to the high speed to achieve robust speed control for the networked dc motor system. The experimental result demonstrated the effectiveness of our proposed method that we want to desired in the research strategy so the robust speed control for the networked dc motor system have been done by using our research strategy to make the simulation and the experimental result that we have already demonstrated in figure.

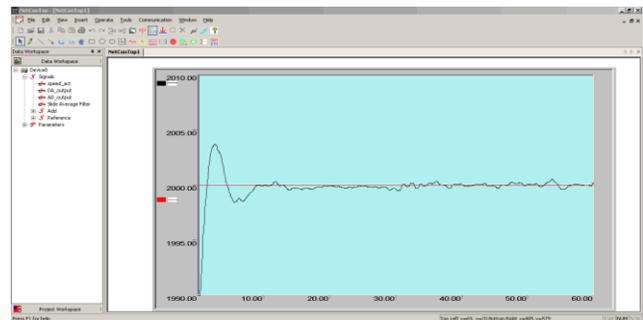


Fig. 11. Experiment Result of Robust Speed Control with Reference Speed 2000 Rpm.

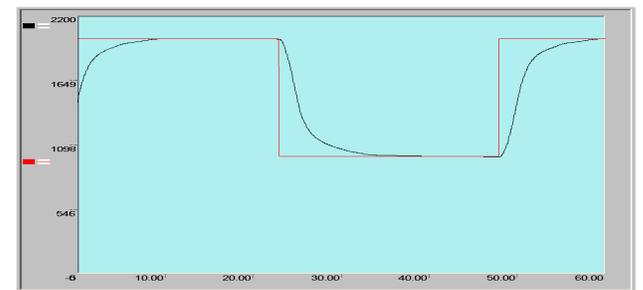


Fig. 12. Experiment Result of Robust Speed Control with Reference Speed 2000 and 1000 Rpm.

```

Variable Name:
Sample Time (s)
Time Value

Device0/Signals/speed_act:
0.020000
0.000000 1999.838623 0.020000 1999.838623 0.040000 1999.838623
0.060000 1999.838623 0.080000 1999.838623 0.100000 1999.838623
0.120000 1999.858154 0.140000 1999.877686 0.160000 1999.877686
0.180000 1999.877686 0.200000 1999.877686 0.220000 1999.877686
0.240000 1999.877686 0.260000 1999.877686 0.280000 1999.877686
0.300000 1999.877686 0.320000 1999.877686 0.340000 1999.877686
0.360000 1999.877686 0.380000 1999.877686 0.400000 1999.877686
0.420000 1999.877686 0.440000 1999.877686 0.460000 1999.877686
0.480000 1999.897217 0.500000 1999.897217 0.520000 1999.897217
0.540000 1999.897217 0.560000 1999.897217 0.580000 1999.897217
0.600000 1999.897217 0.620000 1999.897217 0.640000 1999.897217
0.660000 1999.877686 0.680000 1999.877686 0.700000 1999.877686
0.720000 1999.877686 0.740000 1999.877686 0.760000 1999.877686
0.780000 1999.877686 0.800000 1999.877686 0.820000 1999.877686
0.840000 1999.877686 0.860000 1999.877686 0.880000 1999.877686
0.900000 1999.877686 0.920000 1999.877686 0.940000 1999.877686
0.960000 1999.877686 0.980000 1999.877686 1.000000 1999.877686
1.020000 1999.877686 1.040000 1999.877686 1.060000 1999.877686
1.080000 1999.877686 1.100000 1999.877686 1.120000 1999.877686
1.140000 1999.877686 1.160000 1999.877686 1.180000 1999.877686
1.200000 1999.877686 1.220000 1999.897217 1.240000 1999.897217
1.260000 1999.897217 1.280000 1999.916748 1.300000 1999.916748
1.320000 1999.916748 1.340000 1999.916748 1.360000 1999.916748
1.380000 1999.916748 1.400000 1999.916748 1.420000 1999.916748
1.440000 1999.916748 1.460000 1999.916748 1.480000 1999.916748
1.500000 1999.916748 1.520000 1999.916748 1.540000 1999.916748
1.560000 1999.916748 1.580000 1999.936279 1.600000 1999.958111
1.620000 1999.975342 1.640000 1999.975342 1.660000 1999.975342
1.680000 1999.975342 1.700000 1999.975342 1.720000 1999.994873
1.740000 1999.994873 1.760000 1999.994873 1.780000 1999.994873
1.800000 1999.994873 1.820000 1999.994873 1.840000 2000.014404
1.860000 2000.014404 1.880000 2000.033936 1.900000 2000.033936
1.920000 2000.053467 1.940000 2000.053467 1.960000 2000.053467
1.980000 2000.053467 2.000000 2000.053467 2.020000 2000.053467
2.040000 2000.053467 2.060000 2000.053467 2.080000 2000.053467
2.100000 2000.053467 2.120000 2000.053467 2.140000 2000.072998

```

Fig. 13. Data of Speed Error.

After we did the experimental of the networked dc motor system to demonstrate the robustness, there are always some data speed error in the experimentation. The data speed error has been downloaded by using NetConTop software. The data steady state speed error of the networked dc motor system with iNetCon-PC104 to communicate through networked which the implement of the experimentation result is shown.

In Fig. 13, analysis of the experimental results of data steady state speed tracking error compared with reference input signals. The data have been saved from the iNetConn-PC104 system by using NetconTop APP, and the hols system experiment we have used some software such as VMware Workstation to visual and communicated between controller to the plant, MATLAB R2208a to the designed structure of controller for NCS system and NetConTop to display the signal form input and the output of the system. Finally, the exponentially stabilities of the mean square and steady state speed tracking have been proofed and have been invested in our studying by a newly designed observer to distribution based on  $H_\infty$  output feedback and consider with uncertainty parameter and disturbance.

## V. CONCLUSION

Based on observer design to distribution consider on  $H_\infty$  performant control problems deal with random data dropout, the exponentially stabilities of the mean square in the system of closed-loop, because of the unique advantages and wide potential application outlook have widely researched and developed. Compared to the previous research the NCS system stable result is lower than this work so far. However, the research on NCS system by applying the  $H_\infty$  performant control problems which the LMI toolbox solving is still

develop subsequently, it still has many unsolved problems, but now it has many simulations and experimental results studies on NCS system. This method can be also widely applied to the network-based dc motor system as well.

The creation of robust speed control for the network-based dc motor system, such as robust speed tracking error, the system of closed-loop is exponentially stabilities of the mean square. It has motivated and more inspiring for researcher to develop this kind of technology. Because the NCS system has transmitted data through a network without wiring, as well as to enrich robust speed control, and that is the reason of this paper uses a new observer design to distribution by  $H_\infty$  performant to optimal parameters and improved the previous result as well.

In this paper, we have studied and illustrated observer designed to a distribution deal with random data dropout to performant into the networked dc motor system. Because of the data packets have transmitted in the control system and there is some packets dropout during network-based transmissions. In additions, the robust speed control proposed with observer designed and random data dropout has been considering and modeling to issue that in NCS system is the data dropout phenomenon have solved in the form of mean-square sense. Consequently, the network-based dc motor system can be able obtained by observer design to distribution with  $H_\infty$  performant to optimize the parameters.

## ACKNOWLEDGMENT

This work was supported in part by school of information and Communication Engineering of Guilin University of Electronic Technology and the National Natural Science Foundation of China (Nos. 61562016, 61702129, 61772149, and 61320106008), and by Guangxi Key Laboratory of Intelligent Processing of Computer Images and Graphics (No. GIIP201703).

## REFERENCES

- [1] Ren M, Zhang J, Jiang M, Yu M and Xu J 2015 Minimum( $h, \phi$ )-Entropy Control for Non-Gaussian Stochastic Networked Control Systems and Its Application to a Networked DC Motor Control System IEEE Trans. Control Syst. Technol. 23 406–11.
- [2] Zhang X, Hengster-Movrić K, Šebek M, Desmet W and Faria C 2019 Distributed Observer and Controller Design for Spatially Interconnected Systems IEEE Trans. Control Syst. Technol. 27 1–13.
- [3] Tang X, Deng L, Yu J and Qu H 2018 Output Feedback Predictive Control of Interval Type-2 T-S Fuzzy Systems With Markovian Packet Loss IEEE Trans. Fuzzy Syst. 26 2450–9.
- [4] Xue B, Yu H and Wang M 2019 Robust  $H_\infty$  Output Feedback Control of Networked Control Systems With Discrete Distributed Delays Subject to Packet Dropout and Quantization IEEE Access 7 30313–20.
- [5] Postoyan R, Wouw N van de, Nešić D and Heemels W P M H 2014 Tracking Control for Nonlinear Networked Control Systems IEEE Trans. Automat. Contr. 59 1539–54.
- [6] Liu G, Yang Y and Yuan C 2013 Observer-based  $H_\infty$  control for discrete-time systems with random communication delays 2013 25th Chinese Control and Decision Conference (CCDC) pp 3353–8.
- [7] Liu S, Liu P X and Wang X 2016  $H_\infty$  control of networked control systems with stochastic measurement losses 2016 IEEE International Conference on Information and Automation (ICIA) pp 1691–6.
- [8] Zhang Q, Zhao D and Wang D 2018 Event-Based Robust Control for Uncertain Nonlinear Systems Using Adaptive Dynamic Programming IEEE Trans. Neural Networks Learn. Syst. 29 37–50.

- [9] Ho S, Chen C and Chen B 2017 Analysis for the Robust $H_\infty$  Synchronization of Nonlinear Stochastic Coupling Networks Through Poisson Processes and Core Coupling Design IEEE Trans. Control Netw. Syst. 4 223–35.
- [10] Zhou Y and Hu S 2015 H infinity control for DC servo motor in the network environment 2015 IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) pp 401–5.
- [11] Yu J, Yang C, Tang X and Wang P 2018  $H_\infty$  control for uncertain linear system over networks with Bernoulli data dropout and actuator saturation ISA Trans. 74 1–13.
- [12] Wang D, Liu D, Li H, Luo B and Ma H 2016 An Approximate Optimal Control Approach for Robust Stabilization of a Class of Discrete-Time Nonlinear Systems With Uncertainties IEEE Trans. Syst. Man, Cybern. Syst. 46 713–7.
- [13] Wang Z, Yang F, Ho D W C and Liu X 2007 Robust $H_\infty$  Control for Networked Systems With Random Packet Losses IEEE Trans. Syst. Man, Cybern. Part B 37 916–24.
- [14] Li Z, Chang X and Wang Y 2017 Robust observer-based  $H_\infty$  control for networked control systems with measurement quantization and packet dropouts 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA) pp 739–44.
- [15] Liu Y, Zhu X, Zhang H and Basin M 2018 Improved Robust Speed Tracking Controller Design for an Integrated Motor-Transmission Powertrain System Over Controller Area Network IEEE/ASME Trans. Mechatronics 23 1404–14.

# Forest Fire Detection System using LoRa Technology

Nicoleta Cristina GAITAN<sup>1</sup>, Paula HOJBOTA<sup>2</sup>

Faculty of Electrical Engineering and Computer Science, Stefan cel Mare University of Suceava<sup>1,2</sup>  
Integrated Center for Research, Development and Innovation in Advanced Materials, Nanotechnologies  
and Distributed Systems for Fabrication and Control (MANSiD) Suceava, Romania<sup>1</sup>

**Abstract**—Millions of hectares of forest worldwide are affected annually by fires, which can lead to the loss of human lives, materials, destruction of natural flora and fauna but also can lead to the losses of raw materials. The problem is even greater in forests that are not guarded and do not have communication systems available. Thus, in recent years, have been proposed various systems that use devices based on Internet of Things (IoT) for real-time forest fire detection. In this paper, it is proposed a system capable of quickly detecting forest fires on long wide distance. In the development of this system it is used LoRa (Long Range) technology based on LoRaWAN (Long Range Wide Area Network) protocol which is capable to connect low power devices distributed on large geographical areas being an innovative and great solution for transmissions of a low data transfer rate and a low transmission power on high ranges, and because has a great efficiency.

**Keywords**—LoRa; real-time; long range wide area network; internet of things

## I. INTRODUCTION

Forest fires that take place in the warm season can be caused or started by natural events or human negligence. Natural events such as burning branches or dry leave appear due to the heat generated by the sun. An example of such event is the fire from Siberia in the summer of the year 2019, resulting in the destruction of millions of hectares of vegetation. The fires generated by human negligence can result from multiple factors, such as leaving unattended fires in the forest, throwing a burning cigar etc. An example of a forest fire started due to human negligence took place in the province of South Sumatra, in 2015.

In the modern era, the technology has developed so much that it's become more and more sophisticated, making wireless transmissions real easy. The network of wireless sensors, named WSN (Wireless Sensors Network), based on the LoRa module, can be a great alternative for detecting forest fires in specific areas.

LoRa (Long Range) is a low power technology developed by Semtech and supported by LoRa Alliance. This technology is characterized by a low data transfer rate and a low transmission power on high ranges [1][2]. While the transmission range grows, LoRa keeps its low power characteristics of modulation in frequency (Frequency Shift Keying). This technology modulates symbols with a bandwidth of 125, 250 and 500 kHz (in case of European applications) with different spreading factors [3].

The LoRa Alliance defines two different layers of this technology: LoRa physical layer and LoRaWAN (Long Range Wide Area Network) protocol.

From the physical layer point of view, LoRa is a radio modulation patented technique by Semtech. In this case, the technology is going to work with frequencies under the order of GHz in the unlicensed ISM (Industrial, Scientific and Medical) band [2].

LoRaWAN protocol is standardized by the LoRa Alliance, being defined as a MAC layer protocol and a system architecture that uses LoRa's physical layer. The access control mechanism to environment offered by LoRa's allows for multiple final devices to communicate with a gateway using LoRa modulation (see Fig. 1) [2].

The Low-Power Wide-Area network ensures the connectivity of low power devices distributed on large geographical areas. These networks represent a new model of communication, successfully completing the already existent wireless communication technology such as: Bluetooth, Zig-Bee, LTE, GSM and Wi-Fi [2].

Thanks to its low power requirements and low costs of manufacturing and operating, LoRaWAN is one of the most used LPWA (Low-Power Wide-Area) technologies.

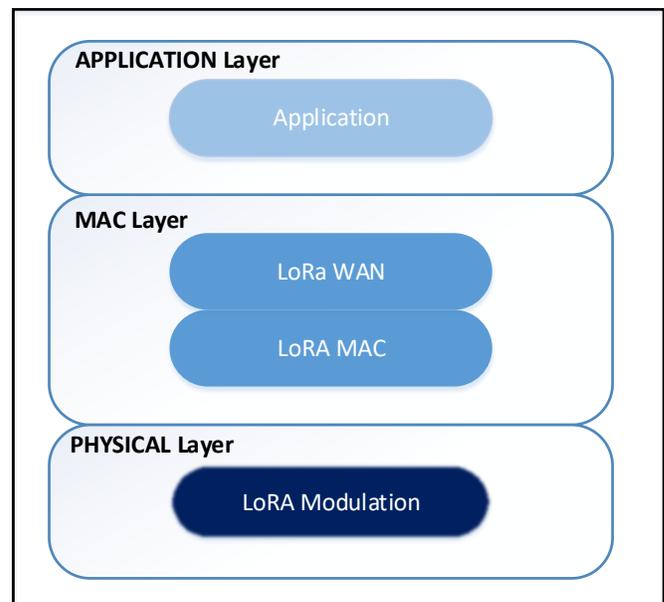


Fig. 1. Long Range Wide Area Network Stack.

The LoRa technology doesn't guarantee a long time development, but there are already available solutions based on it, unlike other technologies that could possibly disturb the global development [4].

Industrial IoT applications are based on long distance communication. The promising protocols in this domain are SigFox, LoRaWAN and the NB-IoT standard [1]. In Table I is presented a comparison between SigFox and LoRa.

TABLE I. SIGFOX VS. LORA COMPARISON

Parameter	SigFox	LoRa
Frequency Band	868, 902 [MHz]	433(US), 863-870 (EU) [MHz]
Data Rate	100 bps	10 kbps
Rural Range	30-50 km	15-20 km
Urban Range	3-10 km	2-5 km

The Internet of Things is a communication model that represents a near future solution, capable of integrating sensors and devices that can communicate directly between them without human intervention. The "things" from Internet of Things include physical devices equipped with microcontrollers, transceivers for digital communications and stacks of protocols for making possible the communication between users [5][6].

With such a great multitude of IoT applications, the LoRaWAN protocol and LoRa devices seems to be very efficient in business, contributing to make people lives well around the globe [4]. Using this technology it expecting to achieve a smart connection of the entire planet. LoRa applications are found in various field, such as, agriculture, smart cities, smart environment, healthcare, smart homes and buildings, industrial control, smart metering, smart supply chain and logistics.

In this article it is presented a study of a system that follows the integration of LoRa technology in forest fire detection.

The rest of the paper it is structured as follows. Section II summarizes some interesting systems focused on fire detectors. The proposed design is presented in Section III. Section IV presents conclusion and future work.

## II. RELATED WORKS

The traditional methods of forest fire detection and prevention are based on observation through satellite images, visual observation by guards, observation by air or video detection on high ranges [7].

Considering all the disadvantages of conventional forest fire detection systems, in the last years different solutions were proposed in order to improve the monitoring systems of the environment and to create a new real time forest fire detection systems using devices based on the Internet of Things [8].

The detection through IoT based devices and online monitoring systems [9]. In this article, the authors proposed a new forest fire monitoring system by detecting temperature variations and analyzing CO2 levels. The forest fire detection system was realized using the Arduino Uno module alongside a

temperature sensor, a smoke sensor and an alarm system. The temperature sensor detects the variations in temperature and the smoke sensors detect the CO2 levels and if it is necessary, it enables the alarm (the buzzer on the Arduino Uno board). Using the IoT technology, the system was connected to a webpage named „FireSecurity System”, created in PHP and controlled by the Arduino programming environment [9].

In the article [10], the authors proposed an implementation of a forest fire detection system in the forest area using modern equipment. The system was not only meant for fire detection, but for alarming the forest officer regarding the started fire in the forest. The system's activities are controlled by a microcontroller, and the used sensors have the role of identifying the fire and its exact location. This system is based on IoT, the activities being continuously monitored, and the data being saved and shared to online websites. The forest officer regularly reviews the newly stored data, which can be verified at any time [10]. The monitoring system is usually based near the forest office or a fireman station.

A Smart Forest-Fire Early Detection Sensory System: Another Approach of Utilizing Wireless Sensor and Neural Networks [11]. This article has as its main purpose the implementation of a fast forest fire detection system using cheap and small sizes sensors that do not require surveillance. The system does not utilize large scale centralization systems, which could affect the system's robustness in vulnerable environments. The used sensors network represents a small scale cell that can be multiplied in order to cover the entire forest [11].

Wireless sensor network for forest fire detection [12]. In order to overcome to loss of thousands of forest hectares, the authors propose a forest fire detection system by using a sensor network. Each node has a microcontroller, a transmitter, a receiver and 3 sensors. The measurement methods consist of temperature measurement, detection of methane levels, hydrocarbons and CO2 levels.

Emerging methods for early detection of forest fires using unmanned aerial vehicles and LoRaWAN sensor networks [13]. In this article, the authors propose 2 different solutions for detecting forest fires. First solution consists of aerial vehicles without pilots (UAVs) equipped with special cameras. In this papers there are presented and analyzed multiple scenarios of drone usages for forest fire detection, including a solution which uses a combination between UAV with fixed-wind and UAV with rotary-wings. The system's basic configuration presented in this article implies using a network of soil mounted cameras through which the constant monitoring of the forest is ensured. The cameras used have double lenses which offer both standard images and infrared images (IR) [13].

In other research paper [14], the authors propose a model of communication based on WSN for detecting forest fires in real time as a much more efficient method of monitoring them than the satellite [15]. Their main purpose is fast detecting of fires in order to reduce the loss of vegetation, flora and fauna. In the realization of this system, it's proposed the usage of some sensors that collect the measured dates and send them to the group of nodes for further processing by building a neuronal network. The neuronal network produces a meteorological

index that measures the probability of the weather favoring an eventual fire. The cluster headers send meteorological indexes to a node manager that has the role of deciding the potential danger for a fire considering the received indexes [15].

### III. SYSTEM DESCRIPTION

In this section of the article it is presented the proposed system and different hardware parts which are going to be implemented.

#### A. General Description

The capacity to detect if a fire is present is the most important part of a fire safety strategy. Without fire detection means, it cannot alert the population, it cannot take safety measure against the fire and it cannot alarm the firefighting service. The people, by our nature, are capable of feeling heat and smoke, to see the flames and to hear the fire burning. Therefore, we could be great potential fire detectors, but we are not always available or trustworthy, and for this reason, we need to use the technology to replace those abilities.

After analyzing the problems, the losses resulted from forest fires and the high costs of detection and monitoring systems, I believe it is necessary to develop some low cost systems capable of evaluating the risk of a forest fire starting, but also its presence.

The system proposed in this article is realized based on a LoPy4 development board connected to a Pycom Expansion Board 3.0 (Fig. 2), an Arduino Mega 2560 module, a temperature and humidity sensor and a flame detector.

#### B. Pycom LoPy and Pycom Expansion Board 3.0

The main component of the proposed forest fire detection system is the LoPy4 development board from Pycom. The LoPy4 is a development platform compatible with MicroPython (WiFi, LoRa, SigFox, and Bluetooth). This board is programmable using MicroPython and the Pymakr plugins for a fast IoT applications development. LoPy4 can be configured in LoRa mode to directly send packages between LoPy4 development boards of the same kind. It represents the best combination of implementing speed and access to the new LPWAN networks from Europe, USA, Africa and India. LoPy4 is CE, FCC, LoRaWAN and SigFox certified [16]. In order for the LoPy4 module to be programmed, we are going to use it along with the PyCom Expansion Board 3.0 development board. This board allows the creation and connection of new IoT projects with WiPy 2.0, WiPy 3.0, LoPy, LoPy4, SiPy, FiPy and Gpy. Pycom Expansion Board 3.0 has different libraries and templates, therefore, the development of an IoT solution becomes much easier and faster [17][20][21][22][23].

#### C. Arduino Mega 2560

Arduino Mega 2560 is a board equipped with a microcontroller and it is based on ATmega2560. The module has 54 pins for digital inputs and outputs (15 of them capable of being used for PWM outputs), 16 analogic inputs, 4 UARTs, a crystal oscillator with a frequency of 16 MHz, an USB connection, an ICSP header and a reset button. The Arduino Mega 2560 board can be programmed with the Arduino Software (IDE) [18]. I chose to use the Arduino board because it is easier to use by students and is even cheaper.



Fig. 2. The Pycom Lopy and Pycom Expansion Board 3.0.

#### D. Temperature Sensor DHT11

The DHT11 is a cheap base sensor, used for measuring temperature and relative humidity at the same time. The measurement of air humidity is ensured by a capacitive humidity sensor, while the temperature is measured using a thermistor, resulting in a digital signal on the data pin. This sensor is a small sized, low power usage, recommended in applications where data acquisition with a high frequency is not required [3]. The sensor operates in cycles of 1 second. In this cycle, the Arduino module and the sensor exchange information about the temperature and humidity. The features of the DHT11 are presented in Table II.

#### E. Flame Detector

A flame sensor is a device used to detect the presence of either a fire source or any lightning source. The flame detection can be done from a distance of 1 meter, and the angle detection is 60 degrees. The flame detectors with infrared or with large band monitor the spectral band in infrared for specific models given by hot gasses. These are recognized using a special camera, using thermal imagistic for countering the fires, which is a thermographic camera [19].

TABLE II. DHT11 FEATURES

DHT11 Specifications	Values
Operating Voltage	3.5V to 5.5V
Operating current	0.3mA (measuring) 60uA (standby)
Output	Serial data
Temperature Range	0°C to 50°C
Humidity Range	20% to 90%
Resolution	Temperature and Humidity both are 16-bit
Accuracy	±1°C and ±1%

#### IV. CONCLUSIONS

As the IoT technology is being developed more and more, the devices capable of communicating remotely from distance without using a lot of energy have become a necessity more than ever. This challenge of the future is currently served by Long-Power Wide-Area Networks. Currently, there are a lot of applications based on the LPWA technologies, but most of them do not have LoRa module incorporated.

The global warming will constantly be contributing to increasing the number of fires and the damages caused by them. Each season, besides thousands of hectares of forest being destroyed, there are a lot of assets and properties being affected as well. Even more, both the firefighters and the civilians' lives are also in danger. This is the main reason I want to address this issue in this article, for the sole purpose of reducing and preventing devastating fires. The desire to create such a system comes from the fact my father is a forest guard and has been so for more than twenty years. One of his most important responsibilities is to watch out for the forest's well growth and make sure nothing is going to harm it.

In this article, in the Related Works section there were presented different systems of forest fires detection based on LPWA communications. The system proposed in the article is based on the LoRa technology and it has as its main component the LoPy development board from Pycom. The system can be used in extreme temperature conditions, but it can be affected if it's situated in the middle of the fire. With the help of the flame detector, the flames can be noticed from a distance of 100 centimeters. It detects and sends information regarding a possible fire starting out.

Although we know that the most environmental monitoring systems are made using Arduino boards and IoT extensions interfaces, it is desire to emphasize that the novelty of this system is the fact that it is tested with the Arduino board. It is easier to test with the Arduino board because there is online software and it is not expensive.

The decision of using LoRa technology in the proposed system is due of it capabilities to connect low power devices distributed on large geographical areas and because represents an innovative solution for long wide distance transmissions with low power on high ranges consumption and also because has a great efficiency on long-distance data transmission and low transfer power.

For future research, the proposed system can be improved, for example by integrating a surveillance camera for a more precise monitoring process and detection.

Another improvement in the development of this proposed system would be the integration of a GPS module.

#### REFERENCES

- [1] T. Polonelli, D. Brunella, A. Girolami, G. N. Demmi și L. Benini, „A multi-protocol system for configurable data streaming on IoT healthcare devices,” 8th International Workshop on Advances in Sensors and Interfaces (IWASI), 2019, pp. 112-117.
- [2] L. Leonardi, F. Battaglia, G. Patti și L. L. Bello, „Industrial LoRa: a Novel Medium Access Strategy for LoRa in Industry 4.0 Applications,” IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, 2018, pp.4141-4146.
- [3] S. Sendra, P. Romero-Díaz, J. Navarro-Ortiz și J. Lloret, „Smart Infant Incubator Based on LoRa Networks,” 15th International Conference on Computer Systems and Applications (AICCSA), Aqaba, 2018, pp. 1-6.
- [4] Semtech Company (LoRa). [Interactiv] 7 February 2020. <https://www.semtech.com/lora/why-lora>.
- [5] A. Zanella, N. Bui, A. Castellani și L. V. a. M. Zorzi, „Internet of Things for Smart Cities,” IEEE Internet of Things Journal , vol.1, pp. 22-32, 2014.
- [6] L. Atzori, A. Iera și G. Morabito, „Internet of Things security: A survey,” Computer Networks: The International Journal of Computer and Telecommunications Networking, vol. LIV, 2010, pp 2787–2805. DOI:<https://doi.org/10.1016/j.comnet.2010.05.010>.
- [7] A. Herutomo, M. Abdurrohman, N. A. Suwastika, S. Prabowo și C. W. Wijiutomo, „Forest fire detection system reliability test using wireless sensor network and OpenMTC communication platform,” 2015.
- [8] J. Toledo-Castro, P. Caballero-Gil, N. Rodríguez-Pérez, I. Santos-González și C. H.-G. a. R. Aguasca-Colomo, „Forest Fire Prevention, Detection, and Fighting Based on Fuzzy Logic and Wireless Sensor Networks”.
- [9] A. K. Sharma, M. F. R. Ansari, M. F. Siddiqui și M. A. Baig, „IoT Enabled Forest Fire detection and Online Monitoring System,” International Journal of Current Trends in Engineering & Research (IJCTER), Volume 3, Issue 5, pp50-54, 2017.
- [10] K. Jayaram, K. Janani, R. Jeyaguru și R. K. a. N. Muralidharan, „Forest Fire Alerting System With GPS Co-ordinates Using IoT,” 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), pp 488-491, 2019.
- [11] H. Soliman, K. Sudan și A. Mishra, „A Smart Forest-Fire Early Detection Sensory System: Another Approach of Utilizing Wireless Sensor and Neural Networks”, in Sensors, 2010 IEEE. IEEE, pp. 1900–1904, 2010.
- [12] M. Hariyawan și A. G. a. E. Putra, „Wireless Sensor Network for Forest Fire Detection,” vol. II, 2013.
- [13] G. Hristov, J. Raychev și D. K. a. P. Zahariev, „Emerging methods for early detection of forest fires using unmanned aerial vehicles and LoRaWAN sensor networks,” 2018.
- [14] L. Yu, N. Wang și X. Meng, „Real-time Forest Fire Detection with Wireless Sensor,” pp1-4, 2005.
- [15] R. Vega-Rodríguez, S. Sendra, J. Lloret și P. R.-D. a. J. L. Garcia-Navas, „Low Cost LoRa based Network for Forest Fire Detection,” 2019.
- [16] „Pycom,” [Interactiv]. Available: <https://pycom.io/product/lopy4/>. [Accesat 15 February 2020].
- [17] „Sparkfun,” [Interactiv]. Available: <https://www.sparkfun.com/products/14675>. [Accesat 9 February 2020].
- [18] „Arduino,” 25 February 2020. [Interactiv]. Available: <https://store.arduino.cc/arduino-mega-2560-rev3>.
- [19] „Wikipedia,” 29 February 2020. [Interactiv]. Available: [https://en.wikipedia.org/wiki/Flame\\_detector](https://en.wikipedia.org/wiki/Flame_detector).
- [20] A. Lavric și A. I. Petrariu, „LoRaWAN Communication Protocol: The New Era of IoT,” Conference: 2018 International Conference on Development and Application Systems (DAS), 2018, pp 74-77.
- [21] A. Lavric, V. Popa, „A LoRaWAN: Long Range Wide Area Networks Study”, 11-th International Conference on Electromechanical and Power Systems, pp. 435-438, 2017.
- [22] O. Khutsoane și B. I. a. A. M. Abu-Mahfouz, „IoT devices and applications based on LoRa/LoRaWAN,” IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society, Beijing, pp. 6107-6112, 2017.
- [23] B. Sarwar, I. S. Bajwa, N. Jamil și S. R. a. N. Sarwar, „An Intelligent Fire Warning Application Using IoT and an Adaptive Neuro-Fuzzy Inference System,” Sensors 19(14):3150, doi:10.3390/s19143150, 2019.

# Detecting C&C Server in the APT Attack based on Network Traffic using Machine Learning

Cho Do Xuan<sup>1</sup>, Lai Van Duong<sup>2</sup>  
Information Assurance Dept  
FPT University,  
Hanoi, Vietnam

Tisenko Victor Nikolaevich<sup>3</sup>  
Systems of Automatic Design  
Peter the Great St. Petersburg Polytechnic University  
Russia, St.Petersburg, Polytechnicheskaya, 29

**Abstract**—APT (Advanced Persistent Threat) attack is a form of dangerous attack, it has clear intentions and targets. APT uses a variety of sophisticated, complex methods and technologies to attack on targets to gain confidential, sensitive information. Currently, the problem of detecting APT attacks still faces many challenges. The reason is APT attacks are designed specifically for each specific target, so it is difficult to detect them based on experiences or predefined rules. There are many different methods that are researched and applied to detect early signs of APT attacks in an organization. Today, one method of great concern is analyzing connections to detect a control server (C&C Server) in the APT attack campaign. This method has great practical significance because we just need to detect early the connection of malware to the control server, we will prevent quickly attack campaigns. In this paper, we propose a method to detect C&C Server based on network traffic analysis using machine learning.

**Keywords**—Advanced Persistent Threat (APT); abnormal behavior; network traffic; machine learning; APT detection; Control Server (C&C Server)

## I. INTRODUCTION

The publication [1, 2] presented the characteristics, procedures and life cycle of APT attack. From the characteristics of the APT attack, it shows that the APT attack has specific, clear objects and goals. Any organization, individual, business or governmental agency can become victims of this attack. In the past, APT attack groups often operated for personal purposes. However, most APT attack groups are now financially supported by government or financial institutions in order to implement political motives. Due to this change, ATP attack groups are increasingly equipped with not only modern attack, hide and cover tracks tools, but also a team of warlike, elite hackers. In the publication [3], the authors have presented some of the characteristics of the attack scenario makes APT attack detection becomes much more difficult than any other threats, such as advanced attack tool, lack of public data, and using standard encryption protocols. From the above presentations, we can see the dangers as well as difficulties in detecting APT attacks. In fact, many APT attacks have taken place over the years, exploiting large amounts of data without the victim's knowledge.

However, although APT attacks are advanced and sophisticated with completely new attack ways, they are all separated into four main stages [3, 4, 5]: spying (collecting

information), attacking and escalating privilege, stealing information, and covering tracks. All four stages have the same role and importance, they support each other in the entire offensive campaign.

In the attacking and escalating privileges stage and stealing information stage, all of these are performed by commands from the C&C Server to the malware that has been exploited in the target machine. Therefore, if the system can detect abnormalities in the connection, it can quickly and accurately detect the signs of APT in the system. To detect the connection to the C & C Server, studies often focus on the issue of monitoring the anomaly of network traffic or rely on a list of C & C Server that has been built before. However, APT malware often easily bypass these traditional approaches. Therefore, in order to improve the ability to detect abnormal connections to a C&C Server, in this paper, we propose a method to analyze abnormal behavior in network traffic based on machine learning techniques. Accordingly, firstly the network traffic data will be analyzed and extracted behaviors relying on domain name or IP address, then these behaviors will be built into a feature vector. Finally, we use a machine learning algorithm to classify them in order to detect the abnormal connection of the C & C server. The science of our paper includes recommending some abnormal features of C&C Server based on Network traffic and using the Random Forest algorithm to detect abnormal connections. The paper is organized as follows. Section II reviews some recent works in the literature on C&C Server detection. The proposed C&C Server detection system using machine learning is presented in Section III. In this section, the new features for the C&C Server detection process are also described in detail. Experimental results and discussions are provided in Section IV. The paper is concluded in Section V.

## II. RELATED WORKS

### A. Several Methods of Detecting APT are based on Abnormal Connections

In [3], the authors proposed a method for APT attack detection based on the analysis of abnormal behaviors of flow in Network Traffic. This method includes the process of extraction, normalization, and analysis of abnormal values of three groups of signals in flow, which are numbytes, numflows, numdst. Andrew Vance et al. [6] used measures non-signature based traffic and involved flow based measurements and applied a statistical for detection APT attack. Weina Niu et al.

[7] introduced a method for APT attack detection based on Mobile DNS Logging using four sets of features, which are DNS request, answer-based features; Domain-based features, Time-based features, Whois-based features. With the selected feature sets, the authors used a number of machine learning methods such as Global Abnormal Forest, k-Nearest Neighbor to detect APT Malware. G. Zhao et al. [8] used five sets of features: Domain-based features, Time-based features, Whois-based features, DNS answer-based features; Active probing features and used J48 decision tree algorithm to detect APT malware command and control domains (C&C Domain). In [9], the authors used three sets of features to detect the domain APT, which are Domain name lexical features; Ranking features; DNS query features and Random Forest machine learning.

### B. Detecting APT Attack using Big Data Technology

The publication [10] listed a number of APT attack detection tools based on analysis and correlation calculations among events such as Splunk, LogRapse, and IBM QRadar. Jisang Kim et al. [11] proposed a method to detect APT attacks based on the process of collecting and processing collected data sources consisting of the network packets are collected; Email logs are traced to accept; the privilege increase logs (Syslog) are traced to accept; Call-back domain blacklist; Internal DNS server; SSL port. However, in this paper, the authors didn't present the technical solutions and the big data processing technology used. Besides, Sung-Hwan Ahn et al. [12] proposed the idea of applying big data technology to APT attack detection. Accordingly, the authors proposed the architecture of big data analysis system consisting of the following stages: collecting data from firewall and log, behavior, status information (date, time, inbound/outbound packet, daemon log, user behavior, process information, etc.) from anti-virus, database, network device and system; preprocessing data; analyzing data; and giving warning results for signs of APT attacks. However, in this paper, the authors didn't present the solution or technology used in bigdata to support the model proposed by the authors. In the paper [13], the authors proposed the APT attack detection model on the big data platform with two main processes: Behavior Rule Generation and Abnormal Behavior Detection. In this proposed model, the authors use the Hadoop MapReduce framework.

### C. Some Commercial Software Detecting APT Attacks

The document [14] introduced a number of commercial products and technologies that support APT attack monitoring and detection, including Symantec, Forcepoint, McAfee, Kaspersky Lab, Fortinet, Cisco, Palo Alto Networks, and FireEye.

McAfee Advanced Threat Defense is designed to detect APT malware and zero-day vulnerability by combining static analysis with dynamic analysis through sandboxing techniques. The analysis results will be provided to the system to detect and alert from within the network to the terminals. However, the disadvantage of this solution is the inability to analyze attachments on emails.

Kaspersky Anti Targeted Attack Platform (KATAP) is a solution that combines machine learning algorithms with sandbox technology to handle information about threats

collected from inside systems and terminals in order to detect signs of APT malware (including known, unknown and APT malware) at any stage in the APT attack's life cycle. However, the disadvantage of the KATAP solution is not providing the monitoring and troubleshooting function after APT attack campaign, and the weakness in preventing data leakage.

FireEye's APT attack prevention solution is a set of solutions that analyze data from multiple sources such as Web, Email, File, Central Management and Malware Analysis. Accordingly, all suspicious files, attachments, files, and URLs are automatically scanned and monitored through the rule sets, then all suspicious signals will be transferred to the sandbox environment to be executed.

Advanced Malware Protection (AMP) solution of Cisco is the solution to detect APT attacks at the stage of spreading or hiding in the system or all 3 phases (consist of before APT attacks, during APT attack, and after APT attack). In Before APT attack phase, AMP uses information about threats worldwide gathered from Cisco's Collective Security Intelligence, Talos Security Intelligence and Research Group, and AMP Threat Grid to prevent known malware attacks.

In During APT attack phase, AMP uses information obtained from known attacks (signature), combined with AMP Threat Grid technology with the ability to automatically analyze malware in order to identify and prevent suspicious, dangerous files that are trying to gain access to the network. In After APT attack phase, AMP not only checks and monitors at the time of the attack, but continues to monitor and analyze all operations and paths of the data (though it was previously considered to be "clean"), and look for signs of dangerous behavior. When detecting a file containing malicious code, AMP provides visual information about the activity in the network, in each terminal of malicious code, AMP also allows quick response and troubleshooting through a simple web interface. However, AMP doesn't provide a function to prevent data leakage.

WildFire is Palo Alto Networks' APT attack detection solution, providing full visibility of all traffic, including APT threats from Web traffic, email protocols (SMTP, IMAP, POP), FTP (regardless of whether it is encrypted or not). The weaknesses of the WildFire solution is that it only focuses on monitoring the network layer, with little interest in the application layer and only focuses on detecting and preventing attacks but doesn't provide troubleshooting.

## III. C&C SERVER DETECTION MODEL DEVELOPMENT

### A. Model Overview

Fig. 1 presents the proposed C&C server detection system using machine learning. The model consists of the following components:

- Network Traffic: Network data that is checked here can be taken directly in real-time from the network card or can be taken from the pcap file.
- Extract features: In this paper, we use the Bro IDS tool to assist in analyzing network traffic into network components. Bro IDS Tool is a network security monitoring tool with fast processing speed. It detects

intrusion based on rule sets and helps to separate features from network traffic at high speed.

- Training: After extracting the necessary information based on Bro IDS log files, the features will be saved in the CSV file. Random Forest algorithm will be used to classify from those features

### B. Select and Extract Features

Table I lists the features of network traffic we use. All features marked “\*” in Table I are newly extracted and selected in this research.

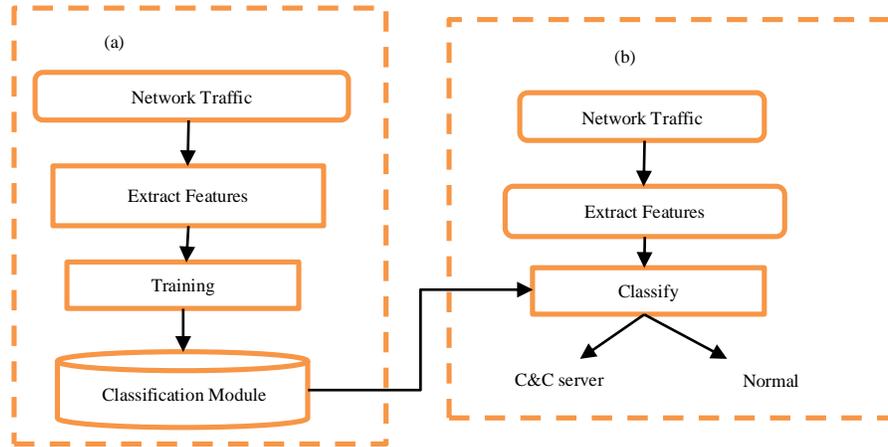


Fig 1. Overview Model.

TABLE I. LIST OF FEATURES USED TO DETECT A C&C SERVER

No	Feature	Type
1	Anomaly Port and Protocol	Bool
2	Ratio of number of packets OUT/IN *	Integer
3	Ratio of number of Bytes OUT/IN *	Integer
4	Ratio of inter-arrival times OUT/IN *	Integer
5	Number of three way handshakes	Integer
6	Number of connection teardowns	Integer
7	Number of complete conversation	Integer
8	Anomaly Data *	Float
9	Number of packets per time *	Integer
10	Number of bytes per time *	Integer
11	Percentage of TCP SYN packets	Float
12	Percentage of TCP SYN ACK packets	Float
13	Percentage of TCP ACK packets	Float
14	Percentage of TCP ACK PUSH packets	Float
15	Command and File System *	Bool
16	Data to computer in LAN	Bool
17	Tor Network *	Bool

The features in Table I are defined as follows:

- Anomaly Port and Protocol: In order to find abnormal ports, which don't run properly service according to the Internet standard, the first step is extracting the strange IP address that the server queries to in the DNS packet. From there, we find the queries that server queries to that IP address. From those records, we extract the protocols and service ports from the server, consider whether they are suitable or not, otherwise, the protocol and the port are abnormal. We define this because when a server provides service out with a specific port, it always listens to requests and returns responses through that port. For example, web services with the HTTP protocol have a default port of 80. Thus, the webserver always listens to requests and responds via port 80.
- Ratio of number of packets OUT/IN: is the ratio of the number of OUT and IN packets.
- Ratio of number of Bytes OUT/IN: is the ratio of the number of bytes of OUT and IN packets.
- Ratio of inter-arrival times OUT/IN: is the ratio of inter-arrival times OUT and IN.
- Number of three way handshakes: is the number of three way handshakes.
- Number of connection teardowns: is the number of the failed connection. Table II shows some cases that occur when a connection fails.
- Number of complete conversation: is the number of successful connections.
- Anomaly Data: The first step is identifying strange IP addresses through DNS records. Then, we take the value of the tcp.len field of all records that their destination IP is a strange IP address. From there, we find the maximum size of the packet and calculate the average size of the packets
- Number of packets per time: is the number of packets in a period such as hours, minutes, seconds, days, weeks, months and years.
- Number of bytes per time: is the number of sizes of packets in a period such as hours, minutes, seconds, days, weeks, months and years.
- Percentage of TCP SYN packets: is the percentage of the SYN flag in the TCP protocol of packets.
- Percentage of TCP SYN ACK packets: is the percentage of SYN ACK flags in the TCP protocol of packets.
- Percentage of TCP ACK PUSH packets: is the percentage of ACK PUSH flags in the TCP protocol of packets.
- Command and File System: The commands that the C&C server sends to malware are always command line, so if traffic has the system command line that is

transmitted to any machine in the system from the external internet, it is very likely from C&C server. In addition, recent attacks such as Sofary Group's Parrallel Attack [15], which is organized by APT 28 in February 2018, the malware will save the data that the malware obtained into the file in the % APPDATA% folder and transfer directly to C&C Server. Similarly, other attacks, the malware also hide information in directories like % TEMP%, etc. and send it directly to the C&C server. Thus, if the network traffic has command lines and system files, it is certainly attacked by APT.

- Data to computer in LAN: In recent attacks, the first computer on the LAN that is hacked will be used to gather all data from other computers on the LAN and data from that machine will be sent to C&C Server via VPN. Typically, the APT15 attack on the US Navy on June 14, 2018 [16].
- Tor network: In order to encrypt the operations and commands of malware when it accessed to the victim machine, APT organizations often use the Tor network to encrypt and to avoid detecting C&C Server addresses.

### C. C&C Server Detection Method

To detect connections from within the network to the C&C server, in this paper, we use the Random Forest algorithm. Random Forest is an ensemble classification method [17]. This algorithm is based on an ensemble of classifiers, which normally are Decision Trees to make the final prediction. The theoretical foundation of this algorithm is based on Jensen's inequality [18]. According to Jensen's inequality applied to the classification problems, it is shown that the combination of many models may produce less error rate than that of each individual model. In the study [19, 20] has proven Random Forest algorithm has many advantages compared to other machine learning algorithms. In this paper, we use the Random Forest algorithm with the number of decision trees of 10 in order to classify and test connections.

TABLE II. THE FAILED CONNECTION CASES

Packet sent	Packet received
Send TCP SYN	TCP Reset
Send TCP SYN or UDP Packet	ICMP unreachable
Send TCP SYN or UDP Packet	Do not receive the packet for 120 seconds

## IV. EXPERIMENTS AND EVALUATION

### A. Dataset and Experiment Environments

In this paper, we collected 61 network traffic files of APT attacks from [21-26]. Table III lists in detail the components of the experimental dataset.

TABLE III. THE COMPONENTS OF THE EXPERIMENTAL DATASET

Source	[21-24]	[25, 26]
The number of PCAP files	28	33
The number of DNS queries	985,595	272
Domain name	50 domain names (consisting of 26 domain names related to APT attacks, and 24 clean domain names)	45 domain names (consisting of 18 domain names related to APT attacks, and 27 clean domain names.)
IP address	921 IP addresses (including 581 public IP addresses, and 71 IP addresses related to APT attacks)	105 IP addresses (including 75 public IP addresses, and 25 IP addresses related to APT attacks)

### B. Metrics

To evaluate machine learning models, in this paper we use the following metrics:

Accuracy: the percentage of correct decisions among all testing samples

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

where: TP (True Positive): is the number of network flows which the model correctly predicts is the APT attack network flow; FN (False Negative): is the number of network flows which the model incorrectly predicts is normal; TN (True Negative): is the number of network flows which the model correctly predicts is normal; FP (False Positive): is the number of network flows which the model incorrectly predicts is the APT attack network flow.

Precision: is the ratio of the number of APT attack network flows that is correctly predicted among those classified as APT attacks network flows.

$$precision = \frac{TP}{TP + FP} \times 100\%$$

Recall: is the ratio of the number of APT attack network flows that is correctly predicted among those that are actually the APT attack network flows.

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

TPR (True Positive Rate): is the rate of network flows which the model correctly predicts is the APT attack network flow.

$$TPR = \frac{TP}{TP + FN}$$

FPR (False Positive Rate): is the rate of network flows which the model incorrectly predicts is the APT attack network flow.

$$FPR = \frac{FP}{FP + TN}$$

TNR (True Negative Rate): is the rate of network flows which the model correctly predicts is normal.

$$FNR = \frac{FN}{TP + FN}$$

FNR (False Negative Rate): is the rate of network flows which the model incorrectly predicts is normal.

$$TNR = \frac{TN}{FP + TN}$$

### C. Experimental Results

Synthesizing APT attack malware data from data set that is divided by the ratio of 70% for training and 30% for testing and obtaining model. Tables IV and V describe the results of APT attack detection using the Random Forest algorithm.

TABLE IV. EXPERIMENTAL RESULTS OF C&C SERVER DETECTION MODEL TRAINING

Precision	Recall	Accuracy	TPR	FPR	FNR	TNR
0.925	0.998	0.955	0.998	0.095	0.002	0.905

TABLE V. EXPERIMENTAL RESULTS OF C&C SERVER DETECTION MODEL TESTING

Precision	Recall	Accuracy	TPR	FPR	TNR	FNR
0.9996	1	0.9998	0.9996	0.0004	1	0

Through the experimental results in Tables IV and V, we can see that C&C server training and detection model brings good results. This shows that the Random Forest classification algorithm and the features that are selected and extracted in the paper have brought good effect. In particular, the experimental part of detecting C&C server is absolutely accurate, which showed that the training model created a very good model for detection. Therefore, from the experimental results in this paper, we can see that the features, which represent the abnormal connection behaviors and are selected and proposed by us, present exactly the difference between normal connections and APT connections. This is very important because most APT attacks will be difficult to detect without the events-stringing system.

### V. CONCLUSION AND FUTURE DIRECTION

The APT attack has been and will be a dangerous attack and a challenge to information security systems. In this paper, based on the Random Forest machine learning algorithm and the unusual behavioral features of network traffic, we successfully detected and alerted C&C servers early. The results of this research can be used in intrusion detection and prevention systems to look for abnormal signs of the network. In the future, we will improve the features of network traffic in order to detect the signs of an APT attack when this attack uses encryption techniques to transmit information

#### REFERENCES

- [1] Adel Alshamrani; Ankur Chowdhary; Sowmya Myneni; Dijiang Huang, "A Survey on Advanced Persistent Threats: Techniques, Solutions, Challenges, and Research Opportunities" IEEE Communications Surveys & Tutorials, 2019, 1, 1–29.

- [2] Eric Code, "Advanced Persistent Threat. Understanding the Danger and How to Protect Your Organization," 1rd ed.; Elsevier, Amsterdam, 2012; 309.
- [3] Mirco Marchetti, Fabio Pierazzi, Michele Colajanni, Alessandro Guido, "Analysis of high volumes of network traffic for Advanced Persistent Threat detection," *Computer Networks*. 2016: 109; 127-141.
- [4] Combating Advanced Persistent Threats – "How to prevent, detect, and remediate APTs", Tech. rep," McAfee Inc. (2011).
- [5] Ivo Friedberg, Florian Skopik, Giuseppe Settanni, Roman Fiedler. "Combating advanced persistent threats: From network event correlation to incident detection," *Computers & Security*, 2015, Volume 48, Pages 35-57.
- [6] Andrew, V, "Flow based analysis of Advanced Persistent Threats detecting targeted attacks in cloud computing," In: 2014 First International Scientific-Practical Conference Problems of Info Communications Science and Technology, pp 173-176, Kharkov, Ukraine, 14-17 Oct. 2014.
- [7] Weina, N., Xiaosong, Z., GuoWu, Y., Jianan, Z., Zhongwei, R, "Identifying APT Malware Domain Based on Mobile DNS Logging" . *Mat. Pro. in. Eng.* 2, 1- 9 (2017).
- [8] Zhao, G., Xu, K., Xu, L., Wu, B, "Detecting APT malware infections based on malicious DNS and traffic analysis," *IEEE Access*. 3, 1132–1142 (2015).
- [9] Do Xuan Cho, Ha Hai Nam, "A Method of Monitoring and Detecting APT Attacks Based on Unknown Domains,". *Pro. Com. Sci.* 150, 316-323 (2019).
- [10] The Radicati Group, Inc. Advanced Persistent Threat (APT) Protection-Market Quadrant 2018. An Analysis of the Market for APT Protection Solutions Revealing Top Players, Trail Blazers, Specialists and Mature Players. February 2018. Pp48.
- [11] Jisang Kim, Taejin Lee, Hyung-guen Kim, Haeryong Park, "Detection of Advanced Persistent Threat by Analyzing the Big Data Log," *Advanced Science and Technology Letters Vol.29 (SecTech 2013)*, pp.30-36.
- [12] Sung-Hwan Ahn; Nam-Uk Kim; Tai-Myoung Chung, "Big data analysis system concept for detecting unknown attacks," In the 16th International Conference on Advanced Communication Technology. 16-19 Feb. 2014. Pyeongchang, South Korea.
- [13] Hyunjo Kim, Jonghyun Kim, Ikkyun Kim, Tai-myung Chung, "Behavior-based anomaly detection on big data," In the 13th Australian Information Security Management Conference, held from the 30 November – 2 December, 2015, Edith Cowan University Joondalup Campus, Perth, Western Australia. pp. 73-80.
- [14] Advanced Persistent Threat (APT) Protection - Market Quadrant 2019. An Analysis of the Market for APT Protection Solutions. The Radicati Group, Inc. April 2019. Pp53. <https://www.symantec.com/content/dam/symantec/docs/reports/2019-radicati-apt-protection-market-quadrant-en.pdf> [access date 3/1/2020]
- [15] Sofacy Attacks Multiple Government Entities. <https://unit42.paloalto-networks.com/unit42-sofacy-attacks-multiple-government-entities/> [access date 3/1/2020]
- [16] MirageFox: APT15 Resurfaces With New Tools Based On Old Ones. <https://intezer.com/blog/research/miragefox-apt15-resurfaces-with-new-tools-based-on-old-ones/> [access date 3/1/2020]
- [17] Leo Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5- 32, 2001.
- [18] Thomas G. Dietterich, "Ensemble Methods in Machine Learning", in *Proceedings of the International Workshop on Multiple Classifier Systems*. (MCS 2000), pp 1-15, Cagliari, Italy, 21–23 June 2000.
- [19] R. Markus., et al., "A survey of network-based intrusion detection data sets, ". *Computers & security*, vol. 8, no. 6, pp. 147–167, 2019.
- [20] Cho Do Xuan, Hoa Dinh Nguyen and Tisenko Victor Nikolaevich, "Malicious URL Detection based on Machine Learning" *International Journal of Advanced Computer Science and Applications(IJACSA)*, 11(1), 2020. <http://dx.doi.org/10.14569/IJACSA.2020.0110119>.
- [21] The CTU-13 Dataset. A Labeled Dataset with Botnet, Normal and Background traffic. <https://www.stratosphereips.org/datasets-ctu13>. [access date 3/1/2020].
- [22] APTNotes - Github Repo. <https://github.com/kbandla/APTnotes>. [access date 3/1/2020].
- [23] APTNotes - Website <https://aptnotes.malwareconfig.com/> Targeted. [access date 3/1/2020].
- [24] Cyber Attacks Logbook (Kaspersky) <https://apt.securelist.com/>. [access date 3/1/2020].
- [25] DeepEnd Research: List of malware pcaps, samples, and indicators for the Library of Malware Traffic Patterns. <https://contagiodump.blogspot.com/2013/08/deepend-research-list-of-malware-pcaps.html>. [access date 3/1/2020].
- [26] Malware dump. <https://contagiodump.blogspot.com/>. [access date 1/4/2020]

# An Abnormal Behavior Detection Method using Optical Flow Model and OpenPose

Zhu Bin<sup>1</sup>, Xie Ying<sup>2</sup>, Luo Guohu<sup>3</sup>, Chen Lei<sup>4</sup>  
School of Mechanical and Electronic Engineering  
Jiangxi College of Applied Technology  
Ganzhou, China

**Abstract**—Abnormal behavior detection and recognition of pedestrian in escalator has always been a challenging task in intelligent video surveillance system. To cope this problem, a method combining optical flow vector of passenger with human skeleton extraction is proposed. At first, adaptive dual fractional order optical flow model is used to estimate the optical flow field under scenes with illumination changes, low contrast and uneven illumination. At the same time, the OpenPose deep convolutional neural network is used to extract body skeleton and persons in image can be located. Then, the optical flow field and the human skeleton are combined to obtain the optical flow vector of the passenger head. After that the optical flow field of the passenger head and the step of escalator under the passenger foot are used for abnormal behavior detection and recognition, random forest is employed to behavior classifier. Experimental results show that our proposed method and its improvement strategy can accurately estimate the optical flow field in real time of low contrast outdoor videos with insufficient illumination, uneven brightness and illumination changes, the accuracy of abnormal action detection and recognition can reach to 97.98% and 92.28%.

**Keywords**—Image sequence analysis; abnormal behavior recognition; fractional order variational optical flow model; random forest

## I. INTRODUCTION

With the rapid development of machine vision, artificial intelligence and other related technologies, intelligent video monitoring system has been widely used in supermarket, campus, medical treatment, government and other public places, which plays a very important social value in improving people's living standards and safeguarding the safety of life and property. Abnormal behavior recognition is one of the core tasks of most intelligent video monitoring systems. It is an effective means of intelligent monitoring the public places in real time.

The traditional abnormal behavior recognition methods can be divided into three steps: target detection, target tracking and behavior recognition. Target detection is to find the interested target methods in image. When the interested target is the escalator passenger, some simple geometric models can be used to match the human body, such as skeleton model [1-2], two-dimensional human body model [3-4] and three-

dimensional human body model [5-6]. Target tracking is to track the interested moving target in real time. Behavior recognition is based on the motion characteristics of the interested area to identify whether its behavior is abnormal. In literature [7], a feature detection method based on dense trajectory descriptors is proposed; in literature [8], a human behavior recognition method based on Fisher vector is proposed; in literature [9], a hierarchical image segmentation method based on time and space is used. However, these methods are easily affected by scene changes, feature selection and feature extraction. In the outdoor moving escalator scene, the scene is particularly affected by the change of light, the difference of light intensity in different time periods is very large, and the difference of light intensity in different elevator positions in the same time period is also very large, at the same time, the target features are easy to be occluded. At present, there is no high-precision behavior recognition method in the moving escalator scene yet.

The neural network based method is a research hotspot in recent years. This kind of method only needs to design neural network according to the input image and its characteristics, and then carry out model training by the training samples labeled abnormal behaviors, after the neural network finished training, we can get whether the image contains abnormal behaviors and the types of abnormal behaviors for each input image. In [10], the image sequence is divided into time stream and space stream, which are processed by different convolution neural networks, and then the results are combined; in [11], a two-layer convolution neural network based on human body region is proposed to distinguish human behavior; in [12-13], different convolution neural networks are designed for tools and actions of pedestrian hands in the scene. However, the training of the network model usually needs the support of a large number of data samples. In the scene of outdoor moving escalator, the collection of abnormal behavior samples such as falling is particularly complex, and due to the large difference of light intensity in different time periods and different elevator positions in the same time period, the required number of training samples is particularly large, even up to tens of millions of levels. But so far, there is no database based on the abnormal behavior recognition for the scene of moving escalator.

Aiming at the problem of abnormal behavior recognition for escalator passenger, we apply adaptive fractional variational optical flow model and OpenPose model to generate a new method. In this algorithm, adaptive fractional order variational optical flow model is used to solve the problem of feature estimation in the scene of illumination change, low contrast and uneven brightness. OpenPose model is used to track escalator passengers in real time, and random forest classifier is used for feature classification and recognition. The advantages of the algorithm are: only extract the optical flow features of a key point that is not easy to be occluded; use image sequence instead of a single image for feature classification; only a small number of samples can complete the training of the classifier.

The innovation of this paper is as follows:

- 1) A method of identifying the abnormal behavior of the passengers in the walking elevator in the outdoor scene is proposed;
- 2) The continuous optical flow feature of image sequence is used to distinguish the abnormal behavior of the passengers in the walking elevator.

The rest of the work is listed as follow: Section 2 shows the flowchart of the algorithm; Section 3 introduces the optical flow model used in this paper; Section 4 presents the skeleton extraction method; and Section 5 is the experimental results and analysis. Section 6 concludes the paper.

## II. FLOWCHART OF THE ALGORITHM

Fig. 1 shows the flow chart of this paper's abnormal behavior detection and recognition algorithm: first, obtain the input video, then extract image sequences from input video, and then carry out Gaussian filtering, light compensation and other preprocessing for each image, and then carry out the optical flow field estimation through the adaptive fractional variational optical flow model; at the same time, use the depth neural network based on OpenPose to extract the human skeleton of passengers in each frame of the image; and combining the human skeleton and the optical flow field at the same position, the velocity and direction of the passenger's nose joint are calculated. Finally, according to the characteristics of the optical flow vector of the passenger's nose joint and the escalator steps, the random forest classifier is used to distinguish whether the passenger's behavior is abnormal, and the abnormal behavior is classified. Because the optical flow vector of the same moving speed pixel point closer to the camera position is larger, the optical flow vector of the key point must be standardized before use. In this paper, the ratio of the vertical distance from the neck joint point of the human skeleton to the connecting line of the left hip and the right hip is taken as the reference standard to standardize the optical flow vector of the key point.

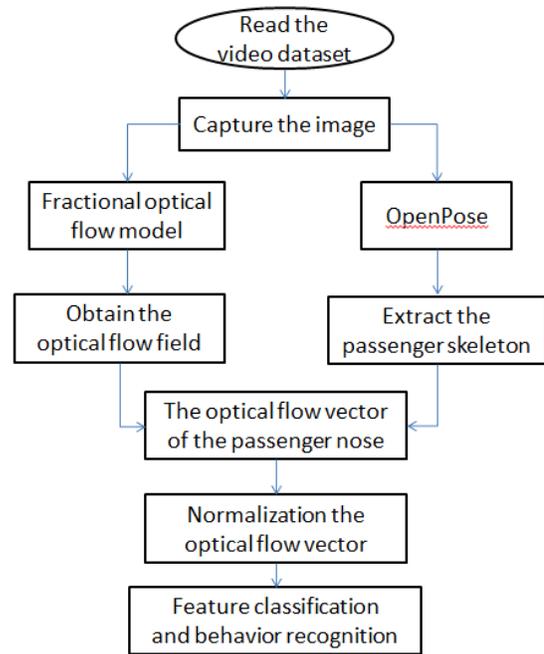


Fig. 1. The Flowchart of the Algorithm.

## III. ADAPTIVE FRACTIONAL OPTICAL FLOW MODEL

The optical flow model is based on DFOVOFM (Dual Fractional Order Variational Optical Flow Model) [14]. In order to enhance the correlation of similar optical flow areas and improve the accuracy of optical flow estimation, the adaptive strategies of fractional order and fractional differential mask are added into DFOVOFM. DFOVOFM is the fractional version of HS (horn Schunck) optical flow model [15], that is, the data and smooth terms of the integral derivative in the original model are replaced by fractional derivative:

$$E(\mathbf{u}) = \int_{\Omega} ((D_x^\alpha I \cdot u^\alpha + D_y^\alpha I \cdot v^\alpha + D_t^\alpha I)^2 + \lambda(|D_x^\beta u|^2 + |D_y^\beta u|^2 + |D_x^\beta v|^2 + |D_y^\beta v|^2)) d\mathbf{x} \quad (1)$$

Where  $E(\mathbf{u})$  represents the energy function of the optical flow model,  $\mathbf{u} = (u, v)$  represents the vector of the optical flow field,  $u, v$  represents the components of the optical flow field in the axial  $x, y$ , respectively,  $D_x^\alpha I$  is the fractional derivative of the brightness function  $I$  in the axial  $x$ , and  $D_y^\alpha I$ ,  $D_t^\alpha I$ ,  $D_x^\beta u$ ,  $D_y^\beta u$ ,  $D_x^\beta v$ ,  $D_y^\beta v$  is the fractional derivative of the function in the corresponding axis,  $\lambda$  is the smoothing parameter,  $\Omega \subset R^3$ , and represents the spatial range of calculating the fractional derivative of the target point  $(i, j, t)$ .

The above model can rewrite using the convolution form of fractional order differential mask and optical flow vector or brightness function:

$$W_1(\alpha) * (I * u_k^{2\alpha-1}) + W_2(\alpha) * (I * u_k^{\alpha-1} * v_k^\alpha) + W_3(\alpha) * (I * u_k^{\alpha-1}) + \lambda W_2(\beta) * u_{k+1} = 0 \quad (2)$$

$$W_5(\alpha) * (I * v_k^{2\alpha-1}) + W_4(\alpha) * (I * v_k^{\alpha-1} * u_k^\alpha) + W_8(\alpha) * (I * v_k^{\alpha-1}) + \lambda W_4(\beta) * v_{k+1} = 0 \quad (3)$$

Where,  $W_1(\alpha)$ ,  $W_2(\alpha)$ ,  $W_4(\alpha)$ ,  $W_5(\alpha)$ ,  $W_7(\alpha)$ ,  $W_8(\alpha)$  represent the dual fractional order differential masks in six different coordinates, respectively.  $\alpha$ ,  $\beta$  represent the order of the fractional order. The concept and derivation process of the dual fractional order differential mask can be found in reference [16],  $I$  and  $u_k^{2\alpha-1}$  are both a  $m * n$  spatial matrix, we let " $\cdot$ " represents the multiplication of the corresponding position elements in the two matrices, " $*$ " represents the convolution operation,  $u_k$ ,  $v_k$  is the historical value of the optical flow vector,  $u_{k+1}$ ,  $v_{k+1}$  is the value of the optical flow to be solved, and  $k$  is the number of iterations.

In the calculation process, the image SNR is used to calculate the fractional order and the size of the fractional differential mask of each pixel point. The obtained optical flow field is segmented by super-pixel, and then the shape of the fractional differential mask is adjusted. Each fractional differential mask is limited in the same super pixel area. The optical flow field is recalculated by the adjusted fractional differential mask. After repeated for 5-10 times, the optical flow field with distinct contour and minimum error rate can be obtained. The detailed process can be found in reference [16].

#### IV. EXTRACTION OF PASSENGER SKELETON

Accurate, real-time and stable tracking and positioning of escalator passengers in complex scenes is the prerequisite for subsequent abnormal behavior detection and recognition. The human body model is the basis of recognizing the human body position in the image. Compared with other human body models [3], the human skeleton features can directly reflect the contour and position of the human body, and the probability of misjudging other targets as human body is small, so the human skeleton is selected to represent the escalator passengers. Compared with the traditional image processing or machine learning methods, the method based on deep learning can extract human skeleton more accurately and stably. It obtains skeleton extraction network through iterative training, processes the input original image through the network, and finally outputs skeleton extraction results. With the continuous development of deep learning methods, skeleton extraction network has been improved and put forward one after another. Among them, the deep convolution neural network based on OpenPose [17] is a kind of extraction method which can extract human skeleton in real time and accurately and stably.

It has gradually become the current mainstream skeleton extraction method and is widely used in the engineering field.

In this paper, the method of literature [18] is used to extract the two-dimensional skeleton of escalator passengers. In this method, firstly, through the trained OpenPose model, we can accurately and stably detect the human body joints and the bones between joints in the complex escalator environment, and then use the part affinity fields (PAFS) to associate the joints of the human body, thus forming the two-dimensional skeleton of the human body. The human skeleton consists of 14 joint points and 13 human bones. The joint points are nose, neck, left shoulder, left elbow, left wrist, left hip, left knee, left ankle, right shoulder, right elbow, right wrist, right hip, right knee and right ankle. Fig. 2 is the sketch map of human skeleton extraction: Fig. 3 is the human skeleton model, in which the red dot represents 14 joint points of the human body, and the 13 green segments represent the human skeleton; Fig. 3(a) is the human skeleton extraction result of the escalator passenger; Fig. 3(b) deleted other image contents, only the human skeleton of the escalator passenger is retained.

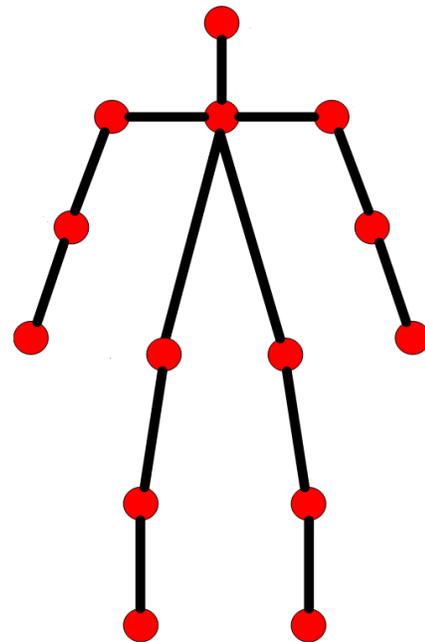


Fig. 2. The Diagram of Human Skeleton.

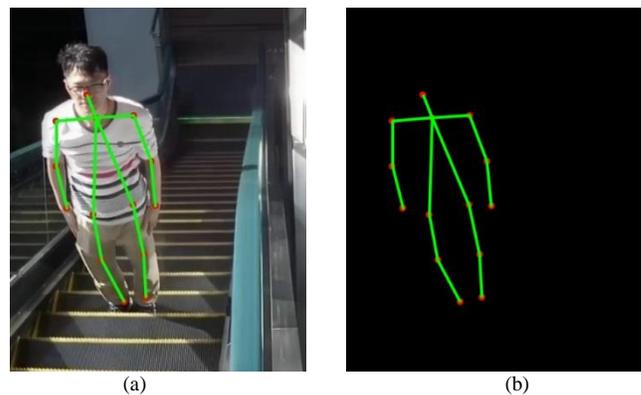


Fig. 3. The Diagram of Skeleton Extraction.

## V. EXPERIMENTAL RESULTS AND ANALYSIS

The experiment is divided into several parts. First, the experiment of estimating the optical flow field of the adaptive fractional order optical flow model is carried out to verify the superiority of the optical flow model in the light abnormal environment and its practicability in the outdoor scene. Second, the relationship between the passenger's behavior and its optical flow vector is explained by demonstrating the optical flow field of the passenger's different behavior. Third, the application of the random forest classifier [19-20] is beneficial. This paper classifies the passenger behavior with the characteristics of optical flow to verify the practicability of the algorithm. All the experiments are run in MATLAB. The selected platform is win 7, Intel 3.3 GHz, 16 GB memory.

### A. The Result of Adaptive Fractional Optical Flow Model

MPI Sintel (scene with insufficient light and low contrast), Kitti (scene with inconsistent light) and outdoor image sequence (scene with inconsistent light and insufficient light) are selected for algorithm evaluation.

The experimental results are shown in Fig. 4. Different letters represent different algorithms, a) is Hast, b) is MDP\_Flow, c) is PH\_Flow, d) is DFMVOFM; the image sequences include: MPI\_Sintel\_cave3, MPI\_Sintel\_shaman1, KITTI; Fig. 5 is the optical flow field of outdoor image sequence. It can be found from the figure that the method in this paper can get clear moving object contour in the abnormal illumination scene, but some texture details will be lost, such as the sharp cone on the girl's hand in cave3; Hast is better than other algorithms in highlighting moving details, but in the large non texture area, the object contour is incomplete, such as the tail of the monster in cave3, the sleeve contour in shaman1; PH-flow will have a big deviation when the light is insufficient; MDP-flow will fail in the low resolution area. However, in the task of abnormal behavior recognition, it is necessary to obtain the complete contour of the target in the case of abnormal light, and do not need to care about the texture details of the target. Compared with the Hast, MDP\_Flow and PH\_Flow with higher average accuracy, the optical flow model selected in this paper is more suitable for escalator riding customer's abnormal behavior identification task.

### B. The Optical Flow Feature for Different Passenger Feature

Fig. 6, Fig. 7, Fig. 8 and Fig. 9 show the color coding diagram and human skeleton extraction diagram of the optical flow field of normal passenger riding, passenger retrograde, passenger falling forward and passenger falling backward, respectively. In order to avoid the transient state of slight shaking of passengers, which is also regarded as abnormal behavior, four frame images, namely three consecutive optical flow features, are used as the basis for abnormal behavior discrimination. The color coding graph of optical flow field uses different colors to represent the direction of optical flow vector, and uses the depth of color to represent the size of optical flow vector.

When the passenger stay on the escalator normally, the optical flow field of the passenger is the same as that of the escalator step under his/her feet, as shown in Fig. 6; when the

passenger is retrograde, the optical flow field of the passenger is opposite to that of the escalator step under his /her feet, in Fig. 7, the color of the passenger is purple, while the color of the escalator step under his / her feet is green and light yellow.

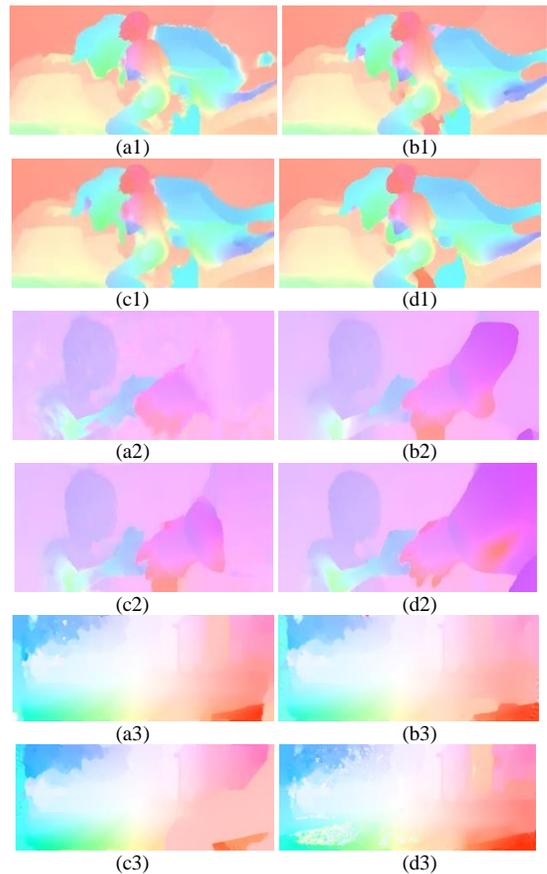


Fig. 4. The Optical Flow Field of Different Algorithms in MPI\_Sintel and KITTI.

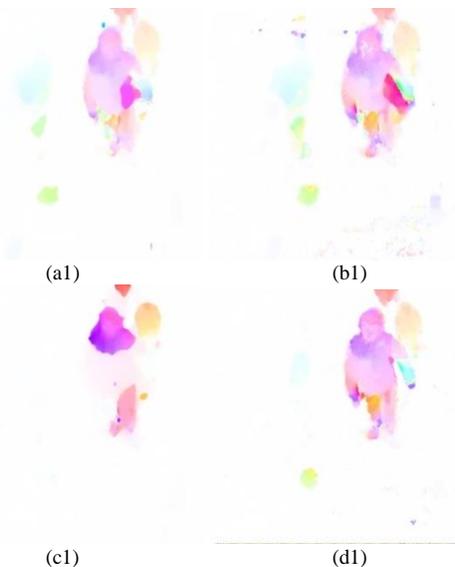


Fig. 5. The Optical Flow Field of Different Algorithms in Outdoor Video.

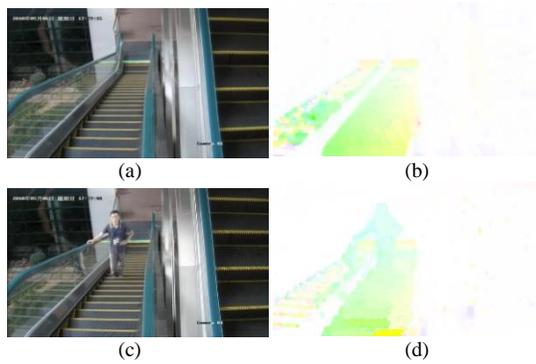


Fig. 6. The Optical Flow Field when the Escalator in Ballast and the Passenger Legitimate using

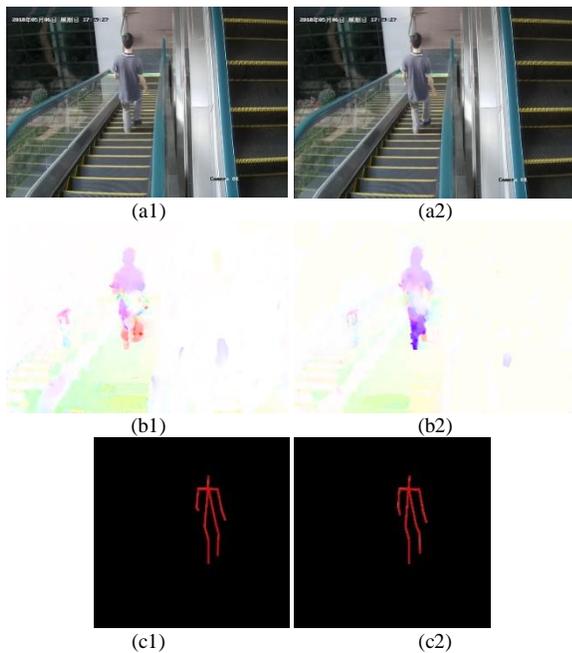


Fig. 7. The Optical Flow Field of Passenger Retrograde and the Diagram of Skeleton.

Fig. 8 shows the characteristics of light flow when the passenger falls ahead. It can be found that the color of the passenger is dark yellow, while the color of the escalator step under his feet is green or light yellow. This is because the movement direction of the passenger is not different from the reverse direction of the escalator, but the movement speed is faster. The fall in the experiment is a simulated fall, and the movement speed of the real fall is faster. Fig. 9 shows the light flow characteristics when the passenger falls back. It can be found that the color of the passenger is red, while the color of the escalator steps under his feet is green and light yellow. Therefore, the experiment shows that different behaviors of passengers correspond to different characteristics of optical flow of passengers, and the difference of optical flow characteristics between passengers and escalators can be used for detection and recognition of abnormal behaviors of passengers.

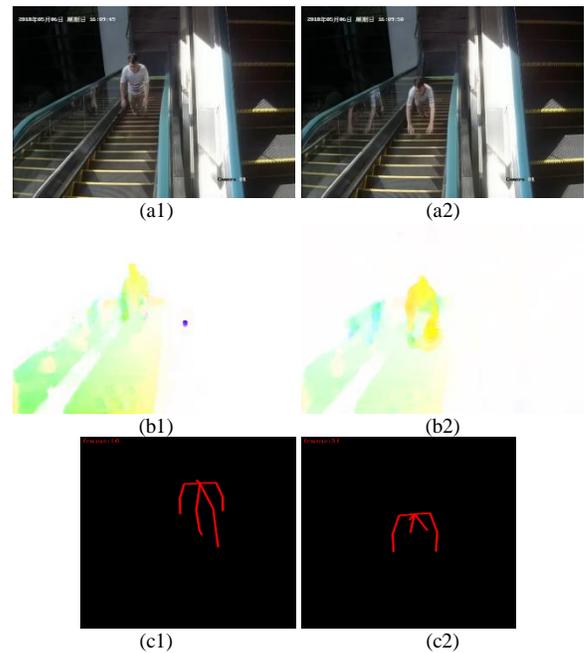


Fig. 8. The Optical Flow Field of Passenger Fall-Ahead and the Diagram of Skeleton.

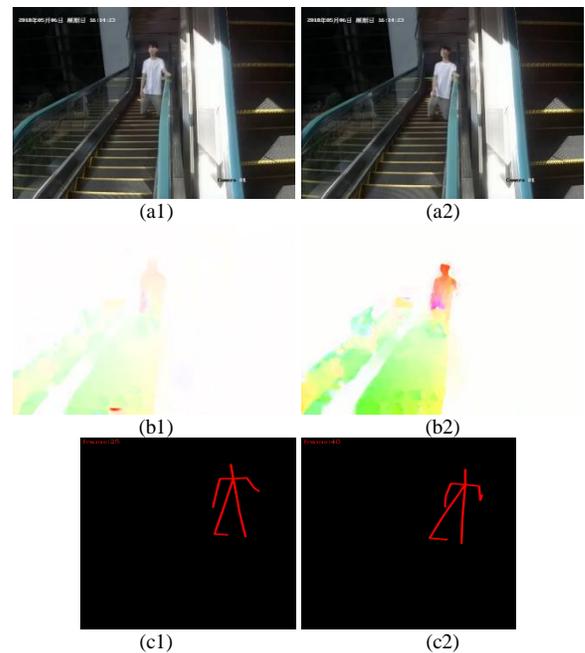


Fig. 9. The Optical Flow Field of Passenger Fall-Behind and the Diagram of Skeleton.

### C. The Classification and Recognition of Abnormal Behavior

1) *The dataset:* In view of the escalator abnormal behavior detection task, the escalator passenger behavior data set is built. The data set consists of continuous image sequence and test video. The image sequence comes from the interval sampling of escalator monitoring video, which is used to train the classification model; the test video is the video segment with key information intercepted from the monitoring

video, which is used to test the detection and recognition of abnormal behavior.

The dataset contains 4000 images and 48 videos. Images are used to train the classification model, in which every four images represent one behavior and construct 1000 passenger behavior sequences, including 100 normal behavior sequences and 900 abnormal behavior sequences. The normal behavior sequence includes passenger walking and riding, passenger standing and riding, slight head up, down, left and right shaking, passenger squatting, etc.; the abnormal behavior includes retrograde, forward and backward falling. The 48 videos include zcdc1-zcdc8, nx1-nx20 and sd1-sd20. Among them, zcdc1-zcdc2, nx1-nx5, sd1-sd5 are the videos intercepted under sufficient and uniform illumination, which are collectively referred to as PA1; zcdc3-zcdc4, nx6-nx10, sd6-sd10 are the videos intercepted under sufficient and uneven illumination, which are collectively referred to as pa2; zcdc5-zcdc6, nx10-nx15, sd10-sd15 are the videos intercepted under insufficient and uniform illumination, which are collectively referred to as PA3; zcdc7-zcdc8, nx16-nx20, sd16-sd20 is the video captured in the scene with insufficient illumination and uneven illumination, collectively known as PA4. All images and videos are captured in the scene of sparse escalator passengers, that is, there is no mutual occlusion between passengers in the image.

2) *The process of classification:* First, the training database is established. The elements in the database are composed of the category code of the behavior sequence and the optical flow vector of the nose joint points and the foot step pixel points of three consecutive images in the behavior sequence. Then, the training database is input into the classification model to train the classification model. In this paper, random forest is selected as the classification model. Finally, the trained classification model is tested in 48 videos.

3) *Experimental results and analysis:* The performance of the algorithm is evaluated by using general indicators, which are precision, recall, F1 score and time. The unit of detection speed is FPS (frame per second).

Table I shows the test results in four different scenarios. Among them, sufficient, uniform and sparse represent whether the lighting in the scene is sufficient (a), whether the lighting is uniform (b). It can be found from the table that the algorithm in this paper has achieved high accuracy in four kinds of videos, which shows that the algorithm can accurately detect abnormal behavior when the passenger flow is sparse. In addition, the decrease of harmonic mean value caused by insufficient and uneven illumination is 1.17% and 0.81% respectively, which shows that the algorithm is robust to the change of environmental factors. The average processing speed of the algorithm in this paper is around 16 FPS, and the real-time detection speed of the algorithm is higher than 20 FPS, so the algorithm in this paper needs to adopt the method of interlace sampling to meet the real-time requirements.

In abnormal behavior recognition, it is easy to have some wrong classification, mainly because of the intersection of different types of feature space. For example, when a passenger falls backward or retrograde, the light flow field at the nose

joint of the passenger is red, the difference is only that the red is deeper when the passenger falls backward; and when the passenger squats down, it will mix with the passenger's falling forward. This is due to the limitations of experimental conditions, the training samples of falls are simulation rather than real falls; the movement speed of real falls is much faster than that of simulation, and the number of training samples is also an important factor affecting the accuracy of recognition. Table II shows the results of abnormal behavior recognition. The abnormal behaviors were divided into retrograde (NX), backward (FB), forward (FA) and other abnormal behaviors. From the table, it can be found that in the sparse crowd scenario, the accuracy of the algorithm in this paper for all kinds of behavior recognition is more than 90%, which shows that the algorithm can accurately identify abnormal behavior when the passenger flow is sparse.

TABLE I. THE PERFORMANCE INDEX FOR THE ABNORMAL BEHAVIOR DETECTION OF ELEVATOR PASSENGER

video	(a)	(b)	PR%	RE%	F1%	TI/fps
PA1	√	√	98.76	100.0	99.4	15.63
PA2	√	×	98.13	99.33	98.52	16.32
PA3	×	√	97.59	99.11	97.96	15.72
PA4	×	×	97.44	98.62	97.48	16.04
AVE			97.98	99.26	98.36	15.92

TABLE II. THE PERFORMANCE INDEX FOR THE ABNORMAL BEHAVIOR RECOGNITION OF ELEVATOR PASSENGER

Abnormal behavior	PR%	RE%	F1%
NX	92.46	93.42	93.16
FB	93.73	92.64	93.34
FA	92.58	91.59	92.18
Others	90.37	91.26	92.95
AVE	92.28	92.23	92.91

## VI. CONCLUSION

In this paper, an algorithm of escalator passenger abnormal behavior recognition based on adaptive fractional variational optical flow model and OpenPose model is proposed, which solves the problem of escalator passenger abnormal behavior recognition under outdoor and sports background. Firstly, the adaptive fractional variational optical flow model is used to estimate the optical flow field in the scene of illumination change, low contrast and uneven brightness, so as to obtain the motion speed and direction of each pixel in the video image; at the same time, the human skeleton extraction method based on OpenPose is applied to solve the passenger positioning problem in the complex scene; then the optical flow field and human skeleton are combined. Finally, according to these two characteristics of optical flow field, a random forest classifier is used to detect and identify the abnormal behavior. Experiments show that this method can detect the abnormal behavior of escalator passengers in real time in outdoor scenes, and has important application value. Future work would focus on recognizing other abnormal behavior and the construction of a public dataset with different abnormal behaviors in moving escalator scenes.

REFERENCES

- [1] Y. Du, Y. Fu and L. Wang, "Representation Learning of Temporal Dynamics for Skeleton-Based Action Recognition," in IEEE Transactions on Image Processing, vol. 25, no. 7, pp. 3010-3022, July 2016.
- [2] R. Li, H. Fu, W. Lo, Z. Chi, Z. Song and D. Wen, "Skeleton-Based Action Recognition With Key-Segment Descriptor and Temporal Step Matrix Model," in IEEE Access, vol. 7, pp. 169782-169795, March, 2019.
- [3] Y. Li, T. Guo, R. Xia and X. Liu, "A Novel Skeleton Spatial Pyramid Model for Skeleton-based Action Recognition," 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 2019, pp. 16-20.
- [4] S. Win and T. L. L. Thein, "Real-Time Human Motion Detection, Tracking and Activity Recognition with Skeletal Model," 2020 IEEE Conference on Computer Applications (ICCA), Yangon, Myanmar, 2020, pp. 1-5.
- [5] P. Fechteler, L. Kausch, A. Hilsmann and P. Eisert, "Animatable 3D Model Generation from 2D Monocular Visual Data," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 2018, pp. 560-563.
- [6] H. Jiang, J. Cai and J. Zheng, "Skeleton-Aware 3D Human Shape Reconstruction From Point Clouds," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 5430-5440.
- [7] Y. Mao, Y. Shen, D. Chao, et al. "Real-Time Simultaneous Pose and Shape Estimation for Articulated Objects Using a Single Depth Camera," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 8, pp. 1517-1532, June 2016.
- [8] V. Morariu, D. Harwood, L. Davis. "Tracking people's hands and feet using mixed network and/or search," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 5, pp. 1248-1262, 2013.
- [9] C. Qiuhui, Z. Chongyang, L. Weiwei, et al. "SHPD: Surveillance Human Pose Dataset and Performance Evaluation for Coarse-Grained Pose Estimation," in 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 2018, pp. 4088-4092.
- [10] W. Nie, W. Wang and X. Huang, "SRNet: Structured Relevance Feature Learning Network From Skeleton Data for Human Action Recognition," in IEEE Access, vol. 7, pp. 132161-132172, 2019.
- [11] Y. Zhao, L. Wen, S. Li, H. Cheng and C. Zhang, "A View-invariant Skeleton Map with 3DCNN for Action Recognition," 2019 Chinese Automation Congress (CAC), Hangzhou, China, 2019, pp. 2128-2132.
- [12] Z. Zhigang, D. Guangxue, L. Huan, Z. Guangbing, W. Nan and Y. Wenjie, "Human behavior recognition method based on double-branch deep convolution neural network," 2018 Chinese Control And Decision Conference (CCDC), Shenyang, 2018, pp. 5520-5524.
- [13] R. Zhang and B. Ni, "Learning Behavior Recognition and Analysis by Using 3D Convolutional Neural Networks," 2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST), Luang Prabang, Laos, 2019, pp. 1-4.
- [14] Zhu Bin, Tian Lianfang, Du Qiliang, et al. An Improved Fractional order optical flow model for motion estimation [J/OL]. Mathematical Problems in Engineering, <https://www.hindawi.com/journals/mpe/2018/6278719>.
- [15] W. Wei, B. Liu, Z. Pan and Z. Wang, "A Simplified HS Algorithm in Optical Flow Estimation," 2016 3rd International Conference on Information Science and Control Engineering (ICISCE), Beijing, 2016, pp. 99-104.
- [16] Zhu Bin, Tian Lianfang, Du Qiliang, et al. Adaptive dual fractional-order variational optical flow model for motion estimation[J]. IET Computer Vision, 2019, 13(3):277-284.
- [17] V. Belagiannis, S. Amin, M. Andriluka, et al. "3D Pictorial structures revisited: multiple human pose estimation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 38, no. 10, pp. 736-742, 2016.
- [18] C. Zhe, T. Simon, W. Shihien, et al. "OpenPose: Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 36, no. 8, pp. 736-742, 2019.
- [19] W. Yisen, X. Shutao, T. Qingtao, et al. "A Novel Consistent Random Forest Framework: Bernoulli Random Forests," in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 8, pp. 3510-3523, 2018.
- [20] H. Yi, Q. Xiong, Q. Zou, R. Xu, K. Wang and M. Gao, "A Novel Random Forest and its Application on Classification of Air Quality," 2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI), Toyama, Japan, 2019, pp. 35-38, doi: 10.1109/IIAI-AAI.2019.00018.

# Clash between Segment-level MT Error Analysis and Selected Lexical Similarity Metrics

Marija Brkic Bakaric<sup>1</sup>, Kristina Tonkovic<sup>2</sup>, Lucia Nacinovic Prskalo<sup>3</sup>

Department of Informatics  
University of Rijeka  
Rijeka, Croatia

**Abstract**—The aim of this paper is to evaluate the quality of popular machine translation engines on three texts of different genre in a scenario in which both source and target languages are morphologically rich. Translations are obtained from Google Translate and Microsoft Bing engines and German-Croatian is selected as the language pair. The analysis entails both human and automatic evaluation. The process of error analysis, which is time-consuming and often tiresome, is conducted in the user-friendly Windows 10 application TREAT. Prior to annotation, training is conducted in order to familiarize the annotator with MQM, which is used in the annotation task, and the interface of TREAT. The annotation guidelines elaborated with examples are provided. The evaluation is also conducted with automatic metrics BLEU and CHRF++ in order to assess their segment-level correlation with human annotations on three different levels—accuracy, mistranslation, and the total number of errors. Our findings indicate that neither the total number of errors, nor the most prominent error category and subcategory, show consistent and statistically significant segment-level correlation with the selected automatic metrics.

**Keywords**—Machine translation; evaluation; error analysis; BLEU; CHRF++; MQM

## I. INTRODUCTION

Machine translation (MT) is used on a daily basis by millions of people and for a range of use cases [1]. Although it will not replace humans any time soon, it can be used as a tool to enhance productivity [2]. Different types of data mean variations in structure, genre, and style, and can result in MT outputs of quite different quality. In order to properly evaluate MT, all data features important for the future use of the translation system need to be covered by the evaluation set, which is comprised of a set of sentences in the source language and their target language translations [2]. After Neural Machine Translation (NMT) took over the world scene from its predecessor phrase-based statistical MT (SMT), there have been a lot of research initiatives which focus on translation error types in an attempt to better describe differences between these two approaches. However, a standard in assessing translation quality does not exist since MT quality evaluation is subjective in its nature and quality depends on the context [3]. A review of translation quality definitions is given in [4].

Unlike the usual automatic and human evaluation metrics which provide only quantitative evaluation, error analysis enables assessing translation in qualitative terms [5]. It refers to the identification and classification of individual errors in a

translated text [6]. Not only that it can reveal strengths and weaknesses of MT engines [6], but it can also show whether a system is superior over any other system regarding one aspect, all aspects, or a subset of them [7]. Multidimensional quality measure (MQM) lists quality issue types which can be used for defining specific metrics for annotation tasks and quality assessment [8]. It is used for evaluating both human and MT translations and, as such, represents a way of connecting the two. A detailed error taxonomy compliant with the hierarchical listing of issue types defined as part of the MQM which is relevant to Croatian is presented in [9]. Translation Error Annotation Tool (TREAT), which employs MQM, is described and tested in [10]. It is worth noting that human evaluation using non-directly expressed judgment-based (non-DEJ-based) metrics is more objective than that of using DEJ-based metrics, and less prone to indirect comparisons of previously assessed segments [4].

A number of metrics for automatic evaluation have been proposed up to date. These metrics are generally benchmarked against manual judgments in terms of system and segment-level correlation. This is typically done in the task of ranking various MT system translations for the same source segment [11]. New automatic MT evaluation metrics constantly emerge. Moreover, there is a metrics-shared task, which is held annually at the Workshop on MT (WMT), where new evaluation metrics are proposed. Metrics can be more or less reliable, depending on the target language, text type and genre, type of MT system, properties of human translation, and the quality aspect measured [11]. High cost and irreproducibility of human judgments can be resolved by automatic evaluation only if the latter matches human evaluation, as acknowledged in [12]. For a detailed overview of metrics and their advantages and disadvantages, we refer the reader to [4]. Metrics based on neural networks have lately shown great potential [14].

Since presenting all of the metrics and calculating their scores cannot be presented in a clear and concise way due to their vast number, only two metrics are selected from the reference-based class of metrics and chosen for the purpose of this research. These metrics always give the same score to the same text, given that all the evaluation parameters stay unaltered [4]. Both of them are based on the lexical similarity between machine translations and reference translations. While the first metric we employ is the *de facto* standard in MT community, i.e. Bilingual Evaluation Understudy (BLEU) [15], the second one – Character n-gram F-score (CHRF++),

shows promising results for morphologically rich languages [16], [17]. BLEU is a precision-based metric [15], which expresses lexical similarity on a 0-1 scale, 0 being the minimum score. While BLEU does not account for the recall directly, but through the brevity penalty, CHRF++ is a combination of precision and recall. Both of these metrics aim to achieve strong negative correlation with human error assessments, unlike error metrics which aim to achieve strong positive correlation with such human assessments. Reference translation-based metrics do require translators, and hence do not allow neither full automation of the process, nor cost and time minimization [4].

In this paper we examine correlation between error analysis results and the selected automatic metrics. The paper is organized as follows. Related work is given in Section 2, followed by the description of the research methodology in Section 3. Error analysis is presented in Section 4, results in Section 5, and discussion is provided in Section 6. Section 7 concludes the paper and gives directions for future work.

## II. RELATED WORK

The evaluation conducted in WMT 2017 examines system level correlation of metrics' scores and manual rankings and segment-level correlation with manual judgments in terms of direct assessment (DA) and UCCA-based MT evaluation (HUME) [12], where UCCA stands for universal conceptual cognitive annotation [13]. A subsequent evaluation in 2019 employs only DA [14]. The authors warn that the metrics results can be overly optimistic when the underlying set of MT systems comprises both well-performing and bad-performing systems. If the sampling of sentences does not provide sufficient number of assessments of the same segment, the evaluation tasks resort to a relative ranking re-interpretation of DA scores (DARR) [12], [14].

CHRF++ is selected for this study since it shows promising results for morphologically rich languages [16], [17]. Best CHRF correlations with human rankings are achieved for 6-grams, both on system and segment level [16]. The results in [17] show that apart from character n-grams, word 1-grams (CHRF+) and 2-grams (CHRF++) also correlate rather well with DAs. The results in [12] confirm that, on average, character-based metrics outperform other metrics. However, segment-level correlations are only around 0.4 or slightly above. BLEU is outperformed not only by character-based metrics, but also by the metric developed by the US National Institute of Standards and Technology (NIST) [18] and Translation Edit Rate (TER) [19] in the scenario where only one reference translation is used instead of the recommended four [15]. The authors in [2] see future efforts in MT evaluation directed toward character-based metrics which show the highest correlation with human judgments at both system and segment levels.

Due to inconsistencies in automatic metrics reported in [20], the study presented in this paper is conducted on three short texts. The authors in [20] base their findings on DA human evaluation of outputs of three different MT systems on three problematic domains. Human judges rate each translation on how adequately it expresses the meaning of the respective reference translation. Large differences are

observed in correlations between automatic metrics and human DA across different domains. A more complex corpus with longer sentences and more complex syntactic structures turns out to exhibit higher correlations between all automatic metrics and human judgments. Reference [11] shows that performance of metrics also significantly varies across different levels of MT quality. The correlations of all evaluation metrics, including BLEU and CHRF3, are substantially lower for low-quality MT output. Not only that metrics are not able to capture nuanced quality distinctions, but they perform poorly when faced with low-quality translations. Moreover, evaluating low-quality translations is challenging even for humans. In addition, metrics prove to be more reliable when evaluating neural MT, as opposed to statistical MT systems. The difference in the evaluation accuracy for different metrics is maintained even when the gold standard scores are based on different criteria.

As expected, there is not much work on MT evaluation involving both German and Croatian. Error analysis based on Vilar's taxonomy is conducted in [21] on the Croatian translation of an essay in German and on the German translation of the same essay in Croatian. Translations are generated by Google Translate (GT). The German-to-Croatian translation direction, which proves to be more difficult for MT, is assessed by two native speakers, while the other direction is assessed by a final year graduate student of German. Incorrect word proves to be the most frequent error type in both directions.

This paper examines correlation between error analysis results and the selected automatic metrics due to several reasons. Beside the fact that DA has already been thoroughly investigated, and that it is often not feasible to obtain multiple judgments for each segment, the correlation between error counts and automatic metrics has been poorly explored. We assess the performance of the selected evaluation metrics in terms of segment-level correlation with human error analysis. We opt for the segment-level evaluation since system-level evaluation is generally an easy task for MT evaluation metrics as the majority of metrics performs extremely well at ranking systems [11].

## III. METHODOLOGY

GT and Microsoft Bing engine are used for obtaining Croatian translations of German sentences, which are then annotated by a human annotator following the MQM Slavic tagset [9] and using the tool TREAT [10]. The authors in [4] call this type of human evaluation a non-DEJ-based evaluation as the judgment is not expressed directly in terms of "better than", i.e. ranking, or "good", i.e. direct assessment. Respective human translations are provided for reference.

Automatic metrics employed in the paper depend on the availability of human reference translations. Since they evaluate outputs of MT systems by comparing them to reference translations, they are also called reference translation-based metrics. BLEU is chosen, despite many of its drawbacks, as it is the *de facto* standard in MT community, and CHRF++ is chosen since it represents a very promising evaluation metric especially for morphologically rich target languages [17].

Pearson correlations between human metric segment-level scores, derived from the total number of errors and the number of errors assigned to the most frequent error category and subcategory, and automatic metric segment-level scores BLEU and CHRF++ are calculated. A stronger negative correlation indicates better performance.

The remainder of the section is divided into several subsections which give descriptions of the evaluation set, the tool used for the manual error analysis, the annotator, MQM issue types used, and the metrics BLEU and CHRF++ employed in the automatic evaluation.

### A. Evaluation Set

Our evaluation set consists of 54 sentences (Table I). This is admittedly a small sample. However, this type of task can quickly become tedious so we did not want to risk inconsistent evaluation or overlooking errors. Under annotation overload, one easily becomes too tired and thus less attentive. In general, when dealing with human evaluation, there is always a necessary trade-off between the size of the sample and the integrity of the results, as acknowledged by [22].

TABLE I. EVALUATION SET DESCRIPTION

	Text		
	Recipe	Manual	News
# of sentences	18	22	14
# of words	268	280	300

The evaluation set consists of three texts in German. Three short texts are taken into consideration instead of just one

longer because of the differences between automatic metrics and human DA which have been detected when dealing with different domains and text types. Texts are chosen randomly, the only requirement being that they are of approximately same size. One text is extracted from a book of recipes, one from a mobile phone manual, and one is a newspaper article. The text from the book of recipes is actually compiled of two recipes. The second text is on battery saving and battery charging. The newspaper article is about the meeting between the German chancellor, Angela Merkel, and the US president, Donald Trump.

The annotator is presented with German source texts, their respective Microsoft Bing and GT translations into Croatian, and reference translations. Since MT engines show a constant improvement over time, the translations obtained only a month later might differ greatly and might be of much better quality.

### B. TREAT Application

The Universal Windows Platform (UWP) application developed and presented in [10] is chosen for the manual MQM annotation task. The user interface is shown in Fig. 1. The error analysis annotation process in TREAT is shown in Fig. 2. The input to the annotation process are three textual files – the source file, the target file, and the optional reference file.

### C. Annotator

The annotator is a native speaker of Croatian with a BA degree in the German language and being very confident about her knowledge of German.

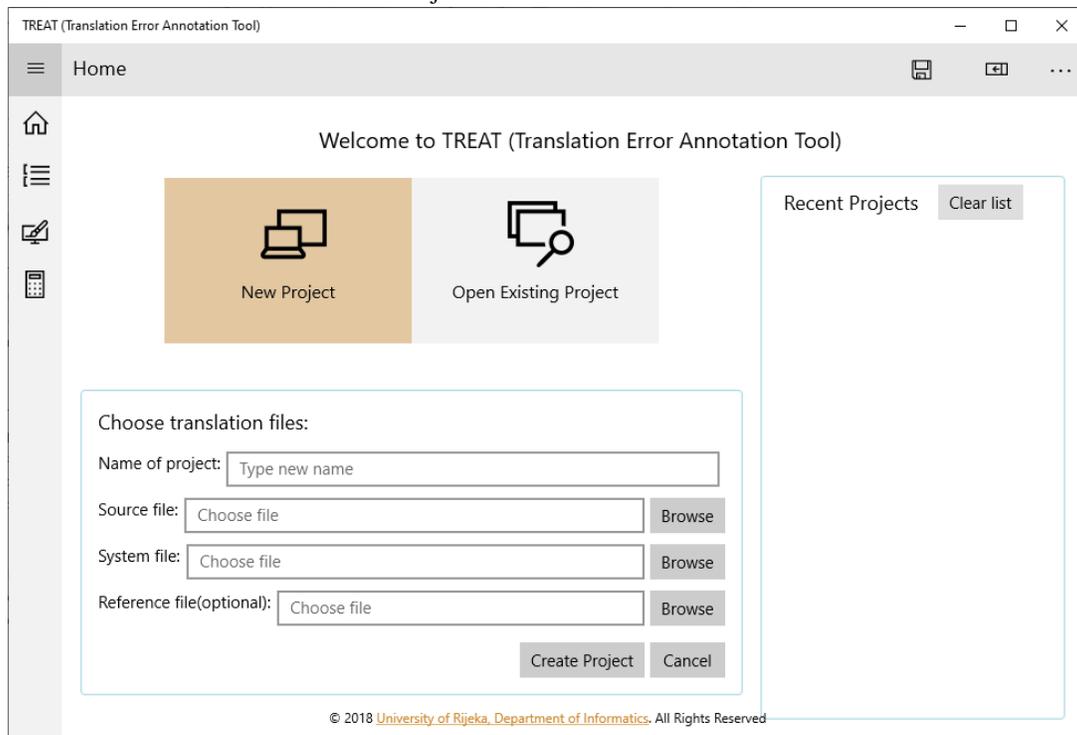


Fig. 1. TREAT user Interface.

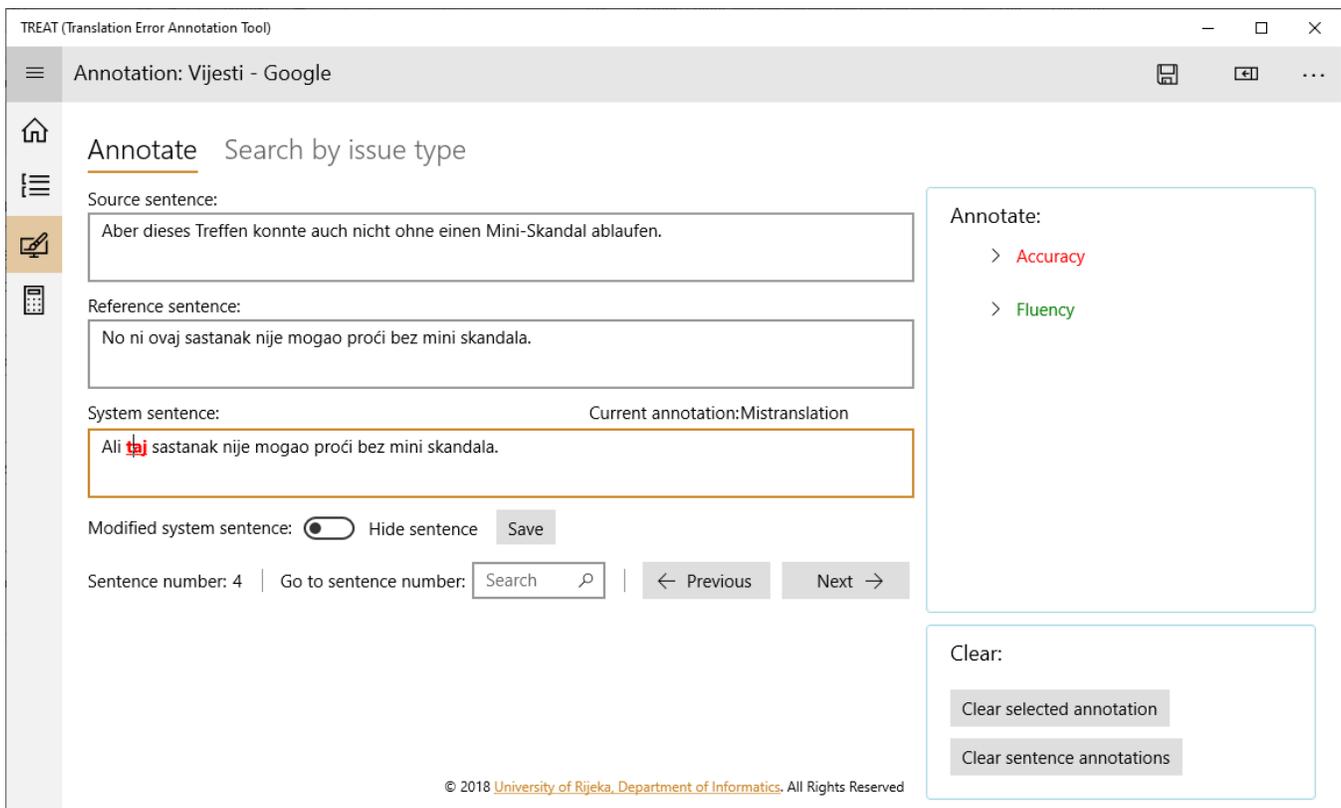


Fig. 2. Error Analysis in TREAT.

#### D. MQM

MQM defines over 100 issue types. The term issue is used to refer to any potential error detected in a text. At the top level there are 10 categories: accuracy, design, fluency, internationalization, locale convention, style, terminology, verity, compatibility, and other. Since it would not be viable to perform the annotation process using the full MQM tag set, it is necessary to choose a smaller subset of interest. The annotator is instructed to use the Slavic tagset [9] of the MQM core with a modification of using typography, as suggested by the core, instead of register suggested by [9]. The Slavic tagset entails higher-level categories accuracy and fluency. While accuracy is included in its original form, the authors suggest three subcategories of the word form issue—part of speech, agreement, and tense/aspect/mood, and three subcategories of function words—extraneous, missing, and incorrect. Typography refers to the issues related to the mechanical presentation of a text (e.g. punctuation is used incorrectly or a text has an extraneous hard return in the middle of a paragraph). This category should be used for any typographical errors other than spelling. If the exact nature of the error cannot be determined and a major break down in fluency occurs, unintelligible mark-up should be used.

Prior to annotation, the annotator is familiarized with TREAT and the official MQM annotation guidelines, which offer detailed instructions for annotation within MQM<sup>1</sup>. The annotator is instructed to avoid guessing by choosing rather a

<sup>1</sup> A decision tree provided to aid the annotation process can be found at <http://www.qf21.eu/downloads/annotatorsGuidelines-2014-06-11.pdf>.

higher-level issue and to use a minimalistic mark-up. Training, evaluation guidelines elaborated with examples, and familiarity with the field to which the text belongs are considered important for evaluation [4].

#### E. BLEU

According to BLEU, the more  $n$ -gram matches with the reference translation, the better the candidate translation is. A modified precision score is calculated for the whole corpus by adding the clipped counts of matches (the total count of each candidate is clipped by the maximum number of times the word occurs in any single reference translation) and dividing the sum by the total number of  $n$ -grams in the candidate. The weighted average of the logarithm of the modified precisions accounts for the exponential decay in precision scores as  $n$ -grams get of higher order (1). The brevity penalty is computed over the entire corpus on best match reference lengths (2), where  $c$  denotes the candidate length, and  $r$  the best match reference length [15]. For calculating BLEU scores we use the NLTK script available at [https://www.nltk.org/\\_modules/nltk/translate/bleu\\_score.html](https://www.nltk.org/_modules/nltk/translate/bleu_score.html).

$$BLEU = brevityPenalty \times \exp(\sum_{n=1}^N w_n \log p_n) \quad (1)$$

$$brevityPenalty = \begin{cases} 1 & \text{if } c > r \\ e^{(1-\frac{r}{c})} & \text{if } c \leq r \end{cases} \quad (2)$$

#### F. CHRF++

CHRF uses character  $n$ -gram F-score, excluding spaces. It is calculated as in (3), where CHRF stands for the percentage of  $n$ -grams in the candidate translation which have a

counterpart in the reference, while CHRR stands for the percentage of  $n$ -grams in the reference which are also present in the candidate. Parameter  $\beta$  assigns  $\beta$  times more importance to recall than to precision. It has been shown that the optimal option for  $\beta$  parameter is the value of 2, and for character  $n$ -grams the value of 6 [16]. CHRF++ score per segment is obtained by adding word 2-grams to the character 6-grams. We use the original python script for calculating the CHRF++ score, available at <https://github.com/m-popovic/chrF>.

$$\text{CHRF}\beta = (1 + \beta)^2 \times \frac{\text{CHRP} \times \text{CHRR}}{\beta^2 \text{CHRP} + \text{CHRR}} \quad (3)$$

#### IV. ERROR ANALYSIS

The most represented error issue in all three texts is mistranslation (Fig. 3). Remaining errors are mostly grammatical, e.g. case concordance, word order or incorrect function word.

An example of a *mistranslation* error is given in Fig. 4. An example of a sentence marked as illegible is given in Fig. 5. Although each error could be marked separately, according to MQM guidelines, the sentence has enough errors to be marked as illegible.

The translation of the manual obtained by Google Translate is surprisingly good. The most represented error category is again mistranslation. It is worth noting that certain parts are even translated into English instead of Croatian. This is probably due to the fact that many terms do not have a standardized equivalent or a standardized equivalent has not entered popular use so the English alternative is acceptable.

Fig. 6 is an example of the news sentence translated by Bing which exhibits several error issues, such as mistranslation, incorrect word form, i.e. tense under the grammar subcategory, and word order issue.

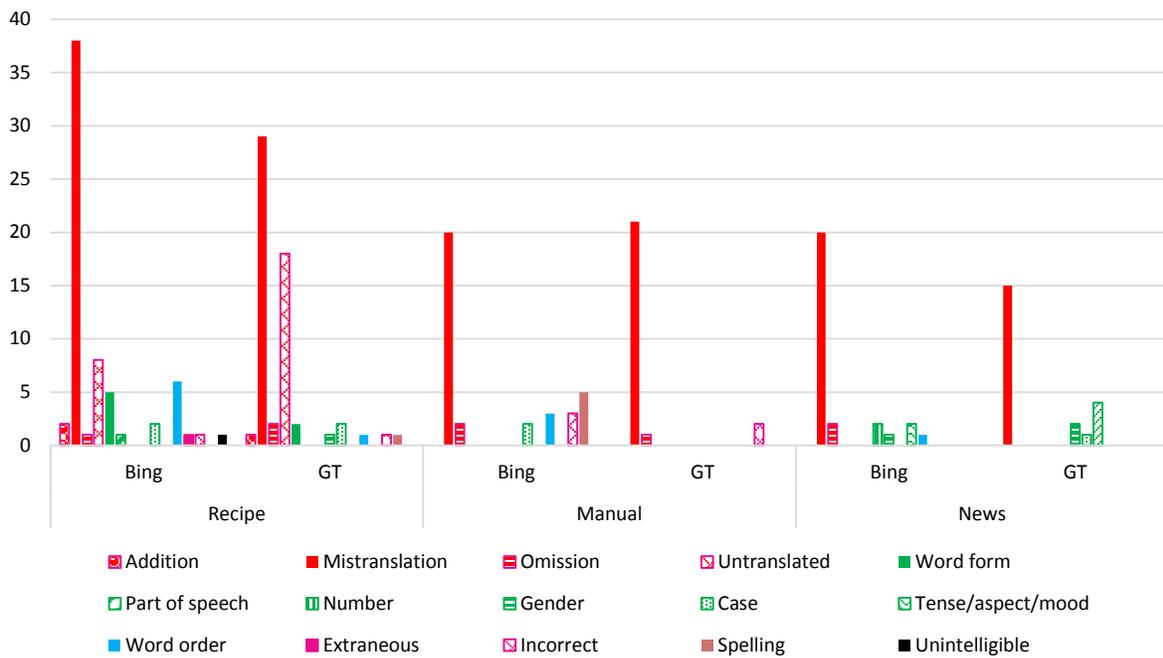


Fig. 3. MQM Issues Found.

Source sentence:

Er kann sowohl aus Germteig, Rührteig oder Biskuiteig hergestellt werden.

Reference sentence:

Može se napraviti od dizanog tijesta, lijevanog tijesta ili biskvita.

System sentence:

Current annotation: Mistranslation

Može se napraviti i od **klice** tijesta, **miješanja** tijesta ili **spužve** tijesta.

Fig. 4. Mistranslation Error in the Recipe Translated by Bing.

Source sentence:

Die geschmacklichen Varianten reichen von Marmorgugelhupf, Gugelhupf mit Rosinen, Kürbiskerngugelhupf bis zu Vollkorngugelhupf oder beschwipsten Versionen wie Glühweingugelhupf und Eierlikörgugelhupf.

Reference sentence:

Različiti okusi kreću se od mramornog kuglofa, kuglofa s grožđicama, kuglofa sa sjemenkama bundeve pa sve do kuglofa s integralnim brašnom ili pijanih verzija kao što su kuglof sa kuhanim vinom ili sa likerom od jaja.

System sentence:

Current annotation:Unintelligible

**Ukusna varijante u rasponu od mramora kuglof, kuglof s grožđicama, bundeve sjeme kuglof na čežitarice kuglof ili hrane verzije kao što su kuhano vino kuglof i punč kuglof.**

Fig. 5. Illegible Sentence in the Recipe Translated by Bing.

Source sentence:

Aber dieses Treffen konnte auch nicht ohne einen Mini-Skandal ablaufen.

Reference sentence:

No ni ovaj sastanak nije mogao proći bez mini skandala.

System sentence:

Current annotation:Word order

Ali **taj** sastanak **ne može isteci** bez mini skandala, **također**.

Fig. 6. A Sentence with Several Issues in the News Text Translated by Bing.

## V. RESULTS

BLEU and CHRF++ scores, as well as the average number of errors per sentence and the total number of errors per text are reported in Table II. Error counts per each category are presented in Fig. 3. For the sake of clarity, when possible, errors of the same superordinate category are shown in the same color with differing patterns. This being said, accuracy errors are presented in red, grammatical errors belonging to a word form subcategory in green, and grammatical errors relating to function words in fuchsia.

It is worth noting that GT manages to translate quite different number of sentences per each text flawlessly– 6% in the recipe, 21% in the manual, and 63% in the news, which indicates its varying performance on different text types. Two variants of CHRF++ score are reported–overall document level (F2), and macro averaged document level F-score (avgF2), which is the arithmetic average of sentence level scores.

The Pearson coefficient is used to calculate segment-level correlations between automatic metrics and total error counts (Table III), and between automatic metrics and two most represented error categories (Table IV). The Pearson coefficient ranges from +1 to -1, where +1 expresses total positive correlation, i.e. by increasing error counts, automatic metric scores increase, and -1 total negative correlation, i.e. by increasing error counts, automatic metrics scores decrease.

A value of 0 denotes that there is no linear correlation between the inspected variables.

TABLE II. THE RESULTS OF AUTOMATIC AND HUMAN EVALUATION OF MT

		Text					
		Recipe		Manual		News	
		Bing	GT	Bing	GT	Bing	GT
BLEU		53.81	52.93	39.49	38.55	47.72	46.79
CHRF++	F2	45.15	47.36	54.85	61.95	62.81	66.47
	avgF2	43.82	47.05	53.11	58.66	63.46	67.50
Error analysis	avg	3.66	3.22	1.59	1.09	2	1.57
	total	66	58	35	24	28	22

TABLE III. SEGMENT-LEVEL CORRELATION BETWEEN AUTOMATIC METRIC SCORES AND TOTAL NUMBER OF ERRORS

	Text					
	Recipe		Manual		News	
	Bing	GT	Bing	GT	Bing	GT
BLEU	0.0016	0.3822	0.3462	0.2126	-0.0011	0.2447
CHRF++	-0.2507	0.5004*	0.0478	-0.3954	-0.6872**	-0.4671

\* statistically significant at 5%

\*\* statistically significant at 1%

TABLE IV. SEGMENT-LEVEL CORRELATION BETWEEN AUTOMATIC METRIC SCORES ON ONE HAND AND ACCURACY CATEGORY AND MISTRANSLATION SUBCATEGORY ON THE OTHER HAND

		Text					
		Recipe		Manual		News	
		Bing	GT	Bing	GT	Bing	GT
BLEU	acc	0.0359	0.4342	0.1239	0.3109	0.1882	0.1432
	mis	0.1087	0.3780	0.1231	0.3341	0.2329	0.1432
CHRF++	acc	-0.2752	-0.5404*	-0.2180	-0.4074	-0.5131	-0.3909
	mis	-0.2311	-0.5661*	-0.1438	-0.4017	-0.4490	-0.3909

## VI. DISCUSSION

As far as the correlation is concerned, segment-level BLEU shows a positive correlation with human judgments (Table III, Table IV). This means that the BLEU score increases with the increase of errors, as if it is an error metric and not a precision-based one. Translations abundant with errors should be scored lower. A negative relationship exists between CHRF++ and human judgments. However, hardly any correlation proves to be statistically significant. Those rare which are significant are related to CHRF++. The abbreviations *acc* and *mis*, used in Table IV, refer to accuracy and mistranslation, respectively.

The examined automatic metrics do not even agree on the ranking of MT engines for the selected genres (Table II). According to BLEU, the most suitable genre for the selected MT engines is the one regarding recipes, followed by news and manual. On the other hand, CHRF++ rates news the best, followed by manual, and lastly recipe. The human annotator attributes the highest number of errors to the recipe, in line with CHRF++, while the other two genres are pretty close regarding the number of errors. However, CHRF++, unlike BLEU, manages to rank them correctly. If taking only GT into consideration, then the second best scoring translation according to the human annotator is the manual, while the best scoring is the news text. While the difference between the systems in terms of BLEU is less than one point, in terms of CHRF++ it ranges from 2 to over 7 points.

Although the authors in [20] show that BLEU best correlates with human judgments in domains containing short and simple sentences, but is surpassed by CHRF in cases with more complex syntactic structures and longer contexts, our evaluation gives advantage to CHRF++ in all three genres. We cannot make any conclusions on the effects of low-quality translations on the correlation since hardly any correlation proves to be statistically significant. Although not even human translations would obtain a score of 1 due to high variability of translations, having multiple references could increase the overall BLEU score and some segment-level scores and affect the correlation results presented in this paper.

## VII. CONCLUSION AND FUTURE WORK

The conducted evaluation is of a black box type. Three texts of different genre are selected in order to examine translations produced by two popular MT services in the German-Croatian language direction.

Automatic metrics employed in the paper depend on the availability of human reference translations. BLEU is chosen, despite of its many drawbacks, as it is the *de facto* standard in MT community, and CHRF++ is chosen as it has potential in dealing with morphologically rich target languages.

MQM Slavic tagset compliant error analysis of translations is performed in TREAT by one annotator who is a native of Croatian and has a BA in the German language. The training is conducted prior to annotation in order to familiarize the annotator with MQM and TREAT.

Pearson correlation coefficients are calculated to check how close automatic evaluation reflects manual judgments. CHRF++ metric, which works on character level and which is enhanced with word *n*-grams, proves to correlate better with human judgments than BLEU, which is a metric that works only on word level. This is valid for all three genres.

A major drawback of this study, beside the fact that only one reference translation is present, is the inevitable trade-off made between the size of the sample and the integrity of the results. Manual error analysis performed in this study is extremely expensive and time consuming. Automatic metrics are, on the other hand, quick and cheap to use. However, they are of no use if they do not correlate with human judgments.

A pairwise comparison of metrics within other classes and between classes is purposefully excluded from this study and left for future work for the sake of clarity and conciseness. Besides enlarging the evaluation set, involving more annotators would be considered beneficial, despite the usual low inter-annotator agreement in such tasks. A particular focus of our future work will be put on the results of neural network-based metrics. Since hardly any statistically significant correlation is detected, neither on the total number of errors, nor on the most prominent error category and subcategory, our future work will also entail a weighing scheme which will give weights to different error issues in the total error counts. Correlation between automatic metrics and a scoring mechanism provided with MQM, which includes weights on a four-level scale, will also be investigated.

## REFERENCES

- [1] Way, "Quality Expectations of Machine Translation," Springer, Cham, pp. 159–178, 2018.
- [2] M. Sepesy Maučec and G. Donaj, "Machine Translation and the Evaluation of Its Quality," in Natural Language Processing-New Approaches and Recent Applications, 2019.
- [3] A. Secara, "Translation Evaluation - a State of the Art Survey," in Proceedings of the eCoLoRe/MeLLANGE Workshop, 2005, pp. 39–44.
- [4] E. Chatzikoumi, "How to evaluate machine translation: A review of automated and human metrics," Nat. Lang. Eng., pp. 1–25, 2019.
- [5] S. Stymne, "Pre- and Postprocessing for Statistical Machine Translation into Germanic Languages," in Proc. 49th Annu. Meet. Assoc. Comput. Linguist. Student Sess., no. June, 2011, pp. 12–17.
- [6] S. Stymne and L. Ahrenberg, "On the practice of error analysis for machine translation evaluation," in 8th Int. Conf. Lang. Resour. Eval., 2012, pp. 1785–1790.
- [7] M. Popovic and A. Burchardt, "From Human to Automatic Error Classification for Machine Translation Output," in Proceedings of the 15th International Conference of the European Association for Machine Translation, no. May, 2011.

- [8] A. Lommel, H. Uszkoreit, and A. Burchardt, "Multidimensional Quality Metrics: A Flexible System for Assessing Translation Quality," vol. 12, no. Tradumàtica 12, pp. 455-463., 2014.
- [9] F. Klubička, A. Toral, and V. M. Sánchez-Cartagenac, "Fine-Grained Human Evaluation of Neural Versus Phrase-Based Machine Translation," *Prague Bull. Math. Linguist.*, no. 108, pp. 121-132, 2017.
- [10] S. Majcunic, M. Matetic, and M. Brkic Bakaric, "Translation Error Analysis in TREAT: A Windows App Using the MQM Framework," *Zb. Veleučilišta u Rijeci*, vol. 7, no. 1, pp. 149-162, 2019.
- [11] M. Fomicheva and L. Specia, "Taking MT Evaluation Metrics to Extremes: Beyond Correlation with Human Judgments," *Comput. Linguist.*, vol. 45, no. 3, pp. 515-558, 2019.
- [12] O. Bojar, Y. Graham, and A. Kamran, "Results of the WMT17 Metrics Shared Task," in *Proceedings of the Conference on Machine Translation (WMT)*, 2017, vol. 2, pp. 293-301.
- [13] O. Abend and A. Rappoport, "Universal conceptual cognitive annotation (UCCA)," in *ACL 2013 - 51st Annu. Meet. Assoc. Comput. Linguist. Proc. Conf.*, vol. 1, 2013, pp. 228-238.
- [14] Q. Ma, J. Wei, O. Bojar, and Y. Graham, "Results of the WMT19 Metrics Shared Task: Segment-Level and Strong MT Systems Pose Big Challenges," in *Proceedings of the Fourth Conference on Machine Translation*, 2019, vol. 2, no. Day 1, pp. 62-90.
- [15] K. Papineni, S. Roukos, T. Ward, and W. Zhu, "B LEU: a Method for Automatic Evaluation of Machine Translation," in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 2002, no. July, pp. 311-318.
- [16] M. Popovic, "CHRF: character n-gram F-score for automatic MT evaluation," in *Proceedings of the Tenth Workshop on Statistical Machine Translation*, 2015, no. September, pp. 392-395.
- [17] M. Popovic, "CHRF++: words helping character n-grams," in *Proceedings of the Conference on Machine Translation*, 2017, vol. 2, no. 1, pp. 612-618.
- [18] G. Doddington, "Automatic evaluation of machine translation quality using n-gram co-occurrence statistics," in *Proceedings of the Second International Conference on Human Language Technology Research, HLT '02*, 2002, pp. 138-145.
- [19] M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul, "A study of translation edit rate with targeted human annotation," in *AMTA 2006 - Proc. 7th Conf. Assoc. Mach. Transl. Am. Visions Futur. Mach. Transl.*, 2006, pp. 223-231.
- [20] M. Chinea-Rios, A. Peris, and F. Casacuberta, "Are Automatic Metrics Robust and Reliable in Specific Machine Translation Tasks?," in *21st Annual Conference of the European Association for Machine Translation*, 2018, no. May, pp. 89-98.
- [21] M. Brkic Bakaric, N. Babic, L. Dajak, and M. Manojlovic, "A comparative error analysis of English and German MT from and into Croatian," in *InFuture 2017*, 2017, pp. 31-41.
- [22] R. Fiederer and S. O'Brien, "Quality and machine translation: A realistic objective?," *J. Spec. Transl.*, no. 11, pp. 52-74, 2009.

# Technologies for Making Reliable Decisions on a Variety of Effective Factors using Fuzzy Logic

Yousef Ibrahim Daradkeh<sup>1</sup>

Department of Computer Engineering and Networks  
College of Engineering at Wadi Addawasir, 11991  
Prince Sattam Bin Abdulaziz University, KSA

Irina Tvoroshenko<sup>2</sup>

Department of Informatics  
Kharkiv National University of Radio Electronics  
Kharkiv, Ukraine

**Abstract**—Problem of choosing methods has been presented for increasing the reliability of solutions while reducing the time resource for researching the state space of a particular object. Effective factors that can significantly affect the attainability of decision-making goals in a fuzzy process development environment are identified. Presented the selected factors in a classic form in natural language, which significantly increases the confidence in the decisions made by determining the maximum, minimum values of membership functions and the set of criteria by using manifest conflict situations when deciding. Information technologies and approaches for identifying the completeness of decision-making goals on a variety of effective factors are proposed. The effectiveness of the proposed solutions using the fuzzy inference procedure and the Zade-Mamdani's approach is estimated. Software testing carried of the described methods for evaluating and deciding on a variety of factors out using object-oriented programming. Experimental testing of realized ideas has confirmed an increase in the reliability of making aim management decisions in various applied subject areas.

**Keywords**—Attainability; completeness of decision-making; conflict; factor; membership function; state space

## I. INTRODUCTION

Effective solutions in complex control systems and data processing of artificial intelligence systems of subject areas require the operational consideration of many conflicting factors: the essentially fuzzy nature of the interacting dynamic processes and the spaces of their states, the complex parallel-sequential interaction of processes in the face of uncertainty [1], the significant complexity of the tasks, the significant share of human a factor that largely determines the quality and level of modern solutions, and others.

Implementing control, data processing, and decision-making processes in the computing intelligence systems of distributed complexes operating in a fuzzy environment [2], requires solving several problems [3].

These include analysis [4], identification [5], localization, and elimination of conflict situations during the interaction of processes in a fuzzy state space [6].

It is urgent to propose and develop a systematic approach to solving this problem in the form of constructing new mathematical fuzzy models, an adequate set of formal criteria, modern approaches, information technologies, effective methods and tools for solving a wide class of theoretical and applied problems.

## II. REVIEW OF THE LITERATURE

The problem of goal attainability analysis is one of the main tasks of fuzzy process analysis [5]. In the process of creating computing intelligence systems, the developer has many goals, the quality of implementation, of which determines the degree of development efficiency. The goal will be realized in an environment of interacting processes. With inadequate and/or incorrect presentation and interaction, may not realize some goals. This requires a set of actions to ensure the goals of decisions [7].

This paper discusses the question of analyzing the attainability of the goals of decisions that are formed because of fuzzy processes and can be reflected in a variety of relations "condition-action".

It greatly complicates the task [1] of analyzing fuzzy processes is greatly because of the influence of fuzzy properties on by using develop processes in the structures of knowledge bases of intelligent systems [3] and decision-making based on them [5]. Existing approaches, given the complexity [6] and limited capabilities [7], do not allow us to get effective solutions. This requires using create new approaches and solutions-oriented to solving applied problems.

Qualitative accounting of many important factors [8] significantly affects the effectiveness of systems in practical implementations.

The primary problems of making reliable decisions [9] include ensuring: the attainability of decision-making goals, the consistency of decision-making goals, the completeness of decision-making, using conflict situations, using excess and useless loops, the quality of optimizing the choice of process development alternatives, the quality of optimizing the laws of learning and setting up fuzzy models and their reflection on real objects.

There is a significant amount of scientific research [10], models [11], methods [12] and software products [13] that, with varying degrees of effectiveness, solve some aspects of creating systems [14] and approaches based on computational intelligence [15].

The problems of comprehensively providing an actual reflection of the interaction of certain dynamic fuzzy processes, minimizing resources and ensuring an adequate choice of many alternatives for the possible development of fuzzy processes on

a variety of specific constraints and assumptions are very important.

However, they do not have an analytical interpretation, which has found effective practical application in solving complex problems [16–18] requiring individual or collective decisions.

Such problems have a highly specialized character [19] and are solved on a limited set of subject areas.

The result of the analysis of modern scientific and practical literature [4–7, 15–22] confirmed that research in this area is not efficient enough; there is no clear interpretation and visualization of the interactive features of specific dynamic fuzzy processes and complex systems.

### III. PROBLEM STATEMENT

There are many interacting fuzzy processes  $\{\tilde{P}_i\}$ ,  $i \in I$ .

The state space of interacting processes is defined on set  $\{\tilde{A}_i\}$ ,  $i \in I$ . The initial state of the development of processes is defined by the space of initial states  $\{\tilde{A}_{oi}\} \subset \{\tilde{A}_i\}$ . The final state of development of processes is defined by the space of final states  $\{\tilde{A}_{ki}\} \subset \{\tilde{A}_i\}$ , which defines the set of realized goals  $\{C_r\}$ ,  $r \in R$ .

If the set of processes  $\{\tilde{P}_i, i \in I, \{\tilde{P}_i\} \neq \emptyset$  is defined, the space of initial states  $\{\tilde{A}_{oi}\}, \{\tilde{A}_{oi}\} \neq \emptyset$ , then the attainability of the goals of the decisions made is realized if a subset of  $\{\tilde{A}_{ki}\} \subset \{\tilde{A}_i\}$  is reachable, and it is true:

$$\forall \tilde{P}_i \in \{\tilde{P}_i\}, \forall \tilde{A}_{oi} \in \{\tilde{A}_{oi}\} \left| \{\tilde{A}_{ki}\} \neq \emptyset \right. \quad (1)$$

Justice (1) determines the truth  $Ds_j = true$ ,  $j \in J$ .

Based on the above interpretation of the space of final states  $\{\tilde{A}_{ki}\} \subset \{\tilde{A}_i\}$ , as the set of realized goals  $\{C_r\}$ ,  $r \in R$ , the necessity and sufficiency of (1) to ensure  $Ds_j = true$ ,  $j \in J$  is obvious.

The solution to the problem of analyzing the attainability of decision-making goals includes the study of the state space of interacting fuzzy processes and the manifestation of the validity of conditions (1). All developed provisions should be focused on modern information technologies [9–15].

The aim of this work is to increase the reliability of decisions while reducing the time to analyze the state space of an object by developing information technologies and approaches that allow us to assess the state of many factors and highlight the most important for making effective decisions.

### IV. MATERIALS AND METHODS

Some effective factors will be defined that can significantly affect the attainability of decision-making goals in a fuzzy environment [23] for the development of processes [16].

A subset of the goals of decisions made  $\{\tilde{D}_{s_j}\}$ , for a given space of initial states  $\{\tilde{A}_{oi}\}$ , is unattainable if at least one of the following situations is true:

Situation 1.

$$\exists \tilde{P}_i \in \{\tilde{P}_i\} \left| \mu_{\tilde{P}_i}(k_0) < \mu_{\tilde{P}_i}(k_0)^* \right. \quad (2)$$

Where  $\mu_{\tilde{P}_i}(k_0)^*$  is the minimum permissible value of the membership function for  $\tilde{P}_i$  processes;  $k_0$  – the value of variable  $k$ , which determines the specific value of function  $\mu_{\tilde{P}_i}(k)$ .

Situation 2. There is at least one process  $\tilde{P}_i \in \{\tilde{P}_i\}$ , which cannot be performed and is fair for it

$$\exists \tilde{P}_i \in \{\tilde{P}_i\} \left| \{\tilde{D}_{s_j}\} = false, \{\tilde{D}_{s_j}\} \neq \emptyset \right. \quad (3)$$

Situation 3. There is a state that does not stop the mutual expectation of two or more processes (tasks).

The achieved processes  $\{\tilde{P}_i\}$  are a subset of processes

$\{\tilde{P}_i\} \subset \{\tilde{P}_i\}$ ,  $\{\tilde{P}_i\} \neq \emptyset$  for which none of the conditions (2) and (3) is satisfied.

The analysis of characteristic situations makes it possible to determine the following:

- Situation 1 determines that in order to solve the problem of attainability of goals (1), it is necessary to modify the values of  $\mu_{\tilde{P}_i}(k_0)$  in (2) so that

$$\forall \tilde{P}_i \in \{\tilde{P}_i\} \left| \mu_{\tilde{P}_i}(k_0) \geq \mu_{\tilde{P}_i}(k_0)^* \right. \quad (4)$$

To fulfill (4) in real systems, some additional resources are attracted [17], the nature, volume, and labor of which are determined by a specific subject area;

- Situations 2 and 3 reflect the fact of logical inadequacy of the goals of decision-making processes in the subject area [9], which necessitates the modification of the laws of interaction of fuzzy processes.

The conflict situation in the interaction of fuzzy processes is defined as the presence of at least one of the situations:

- Two or more processes compete for data that is obtained as a result of previous actions;
- Two or more processes compete for resources that ensure the following processes.

The presence of conflicts in the interaction of processes in complex objects leads to a significant decrease in the efficiency of functioning.

If the set of processes  $\{\tilde{P}_i\}, i \in I, \{\tilde{P}_i\} \neq \emptyset$  is defined, the space of initial states  $\{\tilde{A}_{oi}\}, \{\tilde{A}_{oi}\} \neq \emptyset$ , the space of current states  $\{\tilde{A}_i\}, \{\tilde{A}_i\} \neq \emptyset$ , the space of final states  $\{\tilde{A}_{ki}\}, \{\tilde{A}_{ki}\} \neq \emptyset$ , then there are no conflict situations if there is no competition for resources:

$$\begin{aligned} \forall \tilde{P}_i \in \{\tilde{P}_i\}, \forall \tilde{A}_{oi} \in \{\tilde{A}_{oi}\}, \forall \tilde{A}_i \in \{\tilde{A}_i\}, \\ \forall \tilde{A}_{ki} \in \{\tilde{A}_{ki}\} | Conf_k = \emptyset \end{aligned} \quad (5)$$

subject to the fulfillment of (4).

Solving the problem of manifestation of conflict situations requires complex studies of the structure of the model in the space of model states and processes.

## V. EXPERIMENTS

The solution of the problem of increasing the reliability of decisions [19] while reducing the time to analyze the state space of an object, taking into account (1) and the accepted interpretation (2)–(4), can be reflected in the form of technology for detecting the completeness of decision-making goals:

- The simulated processes (phenomena) occurring in the object of modeling and analysis are described by a set of events and conditions by which these events are determined, as well as causal relationships that are established on the set “condition–action”;
- Set of “condition–action” define when control the sequence of occurrence, where is a specific space of potential states of the object of research. The state space

is defined by many conditions. Conditions are formulated in the form of predicates and complex procedures;

- Conditions (predicates) may be fulfilled and not fulfilled. Only the fulfillment of the conditions provides the possibility of events. Conditions, the fulfillment of which is associated with the possibility of events, are “in-conditions”;
- After the event has occurred, other conditions that are in-conditions and the consideration of events with a causal relationship will be satisfied, they are called “out-conditions”.

Accept this interpretation of components of the model:

- The set of input conditions (antecedents) and the initial conditions (consequents) is described by the set of positions  $\{\tilde{p}_j\}, j \in J$ , and the set of events (actions) – by the set of transitions  $\{\tilde{t}_i\}, i \in I$ ;
- The sequence of events is displayed by performing transitions. The fulfillment of any condition is associated with the appearance of marker  $\tilde{M}(\tilde{p}_j) \geq 1$  in the position corresponding to this condition. Transition triggering rules are a way of expressing causal relationships between conditions and events in a system. After all, in-conditions are fulfilled, events may occur;
- The moment of actual realization of events is unknown. Sometimes it is difficult to restore the complete chain of immediate causes and consequences that determine the fact and time of the duration of the event, the identification of the properties of clarity and so on;
- After the event has occurred, out-conditions are implemented. These out-conditions are in-conditions of other incident events that determine the desired result. The root causes may disappear or continue to function. This is because fuzzy models have the property of absorption and the generation of markers.

The dynamics of modeling on fuzzy models is associated with the mechanism for changing position markings and the rules for performing transitions in the model state space.

With a direct simulation procedure and fuzzy inference, transition  $\tilde{t}_i$  is triggered for which

$$\forall \tilde{p}_j \in \{\tilde{p}_j(in)\} | \tilde{M}(\tilde{p}_j) > 0 \quad (6)$$

With the reverse simulation procedure and fuzzy inference, transition  $\tilde{t}_i$  is triggered for which

$$\forall \tilde{p}_j \in \{\tilde{p}_j(out)\} | \tilde{M}(\tilde{p}_j) > 0 \quad (7)$$

The transition in the general case can occur after any finite period  $\tau$  after the transition is triggered. Marking of all input and output positions associated with this transition is

instantaneous, taking into account time delays  $\tau$ , provided that the corresponding predicate is true from  $L$ .

Result of performing transitions with a direct procedure, a given value of function  $\tilde{F}(f)$  and the accepted interpretation, only one marker is removed from each incoming position of this transition  $\tilde{t}_i$  and exactly one marker is added to the markings of each initial position

$$\begin{aligned} &(\forall \tilde{p}_j \in \{\tilde{p}_j(in)\} | \tilde{M}(\tilde{p}_j) - 1) \text{ and} \\ &(\forall \tilde{p}_i \in \{\tilde{p}_i(out)\} | \tilde{M}(\tilde{p}_i) + 1). \end{aligned} \quad (8)$$

Result of performing transitions in the reverse procedure, given the value of function  $\tilde{F}(f)$  and the accepted interpretation from each source position of this transition  $\tilde{t}_i$ , only one marker is removed and exactly one marker is added to the markings of each input position

$$\begin{aligned} &(\forall \tilde{p}_j \in \{\tilde{p}_j(in)\} | \tilde{M}(\tilde{p}_j) + 1) \text{ and} \\ &(\forall \tilde{p}_i \in \{\tilde{p}_i(out)\} | \tilde{M}(\tilde{p}_i) - 1). \end{aligned} \quad (9)$$

Preliminary actions are performed if:

- An acceptable value of fuzziness of processes  $\left\{ \tilde{P}_i \right\}$  is provided, which determines the permissibility of transitions  $\exists \tilde{t}_i \in \tilde{T}$ ;
- The validity of some predicate  $L\{x_u\}$  is satisfied, which displays the properties of sets of factors and functions  $\tilde{F}(f)$ , state spaces  $\{\tilde{M}(f)\}$  of a fuzzy model;
- The truth of some procedure  $\{\tilde{P}r_j\}$  reflects the properties of many factors and state spaces  $\{\tilde{M}(f)\}$  of the fuzzy model.

Analysis of the completeness of decisions made using fuzzy models makes it possible, based on formal approaches, to determine the completeness properties of the goals of decisions made.

## VI. RESULT AND DISCUSSION

Operators of fuzzy relations [1] using in the work make it possible to implement fuzzy inference procedures.

In a comparative analysis, the Zadeh-Mamdani's approach is more promising; he can be represented of rules

$$\begin{aligned} &\{if \ x_i \text{ is } \mu(x) \text{ then } y_i \text{ is } \mu(y)\}, \ i \in I \\ &x_i = x' \\ &----- \\ &y'_i = ? \end{aligned} \quad (10)$$

Where  $x'$  is some received vector;  $y'_i$  – the desired solution for the case  $i \in I$ .

Procedure [8] on the basis of (5) implemented for  $\forall x'_j$ ,  $x'_j \in A$  with  $A \neq \emptyset$  in the form of some composition of rules  $|I|$ , then for (10) have

$$y' = \wedge x' \vee \mu(x, y)$$

Where  $\mu(x, y)$  – fuzzy relationship of antecedent and consequent with [1].

Defuzzification rules the based on the center of mass [1] get for (10) the desired approximate solution  $y'_i$ . The value  $y'_i$  can be refined by adjusting membership functions [8] until achieve the required accuracy standard.

Particular attention is paid to maximizing the values of membership functions of the expected solutions in such a way

$$\Delta \xrightarrow{F} extr \quad (11)$$

Where  $\Delta$  – an error that may exceed some acceptable value;  $F$  – restrictions determining the value of membership function.

Based on the developed methods for evaluating and making decisions on a variety of factors, fuzzy-type big data processing systems, presented in natural language in quantitative and qualitative form, have created a tool for constructing models.

Optimization criteria into account take: maximum clarity of components and processes, minimum time costs, maximum values of reliability indicators under given restrictions, and their derivatives and vector criteria.

Criteria and limitations are defined and can be specified by the features of functional tasks and the subject area.

It is advisable to focus on implementing modeling and analysis of processes on the capabilities of modern computer systems.

The structure of the developed tool provides:

- Tracking marker paths and calculating parameter values, conditions, predicates, and procedures;
- Assignment and processing of accessory functions for model components;
- Analysis of the attainability of the goals of decisions;
- Analysis of conflict situations;
- Analysis of alternatives to decisions (Fig. 1);
- Analysis of the completeness of decision-making goals;
- Analysis of the consistency of decision-making goals;
- Analysis of excess and useless loops;
- Modification of the model according to the results of analysis and modeling;

- Determination of system time for solving applied problems.

Search for alternatives using the develop of dynamic fuzzy processes (Fig. 2) is carried out according to the criteria:

- Minimum time parameters;

- Maximum clarity values in decision-making procedures;

- Other parameters determined by the subject area (the criteria for choosing solutions based on these parameters may be: the sum of their values, average value, minimum, maximum values) (Fig. 3).

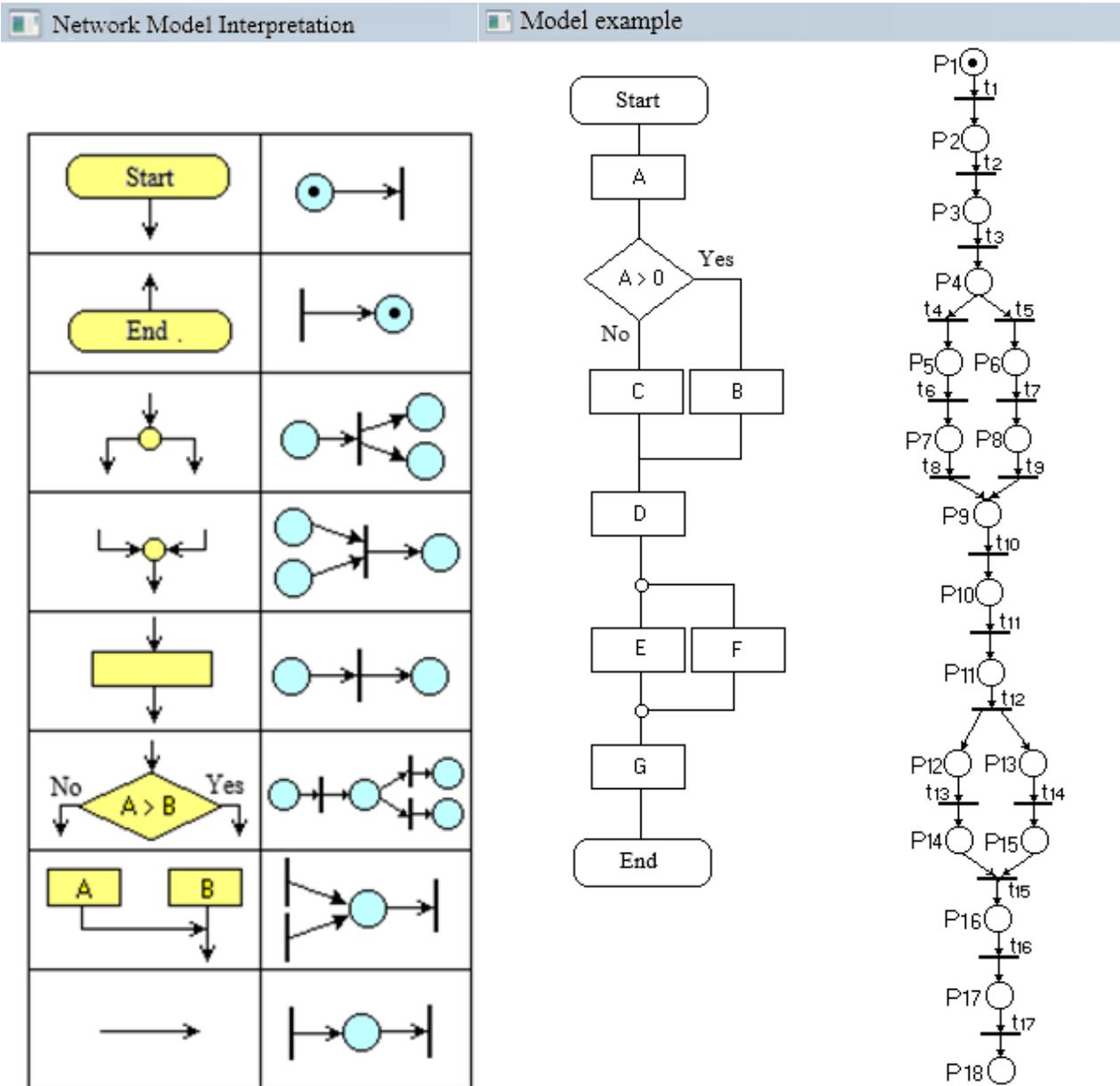


Fig. 1. Interpretation of Components of Network Models and an Example of Constructing a Fragment of the Model.

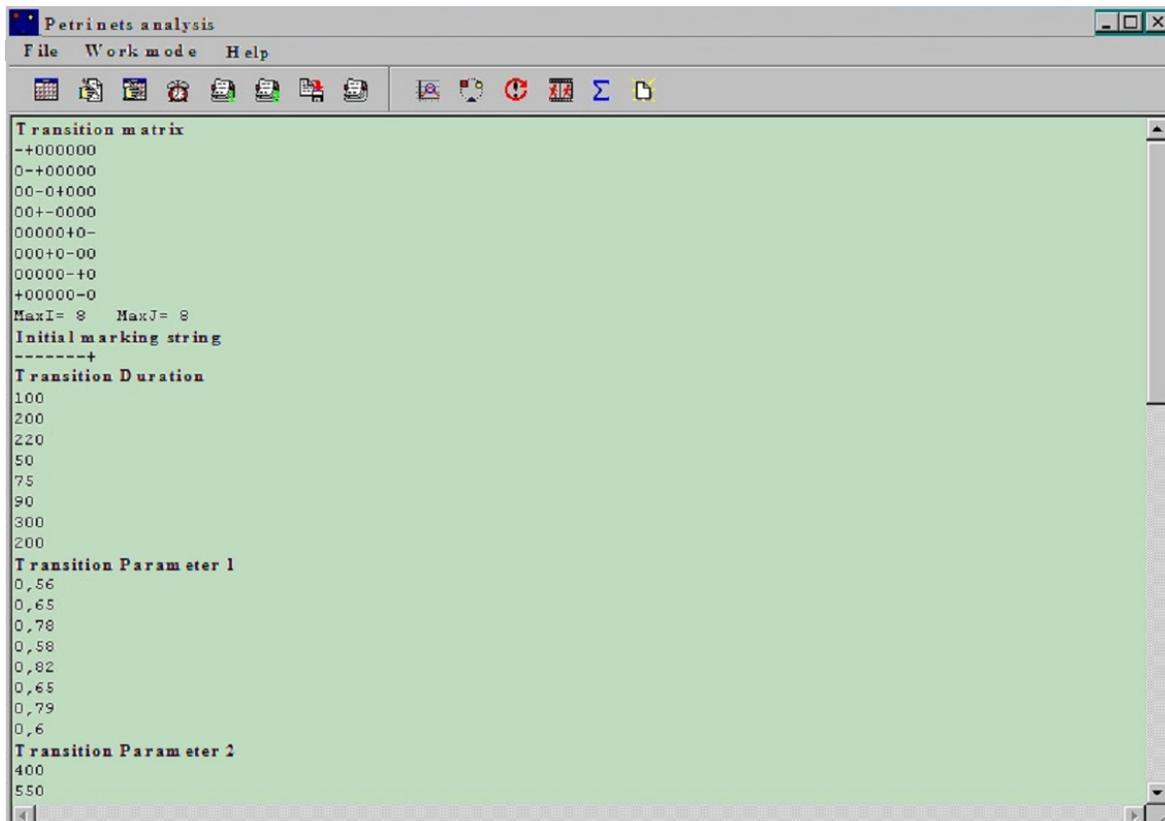


Fig. 2. The Result of Solving the Problem of Finding Alternatives.

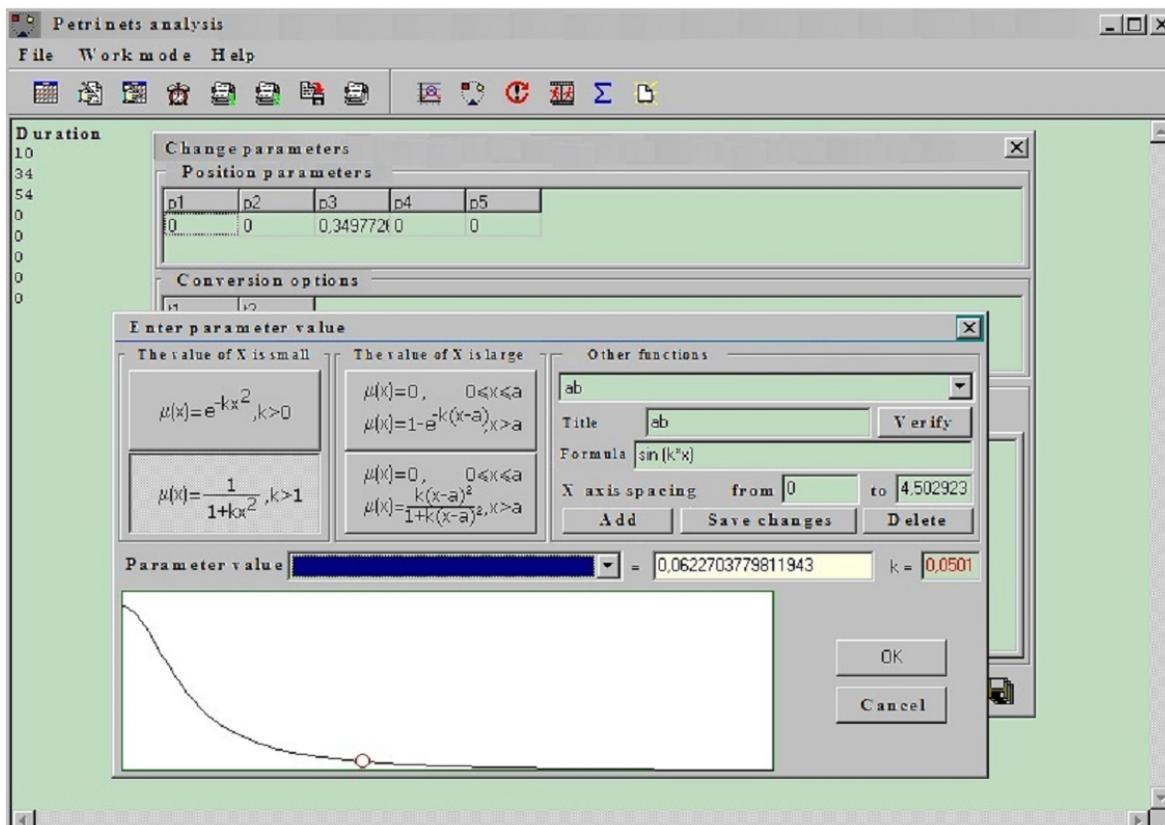


Fig. 3. The Mode of Forming Values of Membership Functions.

For high-dimensional models, decomposition operations, which are often used for detailed research of models and the subject area. Applied problem solved the usage of the composition of the model on the fragment.

In various applied subject areas that work in optimal or extreme situations, it may be necessary to decide under conditions of fuzziness, uncertainty, the unreliability of time and resource constraints or assumptions.

The developed software application allows you to find a reliable solution based on these factors using the apparatus of computational intelligence, which reflects a complex of mathematical models, methods, and means of implementing dynamic interacting processes that operate in conditions of uncertainty.

The implementation of modern trends in the construction of control and data processing systems significantly expands the capabilities of managing facilities, reduces the cost of necessary resources. The traditional and intelligent components of the systems complement each other and function in constant interconnection. Fuzzy components of the description of management processes and decision-making are gaining importance.

## VII. CONCLUSION

Approaches have been developed to identify the completeness of decision-making goals on a variety of effective factors for complex objects operating in fuzzy state space.

The practical application of tools for modeling and analysis of fuzzy processes in solving a wide class of practical problems, identifying, localizing and eliminating the incompleteness and inconsistency of the knowledge bases of dynamic fuzzy systems is proposed. The presented development made it possible to increase the reliability of managerial decisions by 15% according to expert data while reducing by 28% the time resource for researching the state space of a particular object.

The convenience of the proposed approaches also lies because users of the tools can be specialists who do not have in-depth knowledge, skills in computer science and programming.

The software application is developed in an object-oriented programming environment. Testing and repeated testing of the proposed technologies and tools confirmed the high reliability of the adopted objective management decisions on the topic of research.

The proposed approaches determine the effectiveness and feasibility of using knowledge of oriented technologies for decision-making by means of fuzzy logic.

A promising area of new research is the improvement or further expansion of criteria, methods, algorithms and software on a variety of alternatives when making reliable decisions.

## REFERENCES

- [1] I. S. Tvoroshenko, and V. O. Gorokhovatsky, "Intelligent classification of biophysical system states using fuzzy interval logic," *Telecommunications and Radio Engineering*, Vol. 78, No. 14, pp. 1303-1315, 2019. <https://doi.org/10.1615/TelecomRadEng.v78.i14.80>
- [2] M. Ayaz Ahmad, Irina Tvoroshenko, Jalal Hasan Baker, and Vyacheslav Lyashenko, "Computational Complexity of the Accessory Function Setting Mechanism in Fuzzy Intellectual Systems," *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 8, No. 5, pp. 2370-2377, 2019. <https://doi.org/10.30534/ijatcse/2019/77852019>
- [3] S. Ashraf, T. Mahmood, S. Abdullah, and Q. Khan, "Different approaches to multi-criteria group decision making problems for picture fuzzy environment," *Bulletin of the Brazilian Mathematical Society, New Series*, Vol. 50, No. 2, pp. 373-397, 2019. <https://doi.org/10.1007/s00574-018-0103-y>
- [4] K. Zhang, J. Zhan, and W. Z. Wu, "Novel fuzzy rough set models and corresponding applications to multi-criteria decision-making," *Fuzzy Sets and Systems*, Vol. 383, pp. 92-126, 2020. <https://doi.org/10.1016/j.fss.2019.06.019>
- [5] P. H. Thong, and S. Le, "Picture fuzzy clustering: a new computational intelligence method," *Soft computing*, Vol. 20, No. 9, pp. 3549-3562, 2016. <https://doi.org/10.1007/s00500-015-1712-7>
- [6] T. Hanne, and R. Dornberger, "Computational intelligence," *Computational Intelligence in Logistics and Supply Chain Management*, pp. 13-41, 2017. [https://doi.org/10.1007/978-3-319-40722-7\\_2](https://doi.org/10.1007/978-3-319-40722-7_2)
- [7] V. A. Gorokhovatskyi, "Image classification methods in the space of descriptions in the form of a set of the key point descriptors," *Telecommunications and Radio Engineering*, Vol. 77, No. 9, pp. 787-797, 2018. <https://doi.org/10.1615/TelecomRadEng.v77.i9.40>
- [8] S. M. Aristova, Y. I. Daradkeh, and P. M. Korolev, "Processes in Experiences with Uncertainty: How to Approach," *Journal of Materials Science & Nanotechnology*, Vol. 5, No. 1, pp. 106-110, 2017. <https://doi.org/10.15744/2348-9812.5.106>
- [9] Majed Kamel Al-Azzam, Malik Bader Alazzam, and Majida Khalid al-Manasra, "MHealth for decision making support: a case study of health in the public sector," *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 5, pp. 381-387, 2019. <https://doi.org/10.14569/IJACSA.2019.0100547>
- [10] Maruf Ahmed Tamal, Md Saiful Islam, Md Jisan Ahmmed, Md. Abdul Aziz, Pabel Miah, and Karim Mohammed Rezaul, "Heart disease prediction based on external factors: A machine learning approach," *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 12, pp. 446-451, 2019. <https://doi.org/10.14569/IJACSA.2019.0101260>
- [11] W. Pedrycz, A. Sillitti, and G. Succi, "Computational intelligence: an introduction," *Computational Intelligence and Quantitative Software Engineering*, pp. 13-31, 2016.
- [12] I. Perova, and I. Pliss, "Deep hybrid system of computational intelligence with architecture adaptation for medical fuzzy diagnostics," *International Journal of Intelligent Systems and Applications*, Vol. 9, No. 7, pp. 12-21, 2017. <https://doi.org/10.5815/ijisa.2017.07.02>
- [13] B. Xing, and W. J. Gao, "Overview of computational intelligence," *Artificial Intelligence: Concepts, Methodologies, Tools, and Applications*, pp. 12-31, 2017. <https://doi.org/10.4018/978-1-5225-1759-7.ch002>
- [14] Blessing Ekong, Idara Ifiok, Ifreke Udoeka, and James Anamfiok, "Integrated fuzzy based decision support system for the management of human disease," *International Journal of Advanced Computer Science and Applications*, Vol. 11, No. 2, pp. 268-274, 2020. <https://doi.org/10.14569/IJACSA.2020.0110235>
- [15] I. J. Pérez, and et al., "On dynamic consensus processes in group decision making problems," *Information Sciences*, Vol. 459, pp. 20-35, 2018. <https://doi.org/10.1016/j.ins.2018.05.017>
- [16] D. Hamann, N. P. Walz, A. Fischer, M. Hanss, and P. Eberhard, "Fuzzy arithmetical stability analysis of uncertain machining systems," *Mechanical Systems and Signal Processing*, Vol. 98, pp. 534-547, 2018. <https://doi.org/10.1016/j.ymssp.2017.05.012>
- [17] Sumaira Nazir, Nargis Fatima, and Suriyati Chuprat, "Situational factors for modern code review to support software engineers' sustainability," *International Journal of Advanced Computer Science and Applications*, Vol. 11, No. 1, pp. 498-504, 2020. <https://doi.org/10.14569/IJACSA.2020.0110161>

- [18] January D. Febro, "Utilizing feature selection in identifying predicting factors of student retention," *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 9, pp. 269-274, 2019. <https://doi.org/10.14569/IJACSA.2019.0100934>
- [19] P. Pourhejazy, J. Sarkis, and Q. Zhu, "A fuzzy-based decision aid method for product deletion of fast moving consumer goods," *Expert Systems with Applications*, Vol. 119, pp. 272-288, 2019. <https://doi.org/10.1016/j.eswa.2018.11.001>
- [20] M. J. Knowling, J. T. White, and C. R. Moore, "Role of model parameterization in risk-based decision support: An empirical exploration," *Advances in water resources*, Vol. 128, pp. 59-73, 2019. <https://doi.org/10.1016/j.advwatres.2019.04.01>
- [21] J. J. Rainer, S. Cobos-Guzman, and R. Galán, "Decision making algorithm for an autonomous guide-robot using fuzzy logic," *Journal of Ambient Intelligence and Humanized Computing*, Vol. 9, No. 4, pp. 1177-1189, 2018. <https://doi.org/10.1007/s12652-017-0651-9>
- [22] Subashini Ganapathy, Zulkefli Mansor, and Kamsuriah Ahmad, "Investigating factors affecting knowledge management practices in public sectors," *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 11, pp. 205-212, 2019. <https://doi.org/10.14569/IJACSA.2019.0101128>
- [23] I. S. Tvoroshenko, and V. O. Gorokhovatsky, "Modification of the branch and bound method to determine the extremes of membership functions in fuzzy intelligent systems," *Telecommunications and Radio Engineering*, Vol. 20, No. 78, pp. 1857-1868, 2019. <https://doi.org/10.1615/TelecomRadEng.v78.i20.80>. Available online: <http://www.dl.begellhouse.com/ru/journals/0632a9d54950b268,55c1ca0a73cc00be,339e67477ec4f851.html>

# Adaptive Hybrid Synchronization Primitives: A Reinforcement Learning Approach

Fadai Ganjaliyev

School of IT and Engineering, ADA University  
Baku, Azerbaijan

**Abstract**—The choice of synchronization primitive used to protect shared resources is a critical aspect of application performance and scalability, which has become extremely unpredictable with the rise of multicore machines. Neither of the most commonly used contention management strategies works well for all cases: spinning provides quick lock handoff and is attractive in an undersubscribed situation but wastes processor cycles in oversubscribed scenarios, whereas blocking saves processor resources and is preferred in oversubscribed cases but adds up to the critical path by lengthening the lock handoff phase. Hybrids, such as spin-then-block and spin-then-park, tackle this problem by switching between spinning and blocking depending on the contention level on the lock or the system load. Consequently, threads follow a fixed strategy and cannot learn and adapt to changes in system behavior. To this end, it is proposed to use principles of machine learning to formulate hybrid methods as a reinforcement learning problem that will overcome these limitations. In this way, threads can intelligently learn when they should spin or sleep. The challenges of the suggested technique and future work is also briefly discussed.

**Keywords**—*Spinning; sleeping; blocking; spin-then-block; spin-then-park; reinforcement learning*

## I. INTRODUCTION

While multicore architectures bring new opportunities for parallel software, they also present certain challenges, such as the choice of contention management strategy, which is crucial for the performance and scalability of parallel applications. The diversity of computing environments and unpredictability of application behavior makes this issue even more severe.

General synchronization techniques used to provide concurrent access of threads to shared objects are spinning (busy waiting) or blocking (descheduling the waiting thread). The other approaches are some combinations of the two, such as spin-then-block or spin-then-park. When spinning, to get the ownership of the shared resource, a thread continuously polls the resource until it becomes free, while in case of blocking, the thread relinquishes processor, thereby allowing other threads to utilize CPU. Spinning provides very quick lock handoff and is preferred in undersubscribed scenarios. However, in oversubscribed cases, every type of spinning can create scalability bottlenecks because it is highly CPU intensive by design. To mitigate this issue exponential backoff technique [1-3] inserts random delays between consecutive spins and queue-based protocols [4-9] spread contention among different memory locations. Still, in an overloaded system, spinning is inefficient because it wastefully burns

CPU cycles. Blocking, on the contrary, saves processor cycles by descheduling the contending thread even though context switches associated with the lock handoff phase (one to park out and another one to wake up) significantly add up to the critical path. Besides, frequent sleeps and wakeups can make the scheduler very busy and deteriorate its performance.

To balance tradeoffs between spinning and blocking, hybrid, spin-then-block, and spin-then-park strategies are used where threads spin at low and average contentions and block when contention rises [2, 10-13]. These techniques provide quick lock handoff at moderate contentions, meanwhile avoids waste of CPU time because when contention rises, threads suspend themselves, and other threads can do useful work. However, these strategies do not eliminate parking out and waking up from the lock handoff phase. To remove scheduler interaction from the critical path, previous [14] and recent [15] research suggests to maintain system load and to park and wake up threads in bulk as load changes. In summary, these works address two main problems: 1) whether a thread should spin or sleep, and 2) how a thread should make sleeping decisions.

Threads can address these issues more elegantly. Instead of acting in a certain way deemed efficient at particular states of the system, a thread can take action (for example, sleep for a specific duration) and evaluate it by peeking at its outcome, which drives the thread to the goal it is trying to achieve. Eventually, a thread will have a set of state-action pairs that will not only allow it to act optimally at any state but also will help it to predict optimal behavior for future unseen cases.

In summary, this paper makes the following contributions:

- We show that system load cannot serve as the only criteria in sleeping decisions, as previous and recent research states.
- We show that a thread that follows either of the hybrid methods can be treated as an entity capable of learning optimal actions (spin or take a timed sleep) via interaction with the system.
- We formulate both spin-then-block and spin-then-park strategies as a reinforcement learning (RL) problem, which allows a thread to 1) learn when it should spin or sleep 2) adapt its behavior to changes in the system and 3) utilize learned experience to future cases.

The rest of this work is organized as follows. The next section briefly describes the hybrid methods and motivates the

described approach. Section 3 presents the suggested method. Section 4 discusses the opportunities and challenges of the approach and future work. Finally, Section 5 summarizes conclusions.

## II. BACKGROUND AND MOTIVATION

We motivate the need for intelligent learning of sleeping and spinning and provide background on reinforcement learning as applicable to hybrid primitives. Detailed descriptions are out of the scope of this paper, and the interested reader can refer to [10-13] for hybrid methods and to [16-18] for reinforcement learning.

### A. Hybrid Synchronization Primitives

The blocking method extends queue-based spinning protocols in two ways. In the case of the spin-then-block method [10, 11, 19, 20], once a thread enters the system, it either spins or blocks depending on the level contention on the lock. In the other case, a thread may not block right away but may spin for a while and then park itself out, which is the spin-then-park [2, 12-15] strategy. The issue both of the methods are addressing is whether spinning or blocking a better choice at a particular point of time.

Later, researchers suggested [14] to improve it further and move scheduler decisions off the critical path by decoupling contention management from load control. A separate control daemon thread periodically estimates the system load. A set of randomly selected threads is parked out or woken up in response to load change. A more recent work [15] extends this technique for NUMA architectures by maintaining a load metric per socket. In both of the works, system load is the driving factor when deciding about blocking and waking up threads that is done in bulks. Fig. 1(a) and Fig. 1(b) illustrate these approaches. The latter work also addresses memory footprint challenges. But this is not the focus of this work.

Duration of sleep is important. When parked out threads wake up too early, they burn CPU cycles, whereas waking up later than expected lengthens the critical path (a thread still sleeps even though the lock is free). In Fig. 2(a) the lock is released at time  $t'$ . A contending thread that is currently sleeping wakes up at  $t'$  and spins until  $t''$ . Should it predict lock release time more accurately, it could have slept a bit more so that to wake up right before the lock is released, which would minimize the unnecessary burning of CPU time. In Fig. 2(b) a thread sleeps such that it wakes up much later the lock is released. The lock holder frees it at time  $t'$ , but the thread continues to sleep until  $t''$  which lengthens the critical path. This time the issue is different while the reason is the same. To avoid these issues, threads should be able to judge the consequences of their actions. If sleeping for the smallest possible amount of time yields unnecessary sleeping, then a thread should never sleep, no matter what other conditions are, in which case spinning is the only choice. Also, because the number of possible sleep durations is large, threads should be encouraged to transfer learning from one state to other similar states.

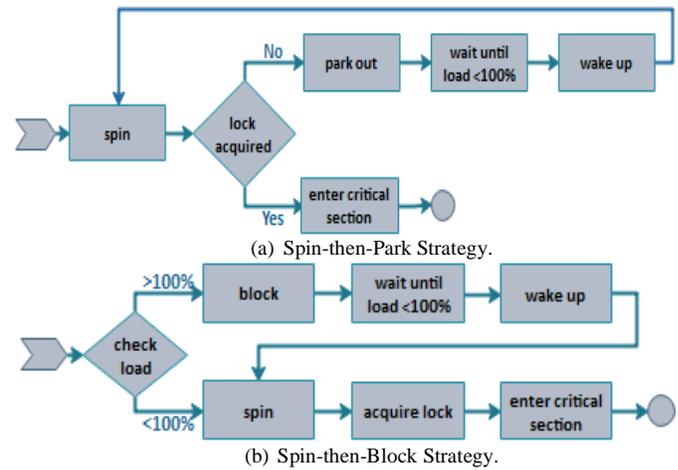


Fig. 1. Spin-then-Park and Spin-then-Block Strategies.

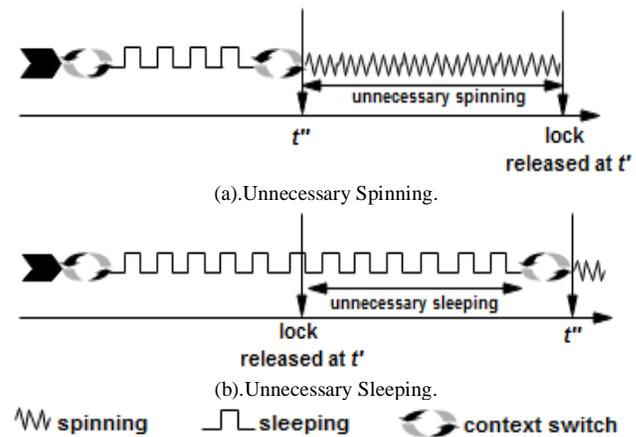


Fig. 2. Spinning and Sleeping Unnecessarily.

According to previous [14] and recent research [15] estimating system load in blocking decision eliminates scheduler interaction from the blocking phase and improves latency of the lock handoff phase. Threads sleep for a duration proportional to the overload metric value. However, there are other factors that should be taken into account when making sleeping decisions, such as, length of the queue of nodes created by contending threads. If a thread decides to take a sleep, then it should also take into account length of the queue. Each successor thread will hold the lock for the time it takes to execute the critical section. This means that for the same overload factor a thread should sleep for different durations depending on the length of the queue created by successor threads. Otherwise, it may result in unnecessary spinning or sleeping or both.

That is, the following conclusions can be made:

- When making sleeping decisions, both system load and number of successor contending threads should be considered. A thread should have both a global and a local view of its environment.
- To decide about spinning or sleeping, threads should be able to evaluate their actions.

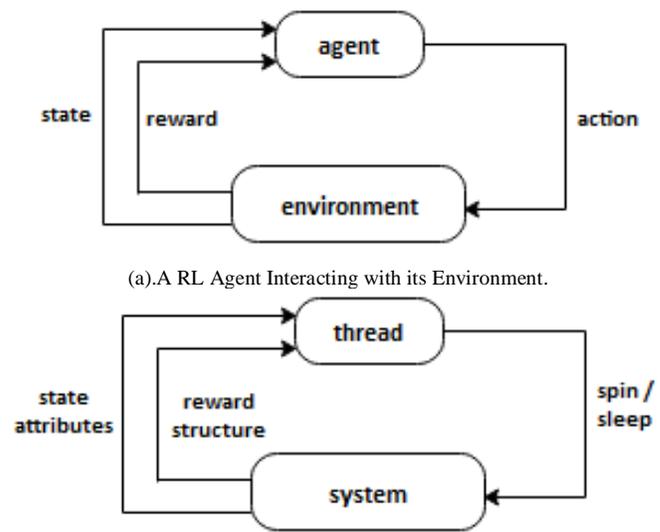
- As the system changes its behavior, a thread should be able to 1) adapt accordingly, and 2) use its experience to act optimally in the future.

### B. Reinforcement Learning and its Applicability to Hybrid Synchronization Methods

Once a thread enters the system, it has to choose one of several options to contend for the lock: spin, spin for some time, and then sleep or take a sleep. A thread prefers one of these options depending on some factors, such as the load of the system. Instead of just heuristically choosing one of the available actions, a thread may also retrieve feedback from it to determine how good was the action it took at this particular state of the environment. The thread may then remember this and use it to act more intelligently and efficiently in the future. By behaving in this manner, the thread collects a set of state-action pairs, and whenever it enters any state, it chooses the action that is the best one at this state. The feedback must correctly reflect the goal that the thread is trying to achieve. Thus, a thread can be treated as a RL agent whose aim is to learn to behave optimally in an uncertain environment by interacting with it.

Situated between supervised and unsupervised learning, RL is an area of machine learning that deals with sequential decision making problems in which the feedback is limited [18]. Basic concepts of RL are agents, environments, actions, states, rewards, and policies. An agent represents the learning decision maker. The environment is where the agent learns by taking actions, which is the set of all possible moves an agent can make. Whenever the agent takes action, the environment responds to it by placing the agent in a certain state which is an instantaneous situation where the agent finds itself i.e., the environment's input is agent's current state and action in that state and the output is the new state and the corresponding reward. The reward, which is an immediate scalar signal, is the feedback the agent receives as a consequence of its action and helps to measure how good the action at that state was. It essentially evaluates the agent's action in the current state. The policy is a function that maps states to actions and effectively is a strategy that the agent exploits to decide about the next action in the current state. The goal of the agent is to learn the best policy i.e., the policy that maximizes the long-term reward. Fig. 3(a) represents the agent-environment interaction.

Fig. 3(b) shows how a thread's interaction with the system fits within this framework. The thread (the agent) decides to spin or sleep. Note that now it does not block but takes a timed sleep since one of the purposes is to remove lock handoff from the critical path. Consequently, the thread receives a reward (for example, the reward can be modeled as the cost of lock acquisition in terms of CPU cycles). As a result, the thread finds itself in a new state where it can take a different action (for example, sleep for a different duration). Eventually, the thread collects a set of state-action pairs which it then can utilize.



(a).A RL Agent Interacting with its Environment.

(b).A Thread as a RL Agent Interacting with its System.

Fig 3. RL Framework and Thread as a RL Agent.

RL problems are formalized using Markov Decision Processes (MDP). MDP is a tuple  $\langle S, A, T, R, \gamma \rangle$  with the Markov property, that is, the current state provides sufficient statistics about the future, and thus, all the past information can be discarded.  $S$  and  $A$  is a finite state and action spaces correspondingly.

$T$  is a transition function, which is the probability distribution over the state space for each states  $s \in S$ , and action  $a \in A$ .  $R$  is an expected reward for taking action in a state and  $\gamma$  is a discount factor. The discount factor guarantees, from one hand, that the algorithm converges, and on the other hand, tells the agent how important are immediate rewards compared to future rewards. The closer it is to 0, the less important future rewards are compared to immediate rewards, whereas values closer to 1 make future rewards count as much as immediate rewards. Transition function and reward together completely define the model of the environment. Fig. 4 shows an example of a non-deterministic MDP state diagram that models an environment with two states  $S_1$  and  $S_2$ , and in both states, the agent has two actions  $A_1$  and  $A_2$ . If the agent, for example, in state  $S_2$ , takes action  $A_1$ , it transitions to state  $S_1$  and receives a reward of 9 with a probability of 0.3 and finds itself in the same state and receives a reward of 4 with a probability of 0.7. It is important to note that the next state is not determined just by the agent's action in the current state but also depends on the behavior of the environment. For example, in the context of hybrid synchronization methods, the next state (we suppose, at least scheduler load can be one of the state attributes) partly depends on the action of the thread in the current state and partly on the scheduling decisions which is not under the thread's control.

When the model of the environment is known, the agent can derive the optimal policy by using the transition and reward functions. However, in the absence of these functions (which is the case with hybrid methods), the agent needs to interact with the environment and observe its responses, in which case the algorithm is referred to as model-free. In this

case, the agent derives the optimal policy without using neither transition function nor the reward function. A popular model-free algorithm used for estimating optimal policy is Q-Learning [17, 18] (more in the next section), which associates a value with each state-action pair and derives optimal policy from these values. Typical issues that the agent needs to address are the following.

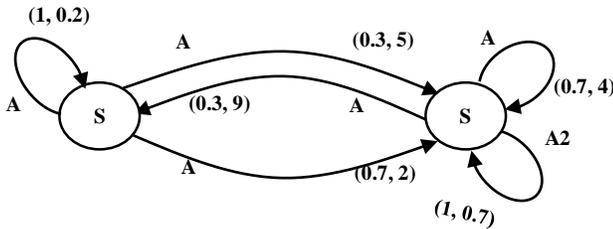


Fig 4. A Non-Deterministic MDP State Diagram with Two States and Two Actions.

**Temporal credit assignment.** The agent tries to collect the largest amount of reward in the long run. For each received reward, it needs to determine whether it leads to the desired outcome. An action that yields high immediate reward may lead the agent to an undesirable state; in other cases, taking an action that yields no reward may seem undesirable first but may be critical to driving the system to the state with the highest reward. For example, some threads may take sleep for a specific duration and get a high reward for this action because it wastes fewer CPU cycles. However, if too many threads do so when the scheduler is very busy, then the scheduler can quickly become a bottleneck. Therefore, to act accurately, the agent needs careful planning.

**Exploration vs Exploitation.** To achieve its goal, the agent needs to interact with its environment to gather more information that may lead to an optimal policy, and at the same time, it needs to exploit at best information it has found so far. An excessive investigation might lead to the best policy, but it causes long learning periods, whereas too little investigation may have the agent to accept a suboptimal policy early on.

**Generalization.** The state space can be exponentially large, and the agent might need to visit a huge number of possible states to try actions and evaluate rewards at those states. Besides, the agent may not have a chance to visit the same state twice over its lifetime. In such cases, the agent has to generalize of experience learned in previous states to new states.

### III. RL-BASED HYBRID SYNCHRONIZATION METHODS

Both spin-then-block and spin-then-park methods can be formulated as a RL problem. As stated above, both methods aim to decide whether spinning or sleeping is a more efficient choice at a particular state of the system. If a thread now acts as a decision maker, then it can take any of the available actions. Therefore, it will be agnostic about which hybrid primitive it follows. It is simply trying to find out best (from a

reward perspective) actions at certain states. For example, if a thread spins for some time and quickly acquires the lock, then it remembers this choice at states. In a different state, it may notice that spinning does not result in high reward and will prefer a timed sleep.

#### A. Formulation of Hybrid Primitives as a RL Problem

Now the reward structure needs to be defined, as well as state and action spaces that threads can use to decide about their optimal behavior. Here, also described the techniques can be used to overcome the challenges mentioned in the previous section are also.

**Action.** A thread can either spin or take a timed sleep but never a combination of both with spinning being first. Initial spinning of the spin-then-park strategy is a waste of CPU cycles now, since if a thread would be scheduled out until it gets the lock, then it should not have spun but rather have taken a timed sleep when it entered the system. To amortize the cost of scheduling out and waking up threads can park out and wake up in batches as previous [14] and recent [15] research suggests.

**Reward.** The goal of a thread is to acquire the lock in the cheapest and fastest way. If a thread manages to acquire the lock solely by spinning, then it receives a positive reward, otherwise, a negative reward, in which case it would be scheduled out by the scheduler. Next time the thread visits the same or a similar state, it should take a timed sleep and consequently receives a positive reward equal to the duration of the sleep if, as a result, the thread did not sleep unnecessarily (still slept while the lock was free). Eventually, it may sleep for even more amount of time to minimize spinning at that state. Otherwise, the thread receives a negative reward of the same magnitude (to discourage it from sleeping for this duration at this state again because it will add up to the critical path which has to be avoided). Hence, the more a thread sleeps, the more reward it receives, given that the sleep does not result in sleeping unnecessarily. To determine whether it has slept more than necessary, the thread can check whether it has at least one failed spin after waking up. If upon waking up, the thread grabs the lock as a result of the first spin that means the lock was free by the time it woke up, in which case the thread is punished. Acquiring the lock solely by spinning (which should always be possible in case of undersubscribed situation and relatively fewer number of successor threads) should yield the highest reward to encourage the thread to prefer spinning in the first place. Reward structure is shown in Fig. 5.

**State.** As recent research indicates [15] system load plays an important role in deciding about sleeping: a thread should sleep only in oversubscribed cases and should never do so in undersubscribed situations. The duration of the sleep should be proportional to the overload factor. However, as explained in the previous section, even in undersubscribed cases, a thread may prefer sleeping if the latter is cheaper than spinning (which could be because of a huge number of threads contending for the lock). Hence, state attributes are system load and the number of successor threads contending for the lock.

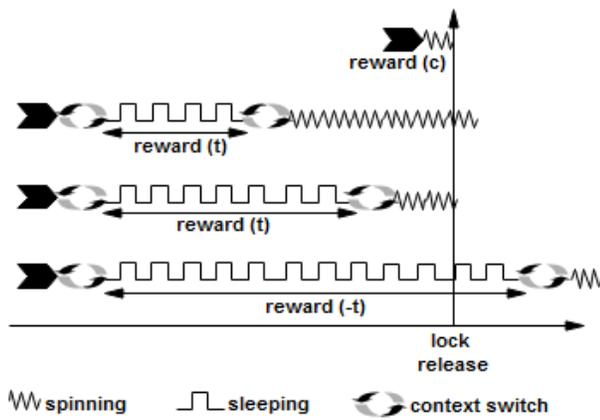


Fig 5. Reward Structure for a Thread as an Agent.

### B. Solving Challenges of the Agent

**Estimating actions.** The temporal credit assignment problem can be addressed by one of the well-known algorithms in RL Q-learning [17, 18], where the agent learns Q-values associated with each state-action pair. One of the key properties of Q-learning is that to derive the optimal policy  $\pi^*$ , the agent needs to determine the best action as defined by Q-value. Q-value of a state-action pair (s, a) under policy  $\pi$  is the reward for taking action a in state s plus the sum of discounted future rewards if the policy  $\pi$  is followed thereafter. The agent learns all Q-values for a particular state. Therefore, if at each state, the agent chooses the action that has the largest Q-value, then it effectively follows the optimal policy.

As threads spin or sleep, they continuously update estimates of Q-values based on the rewards they receive. When a thread executes action  $a_c$  in state  $s_c$  it receives a reward r, transitions to a new state  $s_n$ , and chooses an action  $a_n$ . The Q-value associated with taking action  $a_c$  in state  $s_c$  can then be updated by an error using the SARSA [18] update rule:

$$Q(s_c, a_c) \leftarrow Q(s_c, a_c) + \alpha [r + \gamma Q(s_n, a_n) - Q(s_c, a_c)] \quad (1)$$

Recall that  $\gamma$  is a discount factor and determines the importance of future rewards. The learning rate  $\alpha$  (or step size) determines how quickly the agent learns. Setting it to 0 means Q-values are never updated, that is, the agent does not learn at all, whereas higher values of it facilitate faster learning periods. In the context of hybrid methods, presumably very high values of  $\gamma$  for example,  $\gamma = 0.95$ , should work quite well. This is because a thread is highly agnostic about the effect of its action on the future state of the system. But the value for  $\gamma$  should be experimentally tuned for good performance. The SARSA rule is guaranteed to converge to the optimal policy, assuming that each state can be visited by the agent infinitely often.

**Addressing exploration-exploitation tradeoff.** If a thread never chooses certain actions in a given state, it would not be able to learn the associated Q-values. Even if the optimal policy has been already learnt, the dynamics of the system can make the current policy obsolete. Furthermore, even though threads are not willing to spend much time on learning, they

have to try different actions in a given state to evaluate corresponding rewards. Therefore, threads must continuously explore their environment while at the same time, utilize the best policy they have found so far.

To balance exploration with exploitation, one can implement a widely used  $\epsilon$ -greedy [18] action selection technique. The agent (the thread) takes the actions that are optimal most of the time but to try more actions and potentially find ones with a higher reward, introduces randomness ( $\epsilon$  factor). Threads randomly take actions with probability  $\epsilon$  which intuitively should be set to small values to guarantee that they continue trying different actions in each state, while the majority of the time utilizing the best policy they have derived. It is also important to note that after a while, it is possible to gradually decrease the  $\epsilon$  probability so that exploitation prevails, and the model converges to an optimal policy.

**Efficient generalization and quantization.** Model-free RL techniques assume that Q-values can be stored in a look-up table with one entry for each. However, when the state space is large, the issue is not only excessive storage requirements for storing Q-values but also time to accurately maintain these values. This problem is known as the curse of dimensionality and can be addressed in two ways.

One way consists of discretizing the state space into a smaller number of cells. In this case, all states within each cell are aggregated and linked to a single Q-value. However, if the discretization of the state space is coarse-grained, then some states can be hidden, preventing the agent from learning the optimal policy. On the other hand, a fine-grained discretization may result in too many cells, and the agent may not be able to generalize, and the amount of training data will increase.

CMAC [21] is a computationally efficient generalization and resolution technique with extremely fast learning capability and the special architecture, which makes it effective from an implementation perspective. It takes an arbitrary number of state variables and lays axis-parallel rectangles over them, known as tilings (as shown in Fig. 6). Overlapping partitions are called tiles. The number of the tiles and the widths of the tilings are generally set at design time.

The tilings are offset from each other by the same amount and maintain the weights of each of its tiles. The center of each tile determines which state values activate which tiles. The method computes a value of any given point as a sum of the weights of the tiles, one per tiling. That is, to calculate a Q-value for a given state, all values from tiles sharing the center are summed up. Analogously, an update to a value will heavily affect nearby points and not as heavily farther points. The function approximation is trained by adjusting the weights of each of the involved tiles. In this way, CMAC achieves the quantization of continuous state space into tiles while retaining the capability to generalize by means of several overlapping tilings. Other function approximation methods, such as radial-basis, instance, and case-based approximators [22], are not suitable candidates in the context of hybrid synchronization methods.

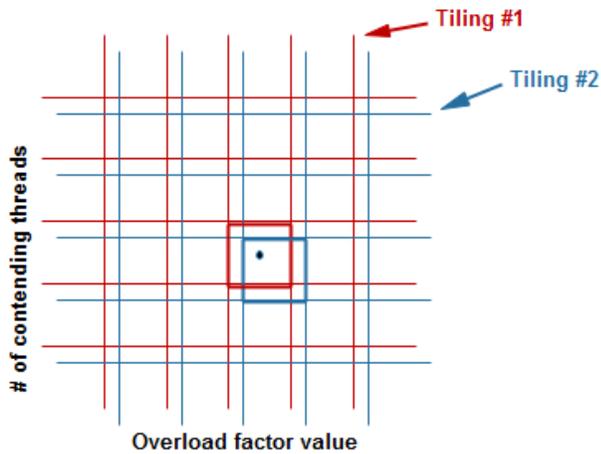


Fig. 6. CMAC using Two Overlapping Tiles for Efficient Resolution. Two Tilings are shown to simplify the Figure.

#### IV. DISCUSSIONS AND FUTURE WORK

Recently, it has been suggested [23] to formulate the spin-then-block method as a reinforcement learning problem. However, it still leaves much room for improvement, which was the focus of this work. First, the reward structure has been refined. Secondly, it is shown that the same idea could be applied to the spin-then-park strategy. Thirdly, it is suggested to take system overload factor as one of the state attributes (rather than the only number of currently running and waiting threads on CPU), and that action can be a decision of a batch of threads. Finally, the techniques and algorithms that are suited well in the context of hybrid methods for overcoming issues faced by the agent are described (the thread). Effectively, the technique presented in [14] and [15] is wrapped into a RL framework, but this idea has been arrived at independently.

Although the approach presented here promises competitive results, there are certain challenges that should be considered. Threads lifetime can be very short and adding policy related updates can lengthen it. One way to overcome this issue is to run updates not for every action of every thread but every few actions or every few time units. Another solution to this problem is to run additional helper threads to monitor, track and optimize threads behavior.

The second issue is related to the number of policies. When many lock instances are involved, the number of policies to be maintained can be huge. Two threads contending for different locks may exploit the same policy if the length of the critical section protected by these locks is the same. That is, the method can significantly reduce the number of policies to be maintained if policies can be clustered on the length of the critical section. Threads then will maintain a fewer number of policies, one per cluster. Future experiments will reveal more details on this.

Currently, no experiments have been set up to evaluate the suggested approach. Nevertheless, the theoretical argumentations presented here allow having a base to claim that the technique, if implemented, will allow application and system programmers to avoid the burden of balancing

spinning and blocking. The method will monitor itself and optimize and adapt to the dynamics of the system.

The primary goal of the near future work is to test the presented method and compare it with the state-of-the-art implementations of hybrid primitives, such as [15]. Also, extensive experiments will reveal the limitations of the suggested technique and cases where it may not be applicable.

Another part of the future work is to apply this idea for the case of NUMA machines, where threads are encouraged to spin locally rather than remotely. For example, the state can be modeled as the load level of the interconnect module. Threads can be given larger rewards if they spin locally and smaller rewards if spinning is remote (remote spinning cannot be eliminated completely, otherwise threads, ultimately, will spin only locally and will not make progress). In this way, threads can learn that local spinning is preferred whenever they are about to start contending for the lock.

#### V. CONCLUSION

Designing a hybrid synchronization primitive that performs well in both under- and oversubscribed scenarios is challenging. In this work, a RL based approach for implementing hybrid synchronization methods (namely, the spin-then-block and spin-then-park strategies) has been presented. It is suggested to make use of principles of machine learning, which a more generic approach. It may release the application developers and system programmers from choosing the appropriate contention management strategy and will optimize and adapt itself to the system as it changes its behavior.

#### ACKNOWLEDGMENT

This work has been carried out at the Center of Data Analytics and Research at ADA University.

#### REFERENCES

- [1] Facebook. A persistent key-value store for fast storage environments, <http://rocksdb.org>, 2012.
- [2] X. Leroy, The open group base specifications, <http://pubs.opengroup.org/onlinepubs/9699919799>, no. 7, 2016.
- [3] Torvalds L. The linux kernel archives, <https://www.kernel.org>, 2017.
- [4] I. Calciu, D. Dice, Y. Lev, V. Luchangco, VJ. Marathe, N. Shavit, "NUMA-aware reader-writer locks," Proceedings of Eighteenth ACM Symposium on Principles and Practice of Parallel Programming, pp. 157–166, 2013.
- [5] M. Chabbi, J. Mellor-Crummey, "Contention-conscious, locality-preserving locks," Proceedings of Twenty-First ACM Symposium on Principles and Practice of Parallel Programming, vol. 22, pp. 1–14, 2016.
- [6] D. Dice, VJ. Marathe, N. Shavit, "Flat-combining NUMA Locks," Proceedings of Twenty-Third Annual ACM Symposium on Parallelism in Algorithms and Architectures, pp. 65–74, 2011.
- [7] D. Dice, VJ. Marathe, N. Shavit, "Lock cohorting: a general technique for designing NUMA locks," Proceedings of Seventeenth ACM Symposium on Principles and Practice of Parallel Programming, pp. 247–256, 2012.
- [8] JP. Lozi, F. David, G. Thomas, J. Lawall, G. Muller, "Fast and portable locking for multicore architectures," ACM Transactions on Computer Systems, no. 33, vol. 4, pp. 1–62, 2016.
- [9] V. Luchangco, D. Nussbaum, N. Shavit, "A hierarchical CLH queue lock," Proceedings of Twelfth International Conference on Parallel Processing, pp. 801–810, 2006.

- [10] L. Boguslavsky, K. Harzallah, A. Kreinen, K. Sevcik, A. Vainshtein “Optimal strategies for spinning and blocking,” *Journal of Parallel and Distributed Computing*, n. 21, vol. 2, pp. 246-254, 1994.
- [11] H. Franke, R. Russell, MK. Fuss, “Futexes and furwocks: fast userlevel locking in linux,” *Proceedings 2002 Ottawa Linux Summit*, pp. 479-495, 2002.
- [12] I. Molnar, Linux rwsem, <http://www.makelinux.net/ ldd3/chp-5-sect-3, 2006>
- [13] I. Molnar, D. Bueso, “Generic mutex subsystem, <https://www.kernel.org/doc/Documentation/locking/mutex-design.txt>,” 2016.
- [14] FR. Johnson, R. Stoica, A. Ailamaki, TC. Mowry, “Decoupling Contention Management from Scheduling,”. *Proceedings of Fifteenth ACM International Conference on Architectural for Programming Languages and Operating Systems*, pp. 117-128, 2010.
- [15] S. Kashyap, C. Min, T. Kim, “Scalable NUMA-aware blocking synchronization primitives,” *Processdings of the USENIX Conference on Usenix Annual Technical Conference*, pp. 603-615, 2017.
- [16] D. Bertsekas, *Neuro Dynamic Programming*: Athena Scientific, 1996.
- [17] T. Mitchell, *Machine Learning*: McGraw-Hill, Boston, 1997.
- [18] R. Sutton, A. Barto, *Reinforcement Learning*: MIT Press, 1998.
- [19] J. Mauro, R. McDougall: *Solaris Internals: Core Kernel Components*: Sun Microsystems Press, 2001.
- [20] JK. Ousterhout, “Scheduling techniques for concurrent systems,”. *Proceedings of Third International Conference on Distributed Computing Systems*, pp. 22-30, 1982.
- [21] R. Sutton, “Generalization in reinforcement learning: Successful examples using sparse coarse coding,”, *Proceedings of Eighth International Conference on Neural Information Processing Systems*, pp. 1038-1044, 1996.
- [22] JC. Santamaria, RS. Sutton, A. Ram, “Experiments with reinforcement learning in problems with continuous state and action spaces,”. *Adaptive Behavior*, n. 6, vol. 2, pp. 163–217, 1997.
- [23] F. Ganjaliyev, “Spin-then-sleep: A machine learning alternative to queue-based spin-then-block strategy,” *International Journal of Advanced Computer Science and Applications*, n. 10, vol. 3, pp.605-609, 2019.

# Creating Knowledge Environment during Lean Product Development Process of Jet Engine

Zehra C. Araci<sup>1</sup>, Muhammad Usman Tariq<sup>3</sup>  
Abu Dhabi School of Management  
Abu Dhabi, UAE

Jan H. Braasch<sup>4</sup>  
Dubai Airports  
Dubai, UAE

Ahmed Al-Ashaab<sup>2</sup>  
Cranfield University  
Cranfield, UK

M. C. Emre Simsekler<sup>5</sup>  
Department of Industrial and Systems Engineering  
Department, Khalifa University of Science and Technology  
Abu Dhabi, UAE

**Abstract**—Organizations invest intense resources in their product development processes. This paper aims to create a knowledge environment using trade-off curves during the early stages of the set-based concurrent engineering (SBCE) process of an aircraft jet engine for a reduced noise level at takeoff. Data is collected from a range of products in the same family as the jet engine. Knowledge-based trade-off curves are used as a methodology to create and visualize knowledge from the collected data. Findings showed that this method provides designers with enough confidence to identify a set of design solutions during the SBCE applications.

**Keywords**—Knowledge creation and visualization; knowledge management; new product development; lean product development; set-based concurrent engineering; trade-off curves; aircraft engine noise reduction

## I. INTRODUCTION

Due to rapid technological changes, organizations are under pressure to be agile enough in order to respond to the fast-changing demand [8]. This agility can be gained by improving its product development activities. However, designers face several challenges, especially during the early stages of developing a new product [1]. These challenges could be addressed by the lean product development (LeanPD) approach [22]. During the LeanPD process, it is essential to have the right knowledge environment in order to achieve a robust optimal design [27]. Trade-off curves (ToCs) provide this environment by creating and visualizing the knowledge that is based on the physical insights of the product as well as experienced data (e.g., outcomes of R&D, data from successful or failed projects) [13].

There are several challenges that the manufacturing industry faces during their product development processes [26]. Some of these challenges are reworking, late design changes, communication challenges between departments, and, most importantly, lack of knowledge [15]. Having the right knowledge environment supports designers or product developers to increase the project success rate, to reduce rework during product development, and to reduce manufacturing costs that are caused by inaccurate design solutions [4]. In order to create such a knowledge

environment, trade-off curves are effective LeanPD tools to be used throughout the product development processes [3].

Trade-off curves are primarily used by Wright Brothers, who succeeded in operating an aircraft for the first time [13]. Air transportation has gained significant popularity, and the form of aircraft considerably improved since then. The efficiency of aircraft production has increased due to technology changes. But most importantly, because of the knowledge-gained throughout all these years [9]. On account of the developments, recently, many international and even domestic flights are preferred as an alternative means of transportation compared to road transport [10]. However, this new habit causes challenges from environmental aspects such as high noise levels [2]. Environmental concerns and statutory regulations push airlines to operate low noise aircraft at takeoff, especially during the night flights [16]. The jet engine is known as the major source of an aircraft noise [17]. Although there are studies on reducing the jet engine noise, there is limited progress achieved so far [18].

This paper aims to create an initial knowledge environment for designers by using trade-off curves in order to identify a set of possible design solutions from the previous projects of an aircraft jet engine [19]. Having such a design-set provides designers with sufficient knowledge during the SBCE applications of achieving a final optimized solution, which is expected to have the lowest noise level available in the market [2].

This paper consists of different research steps. First, the authors reviewed the literature to identify the role of trade-off curves within SBCE processes and to identify the possible causes of the high noise levels in an aircraft engine, specifically civil aircraft. The literature review findings are presented in Section 2. After the literature review, the authors conducted a case study of a family of civil aircraft jet engines as presented in Section 3. Publicly available data was collected and reflected on the form of trade-off curves by using the process of generating trade-off curves [3]. Finally, analysis of the trade-off curves supported authors in developing a set of possible design solutions to be used in the SBCE process. Further stages of the SBCE applications are

not the scope of this paper. Following the case study, discussion, and future work are presented in Section 4, and the research is concluded in Section 5.

## II. LITERATURE REVIEW

### A. Trade-off Curves within SBCE Concept

Open, global markets have been a key driver of growth and profit for manufacturing companies over the last 75 years. This trend can be expected to continue despite recent political developments in some countries. However, with access to international means of production and markets also comes international competition [4]. Combined with the increasing digitalization that lowers entry barriers, this has created pressure on companies to provide high-quality products and services in an environment of often rapidly changing demand. This need for flexibility and short time-to-market timeframes makes an efficient product development (PD) process a key success factor [29]. Demand cycles, especially in consumer markets, are now often characterized by extremely short durations while, at the same time, carrying huge revenue potentials. Sustaining market share (or even improving) it depends on the timely development of products that service this short-lived demand, and companies with such capabilities have a distinctive, differentiating competitive advantage. Efficient PD capabilities rely on several management systems, tools, and techniques that allow companies to leverage organizational knowledge and continuously improve processes [21]. For companies, the efficient development of new products, as well as access to organizational knowledge, have become important assets [30].

In more recent times, Toyota has used trade-off curves as a knowledge visualization tool in order to facilitate their SBCE application [25]. There, trade-off curves are part of "jidoka," which refers to a visual management technique that Toyota integrated into their PD process from lean manufacturing [20]. Now, they visually display subsystem knowledge in a graph so that engineers can explore the design space [31] and evaluate design alternatives [14]. Moreover, in a lean product development context, trade-off curves avoid the reinvention of previously considered design solutions during prototyping [32]. Hence, engineers save time that they can spend on new and innovative solutions [11].

Previous research exhaustively demonstrated how trade-off curves could be generated and utilized throughout the stages of set-based concurrent engineering [4]. This paper aims to show an application of trade-off curves in the early stage of the design of a complex product, which is a turbofan jet engine [1].

### B. Civil Aircraft Noise Challenge

Aircraft noise is a significant issue, and it has a direct effect on human hearing [6]. It is a well-established fact that it can cause hearing problems in humans [7]. Unwanted noise can create problems that can distract communication, reduce the quality of communication, and increase stress. Aircraft noise compatibility has been a serious issue that reduced the growth of commercial aviation [12]. Already several European airports have reached their maximum environmental load capacity before starting the use of runway and other

infrastructure. One of the important challenges faced by environmental management authorities and the advisory council for Aeronautics research is to reduce the current noise of aircraft by 50% (-10db/operation). Different solutions have been tested to control the overall noise at airports. However, the noise in the surroundings of the airports has been trouble and remains high at takeoff and landing time [7].

### C. Work Principles of a Turbofan Jet Engine

A review of the related literature showed that certain parameters influence the reduction of takeoff noise. Additionally, understanding the physical/technical details about a jet engine facilitated identifying parameters to focus on [28]. A jet engine is a key component of most modern aircraft, as it provides, by jet propulsion, required to reach speeds that enable heavier-than-air flight. The jet engine's most common form is the turbofan engine [23]. A propulsive force is generated by accelerating the entering gas (air) between the entrance and the exit of the engine [24]. The "General Thrust Equation" defines thrust as the difference between the product of mass flow at the exit ( $m_e$ ) and the gas speed differential between exit and entry ( $V_e - V_0$ ), and the product of mass flow at the entrance ( $m_0$ ) and the gas speed differential ( $V_e - V_0$ ). By definition, all air entering the engine must also leave it, from which follows that  $m_e = m_0$  at all times.

General Thrust Equation:

$$F_{Thrust} = m_e(V_e - V_0) - m_0(V_e - V_0) \quad (1)$$

The essence of the General Thrust Equation is that additional thrust can be generated in two ways:

- 1) Increasing the mass flow rate  $m_e = m_0$
- 2) Increasing the speed differential of the gas ( $V_e - V_0$ )

## III. CASE STUDY: JET ENGINE NOISE REDUCTION

### A. Customer and Design Requirements

The knowledge is gained through the literature review and technical aspect of the product, as mentioned above. Based on this knowledge, the authors defined essential requirements as decision criteria:

- 1) *Low noise*: The takeoff noise level of the new product should be lower than the noise levels of existing products.
- 2) *Reliability*: The new product should operate 24/7 without significant downtime.
- 3) *Durability*: The new product should be durable enough to be able to operate on an aircraft capable of carrying 150 passengers. As all aircraft must be able to fly with only half their engines operating, each engine on a twin-engine aircraft must be capable of carrying all passengers.
- 4) *Cost*: Fuel consumption should not be higher than the consumption of existing turbofan jet engine solutions.

In order to visualize the requirements by using trade-off curves, the authors also identified the parameters related to the requirements as follows: Take-off noise, maximum takeoff mass (MOTM), bypass ratio, and thrust. Table I displays the parameters and their conflicting relationships based on experts' opinions.

**B. Key Design Parameters**

Data for the identified parameters were collected from publicly available sources. Fifty-five different types of jet engines are demonstrated in the trade-off curves in this section. It is worth to mention that the collected data set is real and experienced from successfully finished projects rather than computer-generated data.

Trade-off curves were generated by using Minitab as analysis software. Data analysis has been performed in order to see the correlations between the different parameters, as indicated in Table I.

**C. Takeoff Noise vs. Thrust, Bypass Ratio, MTOM**

The metric for takeoff noise level is EPNdB, which means effective perceived noise in decibel, and the metric for thrust is defined as newton (N). Fig. 1 shows a positive correlation between the noise and thrust, which means that higher thrust causes higher takeoff noise. However, there is one design solution found with a high thrust (284,500 N) but relatively low noise (90.1 EPNdb) compared to other design solutions.

Simply, bypass ratio is the mass flow of air going into the engine core and burnt divided by the mass flow of air going around the engine core and exiting the engine as it is [5]. For example, a bypass ratio of 10:1 means that 11 units of air are drawn into the engine, ten units go around (bypass) the engine core and exit, but 1 unit goes through the engine core and burn. A high bypass ratio is a desirable factor in this case as it is believed that high bypass ratio results with a low noise engine. However, when we analyzed the data for bypass ratio and takeoff noise in Fig. 2, it was found that there is no significant relationship between these two parameters. A design solution has been defined with a high bypass ratio but relatively low takeoff noise.

TABLE I. CONFLICTING RELATIONSHIPS BETWEEN THE DESIGN PARAMETERS OF A LOW NOISE JET ENGINE

No.	Parameters and Relationships	Conflicts between the design parameters
1	Thrust vs. Takeoff Noise Level	Engine noise was defined as 100% of the aircraft takeoff noise. As aircraft take off with full power, thrust and fuel consumption are at a maximum. It was surmised that the noise level is related to the amount of thrust generated.
2	Bypass Ratio vs. Takeoff Noise Level	In order to achieve high thrust but low noise, the bypass ratio of the engine can be increased. In order to increase the bypass ratio, the fan diameter should be increased so that the air intake increases. However, a larger fan results in the engine being heavier, and this leads to a higher bypass ratio with a higher thrust but heavier aircraft. Consequently, the possibility of reducing aircraft engine noise becomes challenging.
3	Maximum Takeoff Mass (MTOM) vs. Takeoff Noise Level	As mentioned above, in general, larger and heavier aircraft produce more noise than lighter aircraft. By increasing the bypass ratio by increasing the fan diameter, the engine weight will also increase, which will increase the MTOM.

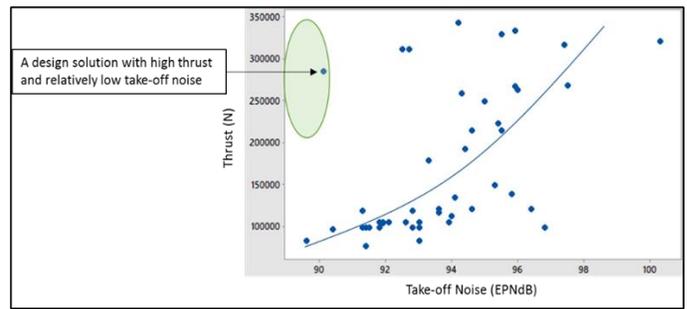


Fig. 1. Correlations between Takeoff Noise and Thrust.

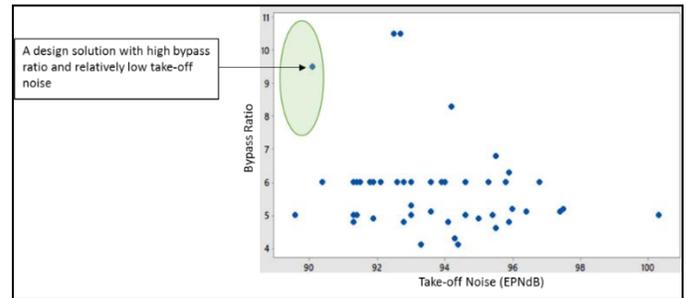


Fig. 2. Correlations between Takeoff Noise and Bypass Ratio.

Three-dimensional trade-off curve has been generated in order to understand if all these three parameters are related to each other; and, consequently, whether they are giving the same feasible design solution. As shown in Fig. 3, it was found that one engine has the potential to be considered within the design set in the early stage of product development. Further discussions are provided in the following section.

The authors carried some more investigations to understand the relationships between takeoff noise, thrust, and maximum takeoff mass (MTOM). MTOM is the weight of the aircraft in kg with an assumption that it operates full capacity (passengers and fuel). Fig. 4 shows that there is a strong positive relationship between thrust and maximum takeoff mass, which eventually means heavier aircraft requires more thrust than light aircraft. The effect of MTOM on takeoff noise has been investigated and displayed in Fig. 5 as a three-dimensional trade-off curve. It appears that MTOM does not have a significant impact on the noise; however, it may facilitate noise reduction indirectly. The next section provides more details about the findings and discussions.

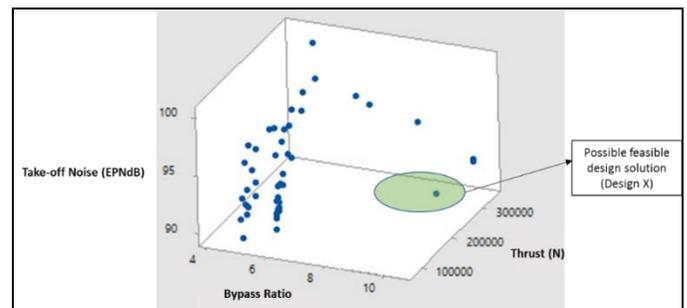


Fig. 3. Correlations between Takeoff Noise, Thrust and Bypass Ratio.

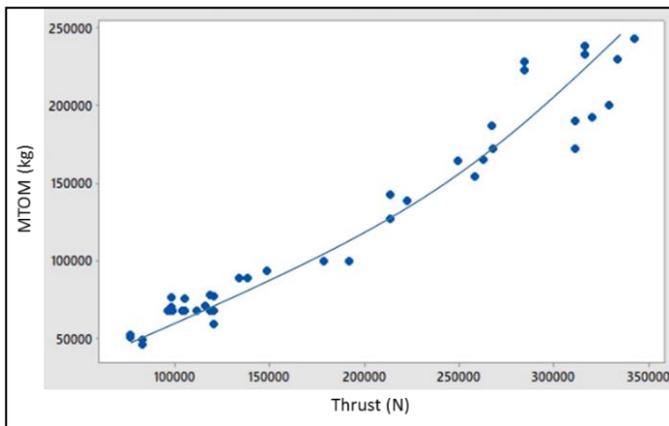


Fig. 4. Correlations between Thrust and MTOM.

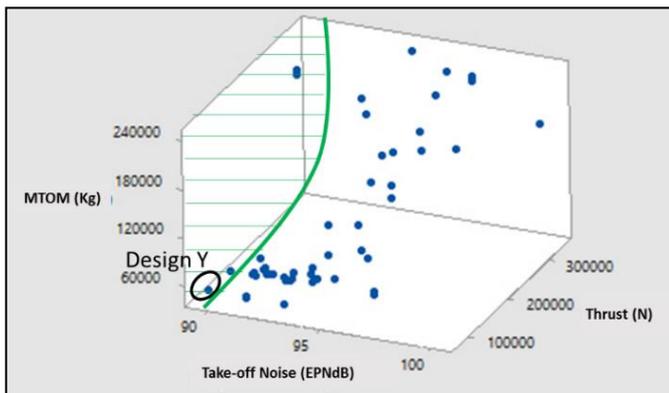


Fig. 5. Correlations between Takeoff Noise, Thrust, and MTOM.

#### D. A Set of Possible Design Solutions

After analysis of trade-off curves and as illustrated in Fig. 3 and 5, two design solutions (Design X and Design Y) are selected in order to evaluate whether they are eligible to be considered within the design set. Selected design solutions were investigated, and the data of these designs were analyzed. It was determined that some designs might be used to create new, potentially feasible designs by combining several characteristics from several existing designs. In order to be a potentially feasible solution, existing designs might need to undergo one of the following actions: (1) Minor modifications, (2) Major modifications, and (3) Complete re-design.

#### IV. FINDINGS AND DISCUSSION

After the authors focused on the data that shows the lowest takeoff noise level, which is 90.6 EPNdB for Design X. Then, noise-related vital features and characteristics of Design X have been investigated. Characteristic features of Design X are:

- Ultra-efficient, swept fan blades enable the quieter operation and optimal engine core protection.
- Full takeoff power is 3 dB quieter than the previous generation engine.

While it was understood from the ToCs that Design X could not be used as a whole system concept, the fan design might be an inspiring idea for a new design solution.

On the other hand, the authors focused on the data that shows the lightest engine with low noise in Fig. 5 (Design Y). A characteristic feature of Design Y is that it has a lightweight, hollow titanium wide chord fan for low noise and high efficiency. If the same material used in Design Y can also be used in a new design, this might decrease the engine weight of the new product design. Design Y engines power the Airbus A340 aircraft. This aircraft is configured with four engines. Four engines will be noisier than two engines. Besides, the total weight of an aircraft with four engines would be higher than an aircraft with two engines. Heavier aircraft also emit more noise. Therefore, Design Y could be reused if the fan diameter is increased (which reduces engine noise), and the number of engines is reduced from four to two. Furthermore, the passenger capacity of Design Y is more than 300 passengers, which is more than the customer requirement for passenger capacity in this study (150 passengers). Hence, it can be investigated if noise decreases when the Design Y engine is simulated for 150 passenger capacity.

Two design solutions, Design X and Design Y, can be considered as the basis of future designs. As explained above, a combination of their characteristics may lead to the emergence of a viable solution that meets customer requirements. Converting these designs to a useful solution requires the use of the SBCE process model, which is not the scope of this paper but can be considered as future work. Several variables have an impact on the noise level of a jet engine. Due to the available data and sources, only four variables were investigated in this research. Fan diameter might be another key design parameter to be investigated in the future.

#### V. CONCLUSION

Previous projects – successful or failed – are vital elements of the intellectual capital of an organization. It is inevitable that the right knowledge environment, based on real data, supports designers' decision-making throughout their product development activities. Utilizing trade-off curves creates such a knowledge environment. This paper demonstrated an application of trade-off curves, how to create an initial right knowledge environment at the very early stages of the product development of an aircraft jet engine. Without any prototyping or investing great resources, two feasible design solutions out of 55 were suggested for further investigation. The existing design solutions Design X and Design Y can be considered hypothetically to be reused, after modifications, in order to develop a design-set for the set-based concurrent engineering application of a low noise turbofan jet engine.

#### REFERENCES

- [1] Addo-Tenkorang, R., Helo, P., & Kantola, J. (2016). Engineer-To-Order Product Development: A Communication Network Analysis for Supply-Chain's Sustainable Competitive Advantage. In *Supply Chain Strategies and the Engineer-to-Order Approach* (pp. 43-59). IGI Global.
- [2] Antoni, J., Griffaton, J., André, H., Avendaño-Valencia, L. D., Bonnardot, F., Cardona-Morales, O., ... & Acuña, D. Q. (2017). Feedback on the Surveillance 8 challenge: Vibration-based diagnosis of

- a Safran aircraft engine. *Mechanical Systems and Signal Processing*, 97, 112-144.
- [3] Araci, Z.C., Al-Ashaab, A. and Maksimovic, M. (2016) 'Knowledge Creation and Visualisation by Using Trade-off Curves to Enable Set-based Concurrent Engineering', *The Electronic Journal of Knowledge Management*, 14 (1), pp. 75-88.
- [4] Araci, Z.C., Al-Ashaab, A., Lasisz, P.W., Flisiak, J.W., Mohd Maulana, M.I.I., Beg, N. and Rehman, A. (2017) 'Trade-off curves applications to support set-based design of a surface jet pump', *Procedia CIRP*, 60 (2017), pp. 356-361.
- [5] Balli, O. (2017). Exergy modeling for evaluating sustainability level of a high bypass turbofan engine used on commercial aircrafts. *Applied Thermal Engineering*, 123, 138-155.
- [6] Berton, J. J., Jones, S. M., Seidel, J. A., & Huff, D. L. (2018). Noise predictions for a supersonic business jet using advanced takeoff procedures. *The Aeronautical Journal*, 122(1250), 556-571.
- [7] Chandiramani, K.L.. (1974). Diffraction of Evanescent Waves; Applications to Aerodynamically Scattered Sound and Radiation from Unbaffled Plates. *Journal of the Acoustical Society of America*. 55. pp. 19-29.
- [8] Cooper, R. G. (2016). Agile–Stage–Gate Hybrids: The Next Stage for Product Development Blending Agile and Stage-Gate methods can provide flexibility, speed, and improved communication in new-product development. *Research-Technology Management*, 59(1), 21-29.
- [9] Gagliardi, P., Teti, L., & Licitra, G. (2018). A statistical evaluation on flight operational characteristics affecting aircraft noise during takeoff. *Applied acoustics*, 134, 8-15.
- [10] Gjestland, T., & Granøien, I. L. (2016, August). Noise surveys at five Norwegian airports. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings (Vol. 253, No. 8, pp. 201-207)*. Institute of Noise Control Engineering.
- [11] Hebly, S. J., & Visser, H. G. (2015). Advanced noise abatement departure procedures: custom-optimised departure profiles. *The Aeronautical Journal*, 119(1215), 647-661.
- [12] Iglesias-Merchan, C., Diaz-Balteiro, L., & Soliño, M. (2015). Transportation planning and quiet natural areas preservation: Aircraft overflights noise assessment in a National Park. *Transportation Research Part D: Transport and Environment*, 41, 1-12.
- [13] Kennedy, B. M., Sobek, D. K. and Kennedy, M. N. (2014) 'Reducing rework by applying set-based practices early in the systems engineering process', *Systems Engineering*, 17(3), pp. 278–296.
- [14] Kerga, E., Taisch, M., Terzi, S., Bessega, W. and Rosso, A. (2014) 'Set-based concurrent engineering innovation roadmap (SBCE IR): a case on Adiabatic Humidification', *International Journal of Design Creativity and Innovation*, 2(4), pp. 1–32.
- [15] Khan, M. S., Al-Ashaab, A., Shehab, E., Haque, B., Ewers, P., Sorli, M. and Sopelana, A. (2013) 'Towards lean product and process development', *International Journal of Computer Integrated Manufacturing*, 26(12), pp. 1105–1116.
- [16] Kopsch, F. (2016). The cost of aircraft noise—Does it differ from road noise? A meta-analysis. *Journal of Air Transport Management*, 57, 138-142.
- [17] Kyprianidis, K. G., & Dahlquist, E. (2017). On the trade-off between aviation NOx and energy efficiency. *Applied Energy*, 185, 1506-1516.
- [18] Lawton, R. N., & Fujiwara, D. (2016). Living with aircraft noise: Airport proximity, aviation noise and subjective wellbeing in England. *Transportation Research Part D: Transport and Environment*, 42, 104-118.
- [19] Lefèvre, M., Carlier, M. C., Champelovier, P., Lambert, J., Laumon, B., & Evrard, A. S. (2017). Effects of aircraft noise exposure on saliva cortisol near airports in France. *Occupational and environmental medicine*, 74(8), 612-618.
- [20] Morgan, J. M. and Liker, J. K. (2006) *The Toyota product development system: integrating people, process, and technology*. New York: Productivity Press.
- [21] Nonaka, I., Kodama, M., Hirose, A. and Kohlbacher, F. (2014) 'Dynamic fractal organizations for promoting knowledge-based transformation – A new paradigm for organizational theory', *European Management Journal*, 32(1), pp. 137–146. doi: 10.1016/j.emj.2013.02.003.
- [22] Rauch, E., Dallasega, P., & Matt, D. T. (2016). The way from lean product development (LPD) to smart product development (SPD). *Procedia CIRP*, 50, 26-31.
- [23] Saadon, S., & Talib, A. A. (2016, October). An analytical study on the performance of the organic Rankine cycle for turbofan engine exhaust heat recovery. In *IOP Conference Series: Materials Science and Engineering (Vol. 152, No. 1, p. 012011)*. IOP Publishing.
- [24] Sandberg, M., Tyapin, I., Kokkolaras, M., Lundbladh, A., & Isaksson, O. (2017). A knowledge-based master model approach exemplified with jet engine structural design. *Computers in Industry*, 85, 31-38.
- [25] Sobek, D. K., Ward, A. C. and Liker, J. K. (1999) 'Toyota's principles of set-based concurrent engineering', *Sloan management review*, 40(2), pp. 67–84.
- [26] Schuh, G., Gartzten, T., Soucy-Bouchard, S., & Basse, F. (2017). Enabling agility in product development through an adaptive engineering change management. *Procedia CIRP*, 63, 342-347.
- [27] Tortorella, G. L., Marodin, G. A., Fettermann, D. D. C., & Fogliatto, F. S. (2016). Relationships between lean product development enablers and problems. *International Journal of Production Research*, 54(10), 2837-2855.
- [28] Trojanek, R., Tanas, J., Raslanas, S., & Banaitis, A. (2017). The impact of aircraft noise on housing prices in Poznan. *Sustainability*, 9(11), 2088.
- [29] Tyagi, S., Choudhary, A., Cai, X., & Yang, K. (2015). Value stream mapping to reduce the lead-time of a product development process. *International journal of production economics*, 160, 202-212.
- [30] Wang, Z. and Wang, N. (2012) 'Knowledge sharing, innovation and firm performance', *Expert Systems with Applications*, 39(10), pp. 8899–8908. doi: 10.1016/j.eswa.2012.02.017.
- [31] Ward, A. C. and Sobek, D. K. (2014) *Lean product and process development*. 2nd edn. Cambridge, MA, USA: Lean Enterprise Institute.
- [32] Womack, J. (2006) 'A lesson to be learned', *Manufacturing Engineer*, 85(2), pp. 4–5.

# The Prototype of Thai Blockchain-based Voting System

Krittaphas Wisessing<sup>1</sup>, Phattaradon Ekthammabordee<sup>2</sup>, Thattapon Surasak<sup>3</sup>,  
Scott C.-H. Huang<sup>4</sup>, Chakkrit Preuksakarn<sup>5</sup>

Department of Computer Engineering, Kasetsart University, Nakorn-Pathom, Thailand<sup>1,2,5</sup>  
Faculty of Information Technology, Thai-Nichi Institute of Technology, Bangkok, Thailand<sup>3</sup>  
Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan<sup>3,4</sup>

**Abstract**—In this paper, the prototype of Thai voting system using blockchain technology (B-VoT) has been successfully designed and developed. Hyperledger Fabric was chosen to be a major blockchain infrastructure. The web-application was developed to allow voters to vote. In addition, real-time voting results will be shown on the same website after election period has passed. We connected the web-application and blockchain database together in order to store the votes as a blockchain transaction. With our Blockchain Internet Election, the voters can easily elect on the website and their votes is automatically stored in the blockchain database as a single blockchain transaction. This voting prototype can assure the data integrity because no one can modify any information or voting results. Therefore, our system could have a huge impact on the voting reliability as well as rebuild public trustworthiness in Thailand election.

**Keywords**—Blockchain; internet election; hyperledger fabric; data integrity; voting reliability

## I. INTRODUCTION

Thailand is a Southeast Asian country [1]. It accommodates around 70 millions persons according to the review in 2020 [2]. The country election system is traditionally handled by the Election Commission of Thailand (ECT) [2], [3]. Thai election schemes are based on the paper election form and manual counting system. The ECT regulates the monopoly system to all election-related tasks during the voting process. Their duties include regulations preparation, election places allocation, votes counting, and the result announcement. Unfortunately, many mistakes and problems cannot be avoided with this traditional system, and even dramatically increased in the election in 2019. The fault result announcement was broadcasted across the country without the awareness of miscounting ballots and confusing allocation of polling places. These systematic problems cause a huge trouble to the country and raise the concerns from populations around the world [4]. In order to prevent the problems on the system, an integration of the information technology can be applied. The use of information technology can effectively prevent human errors, which are the habitual mistakes in any systems across the world [5]. The Internet platform has been chosen by a number of countries to use as the basic voting system. For instance, the “Internet voting system” or “I-voting system” has been used by the Republic of Estonia since 2005 [6]. A major advantage of this system is convenient, reduces cost and increases security [7], [8]. However, this system is still a “Centralized system” which is similar to the election system in Thailand. According to the problem mentioned earlier, blockchain technology can

be applied in order to improve security. Blockchain is a promising technology as its system is decentralized [5]. It is a chain of blocks in which each block stores all information of network activities after the block was added to the chain [9], [10]. In other words, blockchain is a traceable database. It allows every user to add data as a transaction. Moreover, every user can review all data in the blockchain, but no one is able to change it [11]. This work proposes a promising solution in which blockchain technology has been integrated to the voting system. The prototype of Thai Blockchain-based voting system (B-VoT) is developed. Our prototype is based on the election regulation. In addition, web-application is created to facilitate the voting process for users as well as encrypt voting data to enhance the system security.

In this paper, Section II defines the blockchain technology and also reviews the works related to the Internet Voting System. Section III explains our proposed blockchain-based voting system. Section IV presents the case study to verify the feasibility of the proposed system. The results are shown on the website. Section V discusses the evaluation and benchmark of our proposed system design. Finally, Section VI describes the conclusion.

## II. BACKGROUND

Blockchain is a decentralized technology which is secure, tamper-proof and immutable. Every user in the same network can communicate with each other and receives a database that keeps all records of transactions, see Fig. 1. This database is known as ledger [12]. It ensures that any users in the system are able to update and monitor blocks, where data is stored. The block itself cannot be changed after the transaction has been transferred to the blockchain. The transaction is verified by consensus algorithm depending on the structure of each blockchain before adding to the block [13]. This method thus guarantees the integrity, trust, and traceability of the system [5]. Individual blocks are linked with a chain. As can be seen in Fig. 2, each block contains a cryptographic hash of the previous block (except the first block) and a transaction list [14]. There are two major types of Blockchain networks divided by authorization which are permissioned and permissionless blockchain. Permission blockchain requires authorization to join the network, while the permissionless blockchain allows everybody to participate in the network [15]. One of the most well-known blockchain applications is Bitcoin [16]. It was developed by Satoshi Nakamoto in 2008 [17]. Bitcoin is a digital cryptocurrency which will come

into the focus in the future [14]. Blockchain technology itself also applied to various applications at this moment, including Internet of Things (IoT), supply chain, and digital credential [18]–[22].

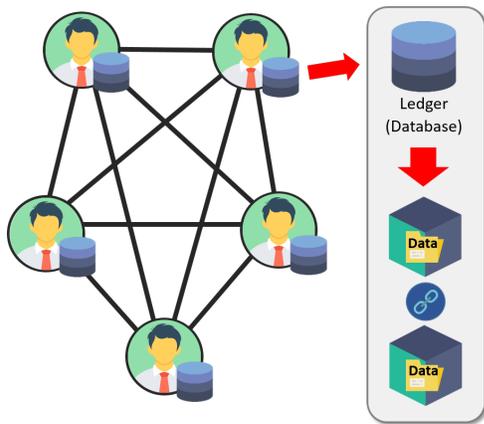


Fig. 1. Blockchain is a decentralized network that allows every user to communicate with each other. The user in the network will contain a database called “Ledger”. The ledger has a role to store data in the form of block and chain. The block has a role to store data. The chain is a cryptographic mechanism.

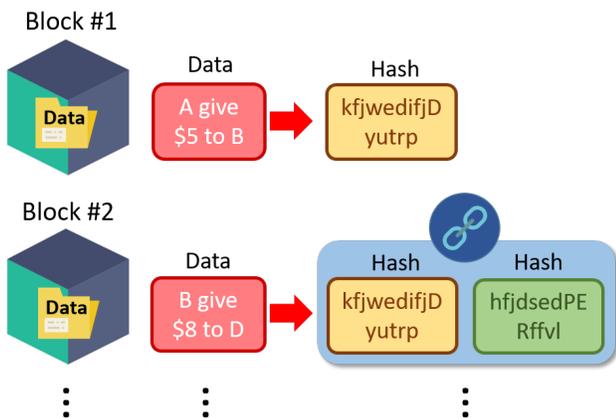


Fig. 2. In each block of the chain must store an individual hashed data with another hashed data from the previous block (the first block of the chain will store only own hashed data).

Hyperledger Fabric is a permissioned blockchain network, initially contributed by IBM and Digital Asset [23]. It offers a decentralized ledger platform with a special role between nodes in the system, allowing pluggable implementations of various functions. A protocol of the Hyperledger Fabric is run by peers [24]. There are two kinds of peers. One is a “validating peer” that runs consensus, executes chaincode, validating transactions, and maintaining the ledger. The validating peers run a Byzantine-fault tolerant (BFT) consensus protocol. Another is “non-validating peer” who maintains blockchain integrity and provides approved transactions to the peers of the network [23], [24]. Hyperledger Fabric is mainly designed for the Distributed Ledger Technology (DLT) development. It is widely used in many prototypes, conceptual models, and practical applications [23]. An overview architecture of Hyperledger Fabric can be seen in Fig. 3.

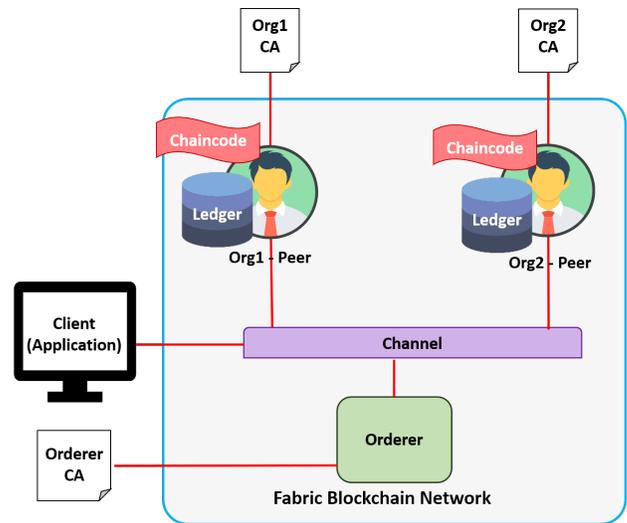


Fig. 3. Hyperledger Fabric Architecture: Every Org (user) contains ledger, which is a data storage, and chaincode, that use to call function in blockchain. Orderer responsible in ordering new block to a blockchain.

In 2005, Estonia introduced the Internet Voting system to their national elections. In this system, voters can identify themselves by using their national ID card to a card reader. After the authentication process, the voter is able to elect a candidate using the web-application. The vote is then encrypted with the server’s public key and signed with the voter’s ID card to identify the voter [7], [8]. The vote results are immediately stored in a storage server during the election period. At the end of the election, all of the votes are transferred from the storage server to the counting server by the election staff. The result is counted in this server. In this process, the voter that signed with the voter’s ID card will drop the voter signature. Then, the vote-counting server will decrypt the vote and count anonymous ballots [7]. Eventually the election result is announced on the web-application. During the voting period, the voter can cast a vote over the Internet several times. However, just the last vote is counted.

A benefit from this system is the speed, election cost reduction and user-friendly. This is because the voter uses only the web application in order to vote for a candidate. These are the evidences proving that the Internet voting system has many problems. With the current system, neither the internet voting system nor paper-based voting can be a promissive system in the future. Unfortunately, this system has only one server storing massive and sensitive data. When the server has a problem or is attacked, the election process certainly fails [7]. In the review, they recommend Estonia to discontinue using their internet voting system as it potentially fails, risks, and unsettles [6]. After Estonia launched the I-Voting system, this system structure has been applied by various countries, including the Netherlands, Switzerland, Norway, Canada, Greece and India [25]–[27]. However, those applications encountered some security issues. For example, some countries faced with DDoS attack software implementation errors [7].

A major problem related to the internet voting is democracy, since it has many factors to consider such as the identity of the voter, availability for the system, authentication, anonymity for a voter, monitoring, voter privacy and coercion-

resistance [8], [28]. These factors make the internet voting system unacceptable by many countries. However, a benefit from the internet voting system makes some countries want to use this system and consider methods that enhance performance of the internet voting system such as multiple encryptions, visual cryptography, biometrics with Raspberry Pi and TFT module [29]–[31].

front of the national ID card while a verification number can be seen on the back of the card. Next, the voters can cast a vote depending on their province and household area registration. After the voters completely casted, the voters cannot cast a new vote. When the voting period ends, the system instantly counts the votes. Finally, the result is announced on the web application in the form of a dashboard.

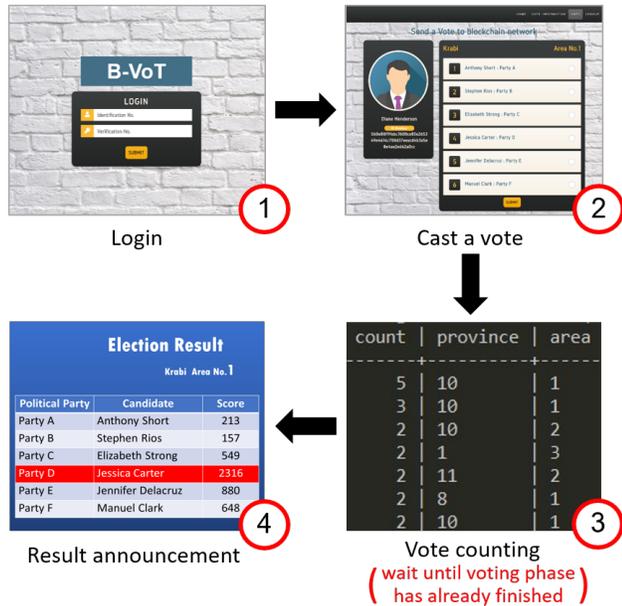


Fig. 4. System Overview: voter login to authenticate themselves and cast a vote, once the voting period is completed, system counts every voting record and announce the result.

### III. PROPOSED METHOD

In order to solve the problem mentioned earlier, this paper proposes the prototype of Thai Blockchain-based Voting System (B-VoT). The blockchain database can assure the data integrity because no one can modify any information or voting results. Therefore, our system could have a huge impact on the voting reliability. Moreover, we are able to decentralize the current Thai voting system using the Hyperledger Fabric blockchain infrastructure. However, the voters are still required to go to the polling places according to their household registrations because verification of voters is needed. The vote is conducted by using the provided devices with a controlled network in order to restrict voters activities. The real-time transaction and database update from the blockchain increase security of the whole system. This paper mainly focused on the B-VoT prototype development. Therefore, the Python program is created in order to generate the random candidates and voters information such as name, the identification and verification number of voter, and party. The generation protocol is designed according to election structure in Thailand.

#### A. System Overview

The B-VoT system can be divided into four main sections as can be seen in Fig. 4. Firstly, the voters access the B-VoT system by using an identification number and a verification number. An identification number is a number that displays in

#### B. System Architecture

1) *User Activity:* In our proposed method, there are three types of users which are voter, administrator, and election staff. These types of users have different functions and allowances.

- Voters have a role to vote for the candidate, by selecting the number of the political party, which belongs to their province and area number. The voters in our proposed system are Thai people over 18 years old (in the year of election). Their main duty is to vote for a candidate in their province and area number according to their household registration. Once the voting period is completed, they are able to see the election result.
- Administrators are Thai government officers in each particular polling place. For the B-VoT prototype, they are responsible for checking the voter information and verification before allowing the voters to cast their votes. In case of emergency, administrators can suddenly report any issues to the election staff in order to solve or prevent problems during the election period.
- Election staff are responsible for validating a list of voters according to their household registration. After that, they are required to add the voters and candidate's information into the database before the election day. During the voting session, their roles conform to the administrators, which is the system failure prevention procedures of our current prototype.

The web application needs the authentication of voters and administrators before performing any activity on the site. They are required to fill in their personal information, identification number, and verification number (stated at the back of the ID card which consists of two alphabets and ten numbers). Actually, the voting features in the web application are unusable before the election date. In addition, the voting period on the day of election is limited to 9 hours, from 8am to 5pm. Once the voting period finishes, the voting feature returns to unusable state. Then, voters and administrators can check the candidate score in the country view or search from province and area number. The candidate's results are also shown in the web-application which every user can assess.

2) *The B-VoT Database:* The B-VoT system requires the election staff to add voters and candidate information to the database prior to the election day. This personal information is encrypted for security reasons and anonymity. The web application is connected to the blockchain database to access the voter's data for user authentication during the voting period. After the user identity is verified, the web application automatically redirects to the voting page. The list of qualified candidates is sorted to display in the browser according to the

political party number. Only one candidate can be chosen for each voter according to the role and permission mentioned in subsection III-B1. Moreover, the voters cannot edit their votes after their votes have been confirmed. Once the vote has been confirmed, the web application calls a blockchain function to receive and store the transactions including the voter identification number (encrypted version), selected political party, and polling place information (province and area number).

3) *Hyperledger Fabric: The Major Blockchain Infrastructure*: The blockchain technology itself is classified into two types: permissioned and permissionless. By the way, the permissioned blockchain, called Hyperledger Fabric, was chosen because we need to strict the network in order to higher the security level. With Hyperledger Fabric, the control layer runs on top of the blockchain can differentiate the actions that performed by each user. Therefore, permissioned blockchain has a better transaction performance because we can set the block size limitation and the validate information by adjusting the Chaincode during the implementation process. Within this type of blockchain, we can assure the security level of the system because every user has a different priority status [5].

#### IV. RESULTS

With this implementation, the Hyperledger Fabric is used as a main infrastructure for the back-end development. Web-application (B-VoT website) is also designed and developed. This website is the only platform that allows voters to cast their votes for our B-VoT prototype. The voting result is finally revealed on the B-VoT website at the end of the election period. The overview of the system from login to result announcement are shown in Section III. In this section, the preliminary results from both back-end and front-end development are shown.

##### A. Back-End Development: Hyperledger Fabric Implementation

After the vote was sent to the blockchain network, vote data including timestamp, voter ID, candidate party's number, province, and area of the voter are stored in the block. This information can be seen as a command line on the Hyperledger Fabric back-end system.

##### B. Front-End Web Development: B-VoT Website

The B-VoT website consists of three main pages, including Login, Home and Vote as shown in Fig. 5, 6, 7. The first page is the Login page that requires the user credentials as mentioned in Section III. After the voters authenticate themselves, the page is redirected to the home page. Home page is the main page of the system. The navigation bar appears on the top of the page. This bar must navigate the voter to other functions such as show vote information function, cast a vote function, and logout function. The home page has a countdown timer in the middle of the page. This timer has a role to countdown to election day. Vote page is separated into two columns. The first one shows voter information such as name, surname, and ID number (hashed). The second one has a role to show a qualified candidate list in the form of a radio button. Each button shows important information about candidates such as the political party name and number, candidate name and surname. The voter can select a candidate and press the submit button to send a vote.

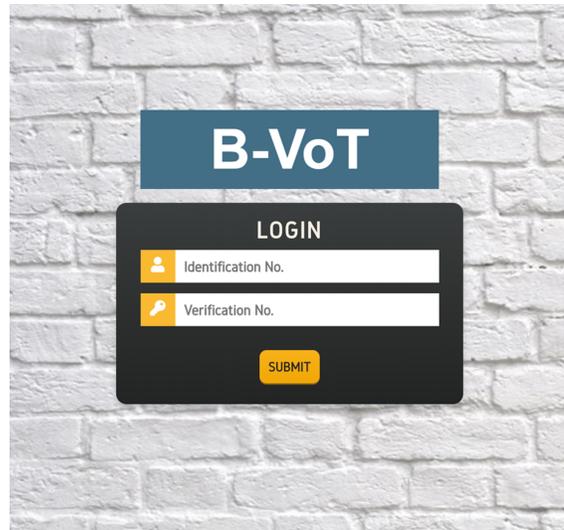


Fig. 5. Login page of B-VoT system: The Login page has a form with two input fields and one submit button. The first input field requires "Identification No.". Another input field requires "Verification No." (the number that displays behind the national ID card).

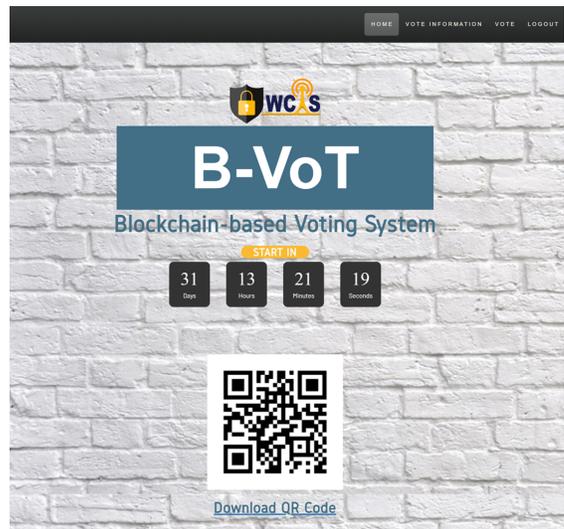


Fig. 6. Home page of B-VoT system: This page is the home page for the users after they authenticate themselves with the information as mentioned in Fig. 5.

#### V. DISCUSSION

After development processes finished, our system automatically counts every voting record in the blockchain network and announces the result. As mentioned in section II, the Estonian Internet Voting system contains all votes in the vote storage server and transfers the votes to the vote-counting server after the election session. This system overcomes the drawbacks of paper-based voting system since the internet voting system reduces time and cost consuming. However, the current Internet voting system contains various risks and fails easily [6], [7]. The objective of the Internet Voting system is to balance between transparency, privacy, and security. Security is the crucial factor to monitor because it relates to trustworthiness of all involving users. Therefore, blockchain is used as a database

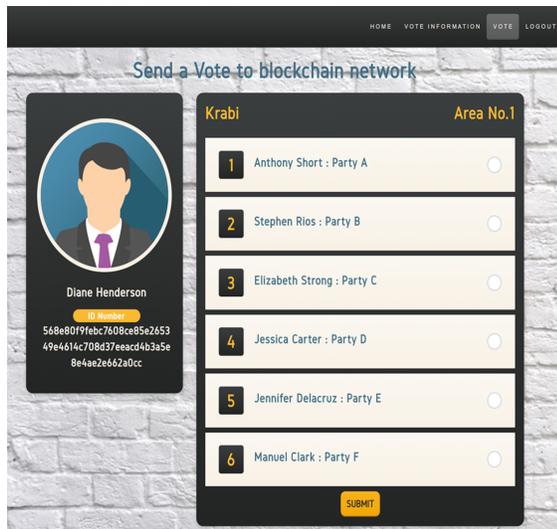


Fig. 7. Vote page of B-VoT system: The voter personal information is shown in the left column, while the right column displays the list of qualified candidates.

for our B-VoT prototype since it can enhance the integrity of recorded data. The security of the system significantly increased due to the hash function, which encrypts all private information. This information can be verified only by a system administrator. The record in blockchain is permanent, thus all activities in the system can be visibly tracked. That means our system contains privacy and transparency simultaneously.

## VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we have integrated the Blockchain technology with the voting system to develop the prototype of Thai Voting System. Thai traditional election procedures are based on the paper ballot election and manual counting system, which leads to many security issues. The permissioned blockchain technology, called Hyperledger Fabric, was chosen because we need to limit the network to strengthen security. Compared with the traditional election, B-VoT reduces cost and time consuming in the whole process. It is also more convenient because voters simply click on the browser to cast a vote. Besides, to compare with a regular internet voting system (without the blockchain integration), the use of Hyperledger Fabric provides an unchangeable database. Moreover, B-VoT is able to work with sensitive data without the block creation fees. These are our significant improvement points against the traditional internet voting system. The users can cast the vote and check the result after all processes completed by using a web application. The system testing according to the real-world implementation is required for future work. Moreover, we also planned to have a re-work on the user interface to higher a very good user experience for our proposed system.

## ACKNOWLEDGMENT

The authors would like to thank the Institute of Communications Engineering and Wireless Communications and Information Technology Research Laboratory at National Tsing Hua University, Taiwan for the supportive idea and resources.

## REFERENCES

- [1] N. Poapongsakorn, M. Ruhs, and S. Tangjitwisuth, "Problems and outlook of agriculture in thailand," *Thailand Development Research Institute Quarterly Review*, vol. 13, 01 1998.
- [2] J. Ockey, "Thailand in 2019," *Asian Survey*, vol. 60, no. 1, pp. 117–124, 2020. [Online]. Available: <https://as.ucpress.edu/content/60/1/117>
- [3] J. Siffin, "The civil service system of the kingdom of thailand," *International Review of Administrative Sciences*, vol. 26, no. 3, pp. 255–268, 1960. [Online]. Available: <https://doi.org/10.1177/002085236002600305>
- [4] C. Prasertdum and D. Wichadaku, "2019 thai general election: A twitter analysis," in *Soft Computing in Data Science*. Singapore: Springer Singapore, 2019, pp. 336–350.
- [5] T. Surasak, N. Wattanavichean, C. Preuksakarn, and C.-H. Huang, "Thai agriculture products traceability system using blockchain and internet of things," *International Journal of Advanced Computer Science and Applications*, vol. 10, pp. 578–583, 10 2019.
- [6] D. Springall, T. Finkenauer, Z. Durumeric, J. Kitcat, H. Hursti, M. MacAlpine, and J. Halderman, "Security analysis of the estonian internet voting system," *Proceedings of the ACM Conference on Computer and Communications Security*, pp. 703–715, 11 2014.
- [7] S. Heiberg and J. Willemson, "Verifiable internet voting in estonia," in *2014 6th International Conference on Electronic Voting: Verifying the Vote (EVOTE)*, Oct 2014, pp. 1–8.
- [8] G. Schryen and E. Rich, "Security in large-scale internet elections: A retrospective analysis of elections in estonia, the netherlands, and switzerland," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 729–744, Dec 2009.
- [9] F. X. Tian, "An agri-food supply chain traceability system for china based on rfid and blockchain technology," *2016 13th International Conference on Service Systems and Service Management (ICSSSM)*, pp. 1–6, 2016.
- [10] A. Shahid, U. Sarfraz, M. Malik, M. Iftikhar, A. Jamal, and N. Javaid, "Blockchain-based reputation system in agri-food supply chain," 02 2020.
- [11] M. Muzammal, Q. Qu, and B. Nasrulin, "Renovating blockchain with distributed databases: An open source system," *Future Generation Computer Systems*, vol. 90, pp. 105–117, 2019.
- [12] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," 06 2017.
- [13] Y. Hao, Y. Li, X. Dong, L. Fang, and P. Chen, "Performance analysis of consensus algorithm in private blockchain," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, June 2018, pp. 280–285.
- [14] D. Vujicic, D. Jagodic, and S. Randic, "Blockchain technology, bitcoin, and ethereum: A brief overview," in *2018 17th International Symposium INFOTEH-JAHORINA (INFOTEH)*, March 2018, pp. 1–6.
- [15] M. Shrivastava and D. Yeboah, "The disruptive blockchain: Types, platforms and applications," 12 2018.
- [16] S. A. Swamy and N. Jayapandian, "Secure bitcoin transaction and iot device usage in decentralized application," in *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*, Oct 2018, pp. 271–274.
- [17] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," <http://bitcoin.org/bitcoin.pdf>, 2008.
- [18] M. Swan, "Anticipating the economic benefits of blockchain," *Technology Innovation Management Review*, vol. 7, pp. 6–13, 10 2017.
- [19] R. Xu, Y. Chen, E. Blasch, and G. Chen, "Blendcac: A smart contract enabled decentralized capability-based access control mechanism for the iot," *Computers*, vol. 7, 07 2018.
- [20] M. Memon, S. S. Hussain, U. A. Bajwa, and A. Ikhlas, "Blockchain beyond bitcoin: Blockchain technology challenges and real-world applications," in *2018 International Conference on Computing, Electronics Communications Engineering (iCCECE)*, Aug 2018, pp. 29–34.
- [21] S. Yakut, . Şeker, E. Batur, and G. Dalkılıç, "Blockchain platform for internet of things," in *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Oct 2019, pp. 1–6.

- [22] S. Madumidha, P. S. Ranjani, S. S. Varsinee, and P. S. Sundari, "Transparency and traceability: In food supply chain system using blockchain technology with internet of things," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 983–987.
- [23] E. Androuraki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich, S. Muralidharan, C. Murthy, B. Nguyen, M. Sethi, G. Singh, K. Smith, A. Sorniotti, C. Stathakopoulou, M. Vukolić, S. W. Cocco, and J. Yellick, "Hyperledger fabric: A distributed operating system for permissioned blockchains," in *Proceedings of the Thirteenth EuroSys Conference*, ser. EuroSys '18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: <https://doi.org/10.1145/3190508.3190538>
- [24] P. Thakkar, S. Nathan, and B. Viswanathan, "Performance benchmarking and optimizing hyperledger fabric blockchain platform," in *2018 IEEE 26th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*, 2018, pp. 264–276.
- [25] N. J. Goodman and J. H. Pammett, "The patchwork of internet voting in canada," in *2014 6th International Conference on Electronic Voting: Verifying the Vote (EVOTE)*, Oct 2014, pp. 1–6.
- [26] A. Delis, K. Gavatha, A. Kiayias, C. Koutalakis, E. Nikolakopoulou, L. Paschos, M. Rousopoulou, G. Sotirellis, P. Stathopoulos, P. Vasilopoulos, T. Zacharias, and B. Zhang, "Pressing the button for european elections: verifiable e-voting and public attitudes toward internet voting in greece," in *2014 6th International Conference on Electronic Voting: Verifying the Vote (EVOTE)*, Oct 2014, pp. 1–8.
- [27] V. P. Singh, H. Pasupuleti, and N. S. C. Babu, "Analysis of internet voting in india," in *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIECS)*, March 2017, pp. 1–6.
- [28] B. Meng, "A formal analysis of coercion-resistance of the internet voting protocol based on dkr formal model," in *2008 International Symposiums on Information Processing*, May 2008, pp. 490–494.
- [29] A. B. Rajendra and H. S. Sheshadri, "Visual cryptography in internet voting system," in *Third International Conference on Innovative Computing Technology (INTECH 2013)*, Aug 2013, pp. 60–64.
- [30] S. M. Jambhulkar, J. B. Chakole, and P. R. Pardhi, "A secure approach for web based internet voting system using multiple encryption," in *2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies*, Jan 2014, pp. 371–375.
- [31] A. M. Jagtap, V. Kesarkar, and A. Supekar, "Electronic voting system using biometrics, raspberry pi and tft module," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, April 2019, pp. 977–982.

# Satellite Image Database Search Engine which Allows Fuzzy Expression of Geophysical Parameters of Queries

Kohei Arai

Faculty of Science and Engineering  
Saga University, Saga City, Japan

**Abstract**—Satellite image database search engine which allows fuzzy expression of geophysical parameters of queries is proposed. A search engine based on knowledge based system which allows fuzzy expression of queries is proposed. A prototype system is created and tested. Whereas conventional search systems had to know in advance the functions of the search system, information about search keys, etc., the search engine proposed here guided the search conditions in a conversational form, by allowing ambiguous expressions (six adverb language hedges) at that time, the user is released from such annoyance. To make this possible, a membership function for each attribute information is defined, and a search condition refinement by fuzzy logic is introduced. The results show that the system accepts a fuzzy expression of query as well as a comprehensive dialogues between users and the system.

**Keywords**—Search engine; fuzzy expression; knowledge base system; membership function

## I. INTRODUCTION

The relational database (relational database) is frequently used for searching the earth observation satellite image database from the viewpoint of its efficiency and system flexibility. A typical example of the database search is a search engine using SQL (Structured Query Language) or QBE (Query by Example) [1], [2]. When these search engines are used, constructing search conditions for the earth observation satellite image database search, it is necessary to input a specific search key in a predetermined format. That is, the search can be performed only after the user sets the search condition according to the determined format. Therefore, it is necessary for the user to know in advance the functions of the search system, information on the search key.

The search system for the Earth Observation Satellite Image Database: EOSID proposed in this paper is based on SQL (since most existing EOSID systems have been developed on the premise of SQL), and allows search requests in natural language. The proposed search engine system translates the search request into a request in the SQL language (agent function [3]) and searches for the desired data. At the search engine is used as EOSID, the system guides the user in a conversational manner (guide system) and also to allow search requests with ambiguous expressions of natural language [4]. Therefore, a membership function for ambiguous expressions is introduced based on the opinions of users in the marine research field (for instance) and enabled

fuzzy search. In order to confirm the effectiveness of the proposed method, a virtual database was constructed in the marine research field. Some experiments are performed with the prototype of the proposed search engine system. As a result, it is found that the search performance is quite good from the viewpoint of easiness of database retrievals.

In the following section, related research works are described. Then, the proposed prediction method and system is described followed by experimental set-up together with experimental results. After that, concluding remarks and some discussions are described.

## II. RELATED RESEARCH WORKS

Vague search of earth observation image database based on Fuzzy theory using physical quantities and spatial features is proposed [5] together with earth observation satellite image database system allowing ambiguous search requests [6]. On the other hand, user friendly and efficient catalog information management for earth observation data, is proposed and well reported [7].

Remote sensing satellite image database system allowing image portion retrievals utilizing principal component which consists spectral and spatial features extracted from imagery data is proposed [8]. Meanwhile, data collection and active database for tsunami warning system is proposed [9]. On the other hand, remote sensing satellite image database system allowing image portion retrievals utilizing principal component which consists spectral and spatial features extracted from imagery data is proposed [10]. Also, numerical representation of web sites of remote sensing satellite data providers and its application to knowledge based information retrievals with natural language is proposed [11].

Image retrieval based on color, shape and texture for ornamental leaf with medicinal functionality, meanwhile, is proposed [12]. Also, comparison contour extraction based on layered structure and Fourier descriptor on image retrieval is proposed and evaluated its effectiveness [13]. Also, image retrieval method utilizing texture information derived from Discrete Wavelet Transformation: DWT together with color information is proposed [14].

Metadata definition and retrieval of earth observation satellite data is proposed [15]. Meanwhile, Geographic Information System: GIS based on neural network for

appropriate parameter estimation of geophysical retrieval equations with satellite remote sensing data is proposed [16]. Image retrieval method based on hue information and wavelet description based shape information as well as texture information of the objects extracted with dyadic wavelet transformation is proposed [17]. Wavelet based image retrievals is attempted [18]. Also, image retrieval method based on back-projection is proposed [19].

### III. PROPOSED METHOD

#### A. System Configuration

Fig. 1 shows the configuration of the proposed earth observation satellite image database system. The system allows search requests in natural language. Since the research on natural language understanding has not been completed, a text search method has been introduced here. That is, the text matching between the search request from the user and the search keyword prepared by the system is determined by Z39.504, [20].

A natural language whose expression is ambiguous has a broad meaning and is ambiguous. For example, a search may be requested in a vague image of the user, such as "slightly warm sea area" or "image with extremely small amount of cloud". Existing database search engines cannot handle such cases. In order to deal with this, the following fuzzy search is proposed.

#### B. Fuzzy Retrieval

A fuzzy search for users in the marine field is shown as an example of an ambiguous search. The search target (object [21] of the constructed virtual database in the marine field is an image ID (Identification No.), and the attribute information (Attributes) serving as a search key is as follows.

- Sea surface temperature
- Chlorophyll a
- Cloud cover
- Observation date
- Sea area
- Image quality
- Sensor name
- Satellite name
- Processing level
- Ground station name
- Pass, row number
- Scene center latitude, longitude

Of these, this time, sea surface temperature, chlorophyll-a and cloudiness were considered to allow search requests with ambiguous expressions.

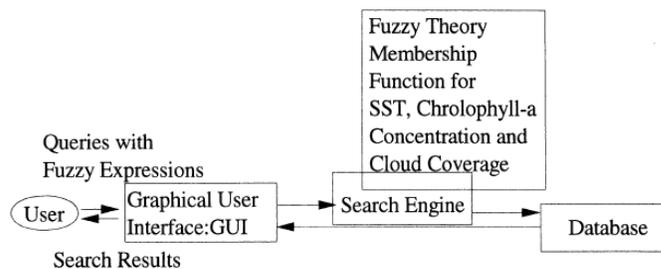


Fig. 1. The Proposed Database System for Earth Observation Satellite Data Retrieval.

#### C. Creating Membership Functions

A fuzzy set (membership function [22], [23], [24]) was created to express the user's ambiguity regarding the above-mentioned sea surface temperature, chlorophyll-a, and cloudiness. At that time, a questionnaire survey was conducted for researchers in the marine field of the Earth Environment Observation Committee of the Earth Science and Technology Agency, and based on the results (shown in Table I).

Equations (1) to (6) show the membership functions corresponding to low, high, high and low chlorophyll-a concentrations and high and low cloudiness, respectively.

$$\mu_A(x) = \int_0^{\infty} e^{-0.015x^2/x} \text{ for Low SST} \quad (1)$$

$$\mu_B(x) = \int_0^{\infty} e^{-0.015(x-30)^2/x} \text{ for High SST} \quad (2)$$

$$\mu_C(x) = \int_0^{\infty} e^{-0.003(x-35)^2/x} \text{ for High Chlorophyll - a} \quad (3)$$

$$\mu_D(x) = \int_0^{\infty} e^{-30x^2/x} \text{ for Low Chlorophyll - a} \quad (4)$$

$$\mu_E(x) = \int_0^{\infty} e^{-0.034(x-17)^2/x} \text{ for High Cloud Coverage} \quad (5)$$

$$\mu_F(x) = \int_0^{\infty} e^{-3.0x^2/x} \text{ for Low Cloud Coverage} \quad (6)$$

where, the units of the physical quantity ( $x$ ) are sea surface temperature ( $^{\circ}$  C.), chlorophyll-a ( $\text{mg} / \text{m}^3$ ), and cloudiness (%), respectively. Next, a linguistic expression (language hedge) that modifies the fuzzy set is defined. As a result, expressions such as "slightly warm" and "very" warm sea surface temperature can be accepted as search requests. The following shows the language hedge used in this prototype system.

- very  $y$
- more or less  $y$
- slightly  $y$
- sort of  $y$
- pretty  $y$
- rather  $y$

where  $y$  is a fuzzy set. logical operations using these combinations are also possible, which allows for modification of complex fuzzy sets, and allows search requests using combinations such as "slightly warm, chlorophyll-a sea areas".

TABLE I. THE RESULTS FROM QUESTIONNAIRE FOR DETERMINATION OF MEMBERSHIP FUNCTIONS OF SEA SURFACE TEMPERATURE, CHLOROPHYLL-A CONCENTRATION AND CLOUD COVERAGE

General Item	Physical Item	Attributes	Range	Res1	Res2	Res3	Res4	Res5	Res6
Physical Quantity	SST	Cold	Highest	15	10		10	0	15
		Warm	Lowest	20	20		25	25	25
		Warm Current	Lowest	15	15		20	20	25
			Highest	30	30		28	28	30
	Cold Current	Lowest	0	5			0	5	
		Highest	10	10		10	19	15	
Chlorophyll-a	Denth	Lowest	2		0	0	0	0	
	Sparse	Highest	1			1	1	1	
Area	Japanese Vicinity	North	Lowest	24	25	20	25		25
			Highest	50	50	46	45		54
		East	Lowest	120	120	122	125		120
			Highest	150	150	148	140		150
	Ocean Area	Coastal	Far	20	10		10	50	20
		Offshore	Nearest	360	100		300	100	100
		Beyond	Nearest	360	100		300	100	500
	Kuroshio Current	North	Lowest	10	28		15		25
			Highest	38	38		38		50
		East	Lowest	120	128		130		120
			Highest	150	150		145		145
	Oyashio Current	North	Lowest	35	35		38		30
Highest			50	48		45		50	
East		Lowest	140	142		140		120	
		Highest	160	150		145		145	
Time	Recently		Far		12		12	6	60
	Season	Summer	From To	July Sep.	June Sep.		June Aug.	June Aug.	July Sep.
		Winter	From To	Jan. Mar.	Nov. Mar.		Nov. Jan.	Nov. Feb.	Dec. Feb.
Other	Cloud Coverage	Slightly	Approx.	25	20		0	10	30
		Dense	From	25	50		50	30	60
			To	75	100			100	100

D. Search Procedure

Fig. 2 shows the search procedure. First,

- 1) A search request containing an ambiguous expression occurs,
- 2) Guide the search conditions in a conversational manner,
- 3) Select the membership function using the knowledge corresponding to the search condition,
- 4) Using the membership function in the database and then calculate the degree of attribution of data attributes,
- 5) Display search results.

where, the knowledge corresponding to the search condition is the semantic network shown in Fig. 3. The determination of the degree of belonging is performed as follows. In other words, if the attribute information of the data in the database is between a certain threshold value and the maximum value of the membership function, it is said to belong, and it is determined that the others do not belong. The

determination of the threshold was determined as the grade in the center of gravity (CG) of the fuzzy set shown in equation (7).

$$CG = \frac{\int_x x\mu A(x)dx}{\int_x \mu A(x)dx} \tag{7}$$

where, A is an attribute item, and x is an attribute value (physical quantity).

E. AND, OR Logical Operation of Search Condition

Logical operations (AND, OR) of a plurality of search conditions are also possible. At this time, the degree of membership is given by equation (8).

$$\mu A \cap B = \min [\mu A (x), \mu B (x)] \tag{8}$$

$$\mu A \cup B = \max [\mu A (x), \mu B (x)] \tag{9}$$

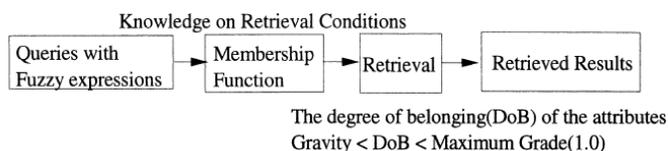


Fig. 2. Search flow of the Proposed Database System which allows Fuzzy Expression of Queries.

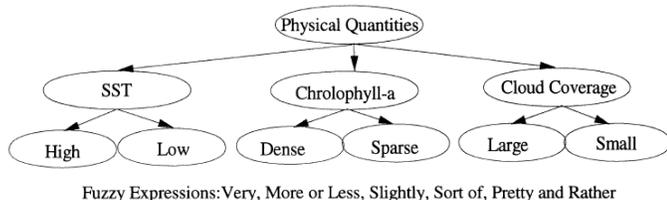


Fig. 3. Semantic Network for the Knowledge of the Proposed Database System.

In other words, when searching for data with a slightly higher sea surface temperature and a very small amount of cloud, use Equation (8) that sets the two search conditions to AND. If any of the conditions is satisfied but data is desired, equation (9) may be used.

#### IV. EXPERIMENTS

Considering that sea surface temperature, chlorophyll-a, and cloudiness are uniformly distributed in the range of 0 to 40 (° C), 0 to 5 (mg / m<sup>3</sup>), and 0 to 100 (%), respectively, 1000 pseudo Generated attribute information of earth observation data. Since the purpose here is to evaluate the performance of the ambiguous search, it is assumed that the search has already been completed for attributes other than these three items. Fig. 4 shows the search flow and search results when the following search request is entered.

In Fig. 4, the sentences with bold and underline characters indicates system response. Firstly, the system asks query, then user account and password are asked. After that, SST, chlorophyll-a concentration and cloud coverage are asked by the system. Then the search result with the table which is shown in Fig. 4 is returned from the system. If the user satisfies the search result, then the system asks whether or not the user would like to save the result. After that, the system asks user's willingness on other search request. If the user has no other search request, then the system is closed.

As a result, candidates meeting these search conditions (in this case, seven candidates) were presented, and it was confirmed that the user could further narrow down from these candidates. Experimental result shows the prediction accuracy of retrieve items are 100% of success rate. In terms of functionality of search engine of the current Earth Observation Satellite Image Database Systems, there is no geophysical parameter search based on knowledgebase system with fuzzy search capability. Therefore, the proposed search engine system has the merit on the geophysical research function.

**Query:** "Slightly Cold, Very Densely Distributed Chlorophyll-a Concentration and Pretty Small Cloud Coverage"

**Welcome to Earth Observation Satellite Data (OCTS) Retrieval System**

**Please key-in your name**

: arai

**Please key-in your password**

:\*\*\*\*\*

**Authentication is completed**

**What is Sea Surface Temperature do you want to retrieve? (You may use the following words: Very, More or Less, Slightly, Sort of, Pretty and**

**Rather)**

: Slightly cold

**What is Chlorophyll-a Concentration do you want to retrieve? (You may use the following words: Very, More or Less, Slightly, Sort of, Pretty and**

**Rather)**

: Very dense

**What is Cloud Coverage do you want retrieve? (You may use the following words: Very, More or Less, Slightly, Sort of, Pretty and**

**Rather):**

Pretty small

**Now knowledge based retrieval is activated**

**Just wait for a moment**

**This is candidate of your retrieved data**

Image ID	SST	Chlorophyll-a	Cloud Coverage.
27	8	3	13
174	9	2	11
276	12	3	5
743	15	3	20
835	10	2	18
849	8	2	9
901	14	3	16

**Do you satisfy this results? (Yes(y) or No(n))**

:y

**Do you want save this results? (Yes(y) or No(n))**

:n

**Do you want to retrieve other data? (Yes(y) or No(n))**

:n

**Now you may exit this system**

Fig. 4. An Example of the Dialogue between users and the Proposed System with the Query of "Slightly Cold, Very Densely Distributed Chlorophyll-a Concentration and Pretty Small Cloud Coverage".

#### V. CONCLUSION

A search engine based on knowledge based system which allows Fuzzy expression of queries is proposed. A prototype system is created and tested. Whereas conventional search systems had to know in advance the functions of the search system, information about search keys, etc., the search engine proposed here guided the search conditions in a conversational form, by allowing ambiguous expressions (six adverb language hedges) at that time, the user is released from such annoyance.

To make this possible, a membership function for each attribute information is defined, and a search condition refinement by fuzzy logic is introduced. The results show that the system accepts a Fuzzy expression of query as well as a

comprehensive dialogues between users and the system. As a result, candidates meeting these search conditions (in this case, seven candidates) were presented, and it was confirmed that the user could further narrow down from these candidates.

## VI. FUTURE RESEARCH WORKS

The proposed system is adopted in the real earth observation satellite database system, and it is a future subject to realize a more usable database system.

## ACKNOWLEDGMENT

This research was developed with reference to some of the results obtained by the joint research with the Remote Sensing Technology Center, which was carried out with the support of the National Space Development Agency of Japan. The author thanks Dr. Shinichi Sobue of the National Space Development Agency of Japan, Mitsuhiro Morishita of the Remote Sensing Technology Center, and other concerned parties for their joint research opportunities.

The author, also, would like to thank Professor Dr. Hiroshi Okumura and Professor Dr. Osamu Fukuda for their valuable discussions.

## REFERENCES

- [1] Shin-ichi Sobue, Kohei Arai, Bunra Yoshida, Osamu Ochiai, Mina Ogawa, Mineo Sekiguchi, Tomotaka Sekiya, Masao Takagi, User-friendly and efficient catalog information management and provision method for earth observation satellite data, Proceedings of Advanced Database System Symposium'94, Information Proceedings of the Symposium on Processing Society of Japan, Vol.94, No.13, 111-116, 1994.
- [2] Personal Correspondence, gA Global Change Data and Information System (GCDIS), Assited Search for Knowledge (ASKh, Committee on Earth Observation Satellite System, Working Group on Data, Catalog System and network System Subgroup Joint Meeting, 1994.
- [3] Ulrich Geske, Optimization and Simulation of Complex Industrial Systems, Proceedings of the 11th International Conference on Application of Prolog, INAP'98, 1998.
- [4] Masao Fukushima, Introduction to Mathematical Planning, Asakura Shoten, 1996.
- [5] Kohei Arai, Manabu Arakawa, Hirofumi Eto, Vague Search of Earth Observation Image Database Based on Fuzzy Theory Using Physical Quantities and Spatial Features, Journal of the Japan Society of Photogrammetry, Vol. 38, No. 4, pp. 17-25, Aug. 1999.
- [6] Kohei Arai, Hirofumi Eto, Tomoko Nishiyama, Earth Observation Satellite Image Database System Allowing Ambiguous Search Requests, Journal of the Japan Society of Photogrammetry, Vol.38, No.4, pp.47-52, Aug.1999 .
- [7] .Shin-ichi Sobue, Kohei Arai, Fumiyoshi Yoshida, User Friendly and Efficient Catalog Information Management for Earth Observation Data, Information Processing Society of Japan, Advanced Database Workshop, pp. 111-116, December 1994, 1994.
- [8] Kohei Arai, Remote sensing satellite image database system allowing image portion retrievals utilizing principal component which consists spectral and spatial features extracted from imagery data, International Journal of Advanced Research in Artificial Intelligence, 2, 5, 32038, 2013.
- [9] Kohei Arai, Data collection and active database for tsunami warning system, Proceedings of the 1st International Workshop on Knowledge Cluster Systems, 2007.
- [10] Kohei Arai, Remote sensing satellite image database system allowing image portion retrievals utilizing principal component which consists spectral and spatial features extracted from imagery data, International Journal of Advanced Research in Artificial Intelligence, 2, 5, 32038, 2013.
- [11] Kohei Arai, Numerical representation of web sites of remote sensing satellite data providers and its application to knowledge based information retrievals with natural language, International Journal of Advanced Research in Artificial Intelligence, 2, 10, 26-31, 2013.
- [12] Kohei Arai, Indra Nugraha Abudullar, Hiroschi Okumura, Image retrieval based on color, shape and texture for ornamental leaf with medicinal functionality, International journal of Image, Graphics and Signal Processing, Vol.6, No.7, June 2014.
- [13] Cahya Rahmad, Kohei Arai, Comparison contour extraction based on layered structure and Fourier descriptor on image retrieval, International Journal of Advanced Computer Science and Applications, 6, 12, 71-74, 2015.
- [14] Kohei Arai, Cahya Rahmad, Image Retrieval Method Utilizing Texture Information Derived from Discrete Wavelet Transformation Together with Color Information, International Journal of Advanced Research on Artificial Intelligence, 5, 10, 1-6, 2016.
- [15] S.Sobue and Kohei Arai, Metadata Definition and Retrieval of Earth Observation Satellite Data, Proceedings of the IEEE Metadata Conference, (1997).
- [16] Kohei Arai, Geographic information system: GIS based on neural network for appropriate parameter estimation of geophysical retrieval equations with satellite remote sensing data, Proceedings of the IEEE Geoscience and Remote Sensing, PID 220128, 2006.
- [17] Kohei Arai, Yuji Yamada, Image retrieval method based on hue information and wavelet description based shape information as well as texture information of the objects extracted with dyadic wavelet transformation, Proceedings of the 11th Asian Symposium on Visualization, ASV-11-08-10, 1-8, 2011.
- [18] Kohei Arai, C.Rahmad, Wavelet based image retrievals, Proceedings of the 260th conference in Saga of Image and Electronics Engineering Society of Japan, 243-247, 2012.
- [19] Kohei Arai, Image Retrieval Method Based on Back-Projection, Proceedings of the Computer Vision Conference 2019.
- [20] ANCI / NISO, Z39.50, Information Retrieval (Z39.50): Application Searvice Dfinition and Protocol Specification, URL: <http://lcweb.loc.gov/z39.50/agency/>, 1995.
- [21] Terano, J., Asai, K., Sugano, M., Introduction to applied fuzzy systems, Ohmsha, 1989.
- [22] Naoya Miya, Kazuo Tanaka, Practical Fuzzy Book in C Language, Russell, 1989.
- [23] Kaoru Hirota, Introduction to Fuzzy Expert System, Ohmsha, 1993.
- [24] Masakazu Saka, Fundamentals and Applications of Fuzzy Theory, Morikita Publishing, 1989.

## AUTHOR'S PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Science Commission "A" of ICSU/COSPAR since 2008 then he is now award committee member of ICSU/COSPAR. He wrote 37 books and published 570 journal papers. He received 30 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA and IJISA. <http://teagis.ip.is.saga-u.ac.jp/index.html>.

# Difficulties in Teaching Online with Blackboard Learn Effects of the COVID-19 Pandemic in the Western Branch Colleges of Qassim University

Fahad Alturise  
Computer Department  
College of Science and Arts  
Qassim University, Alrass, Saudi Arabia

**Abstract**—The global COVID-19 pandemic has compelled educational institutions to shift from face-to-face teaching methods to fully online courses. This was possible with the help of information technology advances, which led to the creation of Blackboard Learn, a Learning Management System (LMS). By transitioning their systems to this newly developed LMS, the western branch colleges of Qassim University in the Kingdom of Saudi Arabia were able to support e-learning. To investigate the influence of online learning e-courses on educational institutions and learning outcomes, this paper intends to perform surveys on both faculties and students. The survey mainly focuses on course objectives, practical skills, faculty member's responses regarding query and discussion, explanations on applied courses, problem-solving, and improving teamwork skills. A comprehensive investigation of the faculties reveals that 59.08% of faculty members believe it is challenging to facilitate course objectives due to the lack of practical lab work and other detailed knowledge exchange on applied courses, which leads to the faculties being unsatisfied with online courses when compared with traditional systems. Moreover, 77.17% of the students think it is difficult to have discussions during online courses in order to solve queries, and this diminishes their problem-solving capability. In addition, with an online course system, there is no way to physically collaborate in teams and work on team projects to improve teamwork abilities.

**Keywords**—COVID-19; blackboard learn; e-learning; learning management system; pandemic; difficulties; Qassim University

## I. INTRODUCTION

During the global COVID-19 pandemic, where most countries went into lockdown, traditional face-to-face teaching in educational institutions has been replaced by fully online, e-learning courses utilizing the Learning Management System (LMS). By using LMS, a student can perform all required activities via e-learning (electronic learning) including lectures, homework, assignments, quizzes etc. on an online platform. The use of Blackboard Learn and e-learning is effective because there is a combined interaction of students and teachers using the same online website with a remote interface. Online communication and e-Learning can be referred to as blended learning, according to authors in [1]. A proper environment, with a user-friendly interface, is a basic objective of the higher authorities [2]. It is worth mentioning that e-learning has greatly helped with the continuation of the educational system, without adverse side effects, during this

tough period dealing with COVID-19. The e-learning system avoids the need for students to gather, which helps in suppressing the spread of the coronavirus, and ensures a reliable and safe educational environment. Without e-learning it would not be possible for the faculty and students to communicate efficiently and continue learning outside the educational institute [3] in this pandemic crisis. For that reason, there needs to be a proper methodology by which students and faculty members are guided and taught about the beneficial outcomes of e-learning in a way that they efficiently adopt to its environment as soon and as effectively as possible.

LMS is comprised of course contents like quizzes, survey, lectures and assignments etc. The course content and details can be accessed remotely from anywhere as long as you have a personal computer or smart phone internet access. The system facilitates easy communication between student and teacher [4]. As already discussed, Blackboard Learn has proved a worthy platform for productive communication between teachers and students, and it also has a justified and transparent grading system. A survey was conducted by authors in [5] on the development of a unified theory of acceptance and the use of technology (UTUAT), model which successfully addresses the lower adaptation of Blackboard Learn in educational institutes. The outcome of the research was in favor of e-learning as it could provide access to courses from anywhere, more efficiently than ever. Therefore, the goal is a detailed study of the Blackboard Learn LMS, specifically its usefulness and shortcomings in the e-learning space of the western branch colleges of Qassim University in the Kingdom of Saudi-Arabia.

## II. E-LEARNING IN DISASTERS

Today, the world is facing the global pandemic of COVID-19, which presents a serious threat to every element of society. This has caused everyone to consider preventative measures against the virus. There has been a notable shift in risk assessment policies made by disaster management authorities due to the recent pandemic [6]. The sections relating to policy management have also been a part of the online e-learning system during these difficult times. Specifically, educational institutes and education in general have been badly affected due to COVID-19 as all worldwide educational institutions are closed [7]. In this challenging situation it is difficult for untrained faculty members and teachers to continue the

provision of education to students. The following section describes the necessary strategies that need to be employed during the current situation.

#### A. Blackboard Learn e-Learning LMS

Online e-learning systems, specifically Blackboard Learn, have provided a suitable platform for both participants and learners to acquire knowledge, either separately or in collaboration. During this current pandemic, Blackboard Learn has provided some very important user benefits in the form of reliable and permanent online content. Also, it permits users and learners to stay at home and continue their education in an efficient way. Furthermore, travelling expenses for both students and teachers have become negligible due to the employment of e-learning, as it is available worldwide and remotely. Since different people are residents in different places, an online, remote interface for communicating and sharing knowledge is greatly needed [8].

Blackboard Learn ensures the efficient delivery of education in a digital format rather than the conventional and traditional education structures. Education and learning in soft form have reduced the use of paper and other writing materials, which also reduces the impact on the environment. In the current, global lockdown situation, people are avoiding unnecessary travel but still need to continue their education. Blackboard Learn is the safest option for acquiring quality education at home. However, Blackboard Learn still has some limitations. The author in [9] researched the limitations of Blackboard Learn, and found that basic knowledge and skillful use of a computer is very important to maximize knowledge acquisition using e-learning. Training and awareness regarding computer knowledge are costly. Furthermore, a teacher cannot explain an idea online as clearly as face-to-face. The author in [10] says that these limitations must not be ignored. Furthermore, due to health issues in this pandemic situation, people may not be able to properly concentrate on online learning, which is also a limitation.

### III. E-LEARNING DURING THE CORONAVIRUS OUTBREAK

COVID-19 is an aggressive, highly infectious virus disease related to pangolins [11], which can be transmitted via physical contact between two human beings. The microscopic image of the virus resembles a crown, hence the common name coronavirus. The fatality rate of COVID-19 is far higher than conventional flu. In some people, a strong immune system and timely quarantine may be helpful in avoiding the spread, but it is particularly life-threatening for older people. To reduce the risks of this disease, and to keep humanity safe from disaster, the only method of protection is to isolate people from one another, and governments worldwide has imposed lockdown and quarantine measures [12]. Lockdown proves to be the most successful way of imposing social distancing, even for state officials and presidents [13].

Many countries have applied lockdown measures including closed markets, government sector organizations and, specifically, educational institutions such as universities and colleges. This has created a need for the adaptation of e-learning such as Blackboard Learn, which ensures social distancing and quality education. Blackboard Learn provides a

platform and interface for bi-directional communication between students and teachers to continue their process of learning during this pandemic [14]. Aside from various fields of education, Blackboard Learn also provides safety measures and guidelines to be taken in this current situation. The absolute advantage of e-learning is a move towards a modern and advanced learning method, avoiding traditional learning techniques.

Most of the top technical, engineering and management institutions, as well as some private learning organizations, have shifted at a rapid rate to the adaptation of e-learning, even though there are still some existing challenges and limitations in this approach. The research carried out on e-learning has not yet been fully established with a proper conclusion, where all pros and cons have been addressed regarding the proper utilization of e-learning [15]. Research has been presented concluding that online learning isn't significantly helpful in lower classes at school level. Still, a good number of higher educational institutions have acquired a great advantage from the adaptation of e-learning in this current global pandemic. Blackboard Learn is a kind of asynchronous and blended teaching system. Blackboard Learn, together with synchronized, on-campus teaching and interactions, are proving more useful than Massive Open Online Courses (MOOC).

#### A. Blackboard Learn e-Learning during COVID-19

It is not known how or when the COVID-19 crisis will end. It could take up to a year, based on the current numbers of deaths and infected people. This brings the likelihood of a full adaptation to e-learning in most parts of the world. To facilitate this, there is an urgent need for all participant to have a personal computer, the internet and a sound knowledge of using both. An uninterrupted internet supply is a huge challenge in underdeveloped and developing countries nowadays. The transition to e-learning is moving very swiftly, and the success of online learning greatly depends on the knowledge of teachers and their way of conveying it to students [16]. The interface needs to be user-friendly for tutors and teachers, and it is challenging to encourage web designers and web developers to adjust websites to make them meet the needs of teachers dealing with lockdown and social distancing. Teaching staff must find a method of understanding the online learning interface, and for that they can seek the help of more experienced colleagues and friends [17]. This is a good opportunity for faculty members to move up a gear and start learning new ways of teaching, in order to excel in their educational and professional careers. In a nutshell, it is clear that it will take a lot of effort for teachers to successfully provide and convey education to students, as they also need to protect their family members, and students need to be restricted by their parents from going outside and ensuring social distancing.

### IV. HISTORY OF BLACKBOARD LEARN

Two professional educational advisors, namely Matthew Pittinsky and Michael Chasen, formed the basics of online learning when they formed the company Blackboard LLC. Later, they partnered with IMS Global Learning. IMS Global was basically a non-profit organization included in the national

Educause program. As mentioned earlier, Blackboard Learn was launched with the theme of connecting faculty members and students so they can communicate anytime, anywhere, from any remote location. Teachers can upload the necessary course material from home, and students access it from home, completing their assignments online. An additional advantage to this is that their time management will improve over time. Furthermore, Blackboard Learn needed a course management company and they selected CourseInfo. Originally, CourseInfo was launched at Cornell University, and went on to serve as a course management company for Blackboard LLC. Blackboard LLC also contracted MadDuck technologies, which were previously very tough competitors of Richmond. Within the period of a single year, Blackboard LLC excelled by purchasing assess solutions from AT&T. In that time, Blackboard LLC acquired competitive enterpriser institute (CEI) special teams from colleges. Blackboard LLC started to purchase their competitor companies, like Promethius, which they bought from George Washington University. To manage financial issues, Blackboard LLC purchased shares in a transaction company called SA cash. Blackboard LLC was publicly launched in June 2004 after acquiring some \$75 million from investing in the stock market. Blackboard LLC partnered with WebCT, and with their help they were able to cover online learning and course management of over 80% of North America. A research concluded that Blackboard LLC is currently being used by almost 70% of universities and colleges in the United States. A study in 2006 found that 60 countries, comprising about 12 million people, employ Blackboard LLC. Blackboard LLC contracts with different educational institutes, offering their services of online e-learning to nearly 2,200 educational institutions using 12 different languages. Blackboard LLC achieved a milestone when its shares on NASDAQ doubled their original price in 2005. Blackboard LLC acquired revenue by offering e-learning to a variety of educational universities. However, a great portion of the revenue was generated from the renewal of licenses of their two-product portfolio [18]. Blackboard LLC is now comprised of two products, namely Networked Transaction Environment (NTE) and Networked Learning Environment (NLE) [19, 20]. Blackboard LLC transactions are carried out by NTE. Its primary aim is to provide an easy interface and server to handle worldwide transactions of Blackboard LLC. Access to new accounts and settings is also being offered by NTE. NTE is an academic suite providing help regarding course and study content. Blackboard LLC can be seen as having NLE at its heart, due to the provision of different services related to education.

#### V. UTILIZATION OF BLACKBOARD SYSTEM

Along with providing an e-learning interface and LMS for communication purposes between teachers and students, Blackboard Learn also provides extra services like email, discussion boards, podcasts etc. Each user derives their own benefit from Blackboard Learn, as every user has a different way of using the application. For example, someone with an in-depth knowledge of e-learning and familiar with its long-term advantages, will gain more from using Blackboard Learn compared with a less capable user [21, 22]. The success of students in acquiring an education from Blackboard Learn is

also dependent on the teaching methodology and awareness of the teacher. The teacher must have a comprehensive knowledge of the usage of Blackboard Learn, and must also be aware of its interface, as that is how he or she can deliver their knowledge to students in the best way [23]. The totally efficient use of Blackboard Learn has not yet been achieved, as a study conducted concluded that only 23% of faulty members have a complete knowledge of e-learning and its integration with books. To address this issue, proper guidance in the shape of tutorials must be provided to everyone involved in e-learning [24]. The study mentioned the advantages of e-learning which include easy communication, a user-friendly interface, an easy way to manage assignments and quizzes, and a transparent grading system particularly suitable for students. Research concluded that Blackboard Learn is currently being used by almost 70 % of universities and colleges in the United States.

As mentioned above, being competent at using Blackboard Learn is very important for both students and faculty members and cannot be ignored, whatever the situation may be. Research conducted by authors in [25] concludes that lecturers and professors prefer to use of Blackboard Learn for optimized course management and online learning purposes. It also provides a good method of intercommunication between teaching staff and students. According to researchers in [26], some professors have difficulties, both personally and with their students, in learning to adapt to the Blackboard Learn interface and have issues with its flexibility. Large numbers of students working online can reduce the amount of time available to interact with their teachers and work out their assigned tasks online on Blackboard Learn [27]. A busy social schedule leads to limited utilization for some students [28]. Students need to work through tutorials for a better understanding of the Blackboard Learn interface, both for their own benefit and also to send feedback to their teachers so they can solve the students' problems remotely.

#### VI. DESIGN METHODOLOGY OF SURVEY

Colleges of the western branch of Qassim University located in the Kingdom of Saudi Arabia carried out a survey based on a questionnaire which included both students and faculty members. The questionnaire was comprised of a course offered by the Deanship about the adaptation and extension of e-learning into every department of the university. The course offered by the Deanship was included in the first phase of this very project. To carry out this project, e-learning was adopted for certain courses, namely Islamic 101, Psychology 101, and Islamic 102. A complete online course was developed by subject specialists with the help of the Deanship, investigate and understand the percentage of students willing to take online courses, and also to get an overall estimate of e-learning success. At the same time, the western branch colleges of Qassim University had already started implementing e-learning by offering six different online courses for students. After the completion of each topic and course, a survey was conducted in order to know whether the e-learning process was proving helpful or not, and to compare the results with conventional teaching methodologies.

A survey questionnaire was created online using Google forms, where all students and concerned personnel could complete the questionnaire online. The survey feedback was recorded and used in subsequent surveys to learn about any progress being made by the implementation of e-learning in the university. A graphical representation of the progress of e-learning was required for easy understanding and visualization of the survey results. This is why the Google spreadsheet was integrated with Google forms to present the complete analytical results graphically.

The survey was completed by both faculty members and students, of which the 22 faculty members comprised eight females and 14 males, and a total of 639 students comprised 271 female and 368 males. The participation process for the survey was made easy for the participants as they were sent an email including the survey link, which they could easily access from anywhere. The survey had different content for different genders, ages and academic positions, and for that purpose the overall survey questionnaire was divided into small subsections. The questionnaire was composed of questions where students and faculty members were directed to give feedback in the form of strongly disagree, disagree, neutral, agree and strongly agree. A comprehensive overview of the feedback received from the faculty members and students is listed in the results and discussion section below.

## VII. RESULTS AND ANALYSIS

Surveys were performed separately on both faculty members and students attending full online courses, and detailed investigation of the feedback achieved from both sets of subjects was analyzed individually. The principal goal was to investigate any difficulties with or limitations of the online e-learning system of the western branch colleges of Qassim University during this pandemic COVID-19 situation. The ultimate goal is to make improvements in the up-coming version of this system in the near future.

### A. Faculty Response to Survey Questionnaire

The faculty questionnaires were composed of six questions regarding any difficulties and limitations of the online course being used during COVID-19. This mainly focused on course objectives, practical skills, training, and handling of the electronic course in such a way that all students become capable of self-learning to present practical work skills. Detailed analysis revealed that about 73.9% of faculty members can handle the full e-course efficiently, and 91.3% can manage student self-learning. Thus, during COVID-19, e-courses make it easy for faculty members to thoroughly guide students attending the course. However, students failed to present practical activities to their respective course professors. A detailed description of the various difficulties being faced are listed below.

Moreover, from Fig. 1, it can be discerned that 18.18% strongly disagree and 40.9% disagree that the developed e-content of the course in this way will facilitate achieving the objectives of the course. However, 13.63% agree and 18.18% strongly agree that course objectives will be fulfilled. Feedback from the remaining 9.09% of faculty members was neutral. Analysis reveals that the majority of faculty members think the

objective of the course may not be facilitated using online e-learning.

From Fig. 2, analysis reveals that 18.18% strongly disagree and 36.36% disagree that the e-content of the course achieves scientific accuracy, however 18.18% agree and 9.09% strongly agree regarding the e-course content accuracy, and the remaining 17.39% of faculty members' feedback was neutral. It is can be seen that a total of 54.54% of the faculty members disagree that the e-course content has scientific accuracy. Whereas, Fig. 3 shows that about 40.9% of the faculty agree that online courses are appropriate for students. It is evident that the majority of the faculty members are against online courses since the basic course objectives can rarely be achieved.

Fig. 4 shows that that 18.18% strongly disagree, 36.36% disagree and a total of 31.91% agree that the course concerned with practical skills related to the course objective. It is evident that the majority of the faculty disagree that the course concerned with practical skills related to the course objective. All the faculty are conscious that fulfilling the practical course objective is challenging.

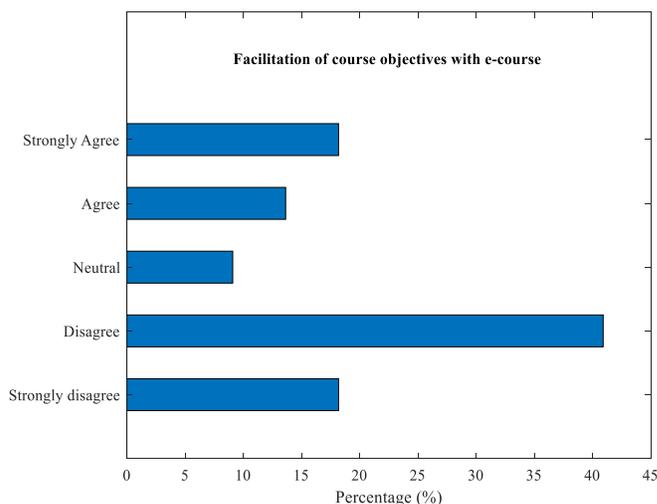


Fig. 1. Faculty Response Regarding Facilitating e-course Objective.

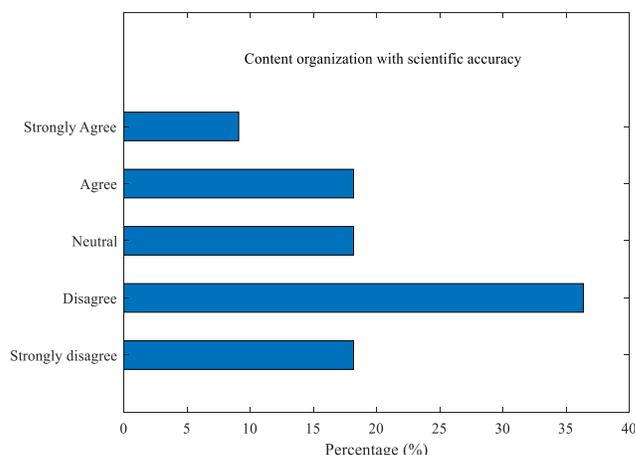


Fig. 2. Response about Content Organization with Scientific Accuracy.

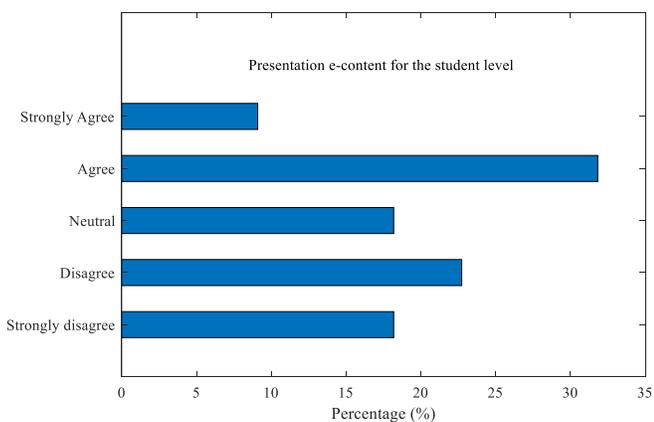


Fig. 3. Feedback about e-content Presentation for Student Level.

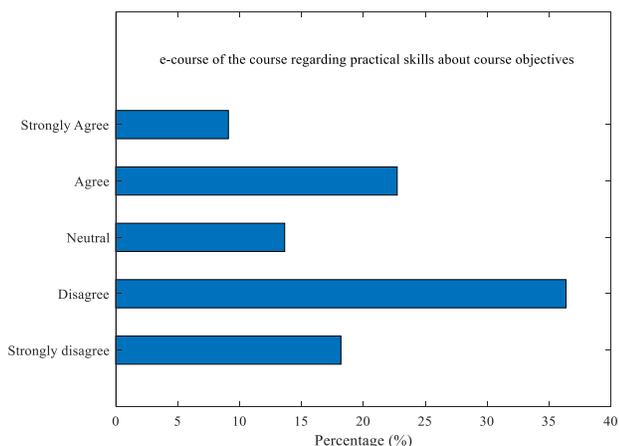


Fig. 4. Response Regarding Practical Skills about Course Objective.

Further Fig. 5 shows the details of the faculty response regarding their satisfaction with online e-learning course. It can be clearly seen that 13.63% strongly agree, 22.72% agree and a total of 49.36% disagree regarding satisfaction with the online course. However, it is worth mentioning that the faculty still feel it is the most reliable and best alternative solution to teaching and learning in this COVID-19 pandemic.

Detailed results of the faculty questionnaire are listed in Table I. Analysis concludes that the majority of faculty members agree that the electronic full online course offered by the western branch colleges of Qassim University, achieve the course objectives with appropriate e-content which are characterized with scientific accuracy, language integrity, with practical skill that can be easily browsed.

**B. Students Response to Survey Questionnaire**

The student questionnaire section was mainly concerned with the faculty response regarding query and discussion, explanation about the applied part of the course, asking questions during online courses, the ability to solve problems and, improving teamwork skills. From Fig. 6, it can be clearly seen that that only 4.85% strongly agree and 4.38% agree that the faculty were interested in responding to discussion and inquires, whereas 34.14% disagree, 43.03% strongly disagree and 13.61% students have neutral feedback. Analysis concluded that students are facing problems and difficulties in

discussions and inquires which arise in their minds during full online courses.

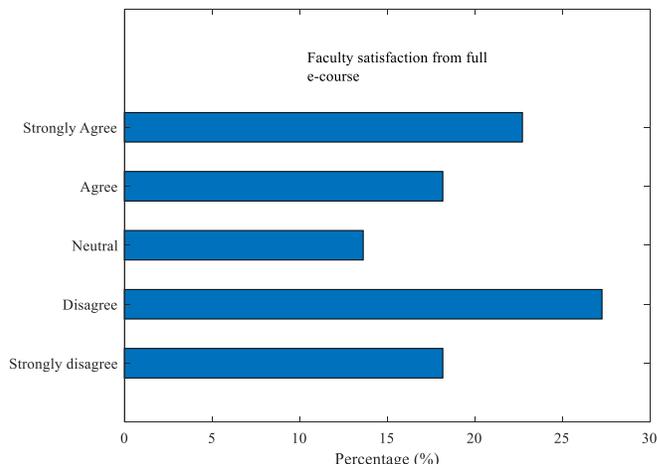


Fig. 5. Faculty Feedback Regarding Satisfaction from Full e-courses.

TABLE I. DETAIL INVESTIGATION OF SURVEY BASED ON FACULTY RESPONSE

Questions	Faculty decision (%)				
	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Facilitation of course objectives with e-course	18.18	40.9	9.09	13.63	18.18
Content organization with scientific accuracy	18.18	36.36	18.18	18.18	9.09
Presentation e-content for the student level	18.18	33.72	18.18	31.81	9.09
e-content of the course regarding practical skills about course objectives	18.18	36.36	13.63	22.72	9.09
Satisfaction from full e-courses	18.18	31.81	13.63	22.72	13.63

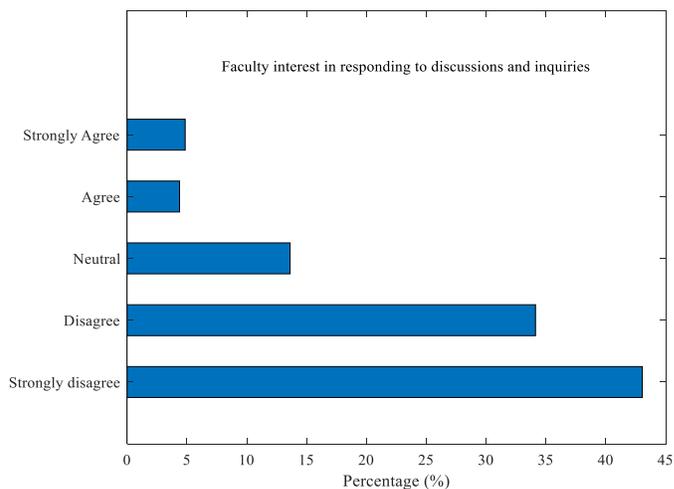


Fig. 6. Students Feedback about Faculty Interest in Responding to Discussion and Inquiries.

From Fig. 7, it is evident that 8.45% strongly agree and 8.29% agree that the faculty provide detailed explanations, whereas about 36.61% disagree and 27.69% strongly disagree with explanations about the applied part. In comparing both agree and disagree responses from students, it can be seen that majority of students have difficulties with the detailed applied knowledge.

In addition, Fig. 8 shows that 11.58% strongly agree and 9.58% agree that the e-course helped them to improve their ability to think and solve problems, whereas 32.70% disagree and 26.60% strongly disagree that helps to improve ability. This analysis concludes that the majority of students face problems with their studies and can't find a way to solve the problem.

Finally, Fig. 9 shows that 11.89% strongly agree and 13.14% of students agree that teaching with the new e-course, based on the full online course of Blackboard Learn helps to improve teamwork skills, whereas 33.64% disagree and 22.84% strongly disagree with this decision. Online courses don't involve any face-to-face meetings, and discussion are rare, therefore greatly influence teamwork skills.

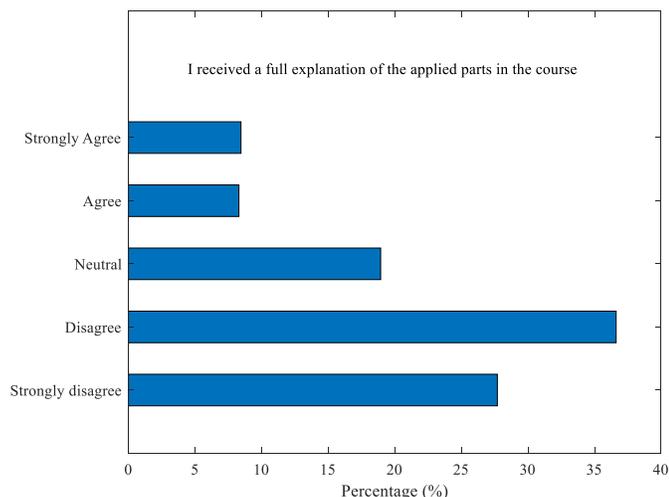


Fig. 7. Faculty Providing Information on Applied Part.

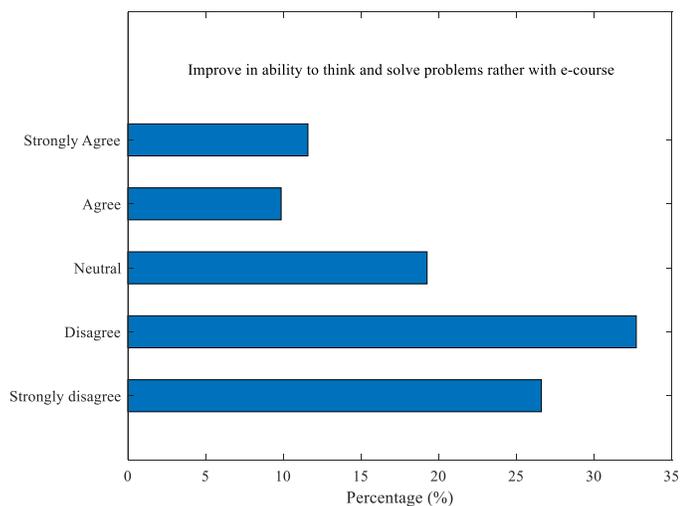


Fig. 8. Improvement to think and Solve the Problem.

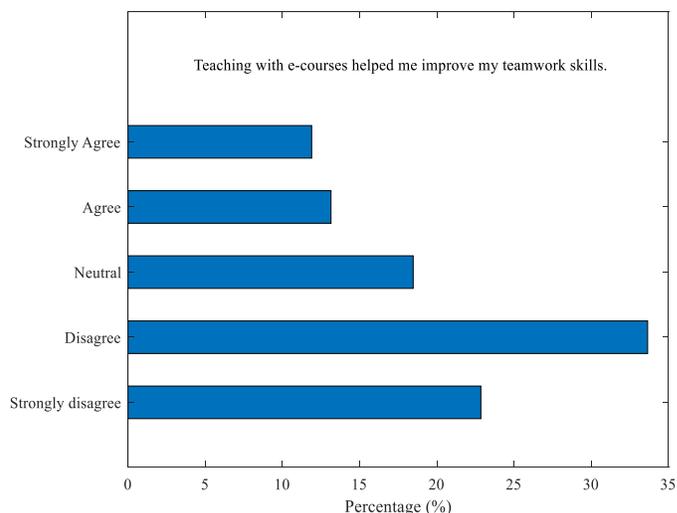


Fig. 9. Quality of the e-course.

Detailed investigations of survey-based student questionnaire responses are listed in Table II. Analysis concludes that the majority of students face a wide range of difficulties with different aspects of the e-course content. Based on the survey feedback, the majority of students suffer from developing knowledge with the applied sections, building teamwork skills, and gaining knowledge regarding problem solving.

TABLE II. DETAIL INVESTIGATION OF SURVEY BASED ON FACULTY RESPONSE

Questions	Faculty decision (%)				
	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Faculty interest in responding to discussions and inquiries	43.03	34.14	13.61	4.38	4.85
I received a full explanation of the applied parts in the course	27.69	36.61	18.93	8.29	8.45
Improve in ability to think and solve problems rather with e-course	26.60	32.70	19.24	9.58	11.58
Teaching with e-courses helped me improve my teamwork skills.	22.84	33.64	18.46	13.14	11.89

### VIII. DISCUSSION

During the COVID-19 pandemic, where the majority of countries worldwide went into a lockdown state, educational institutions are compelled to replace traditional face-to face-teaching with full online courses using LMS such as e-Learning. Different e-learning course systems were implemented based on LMS and visualized for its uses in the educational institutions. Introduction section presents and overview numerous authors studied based on survey regarding the effectiveness of the Blackboard LMS for online courses in

the past years. Here is this paper, with a detailed investigation of difficulties in the new developed full online course system at the western branch colleges of Qassim University in the Kingdom of Saudi-Arabia during COVID-19 where all the classes are shifted to online system. Analysis is carried out based on the survey response from the students and teachers for new online courses and Blackboard system.

Based on the faculty responses, analysis concludes that a total 59.08% of faculty members agree it is hard to facilitate the course objectives by using the online course, while only about 40% think it is achievable. This is because the e-course content differs in online course from the one studied before COVID-19. It also removes teamwork and the ability to practice lab work. Based on the survey response from the faculty members, only 45.46% think that the practical skill is not affected, however the majority believe it is. Since the course objectives are not fulfilled and practical skills are affected, almost 50% of the faculty members are not satisfied with the online courses.

Moreover, a comprehensive overview of student feedback leads to the conclusion that online courses greatly affect teacher-student discussions, since there is no practical lab work and issues arise regarding problem solving. Analysis shows that 77.17% of the students think that it is hard to have discussions in online courses, and to solve the questions that arise in their minds, because online courses lack practical work in the applied parts which are needed to fully understand the phenomena. Furthermore, it is concluded that since the questions remain unanswered, it can't help to develop problem solving capabilities. Finally, since it is online course system, there is no way to collaborate in a team to improve teamwork abilities.

## IX. CONCLUSION

In this paper, the author investigates difficulties in Blackboard Learn, an LMS, through a survey from faculty and students in western branch colleges of Qassim University in the Kingdom of Saudi Arabia on full online courses in the global pandemic COVID-19. The main objective was to investigate the influence of online learning e-courses on educational institutes and learning outcomes. This paper is intended to perform surveys on both faculties and students with key focus on the course objectives, practical skills, faculty response to queries and discussion, explanations about applied courses, the ability to solve problems, and the improvement of teamwork skills. Detailed investigation regarding difficulties in the online e-course during the global pandemic COVID-19 reveals that a total of 59.08% of the faculty members think it is hard to facilitate course objectives, compared with using traditional teaching systems, due to the lack of practical lab work and detailed knowledge on applied courses. . Moreover, 77.17% of the students think it is hard to have discussions in online courses to answer questions, which degrades their problem-solving capabilities. The aforesaid short comings are recommended for improvement in the near future in the updated version of the system.

## ACKNOWLEDGMENT

The authors are grateful to Dr. Ahmed Mohammed Abdelmoteleb, Director of the e-learning unit in the College of Arts and Science Areas, for his support and precious guidance regarding data collection and analysis in western branch colleges of Qassim University.

## REFERENCES

- [1] Poon, J. (2013). Blended learning: an institutional approach for enhancing students' learning experiences. *Merlot Journal of Online Learning and Teaching*, 9(2), 271-288.
- [2] Govender, I., & Mkhize, M. (2015). E-Learning in place of face-to-face lectures: an exploratory study of students' perceptions. *alternation*, 22(1), 183-203.
- [3] Heirdsfield, A. (2011). Blackboard as an online learning environment: What do teacher education students and staff think? *Australian Journal of Teacher Education*, 36(7), 1-16.
- [4] Peter Bradford et al. The Blackboard learning system: The be all and end all in educational instruction? *J. Educational Technology Systems*, Vol. 35(3) 301-314, 2006.
- [5] Devraj M, Irene G. Use of the Blackboard learning management system. *EURASIA J Math Sci and Tech Ed*, 14(7):3069-3082
- [6] Lee, Ho-Dong, Lee, Ho-Dong, "e-Learning system for disaster prevention & emergency management training program in Japan", Korea contents association conference, 2006.11.10, Pages.372-376
- [7] Oranburg, Seth, Distance Education in the time of coronavirus: quick and easy strategies for professors (March 13, 2020).
- [8] Sangrà, A., Vlachopoulos, D., and Cabrera, N. (2012) Building an Inclusive Definition of ELearning: An approach to the conceptual framework, *The international review of research in open and distance learning*, Vol.13, No.2, pp145-159.
- [9] Evans, J.R. and Haase, I.M. (2001) 'Online business education in the twenty-first century: an analysis of potential target markets', *Internet research: electronic networking applications policy*, Vol.11, No.3, pp.246-260.
- [10] Wong, D. (2007) "A critical literature review on e-learning limitations", *UCSI JASA Journal for the advancement of science and arts*, Vol. 2, pp. 55-62 (Science & Technology Issue).
- [11] Lam, Tommy Tsan-Yuk et al. 2020. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature*, <https://doi.org/10.1038/s41586-020-2169-0>
- [12] Harvey, David. 2020. Anti-capitalist politics in the time of COVID-19. *Jacobin*, 20 March 2020, <https://jacobinmag.com/2020/03/david-harvey-coronavirus-political-economy-disruptions>
- [13] Kopecki, Dawn. 2020. WHO officials warn US President Trump against calling coronavirus the "Chinese Virus". *CNBC*, 18 March 2020, <https://www.cnn.com/2020/03/18/who-officials-warn-us-president-trump-against-calling-coronavirus-the-chinese-virus.html>
- [14] Garrison, Randy D. 2011. *E-Learning in the 21st Century. A framework for research and practice*. New York: Routledge. Second edition.
- [15] HELIOS. (2007) e-Learning for innovation, [online], <http://www.menon.org/publications/HELIOS%20thematic%20report-%20Access.pdf>.
- [16] Ekuase-Anwansedo, A., Noguera, J., & Dumas, B. (2017). Transitioning from Blackboard to Moodle amidst natural disaster: faculty and students Perceptions. In *proceedings of the 2017 ACM annual conference on SIGUCCS* (pp. 19-22). ACM.
- [17] Cigdem, H., & Topcu, A. (2015). Predictors of instructors' behavioral intention to use learning management system: a Turkish vocational college example. *Computers in Human Behavior*, 52, 22-28.
- [18] Jayson S (May 2006) Blackboard breaks through. *The Motley Fool*. [www.fool.com/News/mft/2006](http://www.fool.com/News/mft/2006)

- [19] Pittinsky M, Bell T (2005) From the dining hall to the campus bookstore to a networked transaction environment: Overview white paper. Blackboard, Inc.
- [20] Pittinsky M (2004) The networked learning environment: overview white paper. Blackboard, Inc.
- [21] Mioduser, D., Nachmias, R., Oren, A. and Lahav, O. (1999) Web-Based Learning Environments: Current States and Emerging Trends. In: Collis, B. and Oliver, R., Eds., *Ed-Media 1999: World conference on educational multimedia, hypermedia and telecommunications, association for the advancement of computers in education*, Seattle, WA, 753-758.
- [22] Mioduser, D., Nachmias, R., Oren, A. and Lahav, O. (1999) Web-based learning environments (WBLE): Current implementations and evolving trends. *Journal of network and computer applications*, 22, 233-247.
- [23] Moeller, B. and Reitzes, T. (2011) Integrating technology with student-centered learning. Education development center, Inc. (EDC), Quincy, MA. Nellie Mae Education Foundation.
- [24] Bradford, P., Porciello, M., Balkon, N. and Backus, D. (2006-2007) The Blackboard Learning System: The be all and end all in educational instruction? *Journal of educational technology systems*, 35, 301-314.
- [25] Carvendale, D. (2003) Study of Wisconsin professors finds drawbacks to course-management systems. *Chronicle of higher education*, 49, A26.
- [26] Anderson, J.W. (2003) Faculty Perspectives of the Blackboard Course Delivery System.
- [27] Davis, E.A., Hodgson, Y. and Macaualy, J.O. (2012) Engagement of Students with Lectures in Biochemistry and Pharmacology. *Biochemistry and Molecular Biology Education*, 40, 300-309.
- [28] Leeds, E., Campbell, S., Baker, H., Radwan, A., Brawley, D. and Crisp, J. (2013) The Impact of Student Retention Strategies: An Empirical Study. *International Journal of Management in Education*, 7, 22-43.

# Using Geographical Information System for Mapping Public Schools Distribution in Jeddah City

Abdulkader A. Murad<sup>1</sup>, Abdulmuakhir I. Dalhat<sup>2</sup>, Ammar A Naji<sup>3</sup>

Faculty of Architecture and Planning  
Department of Urban and Regional Planning  
King Abdulaziz University, Jeddah, 80210, Saudi Arabia

**Abstract**—Geographical Information System (GIS) remains a unique tool use for school mapping for a clear understanding of the nature, planning, and distribution of educational facilities. The study carried out a GIS analysis for male primary and secondary schools' distribution in Jeddah city, Saudi Arabia, to show the significance of using GIS tools to assist the educational planning authorities. To understand, re-plan and address the location, distribution and availability challenges of the schools in Jeddah city. A Geodatabase for the study area was created, which incorporates education and population data collected from authorities. Spatial and network analyses are utilised to understand the location distribution, students' density, and the accessibility of the schools in the study region. The analyses results identified the services and students'; density, directional growth of the schools, drive-time service areas and served and un-served populace for the authorities in Saudi Arabia to make better planning decisions, address present and future challenges in the provision of primary schools to residents and most importantly to improve educational services. The findings revealed that shorter travel distances found in the denser (central) part of the city and some regions that need more schools.

**Keywords**—GIS; school mapping; educational facilities; geodatabase; spatial and network analysis

## I. INTRODUCTION

Both urban and rural areas are witnessing considerable growth in all aspects, including the education sector. The change requires proper planning and distribution of educational facilities by Authorities to achieve equity and demand. Educational facilities planning is a unique method of planning schools' and facilities distribution, size, and spacing for best utilisation [1] at all levels to accelerate progress and prosperity in political, economic, social and cultural sectors [2]. It includes a broad range of planning and management related to the resources' allocation, location, and distribution of educational services [3]. Planning of schools involves mapping, the planning of schools' location to guarantee competent and impartial distribution, mainly when extensive reforms or considerable growth of educational system take place [4]. School mapping involves the creation of relational geospatial databases for demographic, educational, social and economic information for educational authorities to help decision-makers and planners [5]. It is an essential tool to understand the inequalities from the public authority's investment.

Consequently, public authority's investment determines the pattern of distribution of educational facilities. Also, school mapping is a strategic planning process that tries to address the deficiency in conscious effort to locate educational facilities in a place for maximum utilisation by a targeted population to their advantage [6]. It is, therefore, a dynamic vision of how the current and future education services look like to aid the implementation of policies and to help authorities to achieve better decision-making regarding location and distribution of schools across different geographical areas to ensure optimum utilisation. According to [7], school mapping is essential in promoting accessibility and equity in the distribution to avoid over-localisation or over-concentration of schools in certain areas while other areas are lacking. Another importance of school mapping is to create a conducive condition for achieving universal education [8]. Likewise, it is a means for research, gaging accessibility and aiding good decision-making in school for better policies [9].

## II. BACKGROUND

The Kingdom of Saudi Arabia (KSA) is said to be the second-largest Arab state with an estimated total area of 2,250,000 Km<sup>2</sup> and a population of 34.2 million in 2019, according to the General Authority for Statistics [10]. Positioned somewhere between 16' and 32' S and 35' and 55' N, Saudi Arabia bordering Iraq, Jordan and Kuwait in the North, Yemen in the South, the United Arab Emirates, o Oman, Qatar and the Arabian Gulf to the East and the Red Sea to the West [10]. Besides, the cities of KSA are centres of commerce and culture. Riyadh is the most prominent, central and administrative capital with an estimated population of 4.21 million. While, Jeddah is a commercial city, a gateway to Holy Makkah and Medina with an estimated population of 3.43 million. The third biggest city is the Holy Makkah city, with an estimated population of 1.32 million. The total gross enrollment ratio of school students, according to the Ministry of education report is 98.1%, with an estimated total of 3.8 million students [11]. Table I compares the population of the three major cities kingdom, the area, students' enrollment and schools. Data obtained from the Ministry of Education shows that Riyadh city with the highest population, area, number of students and schools then Jeddah and Makkah cities followed respectively (Table I).

TABLE I. DATA COMPARISON OF MAJOR CITIES IN KSA [11].

City	Population (Million)	Area (Km <sup>2</sup> )	Primary Schools		Secondary Schools	
			No. of Schools	No. of Students	No. of Schools	No. of Students
Riyadh	4.21	1,913	1,896	612,437	711	188,433
Jeddah	3.43	1,765	1060	558,626	438	100,334
Makkah	1.32	1,200	666	213,896	222	71,534

In KSA, challenges particularly in students' enrolment and accessibility due to accelerated urban expansion, population growth, traffic congestion and lack of use of justifiable and suitable planning tool and scientific methods in the distribution planning of schools are currently the main issues in achieving equitable distribution and access to educational facilities in some parts of Jeddah city [12]. Equitable distribution means fairness in the distribution and planning of schools. Lagrab & Aknin [3] stated that the mapping and distribution of the services could provide proper planning and considering a spatial multicriteria (based on geographic information system) which can assist in planning, management, and policymaking. In Educational facilities planning, GIS tools can offer innovative evaluation tools and complete information packages for evaluating facilities and service area distribution [12]. Therefore, the information can aid policymakers, for example, understanding the spatial distribution of existing educational facilities and their catchment areas, both the density schools and students, identifying and selecting schools' locations. GIS as a specialised computer program collects, stores, manipulates, retrieves and analyses spatial information that can assist in understanding the distribution of educational facilities [13]. It is a computer-based data framework fit for incorporating information from different sources to give the data needed for effective decision-making [14]. It has become a useful tool that offers critical understanding, depicts relationships and aids the visualisation of data in significant hotspots within a spatial environment [15]. Environmental Systems Research Institute (ESRI) stated that GIS technology exploits two kinds of data (spatial and attributes data). Spatial data is the information which defines the absolute and relative position of geographic features [16]. At the same time, attribute data known as tabular data describes the quantitative and qualitative characteristics of the spatial features.

According to Musa and Mohammed [17], several measures are available for evaluating spatial accessibility to educational facilities which include; imposing standards (i.e. travel time and catchment area distance) to schools. Point and kernel density, ring buffer and overlay analyses can be employed to present the distribution and accessibility of schools. Additionally, network analysis calculates the schools' proximity access and adequate travel time to various schools' locations within the neighbourhood or study area. However, similar studies carried out applied only spatial analyses like the case of Al-Enazi, Mesbah, & Anwar [12] that applied overlay and buffer analysis in assessing the schools' spatial distribution in Jeddah city. The outcomes provided substantial evidence on the distribution of schools spatially with population density in different districts of Jeddah. Al-Rasheed & El-Gamily [18]

with a similar study in Kuwait presented the distribution of educational facilities by categorising parts of the city with no schools and failing to meet the least of students' needs using only spatial analysis.

Similarly, Aschale [19] presented a study on the assessment of schools' spatial distribution in Debre Markos town of North-Western Ethiopia using GIS to explain the spatial appropriation example of schools and the suitable future areas for the location of schools using various spatial analyses. Another study carried out in Kano, Nigeria by Olubadewo, Abdulkarim, & Ahmed [20] analysed the spatial distribution of primary schools using spatial analysis functions and concluded that most of the schools concentrate where the density of students is high, just like the case of Jeddah. Also, a study carried out by Lagrab and Aknin [3] on the suitability analysis for elementary schools in Mukalla districts of Yemen using spatial analyses; multiple ring buffer, then straight line distance measured, and map algebra of classified raster(s), applied to select suitable locations for the establishment of new schools.

The estimated driving time of students depends on age, mode of transportation, school choice policy, and a long drive to school may influence student's capacity to get to class on time [21]; thus, standard travel-time recommended by education authorities. A study carried out by Urban Institute reported that the average standard travel time to schools in most of the USA cities to be 10 minutes' drive [22]. In the case of Saudi Arabia, the spatial distribution of primary and secondary schools requires the use of the minimum and maximum criteria and standards of planning of schools approved standard catchment area for primary and secondary schools' location in Jeddah city is 500 and 2500 meters, respectively [23]. Moreover, similar previous studies carried out by wang et al. [24] and Sayed-Ahmad et al. [25] stated the international acceptable distance travel time standard to both primary and secondary schools as presented in Table II. The criteria are relative, not constant but depend on several planning variables. This study applies spatial analysis to classify schools, and network analysis to identify the travel time catchment area to schools in Jeddah city. The study also classifies the population that falls inside and outside the resulted travel time catchment area.

TABLE II. STANDARD CRITERIA FOR PLANNING SCHOOLS [23]

Criterion	Primary Schools	Secondary Schools
Catchment Area (Meters)	500m	2500m
Distance from residence to school (Kilometers)	0.5-1km	0.5-2.5 km
Distance from residence to school (Minutes)	10-15mins.	25-30mins.
Number of classrooms/School (Class)	12-24 classes	18-30 classes
Number of student/Class (Students)	20-30 students	20-30 students
Area for each student/Area (Square meters)	15-25m <sup>2</sup>	25-30m <sup>2</sup>

### III. MATERIAL AND METHODS

#### A. Study Area

Jeddah city is the second biggest city in the KSA after the capital city, Riyadh. It is a commercial city with the largest Seaport, located in Makkah Province on the Arabian Red Sea. Murad [26] stated that the Jeddah urban boundary and the total area of 1765 km<sup>2</sup> and 5460 km<sup>2</sup>, respectively. The city population estimated at around 3.43 Million in 2017. According to the UNESCO report, the total school enrollment percentage in Saudi Arabia by 2016 is 116.24%, and the male enrollment percentage is 117.26% compared to 115.19% female [27]. Based on the report, the school enrollment exceeded 100% in Saudi Arabia because of the inclusion of over-aged and under-aged pupils on account of the early or late school admission and grade repetition.

The first Saudi educational system was set up in 1924, at first, with barely any western schools [28]. Moreover, the Education Ministry in 1953 was established, explicitly to improve the educational system for male students to meet international standards [29]. Ministry of Education represents the government in providing and supervising free general education for all residents that consist of the advancement of the national educational plan, issuing strategies, providing training programs and assessing the educational performance of teachers, students and schools at numerous stages [30]. With this, more schools constructed, and public education began to develop in the Kingdom and with an overall principle that the framework offers an equivalent prospect to all students notwithstanding their background. With regards to school management in KSA, the system is highly standardised, centralised and schools' leadership is gendered, where the boys' schools are managed by males and the females for girls' schools [31] due to the Islamic culture of Saudi Arabia.

Additionally, data obtained from the Ministry of Education was analysed using ArcGIS to show schools' categorisation and location of male public schools in Jeddah city (Fig. 1). All public schools are managed and funded by the government, while, the private schools are funded solely or partly by students' tuition and managed by the private body but regulated by the government [11]. Children attend primary schools for six years, a 3-year intermediate and a 3-year secondary cycle with a separate higher education program as stated by the Saudi Arabian Cultural Mission report in the USA [32]. Furthermore, public schools are separated into male and female schools in Saudi Arabia. In KSA, the Ministry of Education sets benchmarks, creates educational plans, and structures course books for all schools everywhere throughout the Kingdom. Consequently, all schools must actualise similar principles, educational plans, and course readings. The Kingdom authorities are improving to address apparent shortcomings in their international standing; it has allowed a multiplicity of schools (private and international) to increase, normally charging dues, while following the authorities educational program [33]. Accordingly, the authorities continue refining rules and benchmarks by offering quality training programs for teachers, improving guidelines for

student assessment, expanding the utilisation of technology and education subsidy [28]. Similarly, King Abdullah bin Abdul-Aziz Project for Public Education Development established in 2007 [31] to transform the kingdom educational framework was to improve the education standard. The project initially set up 25 each of boys' and girls' secondary schools (Tatweer Schools) in 25 regions in Saudi Arabia, including Jeddah under Makkah province.

#### B. The Database

For this study, both analogue and digital data were collected, developed and updated into GIS format to form the database for this application (Fig. 2). Firstly, point data defining the schools-based locations converted to point features with the city district resolution in ArcGIS 10.2.1 by ESRI, Red- lands, CA, USA to present the location of schools in each district. Secondly, polygon and line data showing the city district boundaries/parcels and road network linked with point data and all the GIS converted attributes using ArcGIS application. Therefore, both the non-spatial and spatial data identified and characterised (Table III) which are linked together with detailed features of the study area (population, land use and school data) for different analysis. ArcGIS is then used to carry out various spatial and network analysis that presented the spatial spread of primary and secondary schools in the city Jeddah.

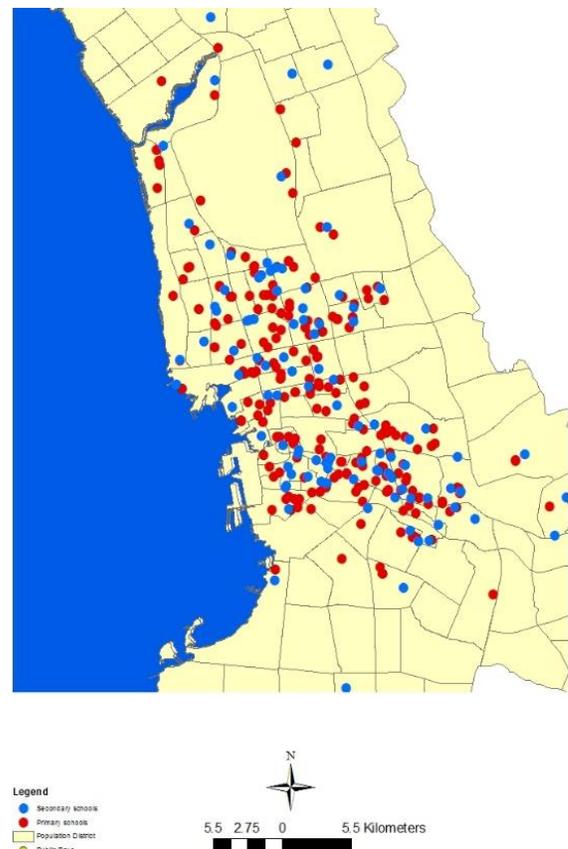


Fig. 1. Male Primary and Secondary Schools Location in Jeddah.

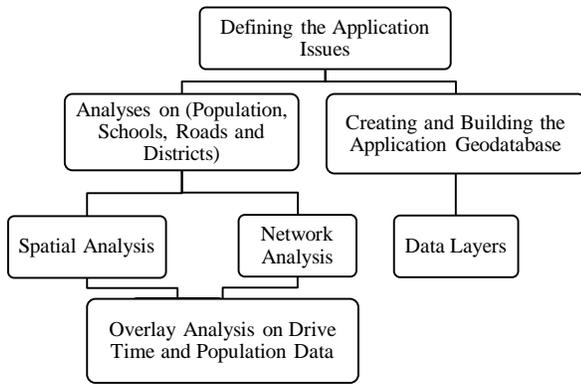


Fig. 2. Methodology Chart.

TABLE III. CHARACTERISATION OF DATA LAYER

Layer Name	Layer Description	Feature Type	Attribute
Land Use	Districts/Neighborhood Boundaries	Vector/Polygon	Name/Area, Population
Schools	School Types	Vector/Point	Supply, demand
Roads	Road Networks	Vector/Line	Length, Time

### C. Spatial and Network Analysis

Spatial analysis can solve multipart location-oriented problems and offer new perspectives to decision making. Overlay spatial analysis that includes schools' accessibility and population information, kernel density, and network analyses carried out using by this paper to provide an insight into the spatial access and distribution of schools in Jeddah. Firstly, Kernel density shows the quantity of the population for every point based on a quadratic formula with the maximum value at the point location and decreasing to zero at the search radius distance [16] (bandwidth) to obtain the density of features in the city districts.

Secondly, network analysis gives network-based spatial analysis tools for explaining complex routing issues via a configurable transportation network data model to represent unique network requirements [16] accurately. It calculates drive-times and locates facilities with poor accessibility. In many GIS software, the network analysis module comprises of a few modelling functions that include finding the shortest path, service area model and allocate model [26]. This paper utilised the Allocate and service area functions in ArcGIS Network Analyst for assessing schools' accessibility. Thus, it defines access to schools within the study area by estimating distances and solving other network problems using network connectivity. The distance along the definite travel direction is measured with the travel impedance between the source and destination and represented based on distance and time.

## IV. RESULTS

### A. Distribution of Schools in Jeddah

Define This study classifies schools based on the number of students using the 500m standard catchment area stated in Table I; in each school specifically to have an explicit nature of

the distribution in each district. Fig. 3 and Fig. 4 GIS analyses clearly show the classification and variation in the number of students in primary and secondary schools, respectively. Both figures indicated an unequal distribution of students, with the central city districts having the highest number and it is justifiably because of the population concentration compared to the outer North and South of the city.

Furthermore, kernel density analysis carried out shows the concentration and relationship between the density of students existing in the study area. It describes the features (students) within each district of the city. A similar study carried out in 2018 on health care centres by Murad [26] stated that kernel density calculates the density of features in a neighbourhood around those features, it can calculate for both point and line features and he further presented, as shown below, the algorithm used to determine the default Kernel search radius (bandwidth) by Murad [26] is:

- 1) Calculate the mean centre of the input points. If a Population field other than None were selected, this, and the values would weight all the following calculations in that field.
- 2) Calculate the distance from the (weighted) mean centre for all points.
- 3) Calculate the (weighted) median of these distances,  $D_m$ .
- 4) Calculate the (weighted) Standard Distance, SD.
- 5) Apply the following formula to calculate the bandwidth: 
$$\text{Search Radius} = 0.9 * \min \left( SD, \sqrt{1 \frac{1}{1n(2)} * Dm} \right) * n^{-0.2}$$

where:

SD is the standard distance

$D_m$  is the median distance

$n$  is the number of points if no population field used, or if a population field supplies,  $n$  is the sum of the population field value (number of students and classes).

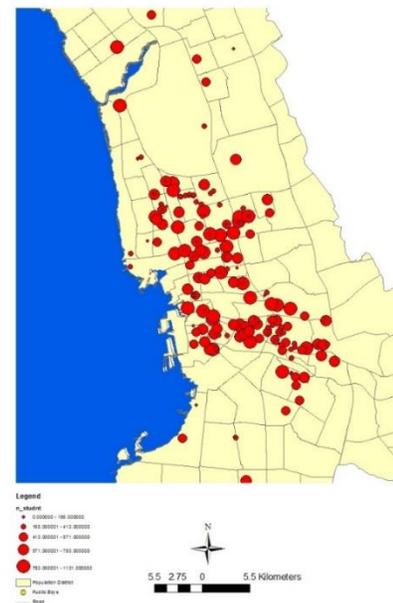


Fig. 3. Classification of Primary Schools based on the Number of Students.

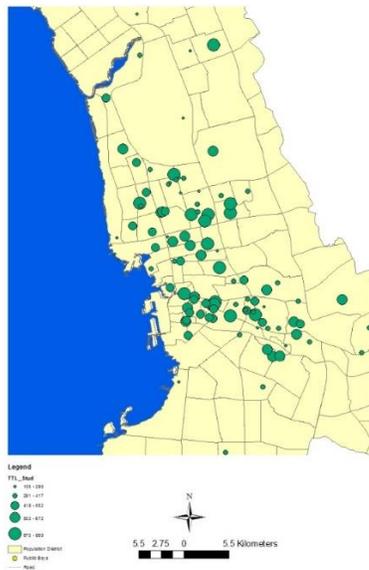


Fig. 4. Classification of Secondary Schools based on the Number of Students.

The presented algorithm calculated the schools' bandwidth based on the number of students in districts. Fig. 5 and Fig. 6 present the kernel density of both primary and secondary schools' students in the districts of Jeddah city and the darker the colour, the higher the density. Both analyses presented, indicated virtually similar results showing that concentration is more abundant in the central part of the city with a lower concentration in the less populated outer part (North, South, and East of Jeddah).



Fig. 5. Kernel Density of Students in Primary Schools.



Fig. 6. Kernel Density of Students in Secondary Schools.

Additionally, GIS analysis can give a sense of direction by drawing highlights on a map, figuring the standard deviational ellipse and clarifies the directional growth of primary schools (Fig. 7) and secondary schools (Fig. 8) in Jeddah city. The tool analyse and clarifies the trend, utilising either the region of the feature or the region impacted by an attribute value related to the features. The latter is labeled as weighted standard deviation. Both analyses indicated that the directional growth is mainly in the concentrated city centre towards the western part where higher number of students are observed. The result could help authorities and planners have an objective understanding of the existing concentration and distribution of students and services in various districts of the city for better decision making and services improvement.

#### B. Accessibility to Schools in Jeddah

Accessibility to services, according to Murad [26] is a multidimensional concept that defines the connection between features of facilities and the features of the facility distribution framework. He further pointed out four main ways of measuring accessibility to health location, and the same way is applied for this paper but in this case, for public schools. So, network distance that calculates the length of the shortest route along the transportation network used considering its actual index accuracy, which produces accessibility regions, road network and travel time among the roads of the city. It gives the actual service areas when the connecting routes are clearly defined. Defining and estimating the adequate travel time to schools in Jeddah city, the 10-15 minutes (primary schools) and 30 minutes (secondary school) acceptable distance travel time standard to school given by the Saudi Arabia Ministry of Municipal and Rural Affairs is applied. Hence, this paper

adopted 15 minutes' and 30 minutes' maximum drive time as the acceptable value for students to travel to primary and secondary schools respectively, taking into consideration the average speeds and speed restrictions on various street types and other physical hindrances to travel.

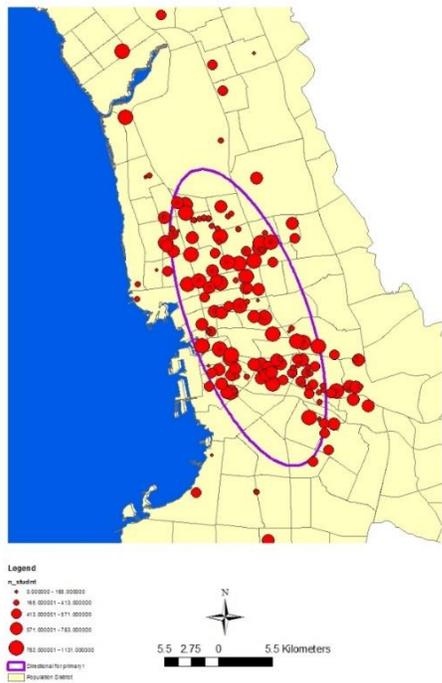


Fig. 7. Directional Growth of Primary Schools.

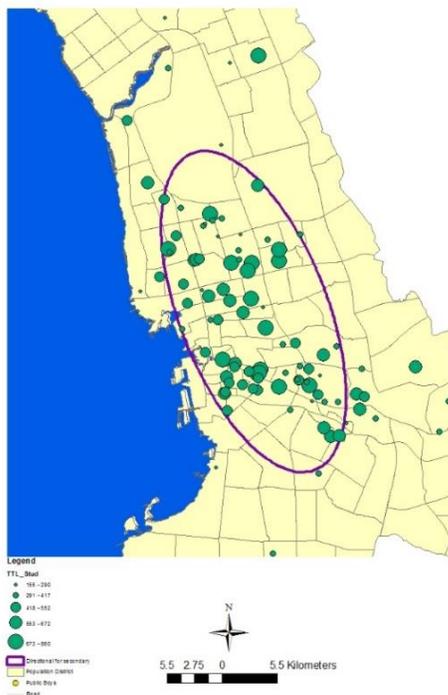


Fig. 8. Directional Growth of Secondary Schools.

Accessibility analysis carried out utilising the ArcGIS Network Analyst extension and created network service areas for existing male public schools' location in the study area. The resultant service zone is a zone which covers altogether the available path which is within 15mins and 30mins travel time to primary and secondary schools respectively. The resulting outcome of the drive-time service areas shown in Fig. 9 and 10 indicated the areas of the city that are well served and regions with low accessibility within the 15 minutes' (primary schools) and 30 minutes (secondary school) standard drive-time. Both the primary and secondary analyses show a similar pattern with the city centre, the north-western and northern part of Jeddah mostly covered. Therefore, most districts are well served (60% of the city population), while the city parts that fall outside the standard drive time service area are within the outer part of the city where most are new districts with less population or parts industrial zones. From the results outcomes, it is apparent that the accessibility to schools in Jeddah city is to a more considerable extent commendable. However, the results give more insight into where the planners and authorities should focus on within the city.

### C. The Population within Access to Public Schools

The population of students that falls within the access zones is obtained provided the service areas are well defined. Therefore, the drive-time analysis carried out in 3.2 presented a chance to define the served and unserved population. Fig. 9 and 10 presented correspondingly the results of the overlay analyses of the 15 minutes' and 30 minutes' drive-time of the population classification served and un-served population within access to the schools in districts of Jeddah (Fig. 13 and 14). The blue areas from both Fig. 11 and 12 indicated parts of the city with a high and served population. However, the red coloured areas in Fig. 13 and 14, respectively indicated the areas of the city with a high un-served population that require additional schools and these areas in the case of primary schools is higher if compared to secondary school that is very insignificant. Hence, the need for planners and authorities to consider increasing the number of schools around the red zones in the city centre to cover the un-served population.



Fig. 9. 15 Minutes' Drive-Time Service Area for Primary Schools.

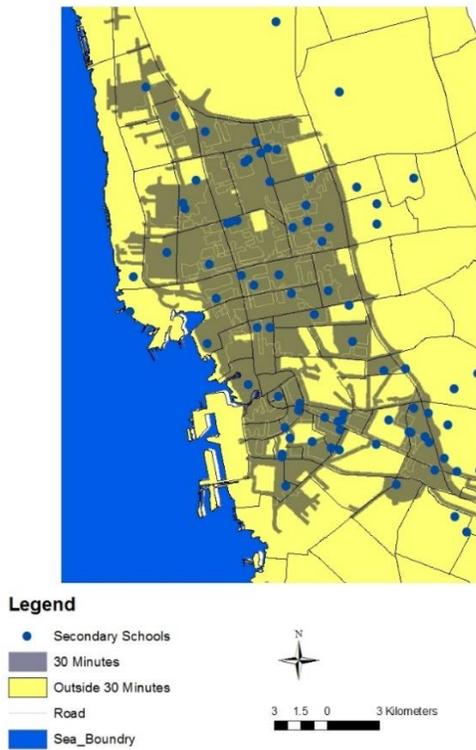


Fig. 10. 30 Minutes' Drive-Time Service Area for Secondary Schools.

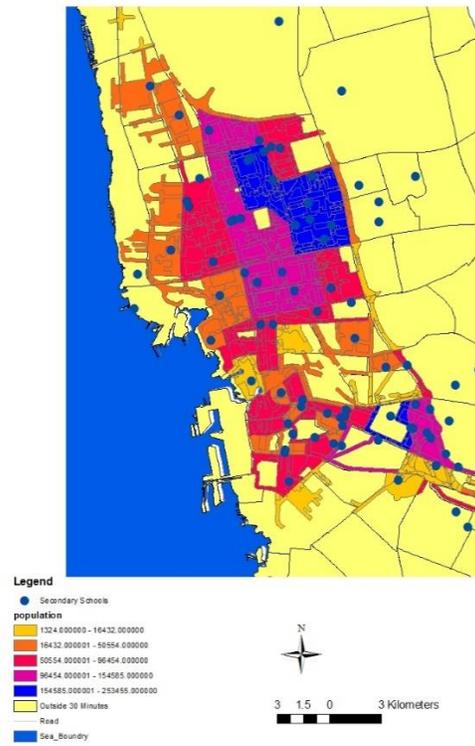


Fig. 12. Classification of the Population Inside 30 minutes' Drive Time Service area for Secondary Schools.

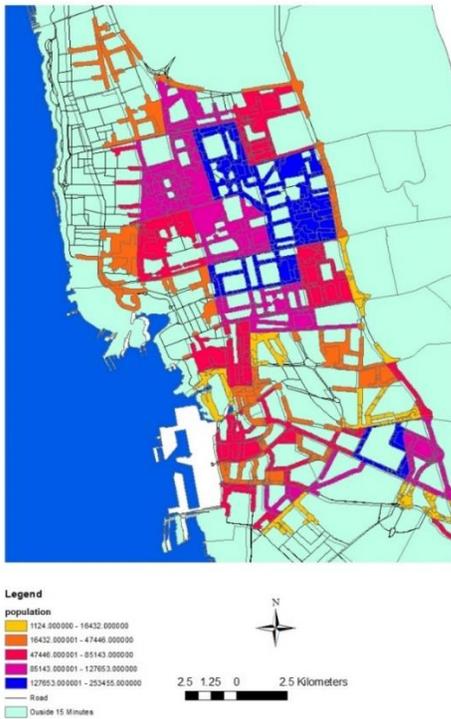


Fig. 11. Classification of the Population inside 15 Minutes' Drivetime Service Area for Primary Schools.

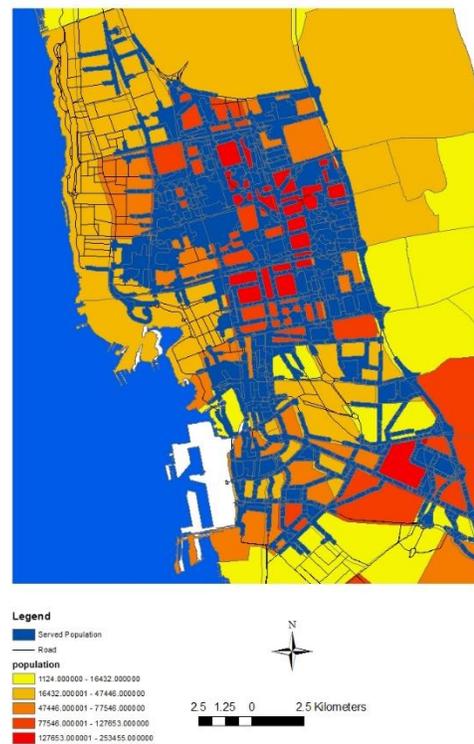


Fig. 13. Classification of Un-Served Population for Primary Schools.

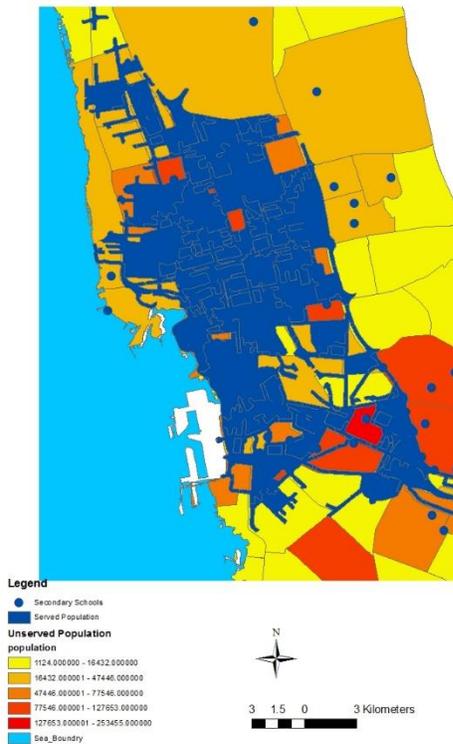


Fig. 14. Classification of Un-Served Population for Secondary Schools.

## V. DISCUSSION

The distribution of educational facilities needs to be a continuous process, and according to Department of Education report [2], functional and practical educational facilities could be planned and distributed based on a comprehensive plan of action to check possible shortages. For this study, there is an attempt to achieve spatial balance in the spreading of the male primary and secondary schools in the districts of Jeddah. While this appears to be achieved in some districts, the result of this study shows that GIS was not used to plan such facilities. However, the use of GIS technology identified the areas with shortages and access problems within the standard drive-time to schools. Meanwhile, the identification of the school locations (Fig. 1), the classifications of schools and students' population data (Fig. 3, 4, 5 and 6), show imbalances between the districts with high concentration and shorter travel distances in the central part of the city compared to the outer and less populated areas. A previous study carried out by Belarem et al. in 2018 mapped and analysed the school distribution and network in Makkah also shown a significant imbalance in the distribution of schools between the districts, lack of accelerated networks and public transport services to schools for ease of accessibility [9], therefore, it shows similar disparities with this study where the city centre has high schools' density compared to peripheral areas. Also, the results of a similar research carried out in Riyadh city [34], shows that the spatial distribution of schools in Orajja, Riyadh is adequate, however, the need to establish more schools to avoid overcrowding from the increasing number of students.

This study could be useful to authorities to address issues, for example, development of micro and macro plan of public schools particularly with the focus on access and dispersion by understudy's distribution by students. The research provided information for policymakers in Saudi Arabia to plan for both short and long term to build the effectiveness of schools' service management and for evaluating the performance of school services. Additionally, it is to note that studies in spatial distribution using GIS technologies have heaved in a brought thought. Unlike this paper that applied both spatial and network analysis for school distribution, most of the comparable studies utilised either spatial or network analysis to present their analyses, not both. For example, a previous study carried out in Jeddah city by Al-Enazi et al. used overlay and buffer analyses to only define the inequalities in school's distribution [12]. Moreover, two studies in Mukalla district in Yemen both used GIS technology also to assess the educational facilities distribution but the analyses did not go further to cover network analysis to assess the population within access to those education facilities [3, 35]. Another similar and recent study carried out in Debre Markos Town in North-Western Ethiopia [19] unlike the presented cases only assess school spatial distribution and identify appropriate zones for the creation of new school sites according to schools' standards but not defining the service areas and drive-time that present a chance to define the served and unserved population.

So, this paper combined spatial and network analyses in analysing male public schools in Jeddah city to help in making meaningful choices, for example, selecting school's best locations and recognising best new schools' locations by authorities. Additionally, the paper found out that the old districts have more population within the standard drive if compared to the new and outer city districts that have low population and fewer services. Students in the old (central) districts have shorter travel distances to schools. This finding is supported by the works of Al-Enazi, Mesbah, and Anwar [12] who argued that the city centres are the areas with the most served population. Additionally, most of the previous studies on educational facilities just found out school distributions using GIS [4, 12, 18, 19, 20] and a study that applied drive-time to classify the served and un-served population in Saudi Arabia is not being reported until this study.

## VI. CONCLUSION

This paper explained that the use of GIS tools in understanding the distribution of educational facilities is one of the best methods as it gives a better understanding of spatial and network situations upon which schools' planning decisions are based. It provided information on the various spatial classification of schools carried-out about students' density and distance travel-time in the districts of Jeddah city. The result identified the imbalances in the planning of male public schools in Jeddah city to help Authorities address the current and future challenges, and most importantly to make the best improvements to raise the level of efficiency in the distribution. Also, the findings indicated the zones of the city with long travel distances. Thus, the utilisation of the GIS tools explains the relationship among various spatial issues by giving a more more apparent and impartial assessment upon which to base planning and implementation decisions, as well as to justify

those decisions to policymakers. Finally, it is recommended that data update is required primarily for private and international schools by relevant authorities or subsequent researches so that more study is to be carried out for implementation, decision-making and services improvement in the education sector.

#### REFERENCES

- [1] Altbach, P. G.; Gumport, P. J.; Berdahl, R.O. American higher education in the twenty-first century: Social, political, and economic challenges. JHU Press: Baltimore, USA, 2011.
- [2] Department of Education. Educational Facilities Manual. Handbook on Educational Facilities: Integrating Disaster Risk Reduction in School Construction. Rev.ed. 2010.
- [3] Lagrab, W.; Akin, N. A Suitability Analysis of Elementary Schools Based Geographic Information System (GIS) A Case Study of Mukalla Districts in Yemen. *Journal of Theoretical and Applied Information Technology* 2017, Volume 95(4), pp. 731-742. Available Online: [www.jatit.org/volumes/Vol95No4/2Vol95No4.pdf](http://www.jatit.org/volumes/Vol95No4/2Vol95No4.pdf) (accessed 5th November 2018).
- [4] Ngigi, M. M.; Musiega, D.; Mulefu, F. O. Planning and Analysis of Educational Facilities using GIS: A Case Study of Busia County, Kenya. *AGSE* 2012. FOSS4G-SEA, pp. 261-269. Available Online: [https://www.researchgate.net/publication/275209527\\_Planning\\_and\\_Analysis\\_of\\_Educational\\_Facilities\\_using\\_GIS\\_A\\_Case\\_Study\\_of\\_Busia\\_County\\_Kenya](https://www.researchgate.net/publication/275209527_Planning_and_Analysis_of_Educational_Facilities_using_GIS_A_Case_Study_of_Busia_County_Kenya) (accessed 7th November 2018).
- [5] Khobragade, S.; Kale, K. School Mapping System Using GIS for Aurangabad City. *International Journal of Innovative Research in Computer and Communication Engineering* 2016. Volume 4. 17110-17119. DOI: 10.15680/IJIRCC.2016.0410002.
- [6] Ekpoh, U. I. School Mapping and Facility Planning. *Educational Planning in Nigeria: Principles and Practices*. University of Calabar Press: Calabar, Nigeria, 2018. Available Online: [https://www.researchgate.net/publication/332804564\\_School\\_mapping\\_and\\_facility\\_planning](https://www.researchgate.net/publication/332804564_School_mapping_and_facility_planning) (Accessed on 1st December 2019).
- [7] Akpakwu, S. O. Principles and techniques of education. Makurdi: Destiny ventures: Makurdi, Benue, Nigeria, 2012.
- [8] Adaja, C. F.; Osagie, R. O. Politics of school mapping and facility provision in higher education in Nigeria. In *Politics of education and national development in Nigeria*, N. M. Abraham, Durosaro, D. O et. al. University of Port Harcourt Press: Port Harcourt, Nigeria, 2015, pp. 81-86.
- [9] Belarem, M., Hamza, M., Jamil, A.; Ajmi, M. Mapping and Analysis of the School Network of Makkah Al-Mukarramah (Saudi Arabia), Jeddah Girls' Secondary Schools as Example. *Current Urban Studies*, 2018, Volume 6, pp. 102-120. DOI: 10.4236/cus.2018.61005.
- [10] General Authority for Statistics. 2019. Available online: <http://www.data.gov.sa/en/general-authority-statistics> (accessed on 9th November 2019).
- [11] Ministry of Education. 2018. Available online: <http://www.data.gov.sa/en/ministry-education-0> (accessed on 9th November 2018).
- [12] Al-Enazi, M.; Mesbah, S.; Anwar, A. Schools Distribution Planning using GIS in Jeddah City. *International Journal of Computer Applications* 2016, Volume 138 (1), pp. 33-36. <http://www.ijcaonline.org/research/volume138/number1/al-enazi-2016-ijca-908693.pdf>
- [13] Steinberg, S. L.; Steinberg, S.J. *GIS Research Method: Incorporating Spatial Perspectives*. Esri Press: Redlands, California, 2015.
- [14] Xia, L. and Yeh, A. G. Modelling sustainable urban development by the integration of constrained cellular automata and GIS, *International Journal of Geographical Information Science* 2000, Volume 14(2), 131-152, DOI: 10.1080/136588100240886
- [15] Graham, S. R.; Carlton, C.; Gaede, D.; Jamison, B. The benefits of using geographic information systems as a community assessment tool. *Public health reports* 2011, Volume 126(2), pp. 298-303.
- [16] ESRI. 2018. Available online: <https://www.esri.com/en-us/arcgis/about-arcgis/overview> (accessed on 19th December 2018).
- [17] Musa, H. D. & Mohammed, B. B. An Analysis of Spatial Distribution of Primary and Secondary Schools in Bida Town, Nigeria. *Abuja Journal of Geography and Development* 2011. Volume 3 (2), pp. 31-33.
- [18] Al-Rasheed, K.; El-Gamily H. GIS as an Efficient Tool to Manage Educational Services and Infrastructure in Kuwait. *Journal of Geographical Information Systems* 2013.
- [19] Aschale, T. M. Assessment of Schools Spatial Distribution and Identifying Suitable Areas by Using GIS Technology: In Case of Debre. *Journal of Resources Development and Management* 2017, Volume 35, pp. 8.
- [20] Olubadewo, O. O., Abdulkarim, I. A., & Ahmed M. The Use of GIS as Educational Decision Support System (EDSS) for Primary Schools in Fagge Local government Area of Kano State, Nigeria. *Academic Research International* 2013, Volume 4(6), pp. 614-624.
- [21] Canfield, J. P., Joseph, N., Dana H., Ashley H., and Willie E. Using a Person-Centered Approach to Examine the Impact of Homelessness on School Absences. *Child and Adolescent Social Work Journal* 2016, Volume 33 (3), pp. 199-205.
- [22] Urban Institute. The Road to School How Far Students Travel to School in the Choice-Rich Cities of Denver, Detroit, New Orleans, New York City, and Washington, DC, 2018. Available online: [https://www.urban.org/sites/default/files/publication/97151/the\\_road\\_to\\_school\\_7.pdf](https://www.urban.org/sites/default/files/publication/97151/the_road_to_school_7.pdf) (accessed on 25th June 2019).
- [23] Saudi National e-Government Portal. 2018. Available online: <http://www.data.gov.sa/en/ministry-municipal-and-rural-affairs> (accessed on 10th November 2018).
- [24] Wang, J. J.; Jing, Y. Y.; Zhang, C. F.; Zhao, J. H. Review on Multi-Criteria Decision Analysis Aid in Sustainable Energy Decision-Making. *Renew Sustain Energy Rev.* 2009; 13:2263-78.
- [25] Sayed-Ahmed, M.; Shehata, E.; Sennah, K. Development of Creep Model for Structural Insulated Timber-Foam Sandwich Walls Under Sustained Soil Pressure in Basements of Residential Buildings. *Proc Al-Azhar University English Conference*, 2010, Volume 5. pp. 603-614.
- [26] Murad, A. Using GIS for Determining Variations in Health Access in Jeddah City, Saudi Arabia. *International Journal of Geo-Information* 2018, volume 7, pp. 245.
- [27] UNESCO. 2019. Saudi Arabia: Education and Literacy. Available online: <http://uis.unesco.org/country/SA> (accessed on 9th May 2019).
- [28] Al-Sadaawi. A. Saudi national assessment of educational progress, *International Journal of Educational Policy and Leadership* 2010, Volume 5, pp. 1-14. <http://journals.sfu.ca/ijep/ijep/index.php/ijep/article/view/202/100>
- [29] Al-Sadan, I. A. Educational assessment in Saudi Arabian schools. *Assessment in Education: Principles, Policy & Practice* 2000, Volume 7(1), pp. 143-155.
- [30] Algarni, F. and Male, T. Leadership in Saudi Arabian Public Schools: Time for Devolution? *ISEA* 2014, Volume 42(3), pp. 45-59.
- [31] Alyami, R. H. Educational Reform in the Kingdom of Saudi Arabia: Tatweer Schools as a Unit of Development. *Literacy Information and Computer Education Journal (LICEJ)* 2014, volume 5(2), pp. 1515-1524.
- [32] Saudi Arabian Cultural Mission. *Educational System in Saudi Arabia*. SACM: Washington D.C, USA, 2006.
- [33] Prokop, M. Saudi Arabia: The politics of education. *International Affairs* 2003, Volume 79, pp. 77-89
- [34] Al Zeer, N. M. (2005). ANALYSIS of the Spatial Distribution of Public Secondary Girls' and Boys' Schools in Riyadh, Saudi Arabia. (Doctoral dissertation, University of Leicester).
- [35] Lagrab, W. and N. Akin, N. Analysis of Educational Services Distribution-Based Geographic Information System (GIS). *International Journal of Scientific & Technology Research* 2015, volume 3. pp. 113-118.

# An Efficient Approach for Storage of Big Data Streams in Distributed Stream Processing Systems

Sultan Alshamrani<sup>1</sup>, Quadri Waseem<sup>2</sup>, Abdullah Alharbi<sup>3</sup>, Wael Alosaimi<sup>4</sup>, Hamza Turabieh<sup>5</sup>, Hashem Alyami<sup>6</sup>  
Faculty of College of Computers and Information Technology, Taif University, Kingdom of Saudi Arabia<sup>1,3,4,5,6</sup>  
Independent Researcher Srinagar, Jammu and Kashmir, India<sup>2</sup>

**Abstract**—Besides, centralized managing, processing and querying, the storage is one of the important components of a big data management. There is always a huge requirement of storing immense volumes of heterogeneous data in different formats. In big data stream processing applications, the storage is given a priority and always plays a big role in historical data analysis. During stream processing, some of the incoming data and the intermediate results are always a good source of future samples. These samples can be used for the future evaluation to eliminate the numerous mistakes of storing and maintaining the big data streams. Hence, a big data stream application requires an efficient support for storage of historical queries. The researchers, scientist and academicians are working hard to develop a sophisticated mechanism that is needed for storage to keep the most useful data for the future references by means of stream archive storage. However, a stream processing system can't store the whole incoming stream data for future references. A technique is needed to get rid of the expired data and free the space for more incoming data in an archive storage. Hence keeping in view, the storage space limitation, integration issues and its associated cost, we try to optimize the stream archive storage and free more space for future data. The proposed enhanced algorithm will help to delete the obsolete data (retention or expired) and free the space for the new incoming data in a distributed platform. Our paper presents an Enhanced Time Expired Algorithm (ETEA) for stream archived storage in a distributed environment for removing the obsolete data based on time expiration and providing a space for the new incoming data for historical data analysis during the skew time (Hot Spots). We also evaluated the efficiency of our algorithm using the skew factor. The experimental results show that our approach is 98% efficient and fast than other conventional techniques.

**Keywords**—Distributed stream databases; storage optimization; stream archive storage; time expiration

## I. INTRODUCTION

Big data management is a way of centralized storing, managing, processing and querying the huge volume of different available data in numerous formats [1-5]. The traditional database technologies had failed to manage and control the flow of data which is overloaded with huge volume, variety, velocity and variability. However, alternative database technologies have played a vital role in solving “big data” issues of managing and processing. Their contribution has played a vital role in overall big data management [6]. As we know, big data computing has a huge demand for storage and processing [7]. Two types of processing can be done on big data. The first one is Batch processing and the Second one is Non-Batch processing of big data (real-time processing) [1] [8]

[9], and the Non-Batch processing includes the real-time OLTP online transaction processing database management DBMS systems. They possess variable workloads, spike in traffic and are always dependent on shared nothing architecture besides using the main memory for the processing and scalability. Hence, they are best to maintain the acid guarantee of the transactions [9-16]. Stream processing engines (SPEs) are the engines that can generate huge and big data streams continuously on the fly in a cluster of commodity servers. Stream computing involves the computations for the analytic purpose. Stream processing engines (SPEs) [17] have achieved broad adoption in research and industry [18-19] and mainly focus on scalable cloud computing. In distributed stream processing systems, most of the stream-based applications are distributed naturally [20]. One of the difficult problems that need to address in distributed stream processing systems is a storage. The large size and a variety of data always creates a hurdle for an efficient big data storage [21-23]. Data Storage is one of the crucial processes of big data analytics for real-world applications. These real-world applications include scientific experiments, social networks, healthcare and e-business. Till now, only Amazon, Google, Apache and some companies had provided the big data storage solutions. However, the available big data storage technologies are not enough efficient in a sense to provide consistent, scalable, and available solutions for the continuously growing heterogeneous data [24]. Distributed Stream Processing Systems (DSPPs) has smartly evolved to store discovered patterns, analyzed data, and extracted knowledge from different data processing stages. The Stored data must be useful data, which must be well controlled, organized and indexed along with metadata or external knowledge. The main purpose of storing the data is to get historical data for future verification and tuning purposes [25-26]. The stored data is often used for later reference. There is always a capacity limitation associated with every archive storage system. None of the systems can store more data than its capacity. The capacity of each system is directly associated with cost. Therefore, a sophisticated mechanism is needed to find the obsolete data (retention or expired), delete that data and free space for more incoming data in archive storage. Our proposed algorithm will provide a solution for the space limitation for the stream archive storage by detecting and deleting the retention data and free the space for the new incoming stream data without adding more storage externally. The algorithm will be beneficial in a way to save the cost of extra storage and its associated issues of integrating. Our proposed algorithm will use the skew factor for the retention policy which will guarantee the strength of the approach.

Hence, we try to optimize the stream archive storage and evaluated the accessibility of our approach by implementing it with the YCSB benchmark. For our experiments, we have used synthetic data as stream data and has implemented it with H-store. In our implementation, we have modified the Time Expired Packets Algorithm (TE) [27] for the optimization of the stream archive storage in a distributed setup. In summary, this paper makes the following contributions as 1) We discuss some of the open issues related to storage of big data streams in distributed stream processing systems and also elaborates storage optimization for archived data in a distributed streams databases (DSDBM's) (Mentioned in Background Section). 2) We presented an Enhanced Time Expired Algorithm (E TEA), for stream archived storage in a distributed set of processing nodes. The algorithm will remove the obsolete data and will provide a space for the new incoming data for historical data analysis. 3) We also maintain the efficiency using Skew factor) in our experiments.

The rest of this paper is organized as follows: In Section 2, is a background, Section 3 is the literature review, Section 4 is related works, Section 5 is the introduction of our proposed algorithm, Section 6 is our detailed algorithm (Enhanced Time Expired Algorithm (E TEA), followed by Section 7 will depict our evaluation and Section 8 which gives our conclusion and future works.

## II. BACKGROUND

Data Storage for stream processing is an important aspect for future enhancement. In big data steam processing applications, the incoming data and intermediate results may need to be stored to enable future analysis [28]. These applications require genuine support for storage and for historical queries. These required efforts help them for future analysis of historical data [29]. A lot of work has been done to optimize the stream archive storage. We have classified the stream archive storage optimization into three main subcategories which include 1) keeping most useful data. 2) Integrate Stored and Streaming Data (join live data). 3) Performance of storage manager.

### A. Need of Historic Data

Storage of Big Data Streams in Distributed Stream Processing Systems plays an important role in today's online stream world. One of its important pillars includes stream archive storage. As we know there is always a need for the storage of the intermediate data of stream processing in distributed setups for the enhancements, verification, future references and for tuning purposes [30]. Hence there is always a needed to keep the most useful data efficiently [28] for future analytics.

We try to highlight some of the open issues related to archive storage optimization.

### B. Open Issues of Storage Optimization

1) *Keeping most useful data:* For distributed stream databases, to keep the most useful data is always a key to storage optimization. An efficient storage algorithm is needed to keep the useful data and delete the obsolete data (retention or expired) from the archived storage. The optimizing protocol

should be simple and automatic in nature. There is a huge need to keep the most useful data in order to get future analysis from the least stored streams [27-28] for enhancements.

2) *Integrate stored and streaming Data (join live data):* Careful management of live and historic data is a basic requirement for archive storage optimization. The routine task for most of the stream processing applications (on-line data mining) is a comparison of live data with historical data. These applications need seamless switching of both past and present data for the purpose of comparison within the same application. Hence, a uniform language is needed to deal with either type of data for seamless integration. Moreover, there is a demand [30-33] for automatically switching from historical to live data without manual intervention.

3) *Performance of storage manager:* The stream databases should have the capability to store, access and modify the state information efficiently and effectively. One of the primary concerns for any stream database is its storage and management. The protocols related to storage managers are regarded as an indictment of determining and preserving the storage for the future. Most of the research works pertaining to stream storage have been on the issues related to the storage manager. On the other hand, storage optimization and related techniques for live stream databases are a comparatively new arena of the storage and that too when its distributed. Thus, maintaining storage becomes a significant mission. Although storage has continued to be studied for decades, its maintenance is still at the infant stage of research. We have studied and mentioned some works related to archive storage optimization and some related work with references for further understanding to shape and enhance the basic protocols of a storage manager, its optimization and related algorithms [8][34].

The capacity limitation of the archive storage system is directly associated with the extra cost and integration issues. Those solutions which are dependent on cost are never considered as an idle solution. Therefore, we provide an alternate solution to storage capacity issue. The feasible solution is to delete the obsolete data (retention or expired) of archives and make a room for the new data for further future analysis. Conventional methods like First in First Out (FIFO) and other related solutions are less efficient to provide the best solution when the stream databases are under skew time (Hot tuples-when most of the users and servers are on max utilization).

## III. LITERATURE REVIEW

There is substantial literature available for the data storage for stream processing related to big data applications. Some of the research work done in the field of data storage management aims to improve storage, integrate the live data with archive data and provide correctness guarantee for big data stream processing.

In [8], Fred Douglass et al. proposed a storage system that optimizes not only reading and writing but the creation and

deletion as well. Efficiency is achieved, by automating deletion based on relative retention values rather than requiring data to be deleted explicitly by an application. It works mostly on retention value functions, which effectively assign each data object a value that changes over time. In paper [28], Kirsten Hildrum et al. proposed an effective scheme for optimizing the placement of data within a distributed storage subsystem employing retention value functions. The goal is to keep the data of the highest overall value, while simultaneously balancing the read load to the file system. In paper [29], Nesime Tatbul, et al. proposed an S-Store which is designed to address the correctness aspect of a streaming application. In the paper, they prove the only way to achieve good performance is by tightly integrating storage management with the streaming infrastructure which supports correctness without serious performance degradation. The paper represents the exactly one processing, exactly one delivery, and transactional workflows. In paper [34], Irina Botan et al. proposed an optimized general-purpose storage management interface based on the parameters from the application requirement at different granularities. Using the interface and SMS (Storage Manager for Streams) can generate a customized storage manager for streaming applications. It uses information about the access patterns of streaming applications to tune and customize the performance of the storage manager. It efficiently handles time-critical tasks such as managing internal states of continuous query operators, traffic on the queues between operators, as well as providing storage support for shared computation and archived data. In Paper [8], Fred Douglass et al. proposed a storage system that optimizes not only the reading and writing but the creation and deletion as well. The papers use the concept of the relative value to retain the data items rather than deleting the data explicitly.

Therefore, keeping in view the necessity and the importance of the archive storage in distributed setups. It seems that there is a need for more research and a lot of work needs to be done for the archive storage and its optimization. We have designed an algorithm for the distributed stream processing engines which will help to optimize the stream archive storage by removing the expired data and free more space for incoming streams for the purpose of historical analysis during the skew time (hot tuples).

#### IV. RELATED WORKS

For research and a commercial purpose, there are many big data storage and analysis models available in a market. The challenges of big data storage are widely understood. At present, a lot of work is going on related to datacenter storage. Storage and processing through different data centers are growing as fast as the big data itself. A huge amount of data is created by a variety of users and devices. For storing and processing big data, the data center is always needed to establish network infrastructure which helps to gather this rapidly generated data [23]. In another work [35], the authors highlighted and classify the components of the network that must be established for better storage and communication in data centers such as the original data network, the bridges and a datacenter. Another related study [36], identifies the issues in using big data through specific locations.

We have categorized some of the related works especially related to resource management and storage optimization for easy understanding.

##### A. Resource Management

A lot of work has been done for storage load management of data stream processing systems. The proposed system uses the up-date queues for minimizing memory consumption and the impact of overload on QoS [37]. Another related research for resource allocation in the stream database includes the minimum spanning tree-based algorithms to discover and allocate the resources to meet real-time constraints [38]. One more connected work aims to maximize the quality of the results in stream processing systems using overload management concept. This concept of overload management related to distribution in stream processing system uses the resource allocation technique. That includes the distributed algorithms for reallocating the system resources (i.e., CPU) based on their utilization [39]. Another related contribution [40] handles the query optimization, scheduling, resource allocation and especially the source availability. The proposed system interacts between resource availability and the approximation.

##### B. Storage Optimization

Similarly, researchers investigated the storage optimization for data streams. One of the works includes a sluggish ladder queue that handles the long-running queries/continuous queries in real-time over a high-volume of data streams. Hence focuses on the latest data and control infinite streams overflow [41]. Some of the related work includes model integration of distributed stream processing system with state management. The model transitionally extends with sub-graphs, integrity constraints and consistency guarantees [42]. Some of the similar contributions include a high-performance stream processing engine I/O architecture which allowed the simultaneous persistence and communication of live and past (retrieved from storage) data streams [32]. Another related work [43] includes a system which constructs models and algorithms for overload prediction for heterogeneous data. The system scales up the performance and reduces the data loss without allocating additional servers.

Moreover, there has been a lot of works that deal with storage and is related to relational databases, data marts, data warehouses, and longer-term storage using Extract, Transform, Load (ETL) or Extract, Load, Transform (ELT) tools [15][21]. Another related work includes the development of an elastic cloud data storage system to support both OLTP and OLAP workloads efficiently within the same storage and processing system [44]. Enterprise Data Warehouse (EDW) traditional environment and its association with the data storage is another related work that needs a genuine consideration to enhance the big data storage [11][45]. Lastly, a review paper [46] has elaborated many data streams processing systems in depth that include their data models, continuous query processing, languages and query optimization.

#### V. PROPOSED ALGORITHM

This section describes the Enhanced Time Expired Algorithm based on a skew time (Hot Spots) by utilizing the

skew factor, generation time and expired time of the stream data.

#### A. Problem Definition

The problem of optimizing the archive storage in distributed stream databases is one of the big issues to tackle. A big concern in-stream archive storages is its limited storage space. One way to get rid of this problem is to increase the archive memory but that too has its limitations of cost and integration issues. Another way is to delete the expired data and make space for the new incoming streams. If there will be more space, more incoming data can be stored for future verification and tuning purposes. The storage space limitation has a direct impact on the storage and effects indirectly to historic data storage. Lots of techniques have been used to do the same work. Even a normal deletion technique would be an idle solution to free up the previous retention data and make space for the new incoming data based on normal retention policies as mentioned in [8]. However, these conventional solutions are not effective when it comes to the peak time (when most of the users are active and huge data is coming in). To our knowledge, none of the techniques had worked under the skew time. So a mechanism is needed which should automatically detect and delete the expired data and free the archive space for more incoming data for future historical data analysis under the skew time (Hot tuples-when most of the users and servers are on max utilization).

The problem of stream archive storage during peak hours needs a genuine optimized solution to automatically manage the storage space for huge incoming data by deleting the retention data and freeing up the space for huge incoming streams for historic analysis.

Thus, we define our algorithm to solve the problem of detecting and deleting the expired data based on the prescribed retention policy (Rp) which we keep 4 in our test environment (quarterly-4). A skew factor as  $Sf = \{\text{states}^*\}$ , which includes the peak time of skews. Its value can be +ve or -ve based on the server and user load. The  $Tg^*$  denotes a Time Generation of the stream and  $Et^*$  denotes the Expire Time based on the retention policy. When retention value is found, the data is deleted and if not, the data is retained, and no action is taken till loop checks and continues to the whole archived storage.

#### B. Enhanced Time Expired Algorithm (ETEA)

The Enhanced algorithm is derived from [27], a skew factor in addition to the generation time and the expired time of the streams is added. The skew factor represents the peak mode and the expired time represents the maximum allowable time for the stream to be available in an archive (based on the retention policy). The algorithm helps to recognize and remove the obsolete data (retention or expired) so as to free the space for more incoming data for archive storage in the future for historical data analysis. Therefore, it optimizes the archive storage. Based on positive and negative value criteria of time retention, the algorithm will discard expired data which is of no benefit and has occupied the archive space. Thus, freeing the archive's storage for more new incoming data.

The overview of an Enhanced Time Expired Algorithm (ETEA) is represented as:

#### Algorithm: Enhanced Time Expired Algorithm (ETEA)

```
1: is skew ← true // check for skew factor Sf
2: Get the “time generation” and generate “expire time” (Tg
   and Et)
3: Calculate the remaining Time Rt //based on retention policy
4: while skew do
5: Remove the retention stream //delete the stream based on
   retention policy
6: Update the value of skew
7: if not skew then
8: return false//no deletion done
9: end if
10: end while
11: return true//the system is still checking
```

#### C. Our Contribution

The main contribution of this paper is that we propose an enhanced algorithm, called Enhanced Time Expired Algorithm (ETEA), to detect and delete the expire data automatically in a stream archive storage for distributed setup. Our algorithm is efficient and effective in detecting and deleting the retention data and free the space for new storage for future verification and tuning purposes.

During a stream processing in a distributed environment (distributed set of processing nodes) commonly known as Distributed Stream Processing Systems (DSPS's) for example OLTP, when a skew factor (Sf) is found, which represents the peak time for incoming data from different types of online servers and users. The proposed algorithm gets the Time Generation and Expire Time (Tg and Et) for each stream using a function: Get Time Generation (Tg) and Get Expire Time (Et). The  $Tg = \text{initial time of the stream}$  and  $Et = \text{Expired time or retention policy time}$  (it might be monthly or quarterly)-in our case we take it as quarterly for a test environment. When Rt Expired is found, the algorithm destroys the data stream in the archive using the following function:  $\text{DELETE} = \text{Stream}$ . If no Rt Expired match is found, no action is taken, then a second stream in the loop (Lp) is checked for the Rt Expired values onwards.

Our main contribution is our proposed algorithm for storage optimization (memory and cost efficient) and works efficiently during the skew time to find (detect) the expired data and delete the expired data based on the retention policy to release the memory for the new incoming stream data.

## VI. ALGORITHM DETAILS

In this section, we explain the working of our proposed algorithm. As we know that many OLTP workloads are heavily skewed to “hot” tuples or ranges of the tuple. So, first, we try to identify the skew mode using the lightweight threshold mechanism based on several streams (transactions or stored procedures). For our experiments, we have assumed that if the maximum no of streams is 1000 per minute, we consider this as the maximum threshold. Once this threshold is achieved, a skew mode is triggered with a value representation. These values can either be 1 or 0 based on a condition whether the skew mode is true or false (+ve or -ve). If the value is found true, a skew factor is assigned with a timesheet which means

more space is needed for the incoming data flow of streams.  $Sf = \{states^*\}=1$  or  $0$ . The initial values for Time Generation decide the Expire Time for the streams based on the retention policy of four (4) months in our experiments, which means  $4*30=120+2=122$  days. The data is deleted automatically once this value is found on matching. The whole process is repeated and cross-checked for any missing values until it checks the whole archive storage.

The Algorithm working is explained through below-mentioned phases:

- Phase 1: The working of the algorithm is based on the concept of checking for the skew factor (+ve/1 or -ve/0).
- Phase 2: Assigning the Time Generation and Expire Time for the streams. When the expired data is found, based on the retention policy, the long-stayed data is deleted automatically.
- Phase 3: The data is checked again for all streams if no match values of Expiration Time is found, the checking proceeds until it checks the whole archive storage.

The logic for (Sf) phase1:  $Sf = \{states^*\}$ , which includes the peak time of skews. Its value can be +ve or -ve based on the server and user load. When the skew factor (Sf) is +ve, which represents the peak time during the online transactions of the distributed environment, the initial values are generated and set as 1 for +ve value and vice versa for 0 for -ve value.

The logic for (Tg and Et) phase2: The  $Tg^*$  gets the Time Generation of the streams i.e. their initial time. Therefore based on it,  $Et^*$  gets the Expire Time from the retention value (which is based on its policy). We are using Expire time (4 months in case of our experiments). Hence in our case it is calculated as  $4*30=120+2=122$  days. When this value (122) is found, the data is deleted automatically and if not, the data is retained, and no action is taken. Therefore, data of more than 122 days is deleted automatically.

The logic for (Lp) phase3: The data is checked again for all streams using a loop function, if no match values of Expiration Time is found, no action is taken, and the checking proceeds till it checks the whole archive storage is checked.

## VII. PERFORMANCE EVALUATION

In this section, we provide an experimental evaluation of the stream archive storage techniques using our proposed algorithm. The main goals of our experimental study are as follows:

- to show working of our Enhanced Time Expired Algorithm (ETEA) for stream archived storage in a distributed environment. The algorithm uses a skew factor based on hot spots and is compared against the state-of-the-art retention policy like conventional approaches First in First Out (FIFO), in terms of storage optimization (memory and cost efficient) for stream archive storage.
- to evaluate how effectively we can manage the memory space during the skew time.

We first investigate the impact of skews of incoming data rates. We found that there is a need for more archive storage for the stream data during peak time. Second, we consider the need for data deletion for skew. If more data will be deleted automatically, more space will be free for the future tuning purpose of historic data in archive stream storage. Thirdly, we detect the data automatically using our retention policy in the proposed algorithm. Fourthly, the expired data was deleted, and more space was available free. We can see from Fig. 1 and Fig. 2 that our algorithm had increased the effectiveness as well as memory optimization for stream archive storage. We observe all differences in performance (in terms of skew time and normal incoming data). Nevertheless, the conclusion is that we start with skew formulation, observe the impact of skews, provide the retention policy, detect the expired data to be deleted and lastly delete the expired data. Finally, we compare our proposed algorithm with one of the conventional approaches First in First Out (FIFO).

In Table I, we can see Time results for data deletion, which keeps increasing with the passage of time and Table II represents the values for Space (memory) results for data deletion).

Within a time of 3ms, the deletion was done with a value of 1 tuple deleted with an efficiency of 30.3 percent. The efficiency keeps increasing and decreasing with the time slots until it verifies the whole deletes of memory chunk (memory block).

In Table II, with the first deletion of 1 tuple, a space of 0.3 bits was released i.e. space(memory). The value of efficiency goes on increasing depending upon the memory released. The rest values of the table represent the respective space released with the efficiency.

In Fig. 1, we can see that in less time, the efficiency has increased, while as in Fig. 2, during the same time, more expired data is deleted, hence releasing the memory(space). In both cases (Fig. 1 and Fig. 2), which are related to efficiency improvement and memory improvement, the frequency is common. The Node values represent the nodes numbers, the time is represented in milliseconds (ms) and the deletion is measured in percentages in aggregate.

TABLE I. BLOCK WISE TIME RESULTS FOR DATA DELETION

		Time	Deletion	Efficiency
Node Value	1.00	3ms	1.00	30.3
	2.00	6ms	2.00	60.3
	4.00	1ms	0.3	10.1
	6.00	9ms	3.00	98.0

TABLE II. BLOCK WISE SPACE (MEMORY) RESULTS FOR DATA DELETION

		Deletion	Space (Memory)	Efficiency
Node Value	1.00	1.00	0.3	30.3
	2.00	2.00	0.6	60.3
	4.00	0.3	0.1	10.1
	6.00	3.00	6.0	98.0

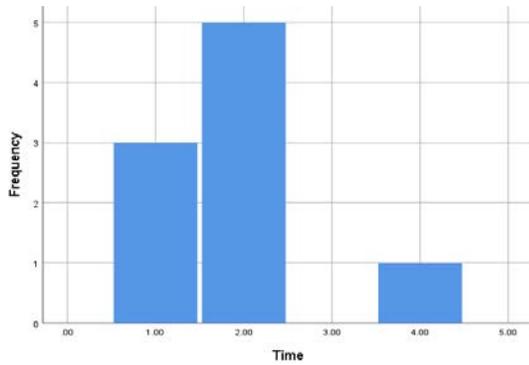


Fig. 1. Efficiency vs Time.

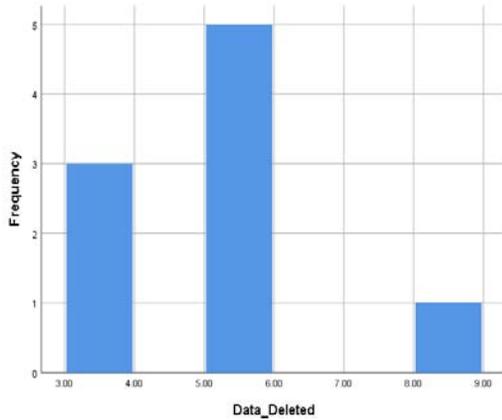


Fig. 2. Efficiency vs Memory (Data Deleted).

Next, we will first describe our experimental setup, then we will present our results.

#### A. Experimental Setup

We implemented our ETEA algorithm as part of the H-store [47], mostly used for the OLTP workload. We monitor the flow of skews using its storage manager component and the memory usage was taken into consideration. We configured the monitoring component of H-store processing system to check the skew mode (hot tuple) and monitor the flow of incoming data into the archive storage.

In all our experiments, we have used a double-node set up for running on a window on an Intel Quad-Core Intel Xeon 3360 2.8GHz processor and 8GB of memory. To understand the impact of skew on an OLTP DBMS, we conducted a basic benchmark using the YCSB workload [48] on four (4) node H-Store clusters.

For this setup, we used a database with 2 million tuples (Each 1KB in size (~2GB in total) that are deployed on two (2) partitions.

1) *Workload*: In all experiments, we have used synthetically generated tuples (Streams). The input rates were set according to the desired level of threshold to be exerted on the system. The input is ordered by a time-based factor that decides the initial time of the streams and accordingly using the initial time. The expired time is calculated based on the retention value. The number of values can be altered and can

range from 122 to any value depending on the experiment. The actual initial value of the streams does not have any significance, it is only what we measure in the experiments. To be able to control the processing cost of this query, we use the retention value to free the space which directly affects the cost of maintaining the more space unnecessarily for the expired data.

2) *Performance Metrics*: We have primarily used two performance metrics in our experiments:

- **Effectiveness during Skews**: Average effectiveness is computed on streams based on the no of tuples deleted and then, an overall average is computed across all the skews, which is measured in seconds.
- **Storage (Memory optimization)**: Maximum memory deleted for the stream archive is recorded between output deliveries. Then we compute a maximum overall memory calculation based on deletion across a given run. Memory deleted is measured in the number of deleted tuples.

#### B. Results

In Fig. 1 and in Fig. 2 we can see that the effectiveness, as well as the memory optimization for the stream archive storage during the skew time. The effectiveness is increased due to a continuous deletion of the expired data. The situation looks improved for all possible scenarios. As expected, during the skew time, the more expired data is removed, the more archive storage memory is released. The reason for this is the response to streams (keep it or delete it) method of our proposed algorithm, which avoids the need for redundancy in rechecking the memory chunk again. This result clearly shows that deleting expired data increase the archive storage for the new incoming streams. Our approach is 98% efficient and fast than other conventional techniques like First in First Out FIFO (which cannot be used for skew workloads). We have benchmarked our algorithm against FIFO in a skew environment. We had concluded with the results which prove our proposed algorithm is well suited for the skews with 98% efficiency and memory optimization. The details of the benchmark are mentioned.

Table III represents the detailed information about the ETEA. We observe the efficiency of ETEA (which works under the skews). We found that in less time, more expired data is deleted, and 98% efficiency is achieved which was not a case with the FIFO showing in Table IV. Fig. 3 and Fig. 4 presents the graphical representation of the efficiencies between the two algorithms.

TABLE III. EFFICIENCY UNDER SKEWS

Enhanced Time Expired Algorithm (ETEA)					
		Time	Deletion	Space (Memory)	Efficiency
Node Value	3.00	2	28.6	28.6	98.0
	5.00	6	71.4	71.4	96.0
	6.00	9	100.0	100.0	98.0

TABLE IV. EFFICIENCY WITHOUT SKEWS

First in First out (FIFO)					
		Time	Deletion	Space (Memory)	Efficiency
Node Value	3.00	2	18.0	28.6	43.6
	5.00	6	71.4	71.4	54.0
	6.00	9	100.0	100.0	54.0

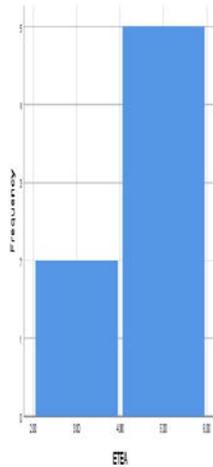


Fig. 3. Efficiency of ETEA.

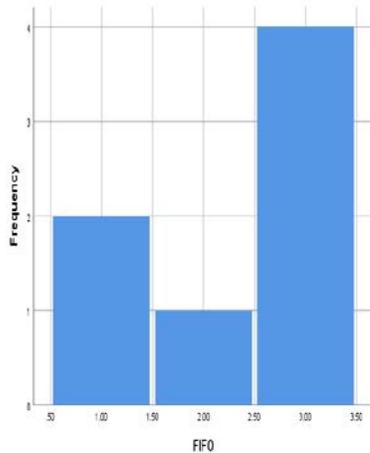


Fig. 4. Efficiency of FIFO.

### VIII. CONCLUSION AND FUTURE WORK

In this paper, we presented an algorithm, Enhanced Time Expired Algorithm (ETEA) for the distributed stream databases, which will help to optimize the stream archive storage and will detect and delete the expired data. Hence it provides the free space for new incoming streams for the purpose of historical data analysis during the skew time (hot tuples) in a distributed platform. Extensive evaluation on four (4) node cluster demonstrates the superiority of the approach compared to normal prior efforts which doesn't include the skew time (conventional retention techniques like First in First Out (FIFO)). Hence, our solution is fully optimized based on effectiveness of memory and cost efficient factors of storage.

In the future, we will try to apply the other retention policies by using a machine learning algorithm (time series) to enhance the effectiveness and the performance of the stream archive storage. In our next series of optimizing the stream archive storage, we plan to design an architecture that will integrate the live data with historic data. Furthermore, we can also try the Pinwheel scheduling algorithm and Reduce-Max algorithm for the best results in our future experiments.

### REFERENCES

- [1] Waseem, Q., Maarof, M. A., Idris, M. Y., & Nazir, A. A Taxonomy and Survey of Data Partitioning Algorithms for Big Data Distributed Systems. *Micro-Electronics and Telecommunication Engineering*, 447: (2020).
- [2] Kaur, P., and Awal Adesh Monga. "Managing Big Data: A Step towards Huge Data Security." *International Journal of Wireless and Microwave Technologies* 2 (2016).
- [3] Emani, C. K., Cullot, N., & Nicolle, C. (2015). Understandable big data: a survey. *Computer science review*, 17, 70-81.
- [4] Hu, Han, Yonggang Wen, Tat-Seng Chua, and Xuelong Li. "Toward scalable systems for big data analytics: A technology tutorial." *IEEE access* 2 (2014): 652-687.
- [5] M. Serafini, E. Mansour, A. Aboulnaga, K. Salem, T. Rafiq, and U. F. Minhas. *Accordion: Elastic Scalability for Database Systems Supporting Distributed Transactions*. *Proc. VLDB Endow.*, 7(12):1035–1046, Aug. 2014.
- [6] Moniruzzaman, A. B. M. "Newsq: Towards next-generation scalable rdbms for online transaction processing (oltp) for big data management." *arXiv preprint arXiv:1411.7343* (2014).
- [7] Kune, Raghavendra, et al. "The anatomy of big data computing." *Software: Practice and Experience* 46.1 (2016): 79-105.
- [8] Douglas, Fred, et al. "Position: short object lifetimes require a delete-optimized storage system." *Proceedings of the 11th workshop on ACM SIGOPS European workshop*. 2004.
- [9] Zhang, Hao, et al. "In-memory big data management and processing: A survey." *IEEE Transactions on Knowledge and Data Engineering* 27.7 (2015): 1920-1948.
- [10] Bakshi, K., 2012. *Considerations for big data: Architecture and approach conference*. s.l., IEEE, pp. (1-7).
- [11] Hartmann, T., Fouquet, F., Moawad, A., Rouvoy, R. and Le Traon, Y., 2019. *GreyCat: Efficient what-if analytics for data in motion at scale*. *Information Systems journal*.
- [12] Hu, H., Wen, Y., Chua, T., & Li, X. (2014). *Toward Scalable Systems for Big Data Analytics: A Technology Tutorial*. *IEEE Access*, 2, 652-687.
- [13] Singh, Dilpreet, and Chandan K. Reddy. "A survey on platforms for big data analytics." *Journal of big data* 2, no. 1 (2015).
- [14] R. Taft. *Elastic Database Systems*. PhD thesis, MIT, Cambridge, 2017.
- [15] Aubrey L. Tatarowicz, Carlo Curino, Evan P. C. Jones, and Sam Madden. *Lookup tables: Fine-grained partitioning for distributed databases*. In *ICDE*. IEEE Computer Society, 2012.
- [16] Phansalkar, Shraddha, and Swati Ahirrao. "Survey of data partitioning algorithms for big data stores." In *2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp. 163-168. IEEE, 2016.
- [17] Huang, Qun, and Patrick PC Lee. "Toward high-performance distributed stream processing via approximate fault tolerance." *Proceedings of the VLDB Endowment* 10.3 (2016): 73-84.
- [18] Fu, Xinwei, et al. "Edgewise: a better stream processing engine for the edge." *2019 {USENIX} Annual Technical Conference ({USENIX}{ATC} 19)*. 2019.
- [19] Zhao, Xinwei, et al. "A taxonomy and survey of stream processing systems." *Software Architecture for Big Data and the Cloud*. Morgan Kaufmann, 2017. 183-206.
- [20] Alshamrani, Sultan, et al. "High availability of data using Automatic Selection Algorithm (ASA) in distributed stream processing

- systems." *Bulletin of Electrical Engineering and Informatics* 8.2 (2019): 690-698.
- [21] Elgendy, N. and Elragal, A., 2014. Big data analytics: a literature review paper. s.l., Springer, cham, pp. 214-227.
- [22] Zhong, R.Y., Newman, S.T., Huang, G.Q. and Lan, S., 2016. Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Computers & Industrial Engineering journal*, pp. 572-591.
- [23] Lv, Zhihan, et al. "Next-generation big data analytics: State of the art, challenges, and future research topics." *IEEE Transactions on Industrial Informatics* 13.4 (2017): 1891-1899.
- [24] Siddiqa, Aisha, Ahmad Karim, and Abdullah Gani. "Big data storage technologies: a survey." *Frontiers of Information Technology & Electronic Engineering* 18.8 (2017): 1040-1070.
- [25] Kamburugamuve, Supun, et al. "Survey of distributed stream processing for large stream sources." *Grids Ucs Indiana Edu* 2 (2013): 1-16.
- [26] Isah, Haruna, et al. "A Survey of Distributed Data Stream Processing Frameworks." *IEEE Access* 7 (2019): 154300-154316.
- [27] Abdullah, Mohammed B., and YazenS Sheet. "Refine Priority Queuing Scheduling Algorithm By Applying Time Expired Packets Algorithm." *AL Rafdain Engineering Journal* 20.2 (2012): 150-163.
- [28] Kirsten Hildrum, Fred Douglass, Joel L. Wolf, Philip Yu, Lisa Fleischer, and Akshay Katta, "Storage Optimization for Large-Scale Distributed Stream Processing Systems," *ACM Transactions on Storage (TOS)* TOS Homepage archive Volume 3 Issue 4, February 2008.
- [29] Nesime Tatbul, Stan Zdonik, John Meehan, Cansu Aslantas, Michael Stonebraker, Kristin Tufte, Chris Giossi, Hong Quach, "Handling Shared, Mutable State in Stream Processing with Correctness Guarantees," *IEEE Data Engg. Bull.*, 2015.
- [30] Antoniu, Gabriel, et al. "The Sigma Data Processing Architecture: Leveraging Future Data for Extreme-Scale Data Analytics to Enable High-Precision Decisions." (2018).
- [31] Stonebraker, Michael, Uğur Çetintemel, and Stan Zdonik. "The 8 requirements of real-time stream processing." *ACM Sigmod Record* 34.4 (2005): 42-47.
- [32] Sebeou, Zoe, and Kostas Magoutis. "Scalable storage support for data stream processing." *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*. IEEE, 2010.
- [33] Tatbul, Nesime, et al. "Handling Shared, Mutable State in Stream Processing with Correctness Guarantees." *IEEE Data Eng. Bull.* 38.4 (2015): 94-104.
- [34] Irina Botan, Gustavo Alonso, Peter M. Fischer, Donald Kossmann, and Nesime Tatbul, "Flexible and Scalable Storage Management for Data-intensive Stream Processing," *ACM Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, Pages 934-945.
- [35] Yi, X., Liu, F., Liu, J. and Jin, H., 2014. Building a network highway for big data: architecture and challenges. *IEEE Network journal*, Volume 28, pp. 5-13.
- [36] H. Eszter, 2015. Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science journal*, Volume 659, pp. 63-76
- [37] Moga, Alexandru. "UpStream: Storage-centric Load Management for Data Stream Processing Systems." *Proceedings of the VLDB 2010 PhD Workshop*. VLDB, 2010.
- [38] L. Chen and G. Agrawal, "Resource Allocation in a Middleware for Streaming Data," in *Proceedings of the 2Nd Workshop on Middle ware for Grid Computing*. NY, USA: ACM, 2004.
- [39] A. Tang, Z. Liu, C. Xia, and L. Zhang, *Distributed Resource Allocation for Stream Data Processing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 91–100].
- [40] Motwani, Rajeev, et al. "Query processing, resource management, and approximation in a data stream management system." *CIDR 2003*. Stanford InfoLab, 2002.
- [41] Tang, Xiang-Hong, et al. "Storage Optimization for Query Processing over Data Streams." *Journal of Chongqing University* 9.2 (2010): 79-92.
- [42] Affetti, Lorenzo, Alessandro Margara, and Gianpaolo Cugola. "FlowDB: Integrating stream processing and consistent state management." *Proceedings of the 11th ACM International Conference on Distributed and Event-based Systems*. 2017.
- [43] Nguyen, Nhan, et al. "Arion: A Model-Driven Middleware for Minimizing Data Loss in Stream Data Storage." *2017 IEEE 10th International Conference on Cloud Computing (CLOUD)*. IEEE, 2017.
- [44] Cao, Yu, et al. "Es 2: A cloud data storage system for supporting both oltp and olap." *2011 IEEE 27th International Conference on Data Engineering*. IEEE, 2011.
- [45] Al-Shiakhli, Sarah. "Big Data Analytics: A Literature Review Perspective." (2019).
- [46] Golab, Lukasz, and M. Tamer Özsu. "Issues in data stream management." *ACM Sigmod Record* 32.2 (2003): 5-14.
- [47] Kallman, Robert, et al. "H-store: a high-performance, distributed main memory transaction processing system." *Proceedings of the VLDB Endowment* 1.2 (2008): 1496-1499.
- [48] Cooper, Brian F., et al. "Benchmarking cloud serving systems with YCSB." *Proceedings of the 1st ACM symposium on Cloud computing*. 2010.

# Comparative Analysis of Methodologies for Domain Ontology Development: A Systematic Review

Abdul Sattar<sup>1</sup>, Ely Salwana Mat Surin<sup>2</sup>  
Mohammad Nazir Ahmad<sup>3</sup>  
Institute of IR4.0  
Universiti Kebangsaan Malaysia 43600  
Bangi, Selangor, Malaysia<sup>1,2,3</sup>  
Department of Computer Science  
KFUEIT 64200, Rahim Yar Khan, Pakistan<sup>1</sup>

Mazida Ahmad<sup>4</sup>  
School of Computing, Universiti Utara Malaysia (UUM)  
Sintok, Kedah, Malaysia

Ahmad Kamil Mahmood<sup>5</sup>  
High Performance Cloud Computing Centre, Department of  
Computer and Information Sciences, Universiti Teknologi  
Petronas, 32610, Bandar Seri Iskandar, Perak

**Abstract**—Interlocking Institutional Worlds (IWs) is a concept explaining the need to interoperate between institutions (or players), to solve problems of common interest in a given domain. Managing knowledge in the IWs domain is complex, but promoting knowledge sharing based on standards and common terms agreeable to all players is essential and is something that must be established. In this sense, ontologies, as a conceptual tool and a key component of knowledge-based systems, have been used by organizations for effective knowledge management of the domain of discourse. There are many methodologies that have been proposed by several researchers during the last decade. However, designing a domain ontology for IWs needs a well-defined ontology development methodology. Therefore, in this article, a survey has been conducted to compare ontology development methodologies between 2015 and 2020. The purpose of this survey is to identify limitations and benefits of previously developed ontology development methodologies. The criteria for the comparison of methodologies has been derived from evolving trends in literature. Our findings give some guidelines that help to define a suitable methodology for designing any domain ontology under the domain of interlocking institutional worlds.

**Keywords**—Knowledge management; interlocking institutional worlds; domain ontology; comparison of methodologies; ontology development

## I. INTRODUCTION

A social reality has elements, and each element is called an institutional fact. For example, in a context A, a physical reality B will be titled as an institutional fact C. The name, marital status and citizenship of a person are examples of institutional facts. Speech act are formal statements created by officials of organizations. These speech acts create and destroy institutional facts. The nomination is a speech act performed by parents who register the birth of a child in a government register. Laws and regulations are created by acts of different branches of government. For example, two institutional facts, wife and husband, are created by the marriage speech act. This speech act destroys two institutional facts such as an unmarried woman and a single man. Speech acts can also be used in creating institutions. For example, any positive physical posture as a speech act may or may not be positive in another area (jurisdiction). Institutional reality may be immaterial, but

it is still real. The financial crisis of the subprime mortgages of 2008 had at least an impact on the world as the undeniably physical tsunami of 2004.

Furthermore, it often happens that two or more parties participate in a speech act, for e.g., a marriage involves two persons, a husband and a wife. Olympic Games as a speech act, requires participation of several institutional facts. International Olympic Committee (IOC) as a supervisory body, which creates committees (national Olympic committees –NOCs) at national level. And each NOC manages Olympic teams and venues as directed by the IOC. An organizing committee (OCOG) for particular Olympic Games is created for a selected area (city) with one or multiple venues. Independent national sports federations follow rules spoken by international sports federations. The IOC provides policies for athletes, NOCs and OCOGs. Finally, it is necessary for sports federations to define sport rules that allow awarding of medals (bronze, silver and gold) for all Olympic Games. This is problematic for tournament-oriented sports like tennis, which usually have a winner and a loser in the final game. Therefore, these sports generally take place even if there is no third place in their games. The bronze medal, in Olympic events, is an element of fixed reality and this may not be true in other institutions.

In this context, if we define the institutional world as the set of types of speech acts and instances that a given institution is authorized to carry out, we can say that the institutional worlds interlock if they are connected in the manner indicated above. This conceptualization that includes both endurants and perdurants, helps in interoperability of information systems.

Semantic web, as a machine-readable web, needs ontologies as its primary and most important component. Ontology describes conceptions and their associations in the domain of discourse [1]. Knowledge based applications require ontologies and they serve as formal models and machine understandable description of a domain. Ontologies help in sharing domain knowledge to other relevant or irrelevant domains. For example, [2] test how sharing of knowledge can improve employee individual performance as well as team performance. Due to the distributed nature of organizational knowledge, the Knowledge based applications, with the help of

ontologies, must be able to integrate knowledge of heterogeneous sources and present an overview of the knowledge available in the organization [3]. In this context, finding a suitable ontology for a domain is one of the bigger research challenges [4]. Ontologies, as a conceptual tool and key component of knowledge-based systems, have been used by organizations, for effective knowledge management of the domain of discourse. In the past decade, investigation and development in ontology has had a large stimulus, the industry has shown an interest in developing novel applications in semantic technology which have resulted in the wide adoption of ontology-based solutions by government, academia and commercial industry. In this context, ontology-based solutions with improved knowledge management help in better decision making. Furthermore, ontology approach makes it easy to share conceptualization of a domain [1], and this sharing offers more opportunities for stakeholders to solve their real-time problems.

A single medium-sized organization can have more than 1000 information systems. Most organizations interact with hundreds of others in the supply chain and other relationships. For example, the UK National Plan for National IT Health Services mentioned in the introduction, attempts to integrate hundreds of thousands of information systems. Our discussion above on interlocking institutional worlds shows that almost all applications involve multiple institutional worlds belonging to different institutions that interlock in complex ways. For now, the construction of a domain ontology for such a large and complex IWs domain (e.g. waste management, semantic Web, Olympic Games, postal codes, preparation of maps and governance of geographic information and flood management) is also difficult, laborious and time consuming.

The analysis of developed ontologies, and methodologies for developing ontologies, help us gain an understanding of ontology engineering. For example, a survey conducted in [5] analyzed ontology related activities of the schema.org community. However, this survey did not discuss the impact of methodologies on ontology development and did not compare their ontology engineering efforts with others. It should be mentioned here that our survey has comparison criteria different than [5]. We compare methodologies based on domain analysis, conceptualization, level of detail, collaborative construction, implementation, evaluation, instantiation, maintenance, documentation, ontology localization, support for reusability, support for integration, support for interoperability, resource estimation human, example of application/project and methodology rooted in established methodologies. Author in [5] included some parameters with respect to the above-mentioned criteria. Therefore, the design of a domain ontology for the IWs domain requires a well-defined ontology development methodology. In this context, we write this paper to survey and compare with criteria, which we think are important, for proposing a well-established methodology for designing domain ontology for the IWs domain. This survey will help us find missing components in developed methodologies for ontology development.

With this in mind, the purpose of this research is to propose a methodology for the development of domain ontology for the IWs domain. Consider a question that arises from the emphasis

on developing ontology for IWs domains: How ontology development methodology can be proposed to design a domain ontology of IWs?

This section deals with an overview of the research, followed by Section 2 which outlines the literature of this study. Section 3 presents the review method, Section 4 is about discussions and lastly, Section 5 is the conclusion of this article.

## II. LITERATURE REVIEW

### A. Interlocking Institutional Worlds (IWs)

Interlocking Institutional Worlds (IWs) is about inter-agency collaboration involving multidisciplinary teams, which is an increasingly popular strategy for innovation, as it offers broader ideas and views. IWs are defined as a set of organizations working together to solve a common problem at the domain level [6, 7 and 8]. The theoretical foundations of the concepts of IWs are strongly anchored in Institutional Facts and Speech Act theories presented by [9]. Examples of IWs domains include waste management, the semantic web, the Olympic Games, postal codes, Maps preparation and Geographic information governance [6 and 10], and flood management [8]. In IWs, when organizations start to collaborate, standard and common terminology must be established to allow for smooth communication to avoid misunderstandings.

The integration or merging of ontologies focuses on explicit specifications. Furthermore, interaction or collaboration of two or more specifications is important. For supporting merging or interoperability, conceptualizations of ontologies should have systems of institutional facts. In this context, interlocking systems of institutional facts help in merging conceptualizations of ontologies. In short, an institutional world is defined as a set of institutional facts in the speech acts of an institution. Furthermore, without interlocking of institutional worlds, it is not feasible to merge specifications of two or more ontologies [6].

The methodology for development of an ontology is a guideline for ontology developers [11]. Different ontology development methodologies have been defined [12, 13, 41, 15, 16, 17, 18, 19, 20 and 21]. The methodology proposed in our previous work [13] was our initial attempt to design a methodology for designing ontology for the waste management domain. Researchers to date, have not agreed on a single methodology for developing ontologies of different domains. Therefore, ontology projects choose or develop their own methodology for the developing ontology of a domain of discourse. This selection or creation of methodology depends on application and the anticipated evolution of the particular ontology [22]. Therefore, the designing of a domain ontology for WM, a type of IWs, requires a well-defined ontology development methodology.

### B. Need for Ontology

In the last decade, ontologies have become widely adopted in a variety of fields ranging from biomedicine, to finance, engineering, law, and cultural heritage as recent review on ontology engineering in [4]. Semantic web, as a machine-

readable web, need ontologies as its primary and most important component. Ontology defines conceptions and their associations in a domain of discourse [1]. Knowledge based applications require ontologies as they serve as formal models and machine understandable descriptions of the domain. Ontologies help in sharing domain knowledge to other relevant or irrelevant domains. For example, [2] tests how sharing of knowledge can improve employee individual performance as well as team performance. Due to the distributed nature of organizational knowledge, the knowledge-based applications with the help of ontologies, must be able to integrate knowledge of heterogeneous sources and present an overview of the knowledge available in the organization [3]. In this context, finding a suitable ontology for a domain is one of the bigger research challenges [4]. Ontologies, as a conceptual tool and key component of knowledge-based systems, have been used by organizations for effective knowledge management of the domain of discourse. In the past decade, the industry has shown interest in providing solutions by developing applications in semantic technology. For example, the HIV Protein Ontology in the medical domain [23]. Artificial intelligence (AI) as an opportunistic area of study, and a challenge for the research community, shows interest in knowledge representation, knowledge management and semantic associations [24]. Truly speaking, keeping in view the advantages of ontology-based approach, there are few ontology-based solutions which benefit from this approach [25]. Here it is worth mentioning that ontology based machine-readable systems help in making better decision support systems with improved knowledge management. Furthermore, ontology approach makes it easy to share conceptualization of a domain [1], and this sharing offers more opportunities for stakeholders to solve their real-time problems.

### C. Ontology Development Methodologies

The main purpose of an ontology proposal was to share domain information between software agents and people. As a result of such interest in ontologies, several ontologies have been designed, each ontology with a different purpose, in different disciplines, such as Gene Ontology (GO) in the biomedicine domain [15].

The study of principles, methods and tools for designing upper or domain ontologies, is a primary focus of the ontology engineering discipline. In this context, a methodology provides guidelines for the development of ontologies. In order to help and support ontology development, several methodologies have been proposed by researchers.

Ontologies help in communication for better decision making, promote sharing of knowledge, facilitate storage of information, and support the reuse of knowledge [3]. An ontology can be developed manually using an ontology editor like Protégé, or automatically using suitable ontology designing algorithms that are implemented in some programming language such as Java. Many researchers design ontologies manually, such as [13], or automatically, for example [26], for particular domains of study.

For designing a domain or upper ontology, it is necessary to follow a set of defined and ordered steps. In order to help and support ontology development, several methodologies have

been proposed by researchers. However, there is a need for a well-documented, mature and widely accepted methodology for ontology engineering. In this article, a survey has been prepared to compare methodologies designed during 2015-2020 for developing domain ontologies in different domains. A methodology is a set of well-designed techniques and methods that guarantee the quality of the results of an ontology design process. Author in [27] describes a series of related concepts related to methodologies for designing ontology.

- Method: The order or a series of steps to develop a product.
- Technique: A procedure for achieving a goal. Therefore, the methodology provides a framework for building ontology for the domain of knowledge.
- Methodology: A set of methods and techniques that guarantee the quality of the results of an ontology design process.

Knowledge Engineering: it is the discipline derived from artificial intelligence and is responsible for the design and development of knowledge-based systems.

As stated by [4], ontology engineering has not changed significantly over the last decade. It is observed by [28] that each methodology follows a different approach. There is no correspondence between the different methodologies. There is no technological support for most methodologies, i.e., they cannot be easily applied in the task of building an ontology.

In order to help and support ontology development, several methodologies have been proposed by researchers [4]. But work on new methodologies for developing ontologies does not seem to have progressed greatly. As a result, a methodology proposed by [29] is the most used or cited methodology for designing a domain ontology. There may be several reasons for this, but one of the main reasons is poor or even lack of documentation about methodologies that have been applied to develop the ontology of a project. Several researchers have made efforts to compare methodologies for designing domain ontologies.

An overview of the developmental aspects of ontology was presented in [30]. According to this survey, there are three methodologies to use, to help developers build a domain ontology. The common methodology described in [31] provides a guideline for each stage of the development of an ontology. In comparison with METHONTOLOGY, this method can identify management activity and would be an effective and applicable approach for building domain knowledge models. Finally, similar to the two methodologies, the methodology proposed by [32] was created based on experience of the TOVE project, which was developed in six phases, beginning with the reasons for the capture, and ending with an established condition. To conclude, ontology designers can design ontologies in their own ways or follow a predefined methodology. Unlike any other development project, ontology is dynamic and flexible, and is not tied to a certain development or progress process to be created.

In [33], a survey is conducted on several aspects of ontology development such as ontology, methodologies for

developing ontologies and automated or semi-automated tools for designing ontologies. However, the survey does not present comparisons of researchers' contributions in terms of proposed or adopted techniques and methods for designing domain ontologies.

Similarly, domains have terminologies, and a well-designed methodology helps in designing ontology for domain terminologies as observed by [34]. Subsequently, based on five criteria, they made comparisons of ontology development methodologies. In this comparison, a conclusion is made that TERMINAE and METHONTOLOGY are the appropriate methodologies for designing domain ontologies. In this context, [35] argue that developing ontology development methodologies is a tedious and complex task. Existing studies show that a methodology developed based on experience of one or a few projects, may or may not be effective for creating domain ontologies. It means, there is a gap in the existing methodologies. The following points are concluded from [34]. First of all, according to established comparison criteria, none of the methodologies is mature enough. Secondly, there is an absence of documentation or poor documentation on techniques, activities and methods used in ontology construction. There are some exceptions, in particular, METHONTOLOGY. Additionally, some methodologies support reengineering concepts and reusability, with few methodologies providing recommendations for these aspects. Similarly, little attention has been paid by researchers to the collaborative construction aspect of domain ontology. In this context, [36], paid attention to collaborative way of ontology construction. As conventional strategies have been used by researchers for collecting ontology conceptions, it is important to propose and use novel techniques and methods for designing domain ontologies. One of the critical problems for the implementation of any ontology that should be addressed, is the problem of selecting a well-established and appropriate methodology for ontology construction. Here, quality of implementation of domain ontology is related to or dependent on selection and adoption of a well-established methodology. Therefore, the study by [37] examined and made a comparison of the best-known ontology development methodologies based on the development life cycle of a common ontology. This study [34] also aimed to define and use, the appropriate and mature methodology for the construction of Semantic Conflicts Detection Ontology (SCDO). Additionally, it also aimed at providing ontology developers with useful information that would facilitate the process of choosing the appropriate methodology for their ontologies. Hence, METHONTOLOGY was chosen and recommended by [34] to be used for the implementation of SCDO.

In [38], an investigation based on specific criteria was conducted on collaborative ontology engineering methodologies and the ontology engineering environment. This study recommends collaborative ontology construction, and collaborative construction needs integrated environments for ontology development. This survey makes it a priority to analyze experience reports and case studies. This analysis could be fruitful for defining future research roadmaps for collaborative ontology development. Finally, this study

recommends the use of an integrated development environment for collaborative ontology construction.

Author in [39] conducted a survey on ontology development methodologies for the public administration domain. In this survey, common ontology development aspects were chosen for comparing methodologies (for example, ontology construction strategies) as well as aspects related to project management (for example, the recommended process model or taking into account the collaborative development of ontologies). Author in [39] summarizes that none of the methodologies is suitable for the development of ontology in the public administration domain.

According to the analysis carried out by [40], the following conclusions were drawn: According to this survey, no methodology meets IEEE standards; although METHONTOLOGY [41] is the most mature. This study found this methodology [32] more formal than others but this methodology was found to be missing re-engineering aspects of ontology. Similarly, the methodology proposed by [31] also missed re-engineering aspects of ontology and omitted well-documented documentation. Similarly, the methodology proposed by [42], omitted detailed documentation, re-engineering of ontology and had no ontology development life cycle. Furthermore, CO4 and (KA) 2 [43] discuss collaborative ontology construction but with few details. This study concludes following:

- 1) None of the methodologies is mature enough because each research develops the ontology in its own way.
- 2) Another important aspect of ontology development is interoperability and its implementation. Domain ontologies developed using the SENSUS [42] methodology have the same high-level ontology structure and can be easily interoperated.
- 3) Most of the surveyed methodologies provide few details except METHONTOLOGY.

It should be mentioned here that our survey has comparison criteria different than [40]. We compare methodologies based on domain analysis, conceptualization, level of detail, collaborative construction, implementation, evaluation, instantiation, maintenance, documentation, ontology localization, support for reusability, support for integration, support for interoperability, resource estimation human, example of application / project and methodology rooted in established methodologies. While [40] included some parameters with respect to the above-mentioned criteria.

### III. REVIEW METHOD

A systematic literature review is a method of finding, assessing and inferring available empirical studies on a topic. There are many reasons for making a systematic review, the most common of which are as follows [44]:

- To identify gaps in the research progress.
- To evaluate current proofs about a technology or method.
- To propose a research roadmap for future research.
- To help in proposing new hypothesis.

Although the work of [45, 46 and 47] has been described in several ways, the steps of the examination method presented by [44] for general applications are relatively similar. In this article, the research process follows the steps as outlined here [47]: research planning, pinpointing of research, choosing primary studies, and finally, classification of selected research items.

### A. Research Planning

Academic research questions are primary concerns while conduction of a systematic literature review must be answered by researchers in their research. Keeping in view the empirical studies, such as [44, 45, and 48], we postulate the following research question discussed in the following subsections to study the design of ontologies: How ontology development methodology can be proposed to design a domain ontology of IWs?

### B. Research Identification

Our search strategy was as follows: we started our research work by finding general search keywords and search terms, and combined searches to identify as many relevant documents as possible. Moreover, we had consulted ontology experts for the identification of the relevant documents. The following electronic databases were used to search using keywords: IEEE Explore; Springer Link; AIS (Association for Information System) Electronic library; ScienceDirect; Google scholar; Business Source Premier; ISI Web of Science; ACM Digital library; Inspec; Scopus; ProQuest Science Journals; DBLP; and Wiley Online Library. Following keywords or phrases has been used for the identification of the relevant articles:

- Ontology development
- Domain ontology
- Methodology for designing domain ontology
- Designing domain ontology
- Methodology for ontology

- How to design ontology
- How to design domain ontology
- Method for designing domain ontology
- Method for designing ontology
- Design of ontology
- Design of domain ontology
- Design and implementation of domain ontology
- Design and implementation of ontology
- Design and development of domain ontology
- Design and development of ontology
- Methodology for domain ontology
- Construction of ontology
- Construction of domain ontology

### C. Selection of Primary Studies

After selecting the relevant articles, we removed the duplicate articles and the titles that were not relevant to ontology development methodologies. With, 90 articles in hand, we read the summaries with following criteria:

- Exclude if the paper was not about domain ontologies.
- Exclude if the paper was about automatic ontology construction.
- Exclude if the paper was not about designing a methodology for ontology development.
- Exclude if the paper was a literature review.

Finally, we made selection of 16 articles for review, from 2015 to 2020. The overall process was substantially in line with Fig. 1, also used by [48].

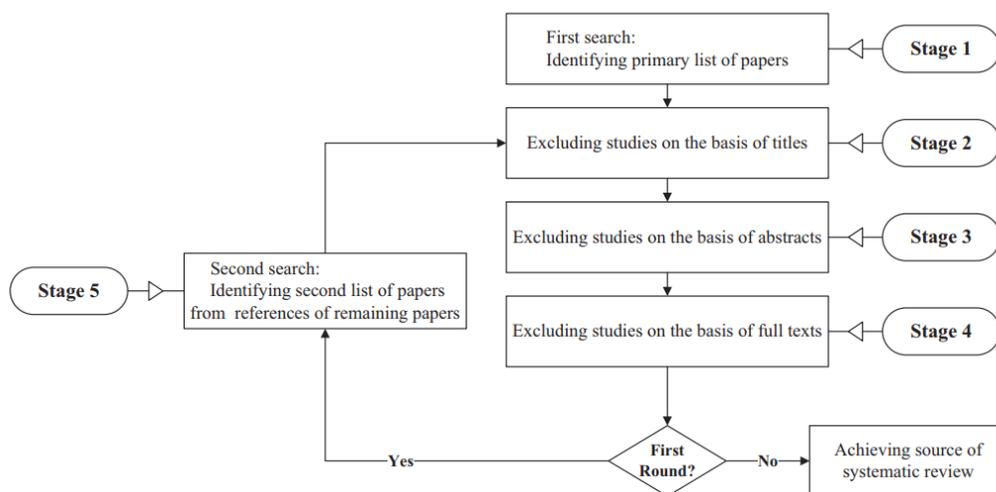


Fig. 1. Process of Article Selection.

#### D. Comparison of Methodologies for Designing Domain Ontologies

The methodology for ontology development, a guideline for ontology developers, discusses about process and methods for ontology development [11]. Different ontology development methodologies have been defined [12, 13, 14, 15, 16, 17, 18, 19, 20 and 21]. The methodology proposed in our previous work [13], was our initial attempt to design a methodology for designing ontology for the waste management domain. In this context, different researches designed their domain ontologies in different ways and there is no single viable method for doing so as discussed by [22].

Additionally, different methodologies focus on development of ontologies differently and include distinct aspects of ontology development. For example, few focused on scope of ontology and domain analysis and there was no design phase and no detail about activities they performed during ontology development. This study made a comparison of several methodologies, along with identification of a criterion, on the basis of which it was necessary to analyze and compare the methodologies. Selected evaluation criteria were defined by observing needs and trends that have evolved over the last decade, which cover sixteen different aspects of methodology for developing domain ontologies. These evaluation criteria will help the reader to understand different ontology development methodologies and will help to select a suitable methodology for designing their domain ontology of the domain of discourse.

#### E. Criteria for Analysis

We define below the criteria used to compare methodologies for ontology development. The last six facets of the criteria are coarse-grained level of a methodology, namely localization of ontology, support for reusability, support for integration, support for interoperability, methodology rooted in established methodologies and estimation of human resources. The first nine facets of the criteria, namely domain analysis, conceptualization, level of detail, collaborative construction, implementation, evaluation, instantiation, maintenance and documentation discuss the technical fine-grained level of a methodology and help the reader in understanding a particular methodology for ontology development. Table I presents a detailed and complete comparison of the methodologies based on the established criteria. This study will compare selected methodologies for ontology development base on following criteria (C1-C16), which are attention-grabbing when it comes to the creation of domain ontologies (few are taken from [49]):

- C1:Domain Analysis (specification, knowledge acquisition). Domain analysis requires a lot of resources and lacks domain-specific recommendations. The situation becomes more complex when ontology integrates knowledge from different fields. The necessary training seminars are important for understanding the terminology of the core of ontologies for team members who are unfamiliar or have a controversial understanding.

- C2:Conceptualization. Does the proposed methodology support or include conceptualization activity, as described in [50]?
- C3:Level of detail. Does the methodology provide details about techniques and methods for various activities? In this case, the level of detail criteria is evaluated on a scale from very few details as 1 to 5 (detailed).
- C4:Collaborative construction. Collaborative construction involves several stakeholders of the ontology development group to work from the same location or different locations, on the same ontology simultaneously, and without affecting the overall efficiency of the ontology development task.
- C5:Implementation (Model to language). Implementation needs a representation language for the implementation of a conceptual ontology model to ontology. In this context, manual construction from a formal conceptual model to ontology is laborious and ontology developers do not prefer it. However, construction of ontology from a formal model requires paying attention to semantic differences between the formal model and the representation language.
- C6:Evaluation (Refinement). The evaluation aspect requires the evaluator to have a good understanding of the ontology and a proper understanding of ontology is a key success factor of the ontology evaluation process.
- C7:Instantiation/ontology population/data generation. Does the methodology apply any technique and method for instantiation? The population of an ontology can be processed easily as compared to the processing of poorly structured sources such as documents, relational tables and XML data.
- C8:Maintenance/Modifications. Ontologies are constantly confronted with the problem of evolution. Due to the complexity of the changes to be made, a maintenance process, manual or semi-automatic or automatic, is increasingly necessary to facilitate this task and ensure its reliability.
- C9:Documentation. Does the methodology provide detailed documentation of methods and activities for ontology construction? Documentation helps in the understanding of the ontology, its use, reuse and future revision of the ontology.
- C10:Ontology localization. Does the methodology support localization of the ontology? Adaption of the ontology into a natural language is called localization of ontology.
- C11: Support for reusability. Ontology development is a tedious and time-consuming task. Usually, it is costly to reuse application dependent ontologies. In this context, use of upper-level or foundational ontologies helps in reducing reusability costs by providing a common conceptual structure in domain ontologies developed using same upper ontology.

- C12:Support for integration and merging. The integration of ontology is dependent on degree of overlapping between two integrating ontologies with same representing language. The quality of integration is measured, in terms of accuracy (correct mapping rate) and recall (discovered mapping rate).
- C13:Support for interoperability. Interoperability between systems requires the same supporting structure (same high-level concepts). The resulting ontologies will hence have a common global conceptual structure and it will help them in sharing their concepts.
- C14:Estimation of human resources. Team needs (people with required skills, experience and team size): ontology engineers, domain experts, technical writer for documentation, software engineers.
- C15:Sample Application/project. Are any samples of ontology designed for the planned project?
- C16:Methodology rooted in established methodologies. Is the developed methodology rooted in any well-established methodology like Design Science Research (DSR)?

#### IV. DISCUSSION

Most of the methodologies studied provide domain analysis, conceptualization, implementation, evaluation, instantiation and provide an example of domain ontology; although most of them have not provided all the details of the techniques and activities involved. With less details, it is difficult to follow a methodology for designing ontologies for a specific domain. There are some exceptions that have provided details, in particular, [19, 51, 53, and 56]. Additionally, most of the methodologies studied (compared to Table I) do not offer support for maintenance, documentation, construction of collaborative ontologies, support for reuse, support for integration, support for interoperability, ontology localization and human resources estimation (for example, ontology experts).

As we know, ontology needs change over time; support for modification and maintenance are mandatory for an ontology. As such, there should be maintenance and support in the domain ontology. Few studied methodologies provide maintenance, collaborative construction of ontologies, documentation, ontology localization, support for reuse, support for integration, support for interoperability and estimation of human resources (for example ontology experts). For example, the maintenance is discussed by [12, 17, 20, 55, and 56]. Likewise, few researchers provide documentation; for example, documentation is provided by [12, 17, 20, 21, 27, 54, and 58]. The documentation provides help in understanding the ontology.

The development of an ontology is an evolutionary process and the availability (in the same place) of domain experts, ontology engineers, technical writers and other human resources is not always possible. The collaborative construction of ontologies will therefore make valuable human resources available worldwide. In this regard, support for the collaborative construction of ontologies will help the correct

execution of the ontology construction plan. [17 and 56] provide the collaborative ontology development.

Additionally, for a large and complex domain, a single large ontology cannot be created. There would be multiple subdomains, and each subdomain needs an ontology. In this context, interoperability of subdomain information systems is needed. Interoperability between systems requires the same support structure (same high-level concepts). The resulting ontologies will hence have a common global conceptual structure and it will help them to share their concepts. The methodologies proposed by [18 and 56] support interoperability. Similarly, ontology integration of subdomain information systems is needed. This can be realized in the integration of the ontology discussed by [12, 17, 21, 55, and 56].

Creating ontology is a tedious and boring activity, and reusability allows you to reuse the existing ontology and add additional components to the target ontology. Few researchers [13, 17, 27, and 55] provide a mechanism to reuse ontology. Similarly, it is important to translate ontology into different natural languages, as it is not possible to conceive a separate ontology for each natural language (English, Malay, Urdu, Hindi, German, French, etc.). Only two [34 and 56] of the studied methodologies provide support for the localization of the ontology. In addition to this, estimating the costs of the ontology building process requires estimating the required human resources, such as domain experts, ontology engineers, and API developers. Only [17, 19, and 56] discuss little about the estimation of human resources. Finally, most of the methodologies examined in the study are not based on any well-established methodology. There are some exceptions, mainly [13], who rooted their methodology in Design Science Research (DSR).

In this context, developing ontology development methodologies, is a tedious and complex task. Existing studies show that a methodology developed based on experience of one or a few projects, may or may not be effective for creating domain ontologies. It means, there is a gap in the existing methodologies and this gap is highlighted by our analysis. The following points are concluded from this study:

- None of the methodologies is mature enough.
- There is an absence of documentation or poor documentation on techniques, activities and methods used in ontology construction. There are some exceptions, mainly NeOn [56].
- Some methodologies support the concept of reusability, localization, integration, interoperability, estimation of human resources and reengineering; but not others.
- Little attention is paid by researchers to the collaborative construction aspect of domain ontology, and collaborative ontology construction is one of important aspects of ontology construction.
- As conventional strategies have been used by researchers for collecting ontology conceptions, it is important to propose and use novel techniques and methods for designing domain ontologies, such as those proposed by [6].

TABLE I. COMPARISON OF METHODOLOGIES FOR CREATING DOMAIN ONTOLOGIES FROM 2015-2020

No.	Methodology	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
1	OntoDI: The Methodology for Ontology Development on Data Integration [21]	Y	Y	4	N	Y	Y	Y	N	Y	N	N	Y	N	N	Y	N
2	An Agile Methodology for Ontology Development [12]	Y	Y	4	N	Y	Y	Y	Y	Y	N	N	Y	N	N	Y	N
3	A Domain Ontology for Software Requirements and Change Management in a Global Software Development Environment [14]	Y	Y	4	N	Y	Y	Y	N	N	N	Y	N	N	N	Y	N
4	Design and Implementation of E-Campus Ontology with a Hybrid Software Engineering Methodology [18]	Y	Y	3	N	Y	N	Y	N	N	N	N	N	Y	N	Y	N
5	An Ontology for the Waste Management Domain [13]	Y	Y	4	N	N	Y	Y	N	N	N	Y	N	N	N	Y	Y
6	Towards a Software Centric Approach for Ontology Development: Novel Methodology and its Application [19]	Y	Y	5	N	N	Y	Y	N	N	N	N	N	N	Y	Y	N
7	Design and Development of a Biocultural Ontology for Personalized Diabetes Self-Management of American Indians [51]	Y	Y	5	N	Y	Y	Y	N	N	N	N	N	N	N	Y	N
8	Towards a Methodology for Reusable Ontology Engineering: Application to the Process Engineering Domain [52]	Y	Y	3	N	Y	Y	N	N	N	N	Y	Y	N	N	Y	N
9	A Methodology for a Criminal Law and Procedure Ontology for Legal Question Answering [53]	Y	Y	5	N	Y	Y	Y	N	N	N	N	N	N	N	Y	N
10	Fine Construction of HIV Protein Ontology [23]	Y	Y	3	N	Y	Y	Y	N	N	N	N	N	N	N	Y	N
11	The Ontology-Based Methodology Phases to Develop Multi-Agent Systems (OmMAS) [54]	Y	Y	3	N	Y	Y	N	N	Y	N	N	N	N	N	N	N
12	The Methodology for Ontology Development in Lesson Plan Domain [55]	Y	Y	4	N	Y	Y	Y	Y	N	N	Y	Y	N	N	Y	N
13	A Lightweight Methodology for Rapid Ontology engineering - UPON Lite [17]	Y	Y	4	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	N
14	The NeOn Methodology for Ontology engineering [56]	Y	Y	5	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
15	YAMO: Yet Another Methodology for large-scale faceted Ontology construction [57]	Y	Y	4	N	N	Y	Y	N	N	N	N	N	N	N	Y	N
16	Methodologies to build ontologies for terminological purposes [34]	Y	Y	2	N	Y	N	N	N	N	Y	N	N	N	N	N	N

### V. CONCLUSION

There are many methodologies that have been proposed by several researchers during the last decade. However, designing a domain ontology for IWSs needs a well-defined ontology development methodology. Therefore, in this article, a survey has been conducted to compare ontology development methodologies between 2015 and 2020. We analyzed that a non-collaborative or centralized ontology development methodology is a major trend in the development of domain ontologies. From this analysis, we recommend collaborative ontology development methodologies for keeping ontologies live, evolved and to increase their reuse potential. An important conclusion is that the collaborative ontology development methodology rooted in a well-established methodology such as Design Science Research, with interoperability, reusability,

merging, localization, detailed documentation, human resource estimation and with at least one sample ontology developed using this methodology, has a good impact on the development of a domain ontology. To conclude, a collaborative methodology rooted in a well-established methodology, with supported collaboration tools for the development, maintenance and advancement of modularized ontologies, has a good impact on the development of a domain ontology.

Finally, proposing a methodology for designing a domain ontology of IWSs is complex. There are criteria that need to be incorporated including C1-C16. Among these criteria, some are extracted from existing methodologies; however, there are key criteria which are missing, and which are crucial for developing a domain ontology of IWSs. These are, collaborative ontology construction, interoperability, reusability, merging,

localization, detailed documentation, human resource estimation and methodology rooted in previously well-established methodology. This paper has presented most of the criteria that we think are important for the research community of domain ontology; and can be further used to establish a well-defined methodology for developing a domain ontology of IWs. In this article, we have not included all ontology development methodologies because there are too many; some are very old, and other did not explain their methodology for ontology development.

#### REFERENCES

- [1] T. Gruber, "A Translational Approach to Por Ontologies," *Knowledge Acquisition*, vol. 5, pp. 199-220, 06/01 1993.
- [2] S.-H. Liao, C.-c. Chen, and D. C. Hu, "The role of knowledge sharing and LMX to enhance employee creativity in theme park work team: A case study of Taiwan," *International Journal of Contemporary Hospitality Management*, vol. 30, pp. 00-00, 03/29 2018.
- [3] M. N. Ahmad, N. Zakaria, and D. Sedera, "Ontology-Based Knowledge Management for Enterprise Systems," *IJEIS*, vol. 7, pp. 64-90, 10/01 2011.
- [4] T. Tudorache, "Ontology engineering: Current state, challenges, and future directions," *Semantic Web*, pp. 1-14, 12/19 2019.
- [5] K. Kotis, G. Vouros, and D. Spiliotopoulos, "Ontology engineering methodologies for the evolution of living and reused ontologies: status, trends, findings and recommendations," *The Knowledge Engineering Review*, vol. 35, 01/31 2020.
- [6] Colomb and Ahmad, "Merging ontologies requires interlocking institutional worlds," *Applied Ontology*, vol. 2, pp. 1-12, 01/12 2007.
- [7] Colomb and Ahmad, "A perdurant ontology for interoperating information systems based on interlocking institutional worlds," *Applied Ontology*, vol. 5, pp. 47-77, 01/01 2010.
- [8] S. Khantong and M. N. Ahmad, "An Ontology for Sharing and Managing Information in Disaster Response: In Flood Response Usage Scenarios," *Journal on Data Semantics*, 2019/12/14 2019.
- [9] Searle, *The Construction of Social Reality*: Free Press, 1995.
- [10] Colomb, "Representation of Action is a Primary Requirement in Ontologies for Interoperating Information Systems," 2013.
- [11] Y. Sure-Vetter, S. Staab, and R. Studer, "Ontology Engineering Methodology," in *Handbook on Ontologies*, 2009.
- [12] A. Abdelghany, N. Ramadan, and H. Hefni, "An Agile Methodology for Ontology Development," *International Journal of Intelligent Engineering and Systems*, vol. 12, pp. 170-181, 04/30 2019.
- [13] M. N. Ahmad, K. B. A. Badr, E. Salwana, N. H. Zakaria, Z. Tahar, and A. Sattar, "An Ontology for the Waste Management Domain," in *PACIS*, Yokohama, Japan, 2018.
- [14] A. Alsanad, A. Chikh, and A. Mirza, "A Domain Ontology for Software Requirements Change Management in Global Software Development Environment," *IEEE Access*, vol. 7, pp. 49352-49361, 01/01 2019.
- [15] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, et al., "Gene Ontology: tool for the unification of biology," *Nature Genetics*, vol. 25, pp. 25-29, 2000/05/01 2000.
- [16] E. Blomqvist, V. Presutti, E. Daga, and A. Gangemi, "Experimenting with eXtreme Design," in *Knowledge Engineering and Management by the Masses*, Berlin, Heidelberg, 2010, pp. 120-134.
- [17] A. De Nicola and M. Missikoff, "A Lightweight Methodology for Rapid Ontology Engineering," *Commun. ACM*, vol. 59, pp. 79-86, 03/01 2016.
- [18] K. Jacksi, "Design and Implementation of E-Campus Ontology with a Hybrid Software Engineering Methodology," *Science Journal of University of Zakho*, vol. 7, pp. 95-100, 09/30 2019.
- [19] S. John, N. Shah, and C. Stewart, *Towards a Software Centric Approach for Ontology Development: Novel Methodology and its Application*, 2018.
- [20] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernández-López, "The NeOn Methodology for Ontology Engineering," in *Ontology Engineering in a Networked World*, M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, and A. Gangemi, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 9-34.
- [21] A. Yunianta, A. H. Basori, A. S. Prabuwo, A. Bramantoro, I. Syamsuddin, N. Yusof, et al., "OntoDI: The Methodology for Ontology Development on Data Integration," *International Journal of Advanced Computer Science and Applications*, vol. 10, 01/01 2019.
- [22] G. Brusa, M. L. Caliusco, and O. Chiotti, "Towards ontological engineering: a process for building a domain ontology from scratch in public administration," *Expert Systems*, vol. 25, pp. 484-503, 2008.
- [23] L. Zhang, H. Chen, Y. Zhang, and C. Tao, *Fine Construction of HIV Protein Ontology*, 2018.
- [24] R. Studer, V. R. Benjamins, and D. Fensel, "Knowledge engineering: principles and methods. *Data Knowl Eng* 25(1-2):161-197," *Data & Knowledge Engineering*, vol. 25, pp. 161-197, 03/01 1998.
- [25] C. Feilmayr and W. Wöß, "An Analysis of Ontologies and their Success Factors for Application to Business," *Data & Knowledge Engineering*, vol. 101, pp. 1-23, 01/01 2016.
- [26] M. Pasha and A. Sattar, "Building Domain Ontologies From Relational Database Using Mapping Rules," *International Journal of Intelligent Engineering and Systems*, vol. 5, No.1, 03/31 2012.
- [27] R. Silva Lopez, M. Bravo, I. Mendez, and V. Arias, "GODEM: A Graphical Ontology Design Methodology," *Research in Computing Science*, vol. 84, pp. 17-28, 12/31 2014.
- [28] S. C and R. Rv, *Ontology Engineering Methodologies: An Analytical Study*, 2017.
- [29] N. Noy and D. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," *Knowledge Systems Laboratory*, vol. 32, 01/01 2001.
- [30] N. Law, M. A, A. Tang, F.-C. Lim, H. Kasim, M. Othman, et al., "A Review of Ontology Development Aspects," *International Journal of Advanced Computer Science and Applications*, vol. 10, 01/01 2019.
- [31] H. Hi, M. Uschold, and M. Grüninger, "Creating Semantically Integrated Communities on the World Wide Web," 03/11 2004.
- [32] M. Grüninger and M. Fox, "Methodology for the Design and Evaluation of Ontologies," 07/08 1995.
- [33] U. Yadav, G. S. Narula, N. Duhan, and V. Jain, "Ontology Engineering and Development Aspects: A Survey," *International Journal of Education and Management Engineering*, vol. 6, pp. 9-19, 05/08 2016.
- [34] M. R. Zambrana, "Methodologies to Build Ontologies for Terminological Purposes," *Procedia - Social and Behavioral Sciences*, vol. 173, pp. 264-269, 02/13 2015.
- [35] R. Iqbal, M. Murad, A. Mustapha, and N. Sharef, "An Analysis of Ontology Engineering Methodologies: A Literature Review," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 6, pp. 2993-3000, 09/01 2013.
- [36] A. Farquhar, R. Fikes, and J. Rice, "The Ontolingua Server: A Tool for Collaborative Ontology Construction. *International Journal of Human Computer Studies*," *International Journal of Human-Computer Studies*, vol. 46, 02/27 2000.
- [37] I. Al-Baltah, A. a. abdul ghani, W. N. Wanabrahman, and R. Atan, "A Comparative Study on Ontology Development Methodologies towards Building Semantic Conflicts Detection Ontology for Heterogeneous Web Services," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 7, pp. 2674-2679, 04/01 2014.
- [38] E. Simperl and M. Luczak-Roesch, "Collaborative Ontology Engineering: A Survey," *Knowledge Engineering Review (to appear)*, vol. 29, 01/01 2012.
- [39] B. Stadlhofer, Salhofer, P. & Durlacher, A, "An overview of ontology engineering methodologies in the context of public administration," presented at the *The Seventh International Conference on Advances in Semantic Processing*, 2013.
- [40] M. Fernández-López and A. Gomez-Perez, "Overview and analysis of methodologies for building ontologies," *The Knowledge Engineering Review*, vol. 17, pp. 129-156, 06/01 2002.
- [41] A. Gómez-Pérez, "Towards a framework to verify knowledge sharing technology," *Expert Systems with Applications*, vol. 11, pp. 519-529, 1996/01/01/ 1996.

- [42] B. Swartout, R. S. Patil, K. Knight, and T. A. Russ, "Toward Distributed Use of Large-Scale Ontologies t," 1997.
- [43] J. Euzenat, "Building consensual knowledge bases: context and architecture," 1995.
- [44] B. Kitchenham, R. Pretorius, D. Budgen, P. Brereton, M. Turner, M. Niazi, et al., "Systematic literature reviews in software engineering – A tertiary study," *Information and Software Technology*, vol. 52, pp. 792-805, 08/01 2010.
- [45] L. Karg, M. Grottko, and A. Beckhaus, "A systematic literature review of software quality cost research," *Journal of Systems and Software*, vol. 84, pp. 415-427, 03/01 2011.
- [46] D. Mellado, C. Blanco, L. E. Sánchez Crespo, and E. Fernández-Medina, "A systematic review of security requirements engineering," *Computer Standards & Interfaces*, vol. 32, pp. 153-165, 06/01 2010.
- [47] J. Valaski, A. Malucelli, and S. Reinehr, "Ontologies application in organizational learning: A literature review," *Expert Systems with Applications*, vol. 39, pp. 7555-7561, 06/01 2012.
- [48] A. Ghapanchi and A. Aurum, "Antecedents to IT personnel's intentions to leave: A systematic literature review," *Journal of Systems & Software*, vol. 84, pp. 238-249, 02/01 2011.
- [49] E. Simperl and C. Dr. Tempich, *Ontology Engineering: A Reality Check*, 2006.
- [50] A. Gomez-Perez, M. Fernández-López, and O. Corcho, *Ontological Engineering: With Examples from the Areas of Knowledge Management, E-Commerce and the Semantic Web*, 1 ed.: Springer-Verlag London, 2004.
- [51] J. Li and S. Alian, *Design and Development of a Biocultural Ontology for Personalized Diabetes Self-Management of American Indians*, 2018.
- [52] N. Trokanas, L. Koo, and F. Cecelja, "Towards a Methodology for Reusable Ontology Engineering: Application to the Process Engineering Domain," in *Computer Aided Chemical Engineering*. vol. 43, A. Friedl, J. J. Klemeš, S. Radl, P. S. Varbanov, and T. Wallek, Eds., ed: Elsevier, 2018, pp. 471-476.
- [53] B. Fawei, J. Pan, M. Kollingbaum, and A. Wyner, "A Methodology for a Criminal Law and Procedure Ontology for Legal Question Answering: 8th Joint International Conference, JIST 2018, Awaji, Japan, November 26–28, 2018, Proceedings," ed, 2018, pp. 198-214.
- [54] A. Yunianta, O. Barukab, N. Yusof, A. Musdholifah, H. Jayadiyanti, N. Dengen, et al., *The ontology-based methodology phases to develop multi-agent system (OmMAS)*, 2017.
- [55] A. Saad and S. Shaharin, "The Methodology for Ontology Development in Lesson Plan Domain," *International Journal of Advanced Computer Science and Applications*, vol. 7, 04/01 2016.
- [56] M. C. Suárez-Figueroa, A. Gomez-Perez, and M. Fernández-López, "The NeOn Methodology framework: Ascenario-based methodology for ontologydevelopment," *Applied Ontology*, vol. 10, pp. 107-145, 09/15 2015.
- [57] B. Dutta, U. Chatterjee, and D. Madalli, "YAMO: Yet Another Methodology for large-scale faceted Ontology construction," *Journal of Knowledge Management*, vol. 19, pp. 6-24, 02/09 2015.
- [58] M. Gawich, A. Badr, A. Hegazy, and H. Ismail, "A Methodology for Ontology Building," *International Journal of Computer Applications*, vol. 56, pp. 39-45, 10/01 2012.

# Performance Analysis of Transient Fault-Injection and Fault-Tolerant System for Digital Circuits on FPGA

Sharath Kumar Y N<sup>1</sup>

Department of Electrical and Electronics Engineering  
Dayananda Sagar College of Engineering, Bengaluru, India

Dinesha P<sup>2</sup>

Department of Electronics and Communication Engineering  
Dayananda Sagar College of Engineering, Bengaluru, India

**Abstract**—A Fault-Tolerant System is necessary to improve the reliability of digital circuits with the presence of Fault Injection and also improves the system performance with better Fault Coverage. In this work, an efficient Transient Fault-Injection system (FIS) and Fault-Tolerant System (FTS) are designed for digital circuits. The FIS includes Berlekamp Massey Algorithm (BMA) based LFSRs, with fault logic followed by one – hot-encoder register, which generates the faults. The FTS is designed using Triple-Modular-Redundancy (TMR) and Dual Modular- Redundancy (DMR). The TMR module is designed using the Majority Voter Logic (MVL), and DMR is designed using Self-Voter Logic (SVL) for digital circuits such as synchronous and asynchronous circuits. The four different MVL approaches are designed in the TMR module for digital circuits. The FIS-FTS module is designed on Xilinx-ISE 14.7 environment and implemented on Artix-7 FPGA. The synthesis results include chip area, gate count, delay, and power are analyzed along with fault tolerance, and coverage for given digital circuits. The fault tolerance is analyzed using Modelsim-simulator. The FIS-FTS module covers an average of 99.17% fault coverage for both synchronous and asynchronous circuits.

**Keywords**—Digital circuits; transient fault; fault injection; fault tolerant; triple modular redundancy; dual modular redundancy; majority voter logic; self-voter logic

## I. INTRODUCTION

Designing the electronic system by concerning the reliability and availability features are used in many critical applications like aerospace, military, transportation, and avionics. These electronics systems provide continuous support while performing real-time applications. If any attacks have occurred on these systems, it affected the overall system performance and led to failure. The fault injection is a process for estimating or evaluating the fault-tolerant system. There are many FIS methods are available based on hardware FIS, simulation-based FIS, and emulation based FIS. The hardware-based FIS injects high Laser beams or ion beams to circuits. The software-based FIS has static and dynamic approaches for analyzing the high computational overheads with high accuracy. The emulation based FIS has hardware reconfiguration, and circuits instrumentations approach. The dependability can tolerate system failures by providing better services, including threats, attributes, and means. Many available attributes include reliability, safety, confidentiality, security, integrity, maintainability, and availability creates

dependability in any of the electronics systems. The faults, errors, failures are a chain of threats, which affect the system or component. The means have methods or techniques which include fault tolerance, fault preventions, fault forecasting, and fault removal [1-2].

The triple modular redundancy technique is used to protect logic circuits against soft errors. The original logic circuit is repeated three times, and the output is obtained from the majority voter. Each replicated circuit works independently from the other circuit. If an error occurs in one circuit, it is masked by TMR by majority voting and thus propagates the error-free output. On the other hand, the dual modular redundancy method, the logic circuit is duplicated and works in parallel. The main goal of the DMR technique is to achieve SET fault mitigation similar to that of the TMR technique, and it consumes less power and area compared to that of the TMR technique.

The FTS is used to extend the dependability of the electronic system. The FIS can be tolerant of any kind of faults in the system with expected and predictive ways. The faults are tolerated by the system using redundancies. There are three redundancies, like space, time, and information, based on system failures. These FIS have also implemented either in software-based and hardware-based approaches, which depend upon the system requirements. The FTS applies to many application fields, including cloud computing, distributed systems, and specific applications. The cloud computing [3] are facing with networking faults, service expiry faults, physical faults, process faults. The distributing system [4] is dependent on availability, safety, reliability, and maintainability. These distributing systems are tolerated by faults using reactive FTS and proactive FTS. Reactive FTS includes retry, checkpointing, message logging, and replication.

Similarly, proactive FTS includes software rejuvenation, self-healing, and preemptive migration. The specific applications [5] include reactive FTS and proactive FTS tolerate montage, inspiral, Sipht, epigenomics, and cyber shake. The FPGA based in-operations faults [6] are occurred due to aging, the stress in terms of heat and voltage, and high energy particle impact along with environmental changes. These FPGA based in-operations faults are recovered by using redundancy approaches like TMR, DMR, and other approaches. The FTS is designed for math circuits [7], which

identifies the critical gates. The critical gates evaluate the reliability in two ways, nominal reliability and practical reliability. The single event upset is prevented using hardware, software, hybrid, adaptive, and adaptive hybrid approaches are available. The hardware approaches include TMR, DMR, and N-Modular Redundancy. Similarly, software approaches use Error Detection – Duplication Instruction (EDDI), which runs on the processor. The hybrid approach is a combination of software, hardware, and Error Correction Codes (ECC) [8-10].

In this work, the transient fault injection and fault-tolerant system module are designed using TMR based majority voter logic, and DMR based self-voter logic for fault analysis with coverage for digital circuits like synchronous and asynchronous circuits. Section II describes the review of the existing approaches of FIS and FTS for different applications. The proposed FIS-FTS module using TMR and DMR is elaborated in Section III. Section IV analyzes the performance evaluation of the FIS-FTS module with synthesis and fault-tolerant results for digital circuits. Finally, conclusion and future work are discussed in Section V.

## II. RELATED WORK

This section describes the review of recent existing works on FIS-FTS modules for different applications for software and hardware approaches. Balasubramanian et al. [11] present combinational circuit based fault tolerance and redundant logic insertion with improvements and estimating fault-masking using the truth-cum fault enumeration table. The fault-tolerant module is designed using 28/32nm CMOS technology by analyzing the figure-of-merit and Power Delay Product (PDP). Schweizer et al. [12] present the FPGA based FIS with fast and accurate analysis for ISCAS-83 based gate circuits. The statistical analysis between RTL/Gate level circuits is mapped for place and route logic in FPGA Design. Li et al. [13] present the CMOS based soft-error sophisticated design for DMR with a low power mechanism. The complementary-DMR is used to mitigate soft errors, which reduces the voting circuit area, error rate, and better time than TMR approaches. The DMR uses separate two circuits and merging circuit to overcome the soft error to improve the area. Sheikh et al. [14] present the DMR based FTS for combinational circuits, which includes C-element (CEL) based DMR to overcome TMR and selective transistor redundancy (STR) modules. The protected C-element provided better fault-tolerance and applied for LGSynth'91 benchmark circuits. The results are analyzed for area-overhead, circuit reliability using DMR, DMR-CEL, and TMR logic modules.

Arifeen et al. [15] describe the approximate-TMR based fault-tolerant voter module to overcome the area –overhead issues in TMR based modules. Approximate-TMR also tolerates the internal faults in voter inputs. Along with full protection to critical parts and partial protection to the rest of the system. The work analysis the PDP, transistors, and reliable calculation using the CMOS approach. The fault-diagnosis system is a module by mandaogade et al. [16] using FIS for discrete systems. The fault-diagnosing system includes a circuit under test (CUT) for 32-bit AND gate, control logic, stimulus generator, and output response. Oliveira et al. [17] present a short circuit based fault- diagnosis (SC-FD) for fault-

tolerant voltage-source-inverter (VSI). The reconfigurable feature is incorporated in VSI like two-auxiliary switches, fast fuse, and fast legs. The VSI based decision system is to isolate the fault conditions on switches using diagnostic variables.

The fault-tolerant system used in many other advanced applications includes System-on-Chip (SoC), Network-on-Chip (NoC), and others. Chekmarev et al. [18] present the On-Chip Debugging (OCD) using the modification of FIS for processor cores of SoC. The fault-injection system includes on-chip bus connection of OCD for processor cores, fault-injector, external interface controller, and controller memory acts as a system under test which is connected externally to memory for fault coverage analysis. Hybrid FIS for NoC is presented by Coelho et al. [19], to improve reliability and NoC performance. The FIS has SET, SEU with timer, and is connected with target FPGA and is interfaced with 4x4 NoC via AXI-Interface. Guruprasad et al. [20] present the fault and congestion-free NoC design where faults are tolerated in NoC using Error Code Correction techniques with better fault coverage. Podivinsky et al. [21] present the hardware-based robot controller with a fault-tolerant mechanism, which evaluates single and multiple fault-injection analysis with electronic failures.

## III. FIS-FTS MODULE FOR DIGITAL CIRCUITS

The fault-injection system, along with the fault-tolerant system, is represented in detail using TMR and DMR methods for digital circuits (both synchronous and asynchronous circuits). The FIS and FTS modules are explained in the following section.

### A. Fault Injection System (FIS)

The fault-injection system is designed to randomly generate the faults using the Berlekamp Massey algorithm based on LFSRs and used in FTS. The fault-injection system using LFSR is represented in Fig. 1. The FIS includes BMA based on two LFSRs, fault logic, and multiplexor. The BMA based LFSR provides better execution time than conventional LFSRs. The two LFSR uses the following polynomials for the random data generations in equations (1) and (2).

$$\text{LFSR}_1 = x^3 + x + 1 \quad (1)$$

$$\text{LFSR}_2 = x^3 + x \quad (2)$$

The BMA based LFSR generates faults based on the above two equations. The control signal load is used in the initial process setup. When the load is activated, the default values assigned to LFSR registers, else the polynomial values are updated in registers. The two LFSR values are XOR'ed in the fault logic process. The fault logic acts as a select line to the multiplexor. The one hot-encoder using shift register generates the encoded data are XORed with feedback register output for the creation of faults in a more random manner. If the fault logic output acts as a select line, if it is 1, then the XORed output is fault data else user data is considered as not a fault data. These faults are considered as transient faults in FIS. These fault data is injected into the FTS for digital circuits, checks the status of the circuits, and analyzes the fault coverage.

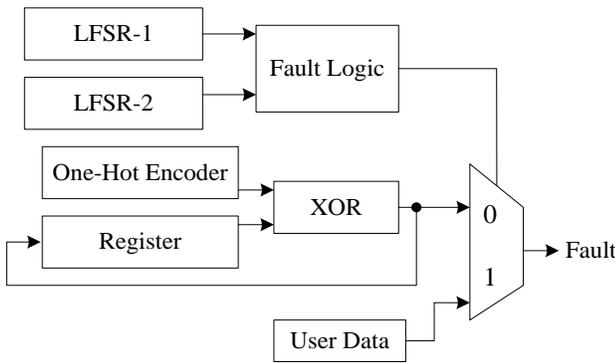


Fig. 1. Fault Injection System using LFSR.

**B. Fault-Tolerant System (FTS) using TMR and DMR**

The fault-tolerant system is used to tolerate the injected faults in digital circuits. The digital circuits include both synchronous and asynchronous circuits that are considered in designs using triple modular redundancy and dual modular redundancy. The FTS has a TMR module using the majority voter logic and DMR based self-voter logic for fault analysis. The FTS using TMR based MVL has represented in Fig. 2. The asynchronous circuits include 4-bit ripple-carry adder (RCA), 4:16 decoder, 4:2 encoders, and comparator has been used in the FTS process.

Similarly, the synchronous circuits include 4-bit data flip-flop, 4-bit shift register, 4-bit counter, and 4-bit read-only memory (ROM) have been considered in the FTS design process. These circuits output's like X, Y, and Z has considered in the four MVL approaches. The MVL is a simple and most popular approach for detecting and tolerating the faults with coverage in the most successful way. The hardware complexity and performance of the system will be improved by using MVL in TMR. So in this FTS, the different and possible TMR based MVL approaches are designed and analyzed for fault coverage.

The TMR module has constructed using four different MVL approaches and analyzed for both synchronous and asynchronous circuits are represented in Fig. 3. The TMR based four MVL design-1, 2, 3, and 4 approaches are also represented in Fig. 3(a-d).

The TMR based MVL Design-1 is a basic approach that includes three AND gate and one OR gate with equation (3) for output O1 function generation as follows.

$$O_1 = (X.Y) + (Y.Z) + (X.Z) \tag{3}$$

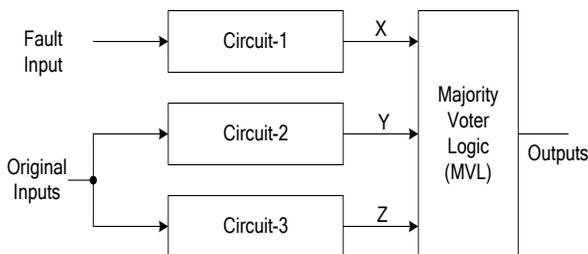


Fig. 2. Fault-Tolerant System using TMR.

The TMR based MVL Design-2 approaches 2 XOR gate, 1 inverter gate, one AND gate, and 1 multiplexor, and its output function  $O_2$  is represented in the following equations (4-6).

$$P = (X^{\wedge}Y), Q = (Y^{\wedge}Z), \text{ and } R = \sim Q \tag{4}$$

$$Sel_1 = P.R \tag{5}$$

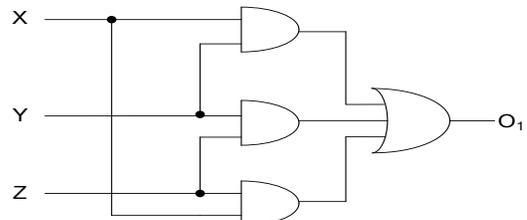
$$O_2 = \text{if } (Sel_1 = 0) \text{ then } X \text{ else } Z \tag{6}$$

Where the representation of basic gates like AND gate is '.' (Dot), OR gate is '+', XOR gate is '^', and the Inverter gate is '~'.

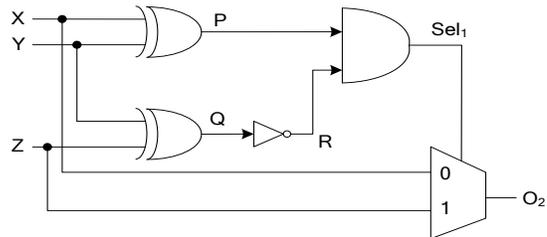
The TMR based MVL Design-3 approach has 1 XOR gate and 1 multiplexor, and its output function  $O_3$  is represented in the following equations (7-8).

$$Sel_2 = X.Y \tag{7}$$

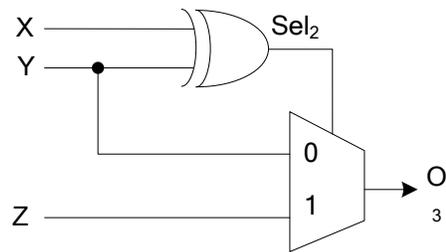
$$O_3 = \text{if } (Sel_2 = 0) \text{ then } Y \text{ else } Z \tag{8}$$



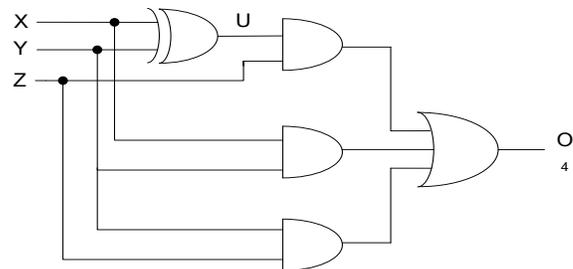
(a). TMR based MVL-Design-1.



(b). TMR based MVL-Design2.



(c). TMR based MVL-Design-3.



(d). TMR based MVL-Design-4.

The TMR based MVL Design-4 approach has 1 XOR gate and 1 multiplexor, and its output function  $O_4$  is represented in the following equations (9-10).

$$U = X.Y \tag{9}$$

$$O_4 = (U.Z) + (X.Y) + (Y.Z) \tag{10}$$

The TMR based MVL design approaches works based on two out of three majority logic. The TMR based MVL Design-two approach uses more basic gates than the other three approaches. The TMR modules triplicate the digital circuits, and the majority voter logic filters out the single event transients. The TMR based modules can control the common-mode SETs failures.

The dual modular redundancy acts as self-voter logic and has two inputs and one output. If any of the two inputs changed, the output would change the state. The DMR based SVL aims to reduce the chip area overhead than TMR based modules. The DMR based SVL has represented in Fig. 4. The DMR module uses only two circuits, assigns the faults to the first circuit, and maintains the original input in the second circuit.

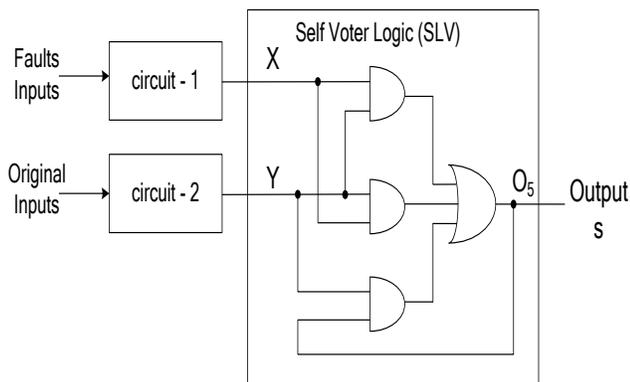


Fig. 4. Fault-Tolerant System using DMR.

The DMR based SVL design approach has three AND gate followed by the OR gate, and its output function  $O_5$  is represented in the following equation (11).

$$O_5 = (X.Y) + (X.Y) + (Y.O_5) \tag{11}$$

The self-check mechanism is incorporated in the DMR module by providing output  $O_5$  data as a feedback input to the SVL module. SET protection applies to clock and data inputs. The FIS-FTS module using TMR and DMR logic provides a single event upset and SET protection for the digital circuit inputs.

#### IV. RESULTS AND DISCUSSION

The Results and discussion of Fault injection system -Fault Tolerant system (FIS-FTS) for digital circuits, which includes synchronous and asynchronous circuits, are analyzed with detailed performance metrics like Chip area, combinational delay, and total power along with fault coverage. The fault-tolerant and coverage are analyzed in a simulator tool based on the number of injected faults. The FIS-FTS module is synthesized using Xilinx ISE 14.7 environment and simulated using Modelsim 6.5f simulator and implemented and prototyped using the Artix-7 FPGA device.

#### A. Synthesis Results

The FIS-FTS module chip-area utilization for Asynchronous Circuits using different voter logic are tabulated in Table I. And graphical representation for the same is shown in Fig. 5. The FIS is designed using TMR module with four different Majority voter logic (MVL) like TMR-D1, TMR-D2, TMR-D3, and TMR-D4. Similarly, FIS is also designed using the DMR module with self-voter logic (SLV).

The FIS-FTS module area utilization is evaluated using LUT's and Gate count (GC). The Utilization of LUT's is <1% in Artix-7 FPGA for all the asynchronous circuits like 4-bit RCA, decoder, encoder, and comparator in different voter logic's includes TMR and DMR. The Gate count (GC) is calculated using many Flip-flops (Registers), XOR-Gates, and Multiplexors, along with Block RAM's (BRAMs) and comparators. The gate count values differ based on the asynchronous circuits used in FIS. The Decoder circuit uses Three BRAMs, and Comparator circuits use six comparisons in FIS based TMR D1 to TMR-D4 modules. Similarly, in DMR based FIS, the Decoder circuit uses two BRAMs, and Comparator circuits use four comparators along with Registers, XOR gates, and Multiplexors. Overall, The DMR based FIS-FTS module using Asynchronous Circuits consumes less gate count (GC) than TMR based FIS-FTS module.

TABLE I. FIS-FTS MODULE AREA UTILIZATION FOR ASYNCHRONOUS CIRCUITS

Asynchronous Circuits	Resources	TMR-D1	TMR-D2	TMR-D3	TMR-D4	DMR
4-bit RCA	LUT's	31	40	31	31	29
	GC	74	79	77	74	66
Decoder	LUT's	41	41	41	41	50
	GC	53	56	55	53	52
Encoder	LUT's	28	33	28	28	36
	GC	59	62	61	59	56
Comparator	LUT's	32	32	32	32	29
	GC	71	77	74	71	64

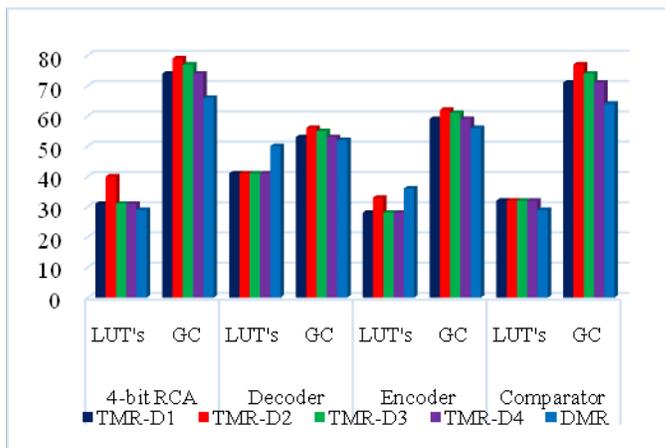


Fig. 5. Graphical Representation of FIS-FTS Module Area utilization for Asynchronous Circuits.

The Combinational delay (ns) are analyzed for the FIS-FTS module using Asynchronous circuits and represented in Fig. 6. The Asynchronous circuits like 4-bit RCA, decoder, and comparator generates better combinational delay using TMR based FIS-FTS module than DMR based FIS-FTS module. The TMR-Design-4 gives better delay for 1.68ns, 1.002ns, and 2.115ns than DMR Based design like 2.26ns, 1.709ns, and 2.517ns for RCA, Decoder, and comparator respectively.

The FIS-FTS module resource utilization for Synchronous Circuits using TMR based MVL designs and DMR based SVL are tabulated in Table II, and graphical representation for the same is shown in Fig. 7. The FIS-FTS module utilizes a < 1% LUTs chip area on Artix-7 FPGA for all the synchronous circuits like Data Flip-flop, Shift register, counter, and ROM using TMR based MVL and DMR based SVL.

In Gate Count Calculation, the counter circuit uses additionally 3 adders for all TMR based MVL designs and 2 adders for DMR based SVL designs. The ROM circuits use 3 BRAMs for all TMR based MVL designs and 2 adders for DMR based SVL design along with Registers, XOR gates, and Multiplexors. The DMR based FIS-FTS module utilizes less GC than TMR Based FIS-FTS module. The LUTs utilization for all the Synchronous Circuits based on TMR and DMR modules on an average is 26.

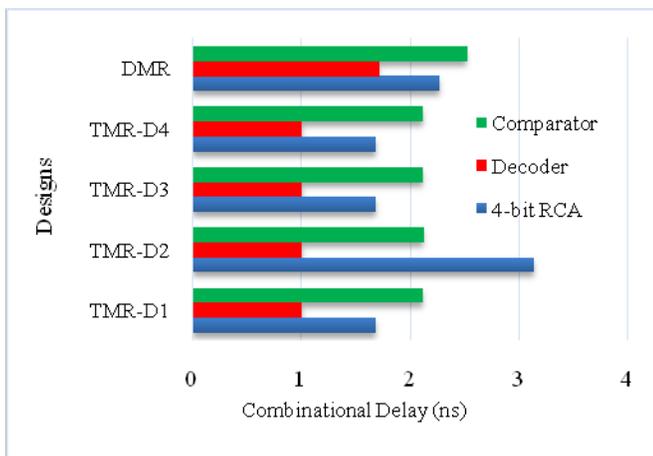


Fig. 6. Combinational Delay Analysis of Asynchronous Circuits in FIS-FTS Module.

TABLE II. FIS-FTS MODULE AREA UTILIZATION FOR SYNCHRONOUS CIRCUITS

Synchronous Circuits	Resources	TMR-D1	TMR-D2	TMR-D3	TMR-D4	DMR
Data Flip-flop	LUT's	25	31	25	25	22
	GC	62	65	64	62	58
Shift Register	LUT's	25	25	25	25	21
	GC	59	62	61	59	56
Counter	LUT's	41	63	41	41	54
	GC	65	68	67	65	60
ROM	LUT's	28	28	28	28	22
	GC	65	68	67	65	52

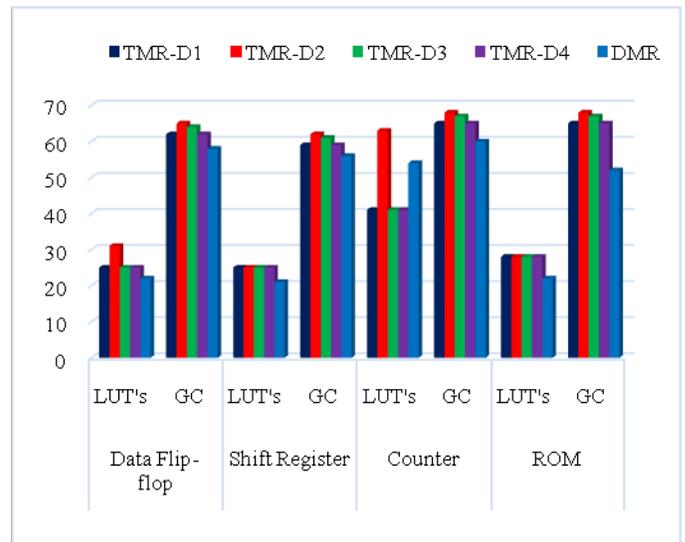


Fig. 7. Graphical Representation of FIS-FTS Module Area utilization for Synchronous Circuits.

The TMR Based FIS-FTS module for Synchronous Circuits and Asynchronous Circuits works at 606.50MHz, and DMR based FIS-FTS module works at 523.80MHz. The DMR Based FIS-FTS design utilizes less chip area than TMR based FIS-FTS designs on Artix-7 FPGA. For both Synchronous Circuits and Asynchronous Circuits.

The Power utilization of digital circuits (Synchronous Circuits and Asynchronous Circuits) in the FIS-FTS module is represented in Fig. 8. The DMR based FIS-FTS utilizes less total power (W) than the TMR based FIS-FTS module. The Xilinx –Xpower analyzer is used for power calculation at 100MHz clock frequency. The TMR Design-2 Approach utilizes more power than all other approaches in the FIS-FTS module. The ROM circuits use an average of 0.086W for all the FTS based design approaches. The power utilization of all digital circuits in the FIS-FTS module is in the range of 0.086 to 0.092W, which is quite less and suitable real-time FTS based digital circuits.

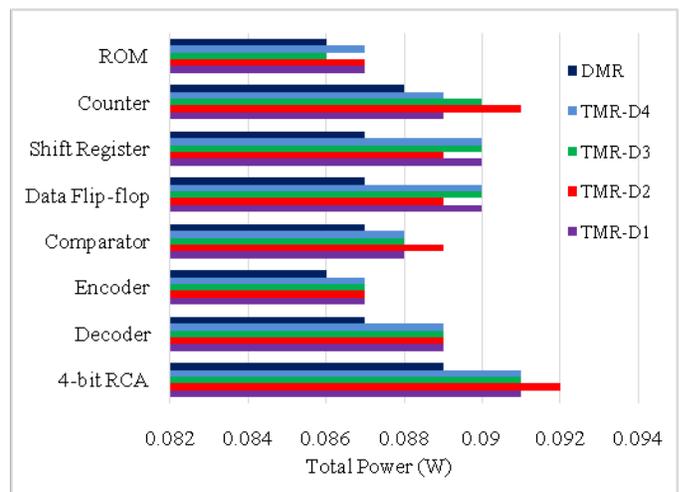


Fig. 8. Total Power utilization of Digital Circuits in FIS-FTS Module using different Voter Logic.

**B. Performance Analysis**

The performance analysis is evaluated using % fault coverage of digital circuits using the FIS-FTS module. The % fault coverage is calculated using the below formula:

$$\% \text{ fault coverage} = \frac{\text{Number of Faults tolerated}}{\text{Total number of faults injected}} \times 100$$

The Number of Transient Faults injected to the FIS-FTS module for all digital circuits is 100. The calculation of tolerated faults or fault coverage is calculated by analyzing the simulation waveform results. Based on simulation analysis for all the digital circuits in FIS-FTS modulation, the number of tolerated faults is calculated and is tabulated in Table III.

The TMR-Design-2 method in the FIS-FTS module tolerates 98 in RCA, 95 in the encoder, 94 in Data Flip-flop, and 92 faults in the counter. Similarly, DMR based FIS-FTS module tolerate 94 faults in data-FFs and 94 faults in the shift register. The TMR based Design-1, 3, and 4 methods are tolerated with 100% fault coverage for all the digital circuits. The TMR Design-2 and DMR based FIS-FTS are tolerated around with fault coverage of 96.42% and 98.5%, respectively.

**C. Comparative Results**

The comparative results are analyzed for Full adder circuits using different FTS methods [22] with the proposed TMR-D3 and DMR methods, tabulated in Table IV. The existing FTS methods and present works are synthesized on the same Spartan-3 FPGA. The Existing TMR and Novel Fault-tolerant (NFT) FTS methods use 5 slices, whereas TMR-D3 uses only 1 slice, and DMR uses 2 slices. The combinational delay (ns) of TMR and NFT applies 13.36ns and 9.97ns, respectively. The proposed TMR-D3 uses a delay of 7.824ns, and DMR uses 8.262ns. The Total power utilization of TMR and NFT consumes 0.135W and 0.129W, respectively. The proposed work TMR-D3 uses total power of 0.28W, and DMR consumes 0.3W. The proposed FTS methods (TMR-D3 and DMR) are better Chip areas (Slices and LUT's), Combinational delay, and Total power than the existing FTS approaches [22].

TABLE III. % FAULT COVERAGE USING FIS-FTS MODULE FOR DIGITAL CIRCUITS

Digital Circuits	TMR-D1	TMR-D2	TMR-D3	TMR-D4	DMR
4-bit RCA	100	98	100	100	100
Decoder	100	100	100	100	100
Encoder	100	95	100	100	100
Comparator	100	100	100	100	100
D-FF	100	94	100	100	94
Shift Register	100	100	100	100	94
Counter	100	88	100	100	100
ROM	100	100	100	100	100
<b>% Fault Coverage</b>	<b>100</b>	<b>96.42</b>	<b>100</b>	<b>100</b>	<b>98.5</b>

TABLE IV. COMPARATIVE RESULTS FOR FULL ADDER USING DIFFERENT FTS METHODS

FTS Method	Slices	LUT's	Delay (ns)	Power (W)
TMR [22]	5	9	13.364	0.135
NFT [22]	5	8	9.977	0.129
TMR-D3 (This work)	1	5	7.824	0.28
DMR (This work)	2	4	8.262	0.3

**V. CONCLUSION AND FUTURE WORK**

In this Manuscript, An efficient Fault-Injection system (FIS), followed by a Fault-Tolerant System (FTS), is designed for Synchronous Circuits and Asynchronous Circuits. The FTS is designed using TMR and DMR methods. The TMR method works based on the Majority voter logic (MVL), and DMR based self-voter logic (SVL). The TMR based four different MVL approaches are designed and analyzed for both Synchronous Circuits and Asynchronous Circuits. The TMR Based Design-1, Design-3, and Design -4 are better than design -2 by concerning the hardware constraints. The DMR based FIS-FTS has better Area (LUTs), Gate-Count, and power utilization than four-TMR based FIS -FTS in both Synchronous Circuits and Asynchronous Circuits. The TMR Based FIS-FTS using Design-1, Design-3, Design-4 approaches are 100% fault-tolerant, while design-2 has 96.42% fault-tolerant, and DMR Based FIS-FTS has 98.5% fault-tolerant for both Synchronous Circuits and Asynchronous Circuits. In the future, the FIS-FTS module is incorporated in Commination and networking applications. The proposed FTS methods are compared with the existing FTS approaches for the Full adder circuit with better resource utilization.

REFERENCES

- [1] Eslami, Mohammad, Behnam Ghavami, Mohsen Raji, and Ali Mahani. "A survey on fault injection methods of digital integrated circuits." Integration (2019).
- [2] Kooli, Maha, and Giorgio Di Natale. "A survey on simulation-based fault injection tools for complex systems." In 2014 9th IEEE International Conference On Design & Technology of Integrated Systems In Nanoscale Era (DTIS), pp. 1-6. IEEE, 2014.
- [3] Kumari, Priti, and Parmeet Kaur. "A survey of fault tolerance in cloud computing." Journal of King Saud University-Computer and Information Sciences (2018).
- [4] Ledmi, Abdeldjalil, Hakim Bendjenna, and Sofiane Mounine Hemam. "Fault Tolerance in Distributed Systems: A Survey." In 2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS), pp. 1-5. IEEE, 2018.
- [5] Singh, Shefali, and Inderveer Chana. "Survey on Fault-Tolerant Techniques in Scientific Applications." In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 618-624. IEEE, 2018.
- [6] Balach, J., & Novak, O. (2010, September). Reconfigurable Fault-Tolerant System Synchronization. In 2010 13th Euromicro Conference on Digital System Design: Architectures, Methods, and Tools (pp. 817-820). IEEE.
- [7] Ban, Tian, and Gutemberg GS Junior. "Critical gates identification for fault-tolerant design in math circuits." Journal of Electrical and Computer Engineering 2017 (2017).

- [8] Hamilton, Nicolas, Scott Graham, Timothy Carbino, James Petrosky, and Addison Betances. "Adaptive-Hybrid Redundancy with Error Injection." *Electronics* 8, no. 11, (2019): 1266.
- [9] Agiakatsikas, Dimitris, Ganghee Lee, Thomas Mitchell, Ediz Cetin, and Oliver Diessel. "From C to Fault-Tolerant FPGA-Based Systems." In 2018 IEEE 26th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), pp. 212-212. IEEE, 2018.
- [10] Lu, Wentao, Ruping Wang, Chenhui Zeng, Chenyan Liu, and Xin Wang. "A General Fault Injection Method Based on JTAG." In 2018 Prognostics and System Health Management Conference (PHM-Chongqing), pp. 604-608. IEEE, 2018.
- [11] Balasubramanian, P., and R. T. Naayagi. "Redundant logic insertion and fault tolerance improvement in combinational circuits." In 2017 International Conference on Circuits, System and Simulation (ICCSS), pp. 6-13. IEEE, 2017.
- [12] Schweizer, Thomas, Dustin Peterson, Johannes M. Kühn, Tommy Kuhn, and Wolfgang Rosenstiel. "A fast and accurate FPGA-based fault injection system." In 2013 IEEE 21st Annual International Symposium on Field-Programmable Custom Computing Machines, pp. 236-236. IEEE, 2013.
- [13] Li, Yan, Yufeng Li, Han Jie, Jianhao Hu, Fan Yang, Xuan Zeng, Bruce Cockburn, and Jie Chen. "feedback-based low-power soft-error-tolerant design for dual-modular redundancy." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 26, no. 8 (2018): 1585-1589.
- [14] Sheikh, Ahmad T., and Aiman H. El-Maleh. "Double Modular Redundancy (DMR) Based Fault Tolerance Technique for Combinational Circuits." *Journal of Circuits, Systems and Computers* 27, no. 06, (2018): 1850097.
- [15] Arifeen, Tooba, Abdus Sami Hassan, and Jeong-A. Lee. "A Fault-Tolerant Voter for Approximate Triple Modular Redundancy." *Electronics* 8, no. 3 (2019): 332.
- [16] Mandaogade, Nitin Nemikumar, and Prashant V. Ingole. "Development of Fault Diagnosis System Using Fault Injection for Discrete Architectures." In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), pp. 1-6. IEEE, 2018.
- [17] Oliveira, Andre BM, Robson L. Moreno, and Enio R. Ribeiro. "Digital short-circuit diagnosis and reconfiguration of a fault-tolerant VSI." In 2017 IEEE Southern Power Electronics Conference (SPEC), pp. 1-6. IEEE, 2017.
- [18] Chekmarev, S. A., V. Kh Khanov, and O. A. Antamoshkin. "Modification of fault injection method via on-chip debugging for processor cores of systems-on-chip." In 2015 International Siberian Conference on Control and Communications (SibCon), pp. 1-4. IEEE, 2015.
- [19] Coelho, Alexandre, Nacer-Eddine Zergainoh, and Raoul Velazco. "NoCFI: A Hybrid Fault Injection Method for Networks-On-Chip." In 2019 IEEE Latin American Test Symposium (LATS), pp. 1-6. IEEE, 2019.
- [20] Guruprasad S.P and Chandrasekar B.S, "Design of an Efficient Fault and Congestion Free NoC Design using Adaptive Routing on FPGA." *International Journal of Recent Technology and Engineering (IJRTE)*, Vol.8, Issue 3, pp.1055-1062, September 2019.
- [21] Podivinsky, Jakub, Jakub Lojda, and Zdenek Kotasek. "An Experimental Evaluation of Fault-Tolerant FPGA-Based Robot Controller." In 2018 IEEE East-West Design & Test Symposium (EWDTS), pp. 1-7. IEEE, 2018.
- [22] Kshirsagar, R. V., and R. M. Patrikar. "A novel fault-tolerant design and an algorithm for tolerating faults in digital circuits." In 2008 3rd International Design and Test Workshop, pp. 148-153. IEEE, 2008.

# Evaluation of the Diffusion Phenomenon using Information from Twitter

Kohei Otake<sup>1</sup>

School of Information and Telecommunication Engineering  
Tokai University, Tokyo, Japan

Takashi Namatame<sup>2</sup>

Faculty of Science and Engineering  
Chuo University, Tokyo, Japan

**Abstract**—Social media services, including social networking services (SNSs) and microblogging services, are gaining prominence. SNSs have a variety of information on products and services, such as product introductions, utilization methods, and reviews. It is important for companies to utilize SNSs to understand the various ways of engaging with them. Against this backdrop, numerous studies have focused on marketing activities (e.g., consumer behavior and sales promotion) using information on the internet from sources such as SNSs, blogs, and news sites. In particular, to understand the dissemination of information on the Internet, various researchers have undertaken studies pertaining to the diffusion phenomenon occurring in the real world. Here, topic diffusion is a phenomenon whereby a certain topic is shared with several other users. In this study, we aimed to evaluate the diffusion phenomenon on Twitter. In particular, we focused on the state of a targeted topic and analyzed the estimation of the topic using natural language processing (NLP) and time series analysis. First, we collected tweets containing four titles of animation broadcasts using hashtags. Approximately 250,000 tweets were posted on Twitter in a month. Second, we used NLP methods such as morphological analysis and N-gram analysis to characterize the contents of each title. Third, using the time series data for the tweets, we created a mixture model that replicated the diffusion phenomenon. We clustered the diffusion phenomenon using this model. Finally, we combined the features related to the content of the tweets and the results of the clustering of the diffusion phenomenon and evaluated them.

**Keywords**—Twitter; diffusion phenomenon; natural language processing; mixture model

## I. INTRODUCTION

In recent years, social media has become highly prominent. Social media is the collection of online communication channels with community-based input, content sharing, interaction, and collaboration [1]. Websites and smartphone applications dedicated to forums, social networking, social bookmarking, and wikis are some of the different types of social media.

In particular, the number of users of social networking services (SNSs), such as Twitter and Facebook, is on the rise. These users can share a variety of information, such as their preferences and favorites related to services, products and so on, with their friends through SNSs. In such a scenario, marketing campaigns using SNSs have received increasing attention from electronic commerce (EC) suppliers. SNSs have a wide range of information about products, such as product introductions, utilization methods, and reviews. In addition,

these pieces of information are posted by the consumers themselves, a phenomenon that did not happen in the past. In general, the information available on SNSs has several features. The representative features are “wide reach” and “high-speed communication”. Using information on SNSs effectively is important for EC suppliers.

Consequently, there have been numerous studies that focused on marketing activities (e.g., consumer behavior and sales promotion) using information from sources such as SNSs, blogs, and news sites [2][3][4]. Several studies have focused on how tweets are shared with users.

To understand the dissemination of information on the Internet, various researchers have conducted studies pertaining to the diffusion phenomenon occurring in the real world. Here, topic diffusion is a phenomenon whereby a certain topic is shared with several other users.

Mane and Borner [5] proposed maps that supported the identification of major research topics and trends through the analysis of a complete set of papers published in Proceedings of the National Academy of Sciences of the United States of America between 1982 and 2001. The authors demonstrated the utilization of Kleinberg’s burst detection algorithm [6], word co-occurrence analysis, and graph layout techniques. In addition to this study, Takahashi et al. [7] proposed a method for measuring bursts of topics estimated by a topic model. The authors analyzed two ways to model information flow in news streams, namely, Kleinberg’s burst modeling and topic modeling such as the dynamic topic model.

Dipak and Bijith [8] proposed a model that predicted movie sales using word-of-mouth publicity by Twitter users. Specifically, the authors classified movie reviews into four types (strongly negative, negative, positive, and strongly positive), and created a model that predicted the intention to watch a movie. As a result, the authors suggested that utilization of online reviews on Twitter will help the movie industry’s marketing strategies.

Matsuzawa et al. [9] proposed a method for analyzing the statistical characteristics of a time series of retweets extracted from the actual logs of tweets. Specifically, the authors suggested a burst detection model, assuming a lognormal distribution based on the results of Sartwell [10] using statistical data about infectious diseases. The time series clustering classification caused most of the retweets to extend over a short time scale of approximately one day.

Ueda and Asahi [11] proposed a model that replicated a sudden increase (boom) in interest among Twitter users. Specifically, the authors created a model that had two features: reflection of the behavior of social media users (including Twitter users) and estimation of the potential interests of the users. Based on these features, the authors analyzed the factors of a boom in interest. From the results of the analysis, the authors evaluated the differences between transient and secondary booms.

From these studies, it was clear that understanding the diffusion phenomenon is very important for understanding social trends. To understand social trends, we believe that it is necessary to use the information from SNSs properly.

The remainder of this paper is organized as follows. Section 2 describes the purpose of this study. Section 3 presents the dataset used in this study. Section 4 discusses the characterization analysis used to target the content of tweets using natural language processing (NLP). Section 5 describes the modeling of the diffusion phenomenon using time series data and presents an evaluation of the diffusion phenomenon based on the results mentioned in Sections 3 and 4. Finally, Section 6 summarizes the paper and discusses future work.

## II. PURPOSE OF THE STUDY

In this study, we attempted to evaluate the diffusion phenomenon on Twitter. We examined the content of tweets using NLP. In addition, based on the results of previous studies, we tried to analyze the tweets using time series data. In particular, we analyzed the diffusion phenomenon by building a mixed normal distribution model using retweeted data, and then evaluated the results. For the analysis, we targeted animations in Japan. Animation information is shared widely across SNSs. Therefore, it was considered appropriate for the analysis.

First, we used posted tweets as data and attempted to characterize the content of the tweeted data by using NLP. From this analysis, we extracted characteristic expressions (characteristic words) included in the tweets for each target animation title. Second, using the time series of the tweeted data, we built a model that reflected the diffusion phenomenon. We also performed clustering of the diffusion phenomenon using the model. Finally, we combined the features related to the tweeted content with the results of the diffusion phenomenon clustering, and then evaluated it.

## III. DATASET

First, we collected tweets pertaining to Japanese animations. In this study, we focused on four animation titles based on broadcasting time and evaluation by ranking site. We collected tweets posted over the course of one month for these four titles. We used hashtags and keywords (animated titles) and collected data using the Twitter application programming interface. Consequently, we collected approximately 250,000 tweets. In addition, at the time the tweets were collected, we also acquired information about the users, such as their number of follows, followers, favorites, and retweets. Table I summarizes the acquired tweeted data, and Table II presents a summary of the contents of each animation title and the number of tweets posted in a month. Data were selected from these datasets and analyzed.

TABLE I. SUMMARY OF THE ACQUIRED TWEETED DATA

Number of tweets and retweets	245,146
Number of unique IDs	224,967
Number of unique tweets (except retweets)	115,104

TABLE II. CONTENT SUMMARY OF EACH ANIMATION TITLE AND NUMBER OF TWEETS

Title	Content summary	No. of tweets and retweets
A	A is an animation created by mixed media of game maker and production. A performs various content development activities such as animation, game applications, and card games. Category: idol group (male); Main target: adult women	48,728
B	B is an animation based on the fifth work of a series of games. Category: serious fantasy; Main target: youth	43,690
C	C is an animation based on a toy project and is also serialized over the same period in magazines for elementary school students. Category: comedy; Main target: children	31,190
D	D is a cartoon from a popular magazine for youth, presented in original animation. Category: serious fantasy; Main target: youth	121,541

## IV. CHARACTERIZATION ANALYSIS TARGETING TWEET CONTENT USING NLP

In this section, we discuss the NLP analysis performed to characterize the content of the tweets. Specifically, the analysis was conducted using the following procedure. From Table II, it can be seen that the number of tweets is highly skewed for each title. In this case, the number of tweets for a title will affect the characteristic extraction. We deemed it necessary to adjust the number of tweets to target to acquire the characteristic amount of the number of tweets more accurately by using NLP. Therefore, we extracted 27,000 tweets per title by random sampling, and we adjusted the number of tweets. Here, 27,000 is the maximum number of tweets that could be acquired for each animation title while excluding duplicate tweets.

- 1) Parts of speech decomposition based on morphological analysis.
- 2) Word weighting by *tfidf* value.
- 3) Calculation of co-occurrence frequency by N-gram (we use the bigram model).

We first decomposed the tweets by morphological analysis into units according to parts of speech. Morphological analysis was used to divide natural language (text data) into columns of morphemes (minimum elements that constitute sentences) and to discriminate parts of speech and the like for each morpheme. Information such as parts of speech for words defined grammatically and in the dictionary, was used for morphological analysis. In this study, a morphological analysis was performed using the R language. In addition, Mecab [12] was used as a Japanese morpheme dictionary. Table III summarizes the frequency of appearance for the parts of speech obtained by the morphological analysis of each title.

Nouns and proper nouns appear frequently in all titles, as seen in Table III. Therefore, in the analysis, we focused on this category.

Next, using the results of the morphological analysis, we calculated the importance of words using the *tfidf* method. The *tfidf* method is a type of index for word weighting, which can be calculated by the product of *tf* (term frequency) and *idf* (inverse document frequency) [13].

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

$$idf_i = \log \frac{|D|}{|\{d:d \ni t_i\}|} \quad (2)$$

$$tfidf_i = tf_{i,j} \times idf_{i,j} \quad (3)$$

where  $n_{i,j}$  is the occurrence frequency of word  $t_i$  in sentence  $d_j$ ,  $\sum_k n_{k,j}$  is the summation of the count of all the words in sentence  $d_j$ ,  $|D|$  is the total number of documents, and  $|\{d:d \ni t_i\}|$  is the number of documents that contain word  $t_i$ . Here, the sentence is one tweet, and the document is the tweet group of each animation title.

Here, *tf* is considered to have a higher degree of importance as the number of occurrences in a sentence becomes larger. In addition, according to *idf*, words that are used across several documents are not important. The reason for calculating the importance of words is that the degree of expression of a characteristic differs based on the title, even though it is the same word. We believe that it is possible to obtain a summary of the title by calculating the importance of words. In addition, in this study, the tweet group of each title was created as a separate document. We calculated the importance of words across the four documents.

Based on the results, words with the top 100 *tfidf* values were selected as characteristic words. We investigated the characteristic words of each title. In addition, to consider the relationship among the characteristic words, we calculated the co-occurrence frequency using the N-gram method. The N-gram is a type of language model that investigates how often N character strings or word combinations appear in a certain character string. In this study, we evaluated the co-occurrence relation with N equal to 2 (a bigram) [14].

Table IV summarizes the analysis results of *tfidf* and the N-gram method for each title.

From the above results, it is clear that the words that have high values of importance differ depending on the title. In addition, some of the features of each title were identified. For example, for title A, the importance of Animation is the lowest compared with other titles. We believe that this is because of the variety of related product development in addition to animation for A. Conversely, for title C, Animation has a higher importance when compared with other titles. C is a toy project, but television broadcasting is its main field of activity. D is a cartoon from a popular magazine for youth in original animation. Therefore, it can be inferred from the appearance of the expected value for the production company that the importance of the production company is high.

TABLE III. FREQUENCY OF APPEARANCE OF PARTS OF SPEECH UNIT

Part of Speech \ Title	A	B	C	D
	Noun, proper noun	651,973	779,838	567,162
Verb	100,167	61,122	108,464	57,962
Symbol	127,261	128,996	96,528	87,153
Adjective	9,468	6,896	17,306	8,113
Total	1,128,610	1,174,366	1,082,699	1,143,288

TABLE IV. SUMMARY OF THE ANALYSIS RESULTS OF TFIDF AND N-GRAM FOR EACH TITLE

Title	Features obtained from NLP results
A	The importance of a word in the content of the card game is high. The importance of Animation is the lowest in comparison with importance of other titles. Words that suggest a product (e.g., Badge, Ice Cream, Commission, Price, Lottery, Cards) have high importance. Similar animation titles have high importance.
B	Words that suggest products have high importance. In particular, Releases, Distribution, Restrictions, and Quantities have high importance in comparison with other titles. The importance of a word with the character name and broadcast contents of animation are high.
C	The importance of Animation is highest in comparison with importance of other titles. The importance of a word containing a character name and broadcast contents of animation is high. Similar animation titles and mixed media product have high importance.
D	The importance of words related to broadcast content of animation is high. The importance of the word that is a character name and Production Company of animation is high. Similar animation titles have high importance.

From the overall trends, it is conceivable that a tweet composed of the following elements represents well the characteristics of the animation:

- Participants in the content: character, producer, production company, and related animation.
- Media mix or related product: product name, product category, and other media (events, radio, and books).
- Review of the animation broadcast: impression, criticism, and next thoughts.

## V. MODELING OF THE DIFFUSION PHENOMENON USING TIME SERIES DATA

In this section, we discuss the construction of a model of the diffusion phenomenon using time series data. Specifically, we focus on the phenomenon of how a tweet spreads across users by being retweeted. We also conducted a clustering of the diffusion phenomenon using the model.

### A. Summary of Data

First, we performed a basic aggregation and estimated the trend. Of the 115,104 unique tweets that were collected, we used 9,378 tweets that had been shared with (retweeted to) others at least once. Fig. 1 shows the distribution of the number of shared tweets. The vertical axis of the figure represents the frequency, and the horizontal axis represents the number of retweets. The maximum number of retweets is 12,317. In

Fig. 1 shows the distribution of the number of tweets according to the number of retweets (up to 40). According to Fig. 1, approximately 40% of tweets were retweeted only once, and the distribution has a very long tail.

There are various definitions of the diffusion phenomenon, but in this study, we focus on 11 original tweets for which the number of retweets is more than 1,000. Tweets with IDs 2, 5 and IDs 10, 11 are transmitted by the same user. Table V shows the number of follows and followers, the targeted title of the animation, and the user attributes for 11 tweets. Based on an investigation of the attributes of the user who posted the 11 tweets, 6 tweets (from 4 accounts) were posted by the official account of a company and the other tweets were posted by personal accounts.

Initially, we aggregated and visualized the transition of the number of retweets for the 11 targeted tweets. We calculated the width of time point  $t$  using the following three approaches:

- 1) Changes in the number of retweets per day from the day of the tweet until the end of the collection period (2016/10/2).
- 2) Changes in the number of retweets per hour from the time of the tweet until the end of the collection period (2016/10/2).
- 3) Changes in the number of retweets per hour that were made within 48 hours of the tweet.

Based on the results of 1, it was clear that the changes in the number of retweets per day did not differ between general and official accounts. We tabulated the change in retweets per hour for each tweet. We investigated the rate of the number of retweets and found that about 95% of the retweets were posted within 72 h (92% of retweets were posted within 48 hours) of the original tweet. From the above results, it can be inferred that the remaining period was approximately 2–3 days when the tweets are spreading.

Based on the results of 2, that is, the changes in the number of retweets per hour, we found multiple peaks. Specifically, there were tweets that peaked quickly after the tweet was posted, and some that reached a peak after some time elapsed. As a whole trend toward zero within 48 h, but the occurrence of the time of the peak differed depending on the tweets themselves.

*B. Analysis of the Diffusion Phenomenon Model*

From the basic analysis, it can be inferred that the state of diffusion is not uniform. In this section, we describe the application of a mixed normal distribution to the time series change in retweets. In a previous study, burst detection used a log normal distribution was performed [9]. However, in this study, a phenomenon occurred whereby the retweeting began a considerable time after the original tweet. In addition, the purpose of this analysis was to evaluate the number of peaks. Consequently, we used a normal distribution in this study. For validation, we also performed burst detection using a log-line

normal distribution. It was found that the BIC (Bayesian Information Criterion) [15] was almost the same. When a log normal distribution was used, it was necessary to estimate the number of peaks after estimating the bursting start point. Therefore, a two-step estimation was necessary. For these reasons, we attempted to estimate the number of peaks using a mixed normal distribution in this study.

The number of tweets was counted every hour.

$$f(t) = \sum_{i=1}^n \pi_i f_i(t) \tag{4}$$

$$f_i(t) \sim N(\mu_i, \sigma_i^2) \tag{5}$$

Here,  $\pi_i$  is the composition ratio of the number of  $i$ . In addition, although the data were collected over 400 hour or more, only the data after the 60 hour following the tweets were used for analysis. The EM (Expectation–Maximization) algorithm [16] was used for parameter estimation. In addition, the number of peaks (double or triple) were selected based on the BIC.

In the estimation results, single and double peaks were observed. The number of peaks for each tweet is described in the next section.

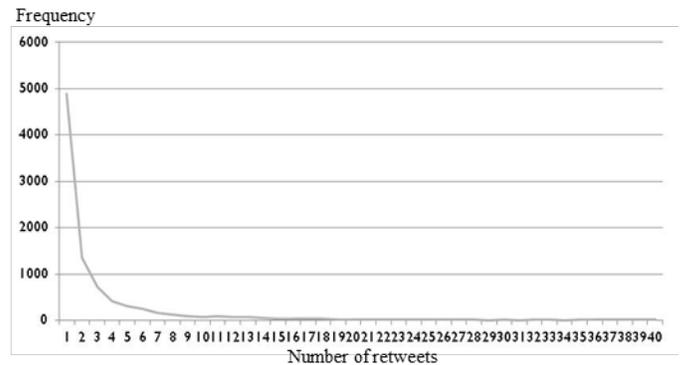


Fig. 1. Distribution of the Number of Shared Tweets (All Tweets).

TABLE V. DISTRIBUTION OF THE NUMBER OF SHARED TWEETS (ALL TWEETS)

User ID	Tweet ID	No. of Follows	No. of Followers	Target Title	User Type
a	1	396	1,637	B	General
b	2,5	0	117,275	D	Official
c	3	310	3,787	A	General
d	4	229	211	A	General
e	6	29,198	305,188	B	Official
f	7	287	2,808	C	General
g	8	182	624,503	D	General
h	9	36,588	566,091	B	Official
i	10,11	9	129,605	B	Official

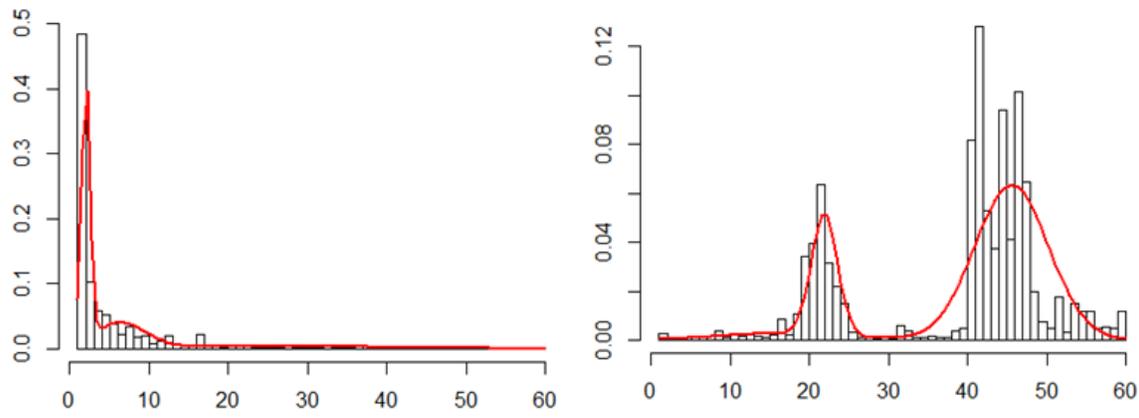


Fig. 2. Example of the Time Trends of Retweets and their Estimation Graphs. Left: “Transient Topic” (ID 13), Right: “Secondary Spread Occurred Topic” (ID 4).

### VI. EVALUATION OF DIFFUSION PHENOMENON

Using the results from Sections IV (NLP analysis) and V (modeling using time series data), we attempted to evaluate the diffusion phenomenon. First, we classified each tweet as a “transient topic” or a “topic with secondary spread”. Here, a “transient topic” refers to a topic that peaked soon after being posted, and then decreased. The peak lasted for approximately 1–3 hour (maximum 6 hour). In contrast, a “topic with secondary spread” refers to a topic where the diffusion phenomenon occurred again, only after the first peak had passed. For the classification, we used the mixture ratio of the estimated results that fit with the mixed normal distribution, and the shape of the established density when it fit with the mixed normal distribution. Fig. 2 shows an example of a “transient topic” and “topic with secondary spread”. Comparing both graphs, it is clear that the states of the diffusion phenomenon are different.

Next, we calculated the characteristic value ( $c_j$ ) for each tweet to evaluate how characteristically the title was expressed. First, we obtained the *tfidf* value ( $w_{i,j}$ ) for word  $i$  of tweet  $j$ . To evaluate that tweet  $j$  expresses the characteristic, the characteristic value ( $c_j$ ) is given by equation (6).

$$c_j = \sum_{i \in d_j} w_{i,j} \tag{6}$$

For the calculation of the characteristic value, we used the *tfidf* value obtained from the document word matrix for each animation title (Section IV). The classification of the diffusion phenomenon and the relative ranking of the characteristic value of the tweets are shown in Table VI.

In Table VI, there is no noticeable relationship between the number of peaks and the ranking of the characteristic values. Therefore, it can be inferred that a tweet has no distinctive characteristic (whether it expresses content) in terms of the presence or absence of reinfection. Furthermore, it can be stated that most of the information transmitted through the official account was classified as “transient topic”. Conversely, approximately 60% of the tweets of general accounts were “topic with secondary spread”.

Next, the specific content of each tweet was described, as summarized in Table VII.

TABLE VI. SUMMARY OF THE STATE OF THE DIFFUSION PHENOMENON AND RELATIVE RANKING OF THE CHARACTERISTIC VALUE

Tweet ID	User type	Title	No. of peaks	Ranking of characteristic value
1	General	B	2	6
2	Official	D	1	4
3	General	A	1	10
4	General	A	2	5
5	Official	D	1	3
6	Official	B	1	11
7	General	C	2	1
8	General	D	1	2
9	Official	B	2	7
10	Official	B	1	9
11	Official	B	1	8

TABLE VII. SUMMARY OF THE CONTENTS OF TWEETS

Tweet ID	Summary of the contents
1	Reviews and impressions
2	New announcement
3	Reviews and impressions
4	Reviews and impressions
5	Event public relations
6	New announcement
7	Reviews and impressions
8	New announcement
9	Mix Media news
10	New announcement
11	Event public relations

In terms of the content of tweets, “reviews and impressions” and “event public relations” had high characteristic values. Thus, the information posted by the official account was found to be related, in terms of content, to

new product announcements such as new animation, information about characters, goods, and mixed media. Generally, information on animation was posted from time to time in official accounts on Twitter and was used as advertising media. When an announcement is made, it can be inferred that it will spread among anime fans, and the diffusion will dissipate over time. Therefore, in the case of official accounts, it can be inferred that there are many “transient topics”.

The tweets from general accounts were reviews conveying the impression made by animation broadcasting, characters, goods, and voice actors/actresses. As mentioned above, 60% of the general account tweets had two peaks. Consequently, further analysis is necessary, but if animation broadcasting and posting reviews and impressions are supported for Twitter users, it can be inferred that subsequently, trends such as continuous spreading until the next broadcast can be evaluated. In addition, according to the next broadcast, past posts are reevaluated and diffused. Based on the results of the analysis, we can evaluate the tendency of the diffusion phenomenon.

## VII. CONCLUSION AND FUTURE WORK

In this study, we attempted to evaluate the diffusion phenomenon on Twitter. We focused on the content of tweets and the diffusion phenomenon. Specifically, we analyzed the content of tweets using NLP. We also analyzed the diffusion phenomenon by generating a mixed normal distribution model using retweeted data, and we evaluated it using the results of the analysis. On the other hand, we modeled and evaluated the diffusion process of specific topics in this study. From the perspective of social marketing, there are also assessment of business relationships in companies [17][18][19] and research focusing on the rise of topics such as flaming phenomena [20]. It is also our important research theme to consider these phenomena based on the diffusion process of this research.

In the future, it will be necessary to increase the number of cases we study. In this study, we targeted animation. However, we are planning to investigate other content in future studies. We intend to study not only the normal distribution but also other distributions for modeling the diffusion phenomenon.

## ACKNOWLEDGMENT

We thank Rooter Inc. for permission to use valuable datasets and for useful comments. This work was supported by JSPS KAKENHI Grant Number 19K01945 and 17K13809.

## REFERENCES

[1] WHITE PAPER Information and Communications in Japan, (2018).  
[2] E. Ioană and I. Stoica, “Social Media and its Impact on Consumers Behavior,” *International Journal of Economic Practices and Theories*, Vol. 4, No. 2, pp. 295-303 (2013).

[3] D. Boyd, S. Golder and G. Lotan, “Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter,” *Proceedings of the 43rd Hawaii International Conference on System Sciences*, pp. 1-10 (2010).  
[4] B. J. Jansen, M. Zhang, K. Sobel and A. Chowdury, “Twitter Power: Tweets as Electronic Word of Mouth,” *Journal of the American Society for Information Science and Technology*, Vol. 60, pp. 2169-2188 (2009).  
[5] K. Mane and K. Borner, “Mapping Topics and Topic Bursts in PNAS,” *Proceedings of the National Academy of Sciences of United States of America*, Vol. 101, Suppl. 1, pp. 5287-5290 (2004).  
[6] J. Kleinberg, “Bursty and Hierarchical Structure in Streams,” *Proceedings of 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 91-101 (2002).  
[7] Y. Takahashi, D. Yokomoto, T. Utshuro and M. Yoshioka, “Analyzing Burst of Topics in News Stream,” *The Special Interest Group Technical Reports of Information Processing Society of Japan*, Vol. 204, No. 6, pp. 1-6 (2011). (in Japanese).  
[8] D. Gaikar and B. Marakarkandy, “Product Sales Prediction Based on Sentiment Analysis Using Twitter Data,” *International Journal of Computer Science and Information Technologies*, Vol. 6, No. 3, pp. 2303-2313 (2015).  
[9] Y. Matsuzawa, S. Saeyor Santi, F. Toriumi Y. Chen and H. Ohashi, “Analysis of Retweets Time Series by Mixture Model of Three-Parameter Lognormal Distribution,” *Proceedings of the 27th Annual Conference of the Japanese Society for Artificial Intelligence*, pp. 1-4 (2013). (in Japanese).  
[10] P. E. Sartwell, “The Distribution of Incubation Periods of Infectious Diseases,” *American Journal of Hygiene*, Vol. 51, pp. 310-318 (1950).  
[11] Y. Ueda and Y. Asahi, “Construction and Verification of Transition Model of Interest of Twitter Users,” *Communications of the Operations Research Society of Japan*, Vol. 59, No. 4, pp. 219-228 (2014). (in Japanese).  
[12] MeCab, <http://taku910.github.io/mecab/> (2017/11/23, author checked).  
[13] R. A. Baeza-Yates and B. A. Ribeiro-Neto, *Modern Information Retrieval: the Concepts and Technology behind Search* (2nd Edition), Addison-Wesley Professional (2011).  
[14] C. E. Shannon, “A Mathematical Theory of Communication,” *The Bell System Technical Journal*, Vol. 27, pp. 379-423 (1948).  
[15] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley (2007).  
[16] G. E. Schwarz, “Estimating the Dimension of a Model,” *Annals of Statistics*, Vol. 6, No. 2, pp. 461-464 (1978).  
[17] H. Tsurumi, J. Masuda and A. Nakayama, “Analysis of Relationship between Communication on Twitter about an Item and its Sales,” *Communications of Operations Research Society of Japan*, Vol. 58, No. 8, pp. 436-441 (2013). (in Japanese).  
[18] G. Mishne and N. Glance, “Predicting Movie Sales from Blogger Sentiment,” in *AAAI 2006 Spring Symposium on Computational Approaches to Analysing*, (2006).  
[19] R. Dijkman, P. Ipeirotis, F. Aertsen and R. van Helden, “Using Twitter to Predict Sales: A Case Study,” *arXiv:1503.04599* (2015).  
[20] N. Takahashi and Y. Higaki, “Flaming Detection and Analysis using Emotion analysis of Twitter,” *IEICE Technical Report*, Vol. LOIS2016-86, pp. 135-141 (2017). (in Japanese).

# Machine Learning and Statistical Modelling for Prediction of Novel COVID-19 Patients Case Study: Jordan

Ebaa Fayyoumi<sup>1\*</sup>, Sahar Idwan<sup>2</sup>, Heba AboShindi<sup>3</sup>

Department of Computer Science and Applications  
Faculty of Prince Al Hussein Bin Abdullah II for Information Technology  
The Hashemite University, P.O. Box 330127, Zarqa 13133, Jordan<sup>1,2</sup>  
Abt Associates, P.O. Box 851275, Sweifieh 11185, Amman, Jordan<sup>3</sup>

**Abstract**—As of December 2019, the world's view on life has been changed due to ongoing COVID-19 pandemic. This requires the use of all kinds of technology to help identify coronavirus patients and control the spread of this disease. In this paper, an online questionnaire was developed as a tool to collect data. This data was used as an input for various prediction models based on statistical model (Logistic Regression, LR) and machine learning model (Support Vector Machine, SVM, and Multi-Layer Perceptron, MLP). These models were utilized to predict potential patients of COVID-19 based on their signs and symptoms. The MLP has shown the best accuracy (91.62%) compared to the other models. Meanwhile, the SVM has shown the best precision 91.67%.

**Keywords**—Novel COVID-19; machine learning; logistic regression; support vector machine; multi-layer perceptron

## I. INTRODUCTION

Coronavirus disease (COVID-19) is an infectious disease caused by newly discovered coronavirus [1]. It is considered as zoonosis which is caused by microbes that are transferred between animals and people [2]. Coronavirus is mysterious since it had other previously reported versions such as SARS-CoV [3] which was transferred from cats to humans and MERS-CoV which is transported from camels to the humans [4].

COVID-19 disease was identified in December 2019 at Wuhan, the capital of China's Hubei province, and spread to the countries of the globe developing an ongoing COVID 2019-2020 pandemic [5]. Based on the World Health Organization (WHO) global COVID-19 outbreak pandemic report, there are 2,858,635 confirmed cases, and 196,295 deaths all over the world till April, 27<sup>th</sup> 2020. While, 165,379 confirmed cases in Eastern Mediterranean. In Jordan, 449 confirmed cases and 7 deaths were reported [1]. The majority people who get COVID-19 displayed mild to moderate symptoms and recovered without special treatment [1]. This virus causes a few noticeable symptoms for its patients such as coughing, high fever and pneumonia which can be utilized to detect the disease in possible patients [6].

Jordan highlighted their surveillance to prospectively have early diagnosis for new COVID-19 cases. Hence, Jordan conducted nationwide unprecedented actions on March 2020

to contain the spread of disease such as large-scale quarantine, extensive controls on travels, social distancing, continuous monitoring of suspected COVID-19 cases, and blocking areas in order to decrease the number of infected cases. Fig. 1 shows the number of COVID-19 confirmed patients in Jordan between March 3rd, 2020 and April 27th, 2020. Yet, it is uncertain whether these subsequent policies have had an impact on the containment of epidemic and what is coming in the future? Accordingly, it is crucial to examine the epidemic progression in the globe by predicting the new COVID-19 cases from the most common symptoms which could effectively control the spread of disease.

Recently, many researchers tackled COVID-19 in their research to prevent the spread of it. Naudé, W. [7] discussed several fields where Artificial Intelligence (AI) can be utilized to influence the fights against COVID-19 such as early warning and alerts, tracking and prediction, data dashboards, diagnoses and prognosis, treatments and cures and social control. He concluded that data shortage or having too much information is considered as an obstacle for using AI against COVID-19.

To prevent the spreading of the COVID-19 many predication models have been utilized by officers and leaders in different countries to generate the appropriate decisions and regulations which help to overcome this pandemic. Authors of [8] explored using of the machine learning for shaping the exponential growth of COVID-19. They concluded that Multi-Layer Perceptron and adaptive network-based fuzzy inference system can be used as a useful tool to handle the epidemic. Other researchers [9] discussed the integration of an improved mathematical modelling in machine learning with the cloud computing. This envisages the expansion of the COVID-19 which follows the exponential dispersion. Machine learning can be developed to predict the extent of COVID-19 by using the high-speed computations in the cloud computing.

Having COVID-19 pandemic in a short period of time raises the need to study the behavior of this virus and the infected patients. The researchers of [10] selected the Prophet Algorithm as the most suitable predictive algorithm. This is done based on what is the analytical question that they need to address in their study and how the predictive algorithms can be approved to get the best results [10].

\*Corresponding Author

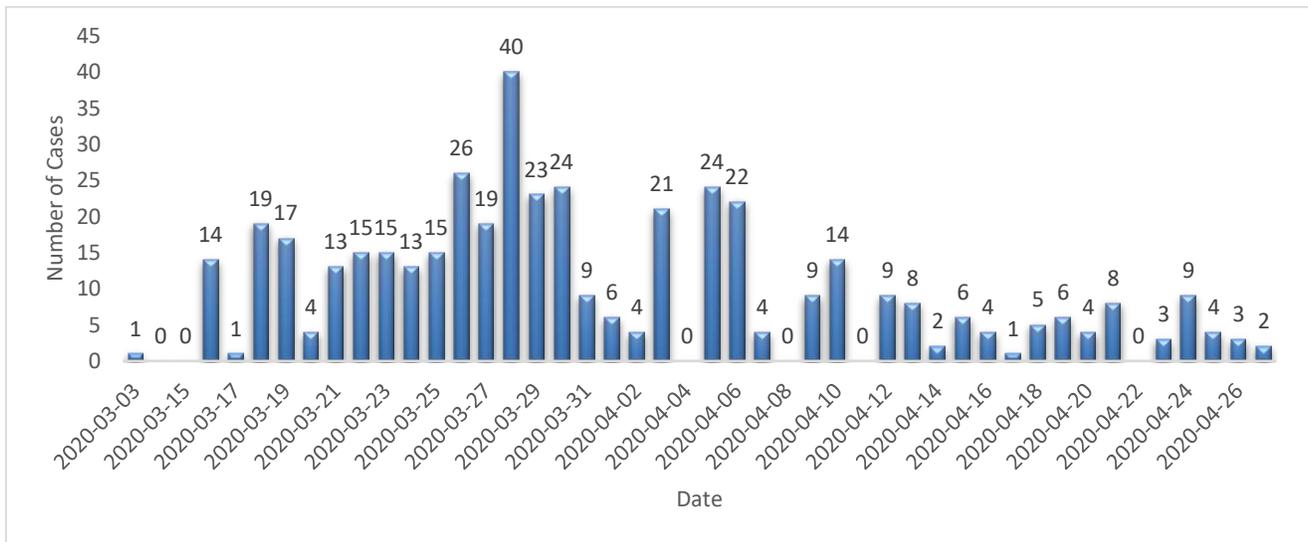


Fig. 1. New COVID-19 Cases between 3-3-2020 to 27-4-2020 in Jordan.

Yadav, D. et al. [11] investigated the foresee of the COVID-19 in different countries including United States of America. The foresee achieved by invoking the machine learning data-driven Prophet time series that analyzed the infected, active and cured cases to outburst the predication.

Large amount of the information is available in the internet and presented numerically and graphically. Thus, utilizing statistical analysis and machine learning are essential to provide better understanding of the results and generate informed decision to meet community, national and international challenges in many fields like medical, business, economics, web search engine, Facebook, spam filters and commerce [12].

To the best of our knowledge, the mathematical model and machine learning model have not been used to predict the infected cases of novel COVID-19 based on the signs and symptoms. There is an urgent need globally and in Jordan particularly to screen patients quickly due to limited availability of the Polymerise Chain Reaction tests. The main objective of this paper is to establish a reliable trusted model to predict the potential patients of COVID-19 by using either statistical or machine learning models. The following procedure had been adopted to achieve the outlined objective:

- Generating a questionnaire to collect data from individual citizens according to their health states during the last two weeks.
- Employing the statistical modelling and machine learning methods individually to assess the obtained data from the questioner in order to predict health condition for people in different cities in Jordan.
- Evaluating the performance of each model to choose the most appropriate one to our domain problem.

This paper is organized as follows: The methodology is presented in Section II, while the experimental results and discussion are reported in Section III. Finally, Section IV draws the conclusion and future works.

## II. METHODOLOGY

A cross-sectional quantitative study was conducted to build up a reliable trusted model to forecast COVID-19 diagnosis from the signs and symptoms that participants had. The signs and symptoms of novel COVID-19 used in this study were obtained from Jordan<sup>1</sup>. Our strategy includes four processing stages, namely data collection, data preprocessing, classification, and performance evaluation. The classification stage can be accomplished either by building statistical model or by invoking machine learning model. In the statistical model we used Logistic Regression (LR), while in machine learning model we invoked Support Vector Machine (SVM), and Multit-Layer Perceptron (MLP). Finally the performance of each classifier is quantified. A block diagram of the proposed work is illustrated in Fig. 2.

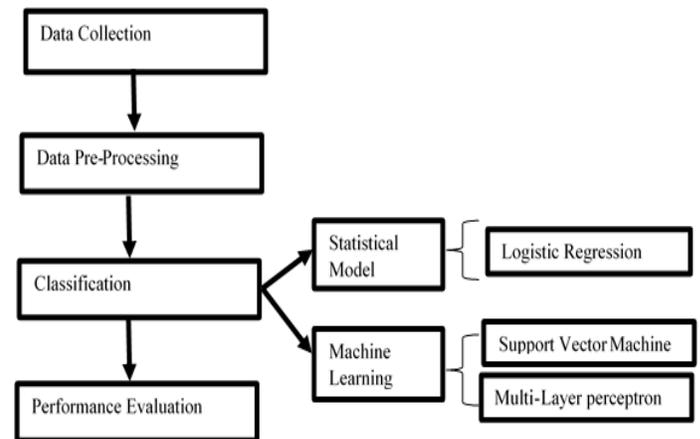


Fig. 2. Block Diagram of Novel COVID-19 Predictor Model.

<sup>1</sup>Jordan is divided geographically into three regions: North, Middle and South. The middle region of Jordan includes four cities: Amman, Zarqa, Madaba and Al-balqa. According to the Jordanian Department of Statistics, population of the middle region of Jordan was estimated as 63.5 % of all population [20]. The questionnaire has been an online survey questionnaire and all eligible participants from all regions can fulfill the survey.

### A. Data Collection

Prior to starting data collection, the ethical approval was obtained from the Institutional Review Board at the Hashemite University (HU-IRB: 2020/2019/7/1). Moreover, the eligible participants' approval to participate in this study was obtained using online consent form before starting the survey. Consent form included information about the purposes of the study, significance, benefits and risks. All participants were aware that participating in this study was voluntary and they could withdraw from it at any time they want without any physical or emotional harm. Also, the consent form has a clear statement that the participation was totally voluntary without any risk of participation or withdrawal from the study. Beside the decision to participate or not in this study would not affect their treatment plan.

Furthermore, participants were informed about the privacy and confidentiality of this study. This is achieved by understanding that data will be used only for the research purposes and no one other than the researcher can access them. Besides understanding that the questionnaire contains code number not their names.

Data were collected using online survey questionnaire on April 2020. The participants who have the willingness to participate in this study was asked if they met the criteria of eligibility<sup>2</sup>, then they signed the consent form before answering the survey. The contact information of the primary researcher was available in the online survey in order to answer or clarify any misunderstanding of questions. As well as, the survey needs as an average 3 minutes to be fulfilled.

### B. Data Pre-Processing

The target population in this study is the potential patient for novel COVID-19. The size of the collected sample was 120. It is worthy to highlight that there were 15 rejected samples due to incomplete or inconsistent (ex. Age: adult) answered survey. The reliability of the accepted sample 105 was 87.50%. The purpose of the questionnaire was to utilize a machine learning and statistical models to predict novel COVID-19 potential patient based on the signs and symptoms they have.

Our real novel COVID-19 data set present imbalanced classification problem, the majority class is referred to as the negative outcome (Negative PCR Test) with 64 out of 105 (60.95%), and the minority class is referred to as the positive outcome (Positive PCR Test) with 41 out of 105 (39.05%).

The collected dataset consists of thirteen attributes. One class attribute, A, and twelve test feature attributes that present the signs and symptoms of the candidate patient for novel COVID-19: age, smoker, positive chest x-ray, fever, sore throat, aches and pain, dry cough, nasal congestion, absence of smell, diarrhea or vomiting, and breathing difficulty. It is worth to mention that all the attributes are binary categorical type except the age attribute which is integer; therefore the age

attribute is normalized to be aligned with other attributes. The descriptive statistics for the collected samples are shown in Table I, while the description of symptoms for novel COVID-19 (N=41) and non-novel COVID-19 (N= 64) are shown in Table II.

### C. Classification

Classification is a process related to categorization, the process in which negative and positive outcomes are recognized and understood. This is usually achieved by either utilizing several statistical models or invoking various machine learning models.

Machine learning is widely used in different applications due to its powerful prediction and high accuracy while statistical analysis show cases emphasis in models that can be interpreted easily with uncertainty and precision [13].

In this paper, we used the following models:

- Logistic Regression (LR): is a predictive analysis which computes the probability of one dependent variable based on the observations of one or more independent variables. It is the most widely used algorithm for solving problems in different scales. It works properly with the minor instances of multicollinearity and in high dimensional datasets [14]. LR provides the direction of the relationship as well as the degree of the significance of the predictor [15]. It is very easy to implement and explain the results using this model.
- Support Vector Machine (SVM): is a simple and effective neural network. It is utilized for prediction and classification in order to increase predictive correctness by excluding over-fit to the data. It generates various classes by establishing the best hyperplane in multidimensional space in order to reduce the error [16]. It is more applicable in large dimensional spaces where the border of the partition between classes are defined clearly.
- Multi-Layer Perceptron (MLP): is a standard type of neural network which is used for prediction and classification problems. This is achieved by building relationships between inputs and outputs, and computing the required patterns. It consists of a set of neurons in different layers with a set of adaptive weights [17]. The number of hidden layers determines whether the machine learning model is deep or shallow.

### D. Performance Evaluation

Evaluation metric plays predominant role in quantifying the performance of the various models. Generally speaking, metrics include comparing the expected class label to the predicted class label. There are many standard metrics that are used for evaluating the predictive models such as accuracy, sensitivity, specificity, and precision. Those metrics are easily calculated from the confusion matrix, since our classification problem is a binary imbalance type as follows [18]:

<sup>2</sup>The eligible criteria for participants are (a) their age should be 18 years or above (b) they can read and understand Arabic Language (c) they did a test of Polymerase Chain Reaction (PCR) before two weeks or more, and finally (d) they have no critical illnesses at the time of data collection (hemodynamically stable).

TABLE I. THE DESCRIPTIVE STATISTICS FOR THE COLLECTED SAMPLES

	Min	Max	Median	Mean	Variance
Positive PCR	0	1	0	0.390	0.240
Age	19	75	40	40.629	128.120
Gender	1	2	1	1.362	0.233
Smoker	0	1	1	0.581	0.246
Positive X-ray	0	1	0	0.095	0.087
Fever	0	1	0	0.286	0.206
Sore Throat	0	1	1	0.819	0.150
Aches and Pain	0	1	1	0.771	0.178
Dry Cough	0	1	1	0.571	0.247
Nasal Congestion	0	1	0	0.238	0.183
Absence of Smell	0	1	0	0.143	0.124
Diarrhea or Vomiting	0	1	0	0.305	0.214
Breathing	0	1	0	0.343	0.227

TABLE II. DESCRIPTION OF SYMPTOMS FOR COVID-19 (N=41) AND NON COVID-19 (N= 64)

Symptoms	COVID-19 n (%)	Non COVID-19 n (%)
Aches and pains	41 (100 %)	40 (62.5 %)
Sore throat	39 (95.1 %)	47 (73.4 %)
Dry cough	36 (87.8 %)	24 (37.5 %)
Difficulty of breathing	27 (65.9 %)	9 (14.1%)
Diarrhea or vomiting	27 (65.9 %)	5 (7.8%)
Fever	24 (58.5 %)	6 (9.4%)
Nasal Congestion	14 (34.1 %)	11 (17.2 %)
Absence smell	15 (36.6 %)	0 (0 %)
Abnormal chest x-ray	7 (17.1 %)	3 (4.7 %)

Accuracy: The percentage of test set tuples that are correctly classified.

$$Accuracy = \frac{\text{Correct Prediction}}{\text{Total Prediction.}}$$

Error Rate: The percentage of test set tuples that are incorrectly classified.

$$Error = \frac{\text{Incorrect Prediction}}{\text{Total Prediction.}}$$

It is also well-known as the complement of classification accuracy as: Error = 1-Accuracy.

Sensitivity: A metric measures the ability to correctly detect patient who do have the disease (The portion of actual positives that are correctly identified). It is well-known as a type I error.

$$Sensitivity = \frac{\text{True positive}}{\text{True Positive + False Negative}}$$

Specificity: A metric measures the ability to reject healthy patient without a condition. (The portion of actual negatives that are correctly identified). It is well-known as type II error.

$$Specificity = \frac{\text{True Negative}}{\text{False Positive + True Negative.}}$$

The domain of our data set is related to diagnostic novel COVID-19 test, so it is important to have a highly sensitive<sup>3</sup> and highly specific<sup>4</sup> test. Therefore; these two metrics can be mutually joint into single score that balances both concerns so called geometric mean (G\_Mean) as follows:

$$G\_Mean = \sqrt{\text{Sensitivity} * \text{Specificity}}$$

Precision: A metric measures the number of positive class predictions that actually belong to the positive class.

$$Precision = \frac{\text{True positive}}{\text{True Positive + False Positive}}$$

### III. EXPERIMENTAL RESULTS AND DISCUSSION

Python programming language has been used to build the various models in this study. It is an interpreted, high-level, general-programming language. Python is described as a “batteries included” language due to its comprehensive standard library [19].

All models were built using 10-fold cross validation and a tolerance value was set to 0.001. All test attributes are used in building each model. In the LR a logit model was built with a confidence interval 95% and the cut point was set to 0.5. In the SVM the Sequential Minimum Optimization algorithm, (SMO), parameters were set as follows: c = 1.0, Epsilon = 1 × 10<sup>-12</sup>, and the kernel type was chosen to be linear. Finally, the MLP is considered as a shallow deep learning model. It was built with one hidden layer of 12 neurons and the activation function was set to sigmoid one because it exists between zero and one.

The performance evaluations for the statistical model LR and the machine learning models SVM and MLP are illustrated in Table III. It is clearly shown that the machine learning model competes the statistical model with respect to the accuracy. The percentage accuracy of the LR model was as high as 85.00%, while it was equal to 90.00% and 91.62% by using the SVM and MLP, respectively. The reason behind this is referred to the fact that statistical model used to characterize the relationship between the test attributes and the class attribute (outcome variable) to assess the model’s legitimacy. Therefore; the accuracy of prediction by using this inference model is not that robust comparable to the machine learning models. The MLP has the best accuracy compared to the other techniques because of its ability to capture very complex features in the hidden layers and the usage of the nonlinear stimulation functions. In general, the machine learning models sacrifice interpretability for the prediction power.

As we know, the metrics of the sensitivity and specificity reflect completely different aspects of the prediction model. As mentioned earlier, the domain of our data set is related to diagnostic novel COVID-19 test, so it is important to have a

<sup>3</sup>Highly sensitive test infrequently overlooks an actual positive PCR.

<sup>4</sup>Highly specific test infrequently registers a positive classification for anything that is not PCR of the test.

highly sensitive and highly specific test. Since sensitivity and specificity are two conflict metrics, researchers have concentrated on either of these metrics. While this is an accepted practice, we feel that a more fair index would be one which considers both of them simultaneously which is known as geometric mean (G\_Mean). It is clearly noted that the shallow deep learning MLP model scored the highest value of geometric mean which was equal to 90.73 while the statistical model LR scored the minimum value of the geometric mean which was equal to 88.53. The SVM model scored 89.53 as geometric mean value. Table III shows the sensitivity, specificity, and geometric mean for each model.

Precision in any prediction model refers to how closely the observed value to the model's prediction. As shown in Table III, the SVM model is the most precise model, it reached up to 91.67%, due to its high correlation shape which can be detected by our data set. On the other hand, the most imprecise model was the LR model, and it reached up to 66.67%. This obviously reducing usefulness of the prediction and making the mistake very costly.

Based on these results, we could conclude that MLP represents a useful model of novel COVID-19 detection based on the sign and symptoms. On the contrary, LR is not an appropriate technique to predict potential COVID-19 patients due to relative error obtained and low precision.

As a summary a machine learning model competes the statistical model in predicting the infected and non- infected cases of novel COVID-19 based on the signs and symptoms. Utilizing the machine learning model will help in screening patients quickly due to limited availability of the Polymerise Chain Reaction tests in Jordan.

TABLE III. EVALUATION THE PERFORMANCE METRICS FOR VARIOUS CLASSIFICATION TECHNIQUES

Metrics	Statistical Method	Machine Learning	
	Logistic Regression (LR)	Support Vector Machine (SVM)	Multi-Layer perceptron (MLP)
Accuracy (%)	85.00	90.00	91.62
Sensitivity (%)	100.00	91.67	87.80
Specificity (%)	78.57	87.50	93.75
G_Mean	88.64	89.53	90.73
Precision (%)	66.67	91.67	90.00

#### IV. CONCLUSION AND FUTURE WORKS

Several measures are utilized to combat the extent of the novel COVID-19 in the world and in Jordan specifically. The obtained data were analyzed to aid the government in predicting potential patients of COVID-19 in order to save lives or to provide best health care. Several predication models were invoked based on either statistical or machine learning approaches to envisage the citizens' health status.

We foresee two avenues for future work. The first avenue is to implement these models in hospitals in Jordan with aim of identifying COVID-19 patients in a quick, safe method

leading to a decrease in the rapid spread of the virus. The second involves applying the studied models on bigger size dataset and study its corresponding parameters.

#### ACKNOWLEDGMENT

Authors would like to thank Institutional Review Board at the Hashemite University. We are also extremely grateful to the anonymous Referees of this paper, for their valuable and constructive comments.

#### REFERENCES

- [1] W. H. Organization, "Coronavirus disease 2019 (COVID-19): situation report," 2020.
- [2] "CDC Works 24/7," 2020. [Online]. Available: <https://www.cdc.gov/>. [Accessed April 25,2020].
- [3] "SARS CORONA Virus," [Online]. Available: <https://www.sinobiological.com/research/virus/sars-coronavirus-overview>. [Accessed April 25,2020].
- [4] I. Al-Turaiki, M. Alshahrani and T. Almutair, "Building predictive models for MERS-CoV infections using data mining technique," Journal of Infection and Public Health (2016), vol. 9, p. 744–748, 2016.
- [5] A. Du Toit, "Outbreak of a novel coronavirus," Nature Reviews Microbiology, vol. 18, no. 3, pp. 123-123, 2020.
- [6] W. Wang, J. Tang and F. Wei, "Updated understanding of the outbreak of 2019 novel coronavirus (2019 - nCoV) in Wuhan, China," Journal of medical virology, vol. 92, no. 4, pp. 441-447, 2020.
- [7] W. Naudé, "Artificial Intelligence against COVID-19: An Early Review," IZA Discussion Papers 13110, 2020.
- [8] S. F. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A. R. Varkonyi-Koczy, U. Reuter, T. Rabczuk and P. M. Atkinson, "COVID-19 Outbreak Prediction with Machine Learning, 2020040311," Preprints 2020, doi: 10.20944/preprints202004.0311.v1.
- [9] S. Tuli, S. Tuli, R. Tuli and S. S. Gill, "Predicting the Growth and Trend of COVID-19 Pandemic using Machine Learning and Cloud Computing," Internet of Things, p. 100222, 2020.
- [10] P. N. Mahalle, N. P. Sable, N. P. Mahalle and G. R. Shinde, "Predictive Analytics of COVID-19 Using Information, Communication and Technologies," Preprints 2020, 2020040257 (doi: 10.20944/preprints202004.0257.v1).
- [11] D. Yadav, H. Maheshwari and U. Chandra, "Outbreak prediction of covid-19 in most susceptible countries," Global Journal of Environmental Science and Management, vol. 6, no. 4, p. 2020.
- [12] S. Das, A. Dey, A. Pal and N. Roy, "Applications of Artificial Intelligence in Machine Learning: Review and Prospect," International Journal of Computer Applications, vol. 115, pp. 31-41, 2015.
- [13] A. Palmer, J. Rafael and E. Gervilla, Data Mining: Machine Learning and Statistical Technique, 2011.
- [14] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, Lulu, 1st edition, 2019.
- [15] J. F. and J. Fang, "Why Logistic Regression Analyses Are More Reliable Than Multiple Regression Analyses," Journal of Business and Economics, vol. 4, no. 7, pp. 620-633, 2013.
- [16] V. Jakkula, "Tutorial on support vector machine (SVM)," School EECS., Washington State Univ, Washington, DC, USA, 2011.
- [17] L. Dencelin and R. Thenmoley, "Analysis of multilayer perceptron machine learning approach in classifying protein secondary structures," Biomedical Research (2016) Computational Life Sciences and Smarter Technological Advancement, pp. 166-S173, 2016.
- [18] J. Han and M. Kamber, "Data mining: concepts and techniques," San Francisco, Morgan Kaufmann Publishers, 2001.
- [19] E. MATTHES, Python crash course: a hands-on, project-based introduction to programming, 2016.
- [20] DOS (2015), "Jordanian Department of Statistics. Retrieved 10 4, 2016, from," [Online]. Available: [http://dos.gov.jo/dos\\_home\\_e/main/linked.html/jordan\\_no.htm](http://dos.gov.jo/dos_home_e/main/linked.html/jordan_no.htm). [Accessed April 2020].

# Analytical Study between Human Urban Planning and Geographic Information Systems: “The Case of the City of Casablanca”

Mohamed Rtal<sup>1</sup>, Mostafa Hanoun<sup>2</sup>  
Laboratory TIM, FSBM  
Hassan II University of Casablanca  
Casablanca, Morocco

**Abstract**—Since the early 1910s, the city of Casablanca has experienced urban and civic expansion so that its population has become a regional center, but this expansion has not been achieved in an organized and equal manner. This has resulted in significant overlap between systematic and structured reconstruction and random reconstruction, and if the geographic researcher is able to study this transformation by observation in the field where he described the phenomena and visualized them with the naked eye, then urbanization of information systems is one of the methods of expression that allows him to study it successfully and more precisely. The study of the stages of urbanization that the city of Casablanca has gone through using urbanization of information systems will give actors and researchers a clear vision of how this expansion can be achieved, it's positive and negative implications, and its prospects, and will thus help to prepare and manage the urban area of the city. This study period extended from 1910 to 2020, and we relied on a set of documents, satellite photos, aerial photos and old maps.

**Keywords**—Casablanca; human urban; urbanization of information systems; urban expansion

## I. INTRODUCTION

Urban planning is a system created at the beginning of the 20th century to establish scientific, economic, social and technical foundations which make it possible to "tell the truth" about the design and development of cities and to discuss, arbitrate and manage challenges, conflicts and urban conflicts. Urban planning has always been at the service of urban policy. It is therefore the main tool for implementing any city policy.

Casablanca is characterized today by the absence of urban policy. Since the “decisive” colonial policy of the beginning of the 20th century, and after the policy of reaction, not to mention reaction, in the 1980s on the basis of forced urban planning, no coherent alternative policy has been proposed.

Colonial urban policy, based on the cultural and social ideology of Lyautey's "calm" ideology, was developed in the first phase of Dual Cities based on the doctrines of modern urban planning and allows, with modern cities to high quality, designed according to the rules of the avant-garde, to preserve traditional urban forms and especially Medina. In the second phase, which corresponds to the management of city affairs and urban planning by Michel Euchar, I tried to cope with the pressure of urbanization and the negative impacts it could

have on the city, especially in terms of social housing and the prevalence of slums.

The politics of the 1980s, for its part, simply confused the causes and the effects of structural socio-economic failure. He considered that urban planning contributed greatly to the imbalances of cities and Moroccan society. This led to the political decision, crucial for the future, and placed it under the supervision of the Ministry of the Interior. Urban planning has thus become, on the one hand, a tool for political and economic organization, and on the other hand, a means of coercion based on very strict urban regulations.

## II. PROVIDE THE FIELD OF STUDY

Casablanca is the largest city in Morocco, located about 90 km south of the capital Rabat. The city stretches on the Atlantic coast. It moved from a small port at the beginning of the twentieth century to a sprawling city that made it take the lead in relation to Moroccan cities in economic, commercial and services activities.

Casablanca occupies the first place in Morocco in terms of population and the third in Africa after Lagos and Cairo, with a population of about five million people.

The economic capital of Morocco is distinguished by its architecture, between the old mixture of the old city, the popular neighborhoods, its buildings and villas that date back to the era of protection, then the architecture of the independence period during which the city witnessed a rapid development in architecture and architecture in its modern and modern shape represented in building high-rise buildings and Luxurious, large commercial centers, classified hotels, and excellent social, touristic and sporting facilities.

## III. WORK METHODOLOGY

To study the urban expansion of Casablanca, we will rely on a set of documents and geographical tools.

Which is mainly manifested in satellite imagery and ancient maps from the year 1910 to 2020 and we use geographic information systems as a tool for work, and we will try to track and draw the urban area of the city in every thirty years, according to the documents available, and track the urban path of the city during these periods.

#### IV. URBANIZATION AND URBAN GROWTH

Urbanization is defined as a process of population concentration that is carried out in two ways, either by increasing the number of population gathering places, or the growth of population gatherings. In this regard, sociologists and the human environment have unanimously agreed that this simple definition is the most obvious and far from ambiguous definition because it distinguishes between the urbanization process itself, and what may be accompanied by positive effects or natural conditions. In line with this convention, cities are defined as centers of urban agglomeration, and with a variation in the minimum number and density of the population that makes a place a city, as urbanization is defined as a process of radiating ideas and experiences from an urban center to the surrounding areas, because this definition means that the city is a source of urbanization, not a result. The process of urbanization does not give an explanation of the emergence and growth of cities, in addition to that there is another definition that states that urbanization means increasing certain problems and characteristics related to living in the present, because this definition confuses cause and effect, or cause and effect because it also assumes the existence of cities before the process of urbanization, and any definition confuses The concept of urbanization and the city is wrong because the urbanization process culminates in the city.

Any human society is the outcome of the interaction of different factors, affecting each other as the process of urbanization itself is caused by the effect of social change on the distribution of the population within this human community, and then the main entrance to the study of urbanization is to study the population distribution. Cities do not automatically emerge into existence, but they are the result of population growth in society, and there are two factors inherent in the process of urbanization, population and technology or scientific progress, and population concentration is the result of the transfer of the surplus number uttered by agricultural land to urban communities, while the scientific progress paves the way for the concentration of large numbers of Population somewhere. That is, overpopulation, scientific progress, and the process of urbanization are interlinked factors, one of which affects the other, although there is nothing to prevent the civilization of a society without reaching a large degree of population density, or technology may advance without being accompanied by a high population density. The definition of the concept of urbanization varies from country to country

#### V. RESULTS AND DISCUSSION

The image of the "new city" promoted by pioneers and settlers at the beginning of the century attracted a group of architects of different origins to Casablanca in the early twenties. The number of engineers in Casablanca exceeded three times their number in Tunisia.

Casablanca's 2020 population is now estimated at 3,752,000. In 1913, the population of Casablanca was 78,000 (Table I). Casablanca has grown by 159,282 since 2015, which represents a 0.87% annual change. These population estimates and projections come from the latest revision of the

UN World Urbanization Prospects (Fig. 1). These estimates represent the urban agglomeration of Casablanca, which typically includes Casablanca's population in addition to adjacent suburban areas.

##### A. Urban Expansion in Casablanca between 1910 and 1930

From 1912, with the signing of the Protection Treaty, the first major constructions appeared outside the walls of the Old City. Speculation on the ground is fierce and the need for regulation is urgent. After the first plans to develop the surveyors, including Tardif, which charts the grip of the new city bounded by the circular avenue (boulevard de la Résistance and boulevard Zerktouni), the first general resident in Morocco, Maréchal Lyautey, "Urban Planner Henri Prost," Director of the Special Service for Engineering Architecture and Town Planning", in February 1914.

Henri Prost, who is still 8 years in Morocco, submits his first plan to develop Casablanca in 1915 (Fig. 2). It will categorically record Casablanca in the history of modern cities, and implement it, the talented "economic capital" of a large port by General Lyautey, organized Original and innovative in matters of city planning (new science).

TABLE I. SHOWS THE SCALE OF CASABLANCA'S EXPANSION BETWEEN 1913 AND 2020

Year	Population
1913	78 000 (of which 31 000 Europeans)
1920	100 000 (of which 40 000 Europeans)
1932	165 000 (of which 50 000 Europeans)
1936	257 000 (of which 80 000 Europeans)
1947	550 000 (of which 100 000 Europeans)
1950	658 000 (of which 158 000 Europeans)
2005	3 002 700
2010	3 197 000
2015	3 460 000
2020	3 752 000

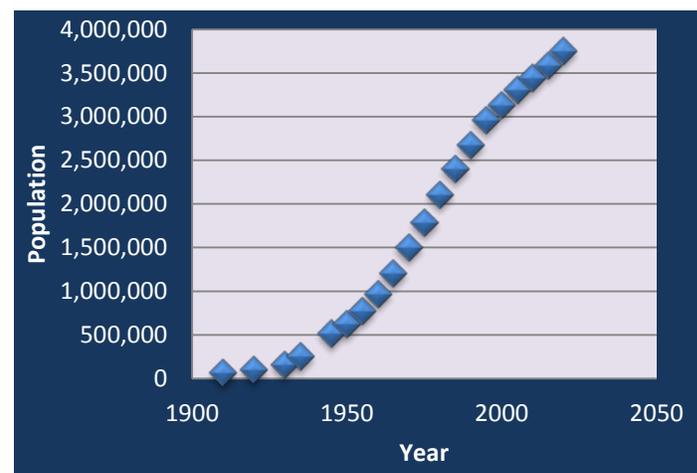


Fig. 1. The Population of Casablanca Developed from 1910 to 2020.



Fig. 2. The Henri Prost Development Project (1915) for Casablanca.

### B. Urban Expansion in Casablanca between 1930 and 1960

The thirties, the break will begin with the use of the new Moorish character, characterized by the abundance of its motifs. The concern for the new generation of architects coming to Casablanca was to employ modern theories that were studied at the School of Fine Arts in Paris. Thus the importance was given to the space and sizes instead of decorating on the level of the facades that left its place for the terraces and prominent windows that took over the space, then extended to the facades of the buildings that took an upper direction. For the luxurious bourgeois oriented buildings, the factor of comfort and luxury was present, as they were equipped with elevators and Household waste fuel, garages, cars, and bathrooms inside the apartments. As artistic architectural masterpieces, it bore the names of its owners as it constituted a heritage reference for this new city, but the architects excelled especially in building villas at the level of architecture, or the comfort and luxury required, and so with the rapid pace of construction in Casablanca and the abundance of workshops, it was described by specialists as the capital of engineering Contemporary architecture.

(Levi Pendion Building) This building was built in 1928 by the architect Marius Puyer, and it is considered the start of the modern movement that characterized the thirties, as it was considered a strong transformation of modern and contemporary architecture in the city.

(Moretti Melon Building) This building is located in the middle of the United Nations square, with its eleven floors. It was built by Pierre Jabban, and was inaugurated in 1934; the luxury of the architecture is represented by its distinctive façade with its vertical and horizontal lines and its high windows in addition to the quality of its equipment and the number of elevators in it.

The regulation of Henri Prost's plans remained in force until the end of the 1940s. Although in 1943 Alexandre Courtois was responsible for their redesign, it was not until the arrival of Michel Ecochard in 1947 that a new development plan will be put in place.

The 1950s witnessed a period of economic prosperity that would be reflected in architecture during this era. Thus, the oval bourgeoisie, influenced by American culture, found the

Californian character of the architecture of its "villas" lost, which resonated with the new generation of architects.

Due to its very modern character, the "villas" of the 1950s acquired the personality of its designers. As a personal masterpiece, the latter dazzled its daring architecture in its details and innovation in its design and engineering. But this difference in style will accompany the difference between social groups. The petty bourgeoisie resorted to building "mixture" villas, inspired by the architectural details of southern France, such as "Sial".

But the 1950s will also define the creation of "housing for large groups". By Eng. Euchar and his team in 1950. The development of public housing programs gave birth to large housing groups with encouraging prices intended for Muslim Jews and Europeans [1].

Thus, for example, the residential city of Karyan Centrale has been programmed to accommodate residents of shanty houses, which offer Muslims traditional residential shops. With regard to "neck" residences, it was intended to house 18,000 Jews from the Old City as well as for the buildings of Burnaziel (1954) intended for European customers of very medium classes.

Parallel to this, collective hobbies emerged that led to the development of private clubs on the Corniche. Building new cinemas such as "Luticia" cinema 1950 or the emergence of service stations and garages, as is the case with the Volvo garage 1950, which is an amazing architectural landmark.

The architectural culture of the fifties will continue after independence until the borders of the eighties, which will define the emergence of postmodern issues.

### C. Urban Expansion in Casablanca between 1960 and 1990

In 1965, Casablanca experienced a riot which now included the urban movement in a new type of social dynamic. The latter depicts new social actors (young people), from the process of modernizing the country, about an equally new cause (education), the fruit of the policy of economic and social development. Originally, young people went on strike over education claims. This urban movement was quickly repressed very quickly, but it led to a reorientation of spatial planning, with a view to better matching between rural and urban planning. The awareness is all the more marked as the lack of consideration of the social costs of urbanization (which prevailed, in part, in the choice of a priority for rural development, shortly after the independence) finds echo in the results of the 1971 census: the average annual growth rate from 1960 to 1971 is 3.5% for Casablanca, when that of the same period for the whole of Morocco reached 4.3%, making increase the urbanization rate from 29.15% (1960) to 35.2% (1971) [2].

The 1970s [3] saw the creation of a ministry of housing and town planning, and this period enabled the Moroccan state to implement a housing policy from its land base. It is intended mainly for the middle strata, which he says carries a driving force for development. But this policy, as ambitious as it is, once its achievements have been translated into the number of homes provided, ultimately only affects part of the

average social category [4]. Why? Despite the fact that at the same time there was a fairly strong access to land offered by the State, it is the first arrivals who are the first served, and the "illegal's" are the inhabitants who could not benefit from lots of land. Then, access to official funding methods concerned only a fraction of the middle layers, the upper middle class. Finally, the official acquisition of land or built housing goes through a long and complex procedure [5], which cannot therefore affect a majority of the middle classes, the lower middle class. The latter and the upper poor class prefer to go through other construction sectors, for reasons of financing, availability of land on the market, or circumvention of the authorization procedure, in particular for migrant workers. The speed of the mode of access to housing, and the ease of the mode of financing demonstrate a certain adaptation to urban socio-economic conditions.

The 1970s was doubly important from the point of view of the housing question since it saw, at the same time as a real estate promotion encouraged by the State, the birth and development of a massive urbanization known as "clandestine" on a short period. The 1970s generally remain the time frame in which a massive phenomenon of state production of housing began and ended. This period corresponded, for the Third World countries resulting from the political independence of the 1950s and early 1960s, to the effort of young states to establish a social base from their action towards the middle strata of the population. Social policies are no longer seen as unproductive, especially since they do not concern the poorest of the city's inhabitants. And if the significant effort that has been made by the State has proved insufficient to avert the housing crisis in Moroccan cities, it is on the one hand the magnitude of the housing deficit to be filled, and on the other hand, certain preferential choices of official town planning which can explain it.

From 1981, Casablanca was administratively divided into five major prefectures (Ben M'sik-Sidi Othman, Ain Chock-Hay Hassani, Casablanca-Anfa, Ain Sebaa-Hay Mohammadi and Mohammedia Zenata [6]), and from 1984, the city is controlled at the urban level by a control and decision center, the Urban Agency of Casablanca. The challenge is ambitious: to create emulation between the different prefectures for better urban management and at the same time to control urban development as absolutely as possible thanks to the "orchestration structure" of the Urban Agency. The ultimate goal of this Casablanca restructuring is to give a new image of the city, consistent with that of the Moroccan political order.

The master plan proposed by the Pinseau team is a document presenting the main options for urban development valid for a period of twenty years. 98,000 hectares have been retained for the growth of the city, according to the linear scheme previously defined by eochard. The most important points of this document are the urban extension towards Mohammedia (rather than towards El Jadida), the improvement of transport, the distribution of central equipment (universities, hospitals), finally the installation of a tool of management and control of urban development.

In the context of the supporting report for the Casablanca master plan, the identification of urban problems is quite

convincing. First of all, it is noted that urban facilities are insufficient. The delay in road equipment (mainly road works and sanitation networks) compared to the extension of the urban perimeter results in a land deficit. The shortage of equipped land is also reflected in a gradual densification of urban space [7]. However, this densification mainly benefits the main producers of housing: landowners, developers and public authorities [8]. The insufficient production of housing and the inadequacy of the method of financing for access to housing [9] are also among the findings noted in the report supporting document.

At the level of urban activities, the master plan notes the deficiency of public transport [10]. With a car for 20 inhabitants and a bus for 4,000 inhabitants, cars and buses each provide 17% of total mobility. And 75% of trips are for work or school. Almost 50% of school trips and 25% of trips to workplaces are pedestrian. This is why the recommendations of the master plan provide for the construction of a light metro in the medium term, an appropriate solution for a city which will have 4 million inhabitants in 2000. With regard to urban planning documents, the report notes their "obsolescence" and "inconsistency", hence the need to reunify urban regulations with the good offices of the Urban Agency.

Finally, concerning the land problem, a central aspect of Casablanca's urbanization, the creation of a Land Agency is envisaged, coupled with the Urban Agency. In this regard, it should be recalled that the master plan had programmed the equipment of 12,000 hectares from 1982 to 2000, including 4000 hectares for the period 1982-1990. However, it should be noted that the Land Agency in question never saw the light of day, its mission having finally been linked to that of the Urban Agency. However, in 1990 (Fig. 3), the Urban Agency had been able to urbanize only a thousand hectares out of the 4000 recommended by the master plan. The bankruptcy in the land sector is obvious. So when agency technicians try to reflect on the means of achieving the objectives set by the master plan, they end up with an annoying observation [11]: the means of public appropriation of the soil (especially the right of preemption) are not effectively legally supported. Although provided for this purpose, a law has still not seen the light of day. By the admission of its main managers [12-14], the agency, however powerful, confines itself to a passive attitude towards speculation. Because in this area, the Urban Agency (which plays the role of Land Agency) has no real authority in the matter. During its existence, it has only had recourse to the declaration of public utility only twice: the first time for the construction of a large avenue to link the Grand Mosque of Casablanca to the business district, a second time to clear a resettlement site outside the Casablanca agglomeration (South-East of Dar Bouazza). But in reality these two expropriation procedures could only take place because of the (par excellence) transcendental nature of the enterprise: the construction of one of the largest places of worship in the Muslim world [15-17].

#### *D. Urban Expansion in Casablanca between 1990 and 2020*

Between 1994 and 2004, the average family size decreased. As a result, the number of families increased by 2.9% per year, while the population increased by only 1.5%

per year. The growth in the number of families is fueling demand for housing, which remains strong as before. This has led to a critical housing situation in the region and the spread of slums and the difficulties of reducing them given the gap between family income and property prices. However, in recent years, a major effort has been made in favor of housing, which has resulted in a record for housing construction, with an average of 24,000 units authorized each year between 2001 and 2004, against 16,000 additional units. Per year in the 80s and 90s, and this revival is the result directly for the “200,000 housing” program, the release of a large amount of land by non-compliance [18].

Despite its drawbacks, urban development in Casablanca remains somewhat well controlled compared to other capitals of the same size in countries with a similar level of development. This translates into the relative agglutination of urbanization, a clear transport network structure, good urban installations, and readable functional organization. This is the result of a long-standing tradition in city planning, which somehow tries to reconcile with the available means and emergencies to be met [19-20].

Fig. 4 presents the Casablanca city development plan between 2010 and 2020.

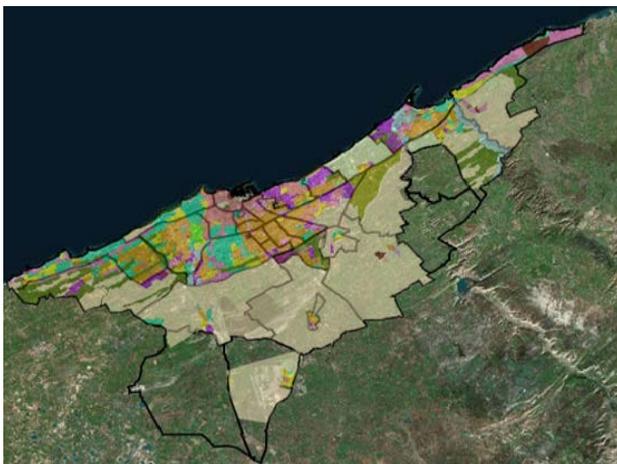


Fig. 3. Casablanca City Development Plan (Source Agence Urbaine de Casablanca).



Fig. 4. Urban Expansion in Casablanca in the late 1990s.

## VI. CONCLUSION

The problem that we addressed in this research is no less important than any other topic, because it is one of the important and thorny issues, especially if applied in a city that needs urban treatment such as the city of Casablanca, which today suffers from the problem of urban growth and urban expansion due to the saturation of its urban fabric and high rate Population growth. This made her a victim of increased demand for housing and land use, and this has led to several results that negatively affected the city as excessive exploitation of the field, random expansion and imbalance in urban fabrics. Our study of urban expansion came as a result of the requirements of the population for the various urban functions of urban life, and from it the need to conduct in-depth studies to analyze the living reality and determine the future needs of housing, facilities and equipment, roads, spaces for the necessary employment as well as the best and optimal direction for city expansion taking into consideration Consider that the city of Casablanca faces a set of barriers that stand in the way of its expansion and the continuity of its urban fabric.

In our study of the reality of urban growth and the phenomenon of expansion in the city of Casablanca, we first analyzed the city from the economic, urban and demographic historical side to determine the characteristics that characterize the city and the problems it suffers from, and an attempt to find solutions for its direct urban expansion based on all the conclusions obtained from the analytical study. We found that reality is completely different from statistics, which conceal a different reality from the reality in which you live. This fact is the result of data inconsistencies between the first stage, the diagnostic phase of the planning plan for the preparation and construction and the last stage of the preparation phase, and therefore it can be judged that it was not programmed tightly to solve the problem of real estate and land consumption.

## ACKNOWLEDGMENT

The authors thank the Casablanca Council and the Casablanca Urban Agency for all the information.

## REFERENCES

- [1] N. Rochd Les Annales de la Recherche Urbaine Année 1990 46 pp. 113-118.
- [2] Population légale du Maroc, Recensement de 1971, p. 5-16.
- [3] M. Ameer, M. Naciri, « L'urbanisation clandestine au Maroc : un champ d'action pour les classes moyennes », in Revue Tiers Monde , t. XXVI, n. 101, janvier- mars 1985, p. 79-92.
- [4] Youssef elmdari, abdelouahed namir and mustapha hakdaoui; analysis of urban growth and sprawl and built up density using remote sensing data: case of casablanca, morocco. Int. J. of Adv. Res. 6 (11). (2018), 1107-1116.
- [5] B. Zyani, « Habitat, contrainte foncière et développement urbain à Casablanca », Habitat, état, société au Maghreb, sous la direction de R. Baduel, C.N.R.S., Paris 1988, p. 213-229.
- [6] M. Pinseau, Schéma Directeur de Casablanca, Rapport justificatif, Royaume du Maroc, Ministère de l'Intérieur, p. 11. avril 1984
- [7] M. Pinseau, Schéma Directeur de Casablanca, Rapport justificatif, op. cité, p. 11.
- [8] A. Amrani, directeur général de la C.G.I. (Compagnie Générale Immobilière), « Grandes opérations d'habitat social à Casablanca », in bulletin de liaison de l'association Réseau, Paris, n. 8, septembre, 4 pages.1990.

- [9] Une sixième préfecture sera créée en 1985, Derb Sultan-El Fida.
- [10] Ainsi la surface urbanisée par habitant est passée de 46 m<sup>2</sup> en 1960 à 39 m<sup>2</sup> en 1982.
- [11] B. Zyani, « Habitat, contrainte foncière et développement urbain à Casablanca », Habitat, État, société au Maghreb, op. cité, p. 220.
- [12] Casablanca, 4 et 5 juin 1990, Agence Urbaine de Casablanca, Royaume du Maroc, Ministère de l'intérieur, 59 pages.
- [13] A. Bouhaya, « Historique de la planification urbaine de Casablanca », Séminaire sur les instruments d'un urbanisme opérationnel pour Casablanca, op. cité, p. 15.
- [14] F. Imanssar, « La mise en oeuvre d'un urbanisme opérationnel pour Casablanca », Séminaire sur les instruments d'un urbanisme opérationnel pour Casablanca, op. cité, p. 57.
- [15] Alnsour, J. A. Managing urban growth in the city of Amman, Jordan. *Cities*, 50, (2016). 93–99.
- [16] Dorning, M. A., Koch, J., Shoemaker, D. A., & Meentemeyer, R. K. Simulating urbanization scenarios reveals tradeoffs between conservation planning strategies. *Landscape and Urban Planning*, 136(3), (2015). 28–39.
- [17] Estoque, R. C., & Murayama, Y. (2015). Intensity and spatial pattern of urban land changes in the megacities of Southeast Asia. *Land Use Policy*, 48, (2015), 213–222.
- [18] Fan, C., & Myint, S. A comparison of spatial autocorrelation indices and landscape metrics in measuring urban landscape fragmentation. *Landscape and Urban Planning*, 121, (2014). 117–128.
- [19] Bae, J., & Ryu, Y. Land-use and land cover changes explain spatial and temporal variations of the soil organic carbon stocks in a constructed urban park. *Landscape and Urban Planning*, 136, (2015). 57–67.
- [20] Bahi, Hicham & Rhinane, Hassan & Bensalmia, Ahmed & Fehrenbach, U. & Scherer, Dieter. Effects of Urbanization and Seasonal Cycle on the Surface Urban Heat Island Patterns in the Coastal Growing Cities: A Case Study of Casablanca, Morocco. *Remote Sensing*. 8. (2016). 829.

# Clustering-Based Trajectory Outlier Detection

Eman O. Eldawy<sup>1</sup>

Faculty of Computers and Information  
Minia University, Minia, Egypt

Hoda M.O. Mokhtar<sup>2</sup>

Faculty of Computers and Artificial Intelligence  
Cairo University, Cairo, Egypt

**Abstract**—The improvement in mobile computing techniques has generated massive trajectory data, which represent the mobility of moving objects like vehicles, animals, and people. Mining trajectory data and especially outlier detection in trajectory data is an attractive and challenging topic that fascinated many researchers. In this paper, we propose a Clustering-Based Trajectory Outlier Detection algorithm (CB-TOD). The proposed algorithm partitions a trajectory into line segments and decreases those line segments to a smaller set (Summary-trajectory  $SS(t)$ ) without affecting the spatial properties of the original trajectory. After that the CB-TOD algorithm using a clustering method to detect the cluster with the smallest number of segments for a trajectory and a small number of neighbors to be sub-trajectory outliers for this trajectory. Also, our proposed algorithm can detect outlier trajectories in the dataset. The main advantage of CB-TOD algorithm is reducing the computational time for outlier detection especially for big trajectory data without affecting the efficiency of the outlier detection results. Experimental results demonstrate that CB-TOD outperforms the state of art existing algorithms in identifying outlier sub-trajectories and also outlier trajectories in real trajectory dataset.

**Keywords**—Data mining; outlier detection; trajectory data processing; clustering

## I. INTRODUCTION

The various advances in GPS devices supported collecting an enormous number of moving objects data easily and rapidly. Therefore, mining of these trajectory data is insistently required to reveal and discover some unknown insights that could be employed to obtain intelligent transportation systems and facilitate smart cities' life. Generally, outlier detection in data mining relates to identifying an object that is incompatible with the other objects [1]. In mining of moving objects database, Trajectory Outlier Detection (TOD) is an important research topic. An outlier trajectory (anomalous) is a trajectory (or a segment of trajectory) that represent different characteristics than the majority trajectories in terms of similarity metrics [2-5]. Outlier segments in a trajectory are different segments from the other segments in the same trajectory as presented in [6], but the outlier trajectory is a trajectory having further few neighbors [4]. The identification of unusual trajectories has great importance in several applications. A popular application of detecting abnormal trajectories is the meteorological monitoring of typhoons. If we can identify unexpected variations in a typhoon path, like a variation in direction, we can announce an early warning for the reduction of casualties and property injuries as quickly as possible [7]. Also, identifying moving objects trends which may be events,

represented by a group of animal moving objects in a specific time that does not conform to a familiar pattern, is essential for detecting animal abnormal habit and attracts the attention of biologists[6]. These applications are behind our motivation work presented in this paper. Outlier detection algorithms can be classified into four categories: distribution-based, distance-based, density-based and clustering-based [8].

Notwithstanding the value of trajectory outlier detection, especially detection sub-trajectory outliers, few research articles discussed this problem. Lee et al. [6] proposed a partition-and-detect framework (TRAOD) for detecting outlying sub-trajectories. TRAOD consists of two phases: partitions trajectories into segments, and then detects the outliers. In the partition phase, TRAOD separates each trajectory into a set of line segments. In the detection phase, density and distance-based measures employed to identify outlying sub-trajectories. Further, Zhang et al. [4] proposed the iBAT algorithm utilizing the isolation mechanism to distinguish outlier trajectories. Also, iBAT utilized a few in number and different than the majority as usual features of abnormal trajectories. However, the outlier trajectories recognized using the iBAT algorithm, but sub-trajectories outliers ignored.

Distinctive from static data, a trajectory may be long and has complicated characteristics. Hence, implementing the computations on the complete trajectory as a fundamental computational unit, it is presumably neglecting to detect local or global outlying partitions that may be essential for various applications.

Example 1: Suppose having five trajectories TR1, TR2, TR3, TR4, and TR5 as shown in Fig. 1. We observe that the thick part in Tr3 is an outlying sub-trajectory as it is different from the remaining partitions in the trajectory. Contrarily, if we compare the whole trajectory with its neighbors we can neglect these partitions because the deviations are averaged over the whole trajectory; so, the overall behavior of the trajectory TR3 appears to be similar to those of the neighboring trajectories.

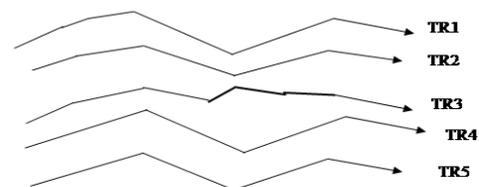


Fig. 1. Example of Sub-Trajectory Outlier.

Our proposed algorithm employs a partition-and-group framework for clustering trajectories [9] with some enhancements to reduce the computational cost. In our methodology, the coresets concept proposed in [10] used but without removing any partitions from trajectory. Basically, after partitioning the trajectories into a collection of line segments, these line segments decreased to a representative less set of lines without adjusting the length of the original trajectory (where the length of trajectory is the summation of the lengths of its line segments). After that, trajectories' partitions clustered employing a density-based spatial clustering of applications with noise (DBSCAN) clustering algorithm [11]. Density-based clustering methods proper for clustering a set of line segments as it identifies clusters of any random shapes. Furthermore, it operates efficiently in a big trajectory dataset [11]. Subsequently, the cluster with the fewest number of line segments for each trajectory in the dataset detected. If this cluster contains line segments that have inadequate neighbors, then the line segments of a trajectory in this cluster recognized outlier line segments for this trajectory. Moreover, if a trajectory contains a considerable number of outlying partitions, then identified it as an outlier trajectory.

In this paper, a Clustering-Based Trajectory Outlier Detection algorithm (CB-TOD) proposed. Our algorithm mainly consists of three phases:

1) *Partitioning and summarization phase*: each trajectory partitioned into several partitions (i.e. line segments); after that these partitions are reduced to a smaller representative set without affecting the information contained in the initial trajectory. Eventually, we get a summarized set of all partitions for all the trajectories in the dataset.

2) *Clustering phase*: similar line segments grouped to a cluster. Consequently, a cluster probably includes line segments from different trajectories.

3) *Outlier detection phase*: after clustering, for each trajectory, we get the cluster which includes the smallest number of segments for that trajectory and a small number of neighbors, then mark this cluster as an outlier cluster for this trajectory and accordingly classify the line segments included in this detected cluster as outlier segments. Moreover, we define an outlier trajectory as the trajectory with a considerable number of outlying partitions.

The main contributions in this paper are the following:

- We employed a novel model that reduces the computational time by decreasing the size of the trajectories dataset and representing each trajectory with the Summary set of line segments that are adequate to define the trajectory behavior without missing the basic motion information.
- A Clustering-Based Trajectory Outlier Detection algorithm (CB-TOD) proposed to detect outlier sub-trajectories as well as whole outlier trajectories utilizing a clustering-based methodology.

- Finally, experimental results are presented and demonstrate that CB-TOD outperforms existing algorithms in detecting both outlying sub-trajectories and outlier trajectories for real trajectory data. Also, the experiments confirm that CB-TOD reduces the computation time of outlier detection without affecting the accuracy of the outlier detection results.

The rest of the paper is structured as follows. Section II presents an overview of related work. Section III describes the problem statement. Our proposed clustering-based trajectory outlier detection (CB-TOD) algorithm presented in Section IV. Section V presents our experimental results. Section VI concludes the work presented in the paper. Finally, in Section VII, we suggest directions for future work.

## II. RELATED WORK

This section categorizes the previous research in trajectory outlier detection into two main directions: detecting sub-trajectories outliers and detecting outlier trajectories.

1) *Sub-trajectories outlier detection*: few research studies were conducted on the problem of detecting sub-trajectories outliers [6, 12-16]. TRAOD is the first approach for detecting outlying sub-trajectories[6]. TRAOD consists of two phases: firstly, partitions the trajectories and then detects the outliers. In the partition phase, TRAOD used the partition method used in TRACCLUS algorithm[9]. Lee et al. [9] presented a TRACCLUS algorithm that includes a partition-and-group framework for clustering trajectory data. TRACCLUS consists of two steps: partitioning and grouping and used for clustering common sub-trajectories. In partitioning step, they applied the Minimum Description Length (MDL) principle[17] for partitioning a trajectory into a set of line segments. In the grouping step, they used a density-based clustering algorithm for grouping similar sub-trajectories. In the detection phase, TRAOD employed density and distance-based measures to detect outlying sub-trajectories. Despite the capability to detect outlying sub-trajectories and outlier trajectories, TRAOD suffered from computational time overhead as well as high complexity of  $O(n^2)$ . Later, Guan et al. [12] proposed R-Tree based Trajectory Outlier Detection (R-TRAOD) and used R-Tree to accelerate the process of outlier detection. Liu et al. [13] proposed a density-based trajectory outlier algorithm (DBTOD) and employed a density-based technique to detect outliers and solve the problems in TRAOD to detect outliers when a trajectory is local and dense. In[14] Daqing Zhang et al. proposed the iBOAT algorithm, which is an improvement on iBAT[4], to work in real-time data. Also, it determines which part(s) of a trajectory is an outlier. iBAT algorithm utilizes the isolation mechanism to identify the outlier trajectory. Despite, it can detect the outlier trajectories and neglect sub-trajectories outlier. In[15] Hao et al. proposed a probabilistic-model called DB-TOD, which models the drivers' behaviors from a historical trajectory dataset and assist in detecting outlier trajectories. DB-TOD used an automatic feature correction mechanism for modeling driving behaviors efficiently. Also, it can identify both complete

outlier trajectories and partial ones. Recently, Yu et al.[16] proposed a TODCSS algorithm that depends upon the common slices sub-sequence for identifying trajectory outlier. Firstly, they compute a direction-code sequence of each segment in each trajectory. Secondly, they used the common slices sub-sequences as a distance measure between two trajectories. Finally, the slice outliers and trajectory outliers discovered based on the new computed distance.

2) *Trajectories outlier detection*: many researchers studied mining in trajectories data to detect outlier trajectories [4, 18-20]. In [18] a framework called ROAM (Rule and Motif-based Anomaly Detection in Moving Objects) was presented. This framework introduces a motion-classifier for trajectory outlier detection. The motifs are a sequence of motion features with values related to time and location. The classifier distinguishes between an anomalous trajectory and a normal one. The main drawback on ROAM framework is that it requires labeled data for the classification process. Sabarish et al. [19] presented a trajectory Outlier Detection algorithm using Boundary (TODB). In TODB algorithm, they used the Convex hull algorithm to generate boundaries for trajectories. Furthermore, they exploit the ray casting algorithm as a classifier to judge a tested trajectory if its inside boundaries or not. The main drawbacks of TODB algorithm, because it used a classification method for categorizing trajectories, is that it required a labeled trajectories dataset that is rarely available. Also, it focuses on the whole trajectory and neglects the detection of sub-trajectories outliers. Moreover, Yong et al. [20] presented TOP-EYE algorithm that employed a decay function to identify the evolving trajectory in an advanced stage. TOP-EYE algorithm computes an outlying score for each trajectory in an accumulating method.

CB-TOD differentiates itself from previous studies by using clustering methodology to detect outlier sub-trajectories and also outlier trajectories. Moreover, the proposed CB-TOD approach decreases the computational time of detecting outliers by reducing the line segments comprising a trajectory and considering only the most representative segments.

### III. PRELIMINARIES

This section presents the preliminary concepts that will be used in the rest of the paper and formalizes the problem statement.

#### A. Definitions

**Definition 1.** A *line segment motion angle*  $\theta$  is a representation of the segment's motion direction and it is measured as follows:

$$\theta = (\arctan\left(\frac{s.end_y - s.start_y}{s.end_x - s.start_x}\right)) * \left(\frac{180}{\pi}\right) \quad (1)$$

where the angle is defined by the two endpoints and the horizontal axis.

**Definition 2.** A *line segment*  $l$  is represented as  $(P_{start}, P_{end}, \theta)$  where  $P_{start}$  is the start point of the segment,  $P_{end}$  is the end

point of the segment, and  $\theta$  is the motion angle of the segment and measured as in Equation 1.

**Definition 3.** A *Trajectory*  $\tau$  is an ordered set of line segments, i.e.  $\tau = \{l_1, l_2, l_3, \dots, l_m\}$ , where  $m$  is the number of line segments in a trajectory  $\tau$ .

**Definition 4.** Given a trajectory  $\tau_i \in S$ , a *Summary trajectory* of  $\tau_i$  is a summarization representation of line segments in  $\tau_i$ . Such that:

- If  $|\tau_i| = m$ , then  $|SS(\tau_i)| = n$ , such that  $n \leq m$
- It mainly divides into two steps:
  - a) *Merge step*:
    - if  $(\theta_{i-1} - \theta_i) < \Phi_1$ (accepted deviation angle) then
    - merge  $(l_i, l_{i-1})$  into one-line segment  $l'_i \in SS(\tau_i)$ . Where  $Len(l'_i) = Len(l_i) + Len(l_{i-1})$
  - b) *Add without merge*
    - if  $(\theta_{i-1} - \theta_i) \geq \Phi_2$ (deviation angle)
    - Then  $l_i \in SS(\tau_i)$

**Definition 5.** *Outlying line segments* of a trajectory  $\tau_i$  called out  $(\tau_i)$  is defined as following:

- Given a cluster  $C$  contains similar line segments depends on a distance measure.
- If  $C$  contains the minimum number of common line segments of this trajectory  $\tau_i$  compared to other clusters (as the trajectory line segments may be divided among different clusters depends on the distance measure), and
- If  $C$  has a small number of similar neighbors' line segments from different trajectories in the dataset of trajectories  $S$ . In another words, if the number of participating trajectories in this cluster (we called it  $Density(C)$ ) is less than a threshold  $P$ .

**Definition 6.** A trajectory  $\tau_i$  is called *outlier trajectory* and added to outliers set if it contains a considerable length of outlying line segments. Such that:

$$\frac{\text{the sum of the lengths of out}(\tau_i)}{\text{the sum of the lengths of } SS(\tau_i)} \geq F \quad (2)$$

where  $F$  is a threshold and its value depend on the length of a trajectory.

#### B. Problem Statement

Given a set of *trajectories*  $S = \{\tau_1, \tau_2 \dots \tau_n\}$ , our goal is to detect the outlying line segments in each trajectory and also detect outliers' trajectories  $Out = \{O_1, O_2 \dots O_{num}\}$  in a given dataset  $S$ . Our objective is minimizing the computation time of detection outliers by reducing the number of line segments in each trajectory to a representative once without losing the basic motion information of a trajectory.

#### IV. CLUSTERING-BASED TRAJECTORY OUTLIER DETECTION (CB-TOD)

In this section, a description of the proposed approach Clustering-Based Trajectory Outlier Detection (CB-TOD) is presented. In CB-TOD we utilize the partition-and-group framework INTRODUCED IN [9]. Our approach is mainly divided into the following phases:

- 1) Trajectory partitioning and summarization phase
- 2) Clustering phase, and
- 3) Outlier detection phase

We explain these phases in the rest of this section. An overview of the proposed approach that abstracts the main steps in our algorithm is shown in Fig. 2. Also, Table I summarizes the main notations used in this paper.

##### A. Partitioning and Summarization Phase

This phase is a preprocessing phase for clustering. The input to this phase is the trajectories dataset  $S$ , then each trajectory in  $S$  is partitioned into a set of line segments by using the minimum description length (MDL) principle as presented in [9]. After that, a summary-trajectory set is created which is a summarization of a trajectory line segments. The coresets method is used for building the summary-trajectory set [10] with some modifications. In [10] the authors added to the coresets a segment with a high impact on the overall trajectory motion pattern and the segments with little effect in trajectory motion pattern are ignored; so, the trajectory-coresets is a *small* representative subset of the trajectory (that highly approximates the trajectory). In contrast, in our proposed approach a summary-trajectory includes segments that affected the motion pattern of a trajectory to a summary-trajectory set. Also, segments with a little effect on the trajectory motion pattern will be merged with the preceding segments to get a single segment with the total length of the merged segments and appended it to a summary-trajectory set.

A thresholds  $\Phi_1$  used for the allowance deviation angle and  $\Phi_2$  controls the deviation angle used in a summary-trajectory set. Thus, given two consecutive segments  $l_1, l_2$  with a motion directions  $\theta_1, \theta_2$  respectively as computed by Equation 1. If  $(\theta_2 - \theta_1) \geq \Phi_2$  (deviation angle), then,  $l_2$  is added to the summarized set, otherwise, if  $(\theta_2 - \theta_1) < \Phi_1$  (accepted deviation angle), then we merge the two line segments ( $l_1, l_2$ ) to get one-line segment ( $l_1'$ ). Thus, we can consider the summary set as a representable set of the original trajectory whose total length is the same as the original trajectory length.

Example 2: A trajectory  $\tau$  consists of the following line segments ( $l_1, l_2, l_3, l_4, l_5, l_6$ ) as shown in Fig. 3, A segments  $l_1, l_2,$  and  $l_3$  have the same motion direction and slope; so, we merge these segments into one-line segment and express it as  $l_1'$  and add it to the summary set of this trajectory. So, a summary-trajectory set will now consist of ( $l_1', l_4, l_5, l_6$ ) line segments. The new set of line segments contains fewer segments which results in decreasing the comparison time for computing the distance between line segments. Furthermore, it does not affect the length of the resulting trajectory as shown in Fig. 4.

TABLE I. LIST OF NOTATIONS USED IN THIS PAPER

Symbol	Definition
$S$	Trajectory dataset
$SS(\tau_i)$	Summary set of line segments for a trajectory $\tau_i$
$Len(l_i)$	Length of line segment $l_i$
$\Phi_1$	accepted deviation angle
$\Phi_2$	Deviation angle
$D$	A set of line segments of all trajectories in trajectories dataset $S$
$Density(C)$	Number of participating trajectories in a cluster $C$
$P$	Threshold of acceptable number of participating trajectories in this cluster
$F$	Threshold for acceptable outlying partition in a trajectory

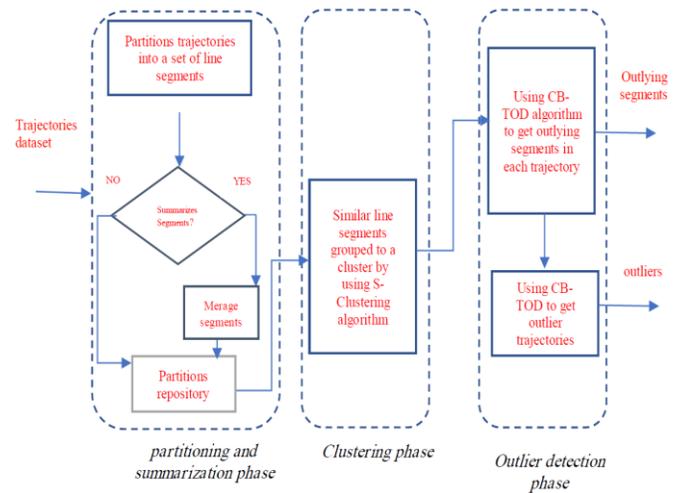


Fig. 2. Overview of CB-TOD.

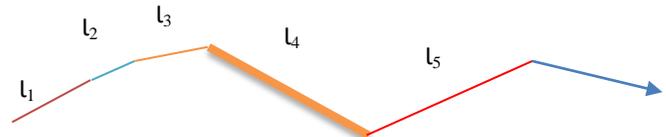


Fig. 3. Initial Trajectory Representation.

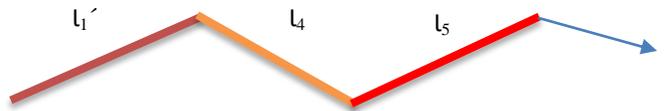


Fig. 4. Trajectory-Summary Example.

Algorithm 1 shows how to create the summary-trajectory set from the original trajectory. The input to the algorithm is the trajectory  $\tau$ , the accepted deviation angle between segments  $\Phi_1$  and the deviation angle between segments  $\Phi_2$ . The algorithm adds segments to the summary-trajectory  $ss(\tau)$  if the absolute difference between its angular value and the preceding segment's angular value is greater than or equal to the deviation angle. Also, if the difference between the angular value of the current line segment and the angular value of the preceding line segment is less than  $\Phi_1$ ; then we extend the preceding line segment to be the result of merging the two segments (replace the end-point of preceding line

segment with the end-point of the current line segment) and then we add this line segment to the summary-trajectory  $ss(\tau)$ .

A summary-trajectory algorithm is used for optimization and speed-up the computations of the distance between line segments.

<p><b>Algorithm 1:</b> Summary-trajectory (<math>\tau, \Phi_1, \Phi_2</math>)</p> <p><b>Input:</b> List of segments in the given trajectory <math>\tau</math>, <math>\Phi_1</math>: the accepted deviation angle between segments' angular values, <math>\Phi_2</math> the deviation angle between segments</p> <p><b>Output:</b> <math>SS(\tau)</math>: List of summary segments in <math>\tau</math></p> <pre> 1: <math>Seg_{Previous} = P[0]</math>; 2: <math>Seg_{Current} = P[1]</math>; 3: <math>SS(\tau).Add(Seg_{Previous})</math> 4: <math>SS(\tau).Add(Seg_{Current})</math> 5: <b>foreach</b> (<math>Seg \in P(\tau)</math>) <b>do</b> 6: <math>Seg_{Current} = Seg</math>; 7: <b>if</b> (<math>( Seg_{Current}.angle - Seg_{Previous}.angle  &lt; \Phi_1)</math>) <b>then</b> 8: <math>SS(\tau).remove(SS(\tau)_{size-1})</math> 9: <math>SS(\tau).add(Seg_{Current})</math> 10: <b>else</b> 11: <b>if</b> (<math>( Seg_{Current}.angle - Seg_{Previous}.angle  \geq \Phi_2)</math>) <b>then</b> 12: <math>SS(\tau).Add(Seg_{Current})</math>; 13: <math>Seg_{Previous} = Seg_{Current}</math>; 14: <b>end</b> 15: <b>output</b> <math>SS(\tau)</math>; </pre>
--

### B. Clustering Phase

In our proposed approach a Density-Based clustering algorithm (DBSCAN) is applied to the summary-trajectory line segments set resulted from the previous phase. DBSCAN is a good choice for clustering large spatial databases [11] as it can discover any cluster with arbitrary shape. Moreover, using DBSCAN in clustering does not require knowing the number of clusters in advance. DBSCAN algorithm uses two parameters  $\mathcal{E}$  and  $MinPts$  (where  $\mathcal{E}$  is a parameter specifying the radius of a neighborhood concerning some point and  $MinPts$  is the minimum number of points required to form a dense region) [11]. In clustering, we used the same distance function as in [9]. Given a set  $D$  of line segments of all trajectories in the trajectory's dataset  $S$ . DBSCAN algorithm is then applied on  $D$  for grouping close line segments according to the distance. Notice that a cluster contains line segments from multiple trajectories to prevent constructing clusters with line segments from only one trajectory [9]. Algorithm 2 illustrates the pseudo code for S-Clustering (Summary-Clustering) algorithm and is used for clustering all line segments  $D$  in our trajectory's dataset  $S$ .

### C. Outlier Detection Phase

In this phase, we get the set of clusters from the previous phase. Each cluster contains line segments that are close to each other. A cluster that includes the smallest number of segments for a trajectory and also has an insufficient number of neighbors is considered as an outlier cluster of this trajectory. Consequently, the line segments introduced in this detected cluster are classified as outlier segments. Moreover, the outlier trajectory is a trajectory that holds an observable length of outlying segments. Algorithm 3 describes a

Clustering-Based Trajectory Outlier Detection Algorithm (CB-TOD). As demonstrated in algorithm 3, CB-TOD algorithm divides into two steps; firstly, we get outliers segments in each trajectory using clustering. Secondly, getting the outliers trajectories in the dataset by using outliers' segments. We sum the lengths of outlier segments of this trajectory and compared them to the total length of the trajectory as described in definition 6.

<p><b>Algorithm 2:</b> S-Clustering (summary clustering algorithm)</p> <p><b>Input:</b> A set of trajectories <math>S = \{ \tau_1, \tau_2 \dots \tau_n \}</math></p> <p><b>Output:</b> A set of clusters contains partitions segments for trajectory dataset</p> <p><math>C = \{ C_1, C_2, \dots, C_m \}</math></p> <pre> 1: <b>for each</b> (<math>\tau \in S</math>) <b>do</b> 2: /* Partitioning Phase*/ 3: Summary-trajectory (<math>\tau, \Phi_1, \Phi_2</math>) /* Fig. 5 */ 4: Get a set <math>SS(\tau)</math> of line segments using the result; 5: Accumulate <math>SS(\tau)</math> into a set <math>D</math>; /* Grouping Phase */ 6: Execute <i>Line Segment Clustering</i> on line segments in <math>D</math>; 7: Output a set <math>C</math> of clusters as the result; </pre>
---

### Algorithm 3: Clustering-Based Trajectory Outlier Detection (CB-TOD)

<p><b>Input:</b> A set of trajectories <math>S = \{ \tau_1, \tau_2, \dots, \tau_n \}</math>, a set of clusters <math>C = \{ C_1, C_2, \dots, C_m \}</math>, <math>P</math> acceptable number of participating trajectories in a cluster, <math>F</math> threshold for acceptable outlying segments length of a trajectory.</p> <p><b>Output:</b> A set of outliers' trajectories <math>Out = \{ O_1, O_2 \dots, O_{num} \}</math> with its outlying segments</p> <pre> 1: <b>for each</b> (<math>\tau \in S</math>) <b>do</b> 2: <b>for each</b> (<math>C_i \in C</math>) <b>do</b> 3: /* Definition 5 */ 4: <math>min = C_1</math> 5: <b>if</b> (<math>min \geq Count(l(\tau), C) \ \&amp;\&amp; \ Density(C) \geq P</math>) <b>then</b> 6: <math>min = Count(l(\tau), C)</math> 7: Insert line segments on this cluster to <math>out(\tau)</math> 8: <b>for each</b> <math>\tau \in S</math> <b>do</b> 9: /* Definition 6 */ 10: <b>if</b> (<math>Len(Out\_seg(\tau)) \geq F</math>) <b>then</b> 11: insert <math>\tau</math> into <math>Out</math> 12: Output <math>Out</math> trajectories with its outlying segment; </pre>
--

## V. EXPERIMENTAL EVALUATION

In this section, the performance of CB-TOD algorithm is evaluated experimentally.

### A. Experimental Setting

CB-TOD algorithm is tested using the same animal movement data set as in [6,9,13] which represents Elk and Deer data. Elk data has 33 trajectories and 15,422 points; Deer data has 32 trajectories and 20,065 points. Our experiments are conducted on Intel core i7 2.7 GHz notebook with 8 GB of

main memory, running on the Windows 10 operating system. We implemented the algorithm using JAVA inside eclipse PHOTON IDE.

### B. Accuracy Evaluation

In this section, we evaluate the accuracy of our proposed algorithm CB-TOD. The accuracy measured by both the number of sub-trajectories outliers and trajectory outliers. In this experiment, we measure the number of anomalous trajectories and sub-trajectories for Elk data and Deer data as shown in Fig. 5 (a and b), respectively. We compare our obtained results with the results in [13], as we used the same datasets with the same parameter values. We observed that the CB-TOD algorithm detects fewer sub-trajectories outliers for both Elk and Deer data respectively, compared to TRAOD [6] and DBTOD [13] algorithms, as shown in Fig. 5(b). That is because we minimize the number of line segments in each trajectory by employing the summary-trajectory technique. Moreover, the CB-TOD algorithm discovers the same number of trajectory outliers compared to the TRAOD algorithm, as displayed in Fig. 5(a) for Elk data. Furthermore, in Fig. 5(a), we observe that our algorithm detects more numbers of trajectory outliers compared to TRAOD and DBTOD algorithms for Deer Data; that is because our algorithm decreases the representative trajectory line set without changing the information contained in the initial trajectory and that accomplished to us the accuracy goal.

Impact of deviation angle ( $\Phi_2$ ). In this experiment, Fig. 6(a, b) displays the effects of varying the deviation angle ( $\Phi_2$ ) on both the number of sub-trajectories outliers and the number of trajectories outliers. We evaluated the changes in the deviation angle ( $\Phi_2$ ) and its effects on the number of outlier segments and the number of outliers trajectories in the dataset. Generally, when we increased the deviation angle ( $\Phi_2$ ), the number of sub-trajectories reduced as it joined more numbers of segments that have the same motion. We observed that the best value for the deviation angle is between 60 and 120 degrees. A constant value for the accepted deviation angle  $\Phi_1$  is used ( $\Phi_1 \leq 30$  degrees).

### C. Performance Evaluation

In this part of the experiments, we evaluate the run-time of the proposed algorithm (CB-TOD).

Computational time. Generally, the processing time of our proposed algorithm CB-TOD is less compared with the competitive outlier detection methods because of summarizing trajectory segments to a smaller set of segments without affecting the length of the original trajectory. We compared the processing time of our algorithm (CB-TOD) with both TRAOD [6] and DBTOD [13] algorithms, as we used the same datasets as in [13]. As shown in Fig. 7, the processing time of CB-TOD algorithm shows the best performance compared to both TRAOD and DBTOD algorithms for the two datasets (Elk and Deer), respectively. This is because using a summary-trajectories technique to reduce the computational time of the outlier algorithm leads to a reduction in dataset size (as it generates a fewer number of segments).

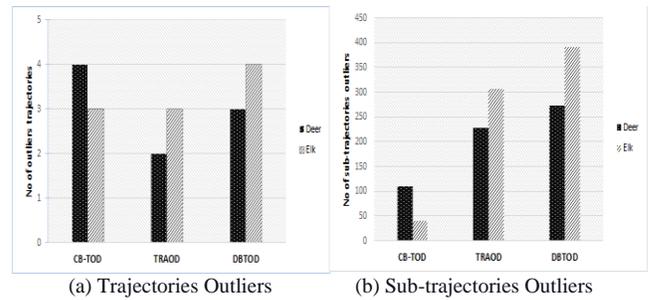


Fig. 5. Comparing between CB-TOD,TRAOD and DBTOD (Accuracy)

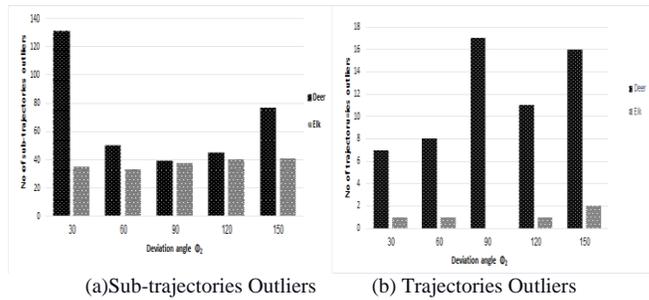


Fig. 6. Effects of Varying the Deviation Angle( $\Phi_2$ )

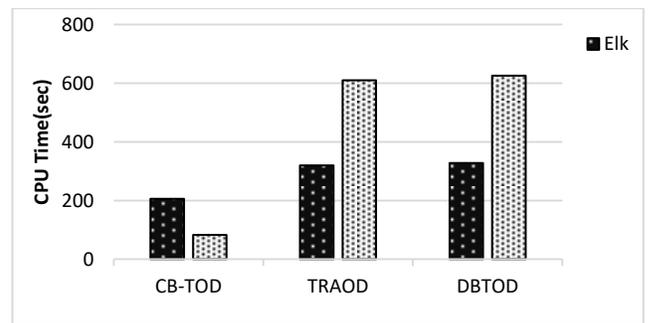


Fig. 7. Comparing between CB-TOD, TRAOD and DBTOD (Performance).

Impact of deviation angle ( $\Phi_2$ ). In this experiment, the effect of varying the deviation angle ( $\Phi_2$ ) on the processing time of CB-TOD algorithm is measured. As shown in Fig. 8, the processing time of CB-TOD decreased by increasing the value of the deviation angle ( $\Phi_2$ ). The intuition behind this observation is that when we increase the deviation angle ( $\Phi_2$ ); we get a smaller number of line segments and consequently the computation time decreases.

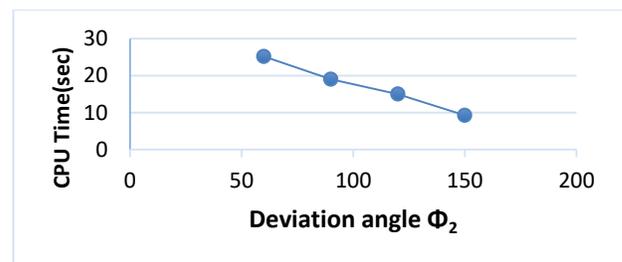


Fig. 8. Effects of Varying the Deviation Angle ( $\Phi_2$ ) on CB-TOD Running Time.

## VI. CONCLUSION

In this paper, we proposed a clustering-based trajectory outlier detection (CB-TOD). Our algorithm summarizes the partitions of a trajectory to the smallest set of partitions without affecting the length of the original trajectory. CB-TOD can efficiently detect outlying sub-trajectory and also outlier trajectory from the trajectory dataset. The main advantage of CB-TOD algorithm is reducing the computational time of outlier detection especially for big trajectory data without affecting the efficiency of the outlier detection results.

## VII. FUTURE WORK

For future work, we aimed to extend our work to maintain bigger datasets. Also, we will use machine learning techniques to predict possible outliers in a big trajectory dataset.

### REFERENCES

- [1] W. I. D. J. M. K. Mining, "Data mining: Concepts and techniques," vol. 10, pp. 559-569, 2006.
- [2] Y. J. A. T. o. I. S. Zheng and Technology, "Trajectory data mining: an overview," vol. 6, no. 3, pp. 1-41, 2015.
- [3] L. Sun et al., "Real time anomalous trajectory detection and analysis," vol. 18, no. 3, pp. 341-356, 2013.
- [4] D. Zhang, N. Li, Z.-H. Zhou, C. Chen, L. Sun, and S. Li, "iBAT: detecting anomalous taxi trajectories from GPS traces," in Proceedings of the 13th international conference on Ubiquitous computing, 2011, pp. 99-108.
- [5] J. D. Mazimpaka and S. Timpf, "Trajectory data mining: A review of methods and applications," Journal of Spatial Information Science, vol. 2016, no. 13, pp. 61-99, 2016.
- [6] J.-G. Lee, J. Han, and X. Li, "Trajectory outlier detection: A partition-and-detect framework," in 2008 IEEE 24th International Conference on Data Engineering, 2008, pp. 140-149: IEEE.
- [7] F. Meng, G. Yuan, S. Lv, Z. Wang, and S. Xia, "An overview on trajectory outlier detection," Artificial Intelligence Review, vol. 52, no. 4, pp. 2437-2456, 2019.
- [8] D. Chen, C.-T. Lu, Y. Kou, and F. J. G. Chen, "On detecting spatial outliers," vol. 12, no. 4, pp. 455-475, 2008.
- [9] J.-G. Lee, J. Han, and K.-Y. Whang, "Trajectory clustering: a partition-and-group framework," in Proceedings of the 2007 ACM SIGMOD international conference on Management of data, 2007, pp. 593-604.
- [10] O. Ossama, H. M. Mokhtar, and M. E. El-Sharkawi, "Clustering moving object trajectories using coresets," in Advances in Wireless, Mobile Networks and Applications: Springer, 2011, pp. 221-233.
- [11] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in Kdd, 1996, vol. 96, no. 34, pp. 226-231.
- [12] B. Guan, Y. Zhang, L. Liu, J. Chen, and R. Guo, "An improving algorithm of trajectory outliers detection," in Advanced Technology in Teaching-Proceedings of the 2009 3rd International Conference on Teaching and Computational Science (WTCS 2009), 2012, pp. 907-914: Springer.
- [13] Z. Liu, D. Pi, and J. Jiang, "Density-based trajectory outlier detection algorithm," Journal of Systems Engineering and Electronics, vol. 24, no. 2, pp. 335-340, 2013.
- [14] C. Chen et al., "iBOAT: Isolation-based online anomalous trajectory detection," IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 2, pp. 806-818, 2013.
- [15] H. Wu, W. Sun, and B. Zheng, "A fast trajectory outlier detection approach via driving behavior modeling," in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 837-846.
- [16] Q. Yu, Y. Luo, C. Chen, and X. Wang, "Trajectory outlier detection approach based on common slices sub-sequence," Applied Intelligence, vol. 48, no. 9, pp. 2661-2680, 2018.
- [17] P. D. Grünwald, I. J. Myung, and M. A. Pitt, Advances in minimum description length: Theory and applications. MIT press, 2005.
- [18] X. Li, J. Han, S. Kim, and H. Gonzalez, "Roam: Rule-and motif-based anomaly detection in massive moving object data sets," in Proceedings of the 2007 SIAM International Conference on Data Mining, 2007, pp. 273-284: SIAM.
- [19] B. Sabarish, R. Karthi, and T. G. Kumar, "Spatial Outlier Detection Algorithm for Trajectory-Data," International Journal of Pure and Applied Mathematics, vol. 118, no. 7, pp. 325-331, 2018.
- [20] Y. Ge, H. Xiong, Z.-h. Zhou, H. Ozdemir, J. Yu, and K. C. Lee, "Top-eye: Top-k evolving trajectory outlier detection," in Proceedings of the 19th ACM international conference on Information and knowledge management, 2010, pp. 1733-1736.

# Measuring the Similarity between the Sanskrit Documents using the Context of the Corpus

Jatinderkumar R. Saini<sup>1</sup>, Prafulla B. Bafna<sup>2</sup>  
Symbiosis Institute of Computer Studies and Research  
Symbiosis International Deemed University  
Pune, India

**Abstract**—Identifying the similarity between two documents is a challenging but important task. It benefits various applications like recommender systems, plagiarism detection and so on. To process any text document one of the popularly used approaches is document term matrix (DTM). The proposed approach processes the oldest, untouched, one of the morphologically critical languages, Sanskrit and builds a document term matrix for Sanskrit (DTMS) and Document synset matrix Sanskrit (DSMS). DTMS uses the frequency of the term whereas DSMS uses the frequency of synset instead of term and contributes to the dimension reduction. The proposed approach considers the semantics and context of the corpus to solve the problem of polysemy. More than 760 documents including Subhashitas and stories are processed together. F1 Score, precision, Matthews Correlation coefficient (MCC) which is the most balanced measure and accuracy are used to prove the betterment of the proposed approach.

**Keywords**—Cosine; dimension reduction; sanskrit; synset; matthews correlation coefficient

## I. INTRODUCTION

The degree of matching between two text pieces based on their statistics as well as semantics is termed as the similarity between text pieces [24]. Statistics of the document means the length of the document, tokens present in the document, etc. The semantics of the document means the understanding meaning of the words present in the document. These documents/ text pieces can be in the form of the word, pdf and so on. There are various measures to calculate the similarity between the two documents. Jaccard, cosine similarity and so on. Cosine similarity is independent of the statistics of the document. Cosine similarity calculates the cosine value of the angle between two vectors. These vectors comprise the frequency of words in a multi-dimensional plane. Each word present in the document represents the dimension/feature [10]. Thus the orientation of the text document gets captured by cosine similarity instead of the magnitude only. Cosine similarity [11] is better than other similarity measures eg. Euclidean distance. The value near to '1' indicates the documents are most similar. Cosine value is always between '0' and '1'. Calculating cosine similarity between English, Hindi [12-14], Marathi [15] text documents [5] is a common task but processing Sanskrit language [33,30,28] and its morphological analysis [35] are critical tasks, as a result finding out the mapping between Sanskrit language texts is challenging. Sanskrit is assumed to be the mother of every language. Panini has introduced this grammar rich language

before 2500 years ago. The Sanskrit language has been the traditional means of communication in Hinduism, Jainism, Buddhism, and Sikhism, still, Sanskrit text mining is an untouched area. Several kinds of literature are available in Sanskrit for eg. stories, subhashits and so on. A subhashita (Sanskrit: 'सुभाषित') can be explained as a legendary kind of Sanskrit concise poems to communicate the message of advice, aphorism and so on. Generally, Sanskrit subhashit or stories are related to all aspects of life. Subhashitas are significant in Indian traditional education and are used to teach values like truthfulness, courage, etc. which are applicable for each phase of life righteousness.

To extract any information from Sanskrit text, various techniques are used. DTMS is one of the techniques using which different operations can be carried out on Sanskrit corpus. Sanskrit documents are placed in rows and significant terms are placed in columns. The entry in the matrix represents the number of times the particular Sanskrit term occurred in the document. The significance of the term is decided based on the frequency of the term. The semantics of the term is considered in DSMS. DSMS uses synset groups in which semantically similar tokens are grouped. Instead of considering term frequency, synset group frequency is considered. It facilitates to solve the polysemy problem means one word used with different senses.

Dimension reduction means the removal of unnecessary features. Several methods are available for dimension reduction like principal component analysis, latent semantic analysis, etc. In the text processing [1][2], Different NLP tasks [5-9] are carried out like removal of stop words [3][4] [32][29] results in dimension reduction. Before stopwords, removal tokenization needs to be carried out for example 'ततो मक्षिका उड्डिय गता.' meaning 'The fly flew away' In this Sanskrit statement 'ततो' meaning 'from there' is removed after separation of tokens. Tokens of the sentence are 'ततो', 'मक्षिका', 'उड्डिय', 'गता', '.'. Lemmatization converts words into their meaningful root form [31]. On the formulated document synset matrix, several applications could be built like plagiarism detection, document clustering, etc. Till now no research is carried out to find Sanskrit document similarity using semantics and context.

To evaluate machine learning algorithms different parameters are available. Eg. Precision, accuracy, Matthews's Correlation coefficient. Matthews Correlation coefficient (MCC) is a quality measure for binary classification. It is a

balanced measure because it considers all false and true positives and negatives. Precision is defined as the ratio of relevant documents to the retrieved documents. Accuracy is the ratio of a number of correctly classified documents to the total number of input documents.

The arrangement of the paper is as mentioned. The existing work carried out by other researchers is written as a literature review in the second section. The research methodology is stated in the third section. The fourth section depicts results and discussions and conclusions are presented as the fifth section to end the paper.

The proposed approach is unique because

- 1) It constructs synset for Sanskrit corpus.
- 2) It extracts the context of the corpus using semantics.
- 3) It builds concept space for the corpus.

## II. LITERATURE REVIEW

Sanskrit is inflectionally strong language and the correct morphological analyzer is required to process Sanskrit text. In spite of being identified as a good analyzed language, morphological analysis of Sanskrit is a challenging task. A morphological analyzer is built which covers wider aspects of language and covers the complexity of words. The applications of the analyzer are used for Sandhi splitting, search engine, etc. [16] Modularity is used while building the morphological analyzer. Other modules like spell check etc. can be added easily. Some modules in the proposed approach are not working, but these modules are based on the grammar which can be easily handled manually. Accuracy of the analyzer is calculated and improvements are discussed [17]. One of the important and unique features of the Sanskrit language is focused known as a dual case. Confusion between plural and dual can be easily removed if grammar rules are followed. Indian treasure is explored to state its different common applications. Various challenges related to NLP along with features of NLP are explored. There are two aspects of similarity implicit and explicit, covering both of these aspects are necessary but critical for finding out the similarity between two texts. This is also termed as the NLP challenge. Different levels that are document or paragraph or sentence or word affect differently and that is why need to be considered to calculate the similarity between two text documents [18]. Sanskrit needs to be handled differently as its morphology is strong. Short texts of Sanskrit [34] are processed semantically using the morphological based approach. Words semantic membership in the sentence is considered assuming that each word has a different significance in the sentence. Ranking of the tokens is carried out using an adaptive measuring algorithm. Sanskrit complex text was processed [19] WordNet is a lexical database that includes vital components like Glosses etc. necessary to identify different NLP features. Domain ontologies could be built using WordNet also semantic relations, polysemy problems can be handled using Wordnet. Sanskrit Wordnet can be used to solve issues like word sense disambiguation (WSD) using gloss. Different techniques that have developed Sanskrit gloss are surveyed to depict different pattern types to solve WSD [20]. Text summarization is carried out to avoid

the efforts and time required to read the document. Natural language processing tools are easily available for the English language. Very less work is explored for Indian languages because of the availability of fewer resources. Existing summarization techniques for Indian languages are surveyed and opportunities for research are discussed [21]. Sentiment analysis is useful for researchers to identify the views of individuals for various services, products, etc. The Internet has allowed exploring NLP tasks by providing a huge amount of data. Machine learning techniques facilitate to provide analysis of this data. The need for domain experts is been reduced for verifying the results due to deep learning algorithms. The sentiment analysis can be done using deep learning techniques and effectively than traditional machine learning techniques for resource-scarce languages. Major challenges like word embedding, ontology building are surveyed which can act as a stepping stone for NLP research [22]. Identifying the language present in the given text is called Language Identification (LI). This could be the important initial preprocessing step to carry out NLP tasks Automatic language recognition is the challenging process. India is a country of multiple languages and there is good scope for language identification problem, It will bridge the digital gap Indian and other languages. Hindi and Sanskrit text is separated using the N-gram approach. The languages which share the same scripts, the technique can be applied [23].

## III. RESEARCH METHODOLOGY

The proposed approach is already being tested for Marathi and termed as DSMM [26] and are now experimenting with Sanskrit. Sanskrit parsers are not developed fully and still are in research phase. Same way the udmodel which is being used in the proposed approach a few times gives incorrect results for a few NLP tasks it impacts other sequential NLP processes. Knowledge of Sanskrit language is necessary and manual intervention, expertise is involved to get correct results unlike Document Synset Matrix Marathi [26]. Also, MCC is used to validate the results which were not used to asses DSMM. Also, the literature available for Sanskrit text similarity is very limited as compared to Marathi text similarity.

R programming is used which provides different packages like tm, quanteda, libraries like udpipe which provides a morphological analysis of the text, for example, identifying the part of speech (PoS), tense, gender and so on. Also functions like documentTermMatrix\_tfidf(), udpipe\_annotate (udmodel\_ Sanskrit) and so on are provided by R programming. Wordnet provided by CFILT, IIT Bombay [27] is used to identifying the semantic relationship between different tokens. Fig. 1 shows the research methodology steps.

### A. Collection of a Dataset, Creation of Corpus and Preprocessing

Sanskrit corpus is not available. The first step is to collect Sanskrit subhashits and stories [25]. The corpus belongs to 340 subhashits and 421 stories. The data is about 15 MB. Different preprocessing steps are carried out. In the first step, tokenization is carried out. Total tokens are 1, 11,123. Stop words are removed, total stop words counted are 34,287. Lemmatization is carried out on the remaining 76836 words.

Unique terms/ tokens are generated which are 56,198. The tokens having similar meanings are grouped and termed as synset groups. Synsets groups are formed and a total number of terms in the form of synset groups are 46,345.

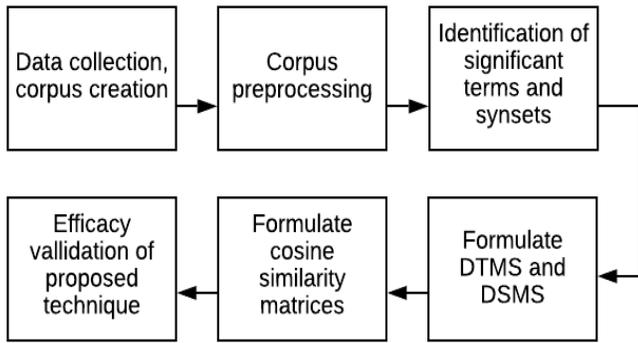


Fig. 1. Steps in Research Methodology.

### B. Identification of Significant Terms and Formulation of DSMS

The frequency of each term is calculated and then the group frequency of each synset group is calculated. Group frequency means, the addition of the frequencies of all terms present in one synset group. Significant synset groups are identified based on a threshold. The synset groups having more than 75% threshold are considered while building DSMS. It means that, if maximum synset frequency is 100 then all those synsets having frequency more than 75, are considered as significant synsets/terms. Documents (subhashits and stories) occupy the rows of the matrix and synsets act as columns in the matrix. The entry in the table/matrix shows the frequency of that particular term/synset in the corresponding document. Due to synset formation, dimensions are reduced, also the context of the term gets considered. For example ‘पक्षे’ means in favor of as well as on the other hand, if ‘कृपा’ (kripa) and ‘पक्षे’ (pakshe) come

together, it means its meaning is ‘in favor of’ and thus the semantics of the words gets identified. This problem of having multiple meanings of a single word is also known as word sense disambiguation or polysemy which is solved by identifying the sense in which it is used in the corpus. Also with the help of other identified significant terms / synsets, the context of the corpus is understood.

### C. Calculate Document Similarity using Cosine Measure

Thus the proposed technique is different than the traditional one which considers only significant terms as column heads and formulates DTMS. DTMS not only lacks to identify semantics and context of the corpus but also dimensions are more than DSMS. On DSMS cosine measure is applied and the similarity of the document is calculated. Two documents are said to be similar if they have cosine measure greater than 0.65 that is the threshold of cosine measure considered for document similarity is 65%. The value of the threshold is considered after evaluating thresholds ranging from 10% to 90%, precision calculated WAS MAXIMUM FOR 65% THRESHOLD.

## IV. RESULTS AND DISCUSSIONS

Table I presents the morphological analysis of the entire corpus. It includes all the documents with their index with respect to the corpus. The paragraph code is assigned for every document. It represents the position of paragraph in the document. The sentence rank of each sentence of a paragraph is represented by sentence code, for example in the first documents, in the first paragraph, the second sentence is shown at the 4<sup>th</sup> Sr. No. The table also depicts lemma, part of speech and other information like gender, the tense of verbs, singular, plural and so on. For example for term ‘असौ’, its lemma is ‘अदस्’, its part of speech is a pronoun and its gender is masculine whereas ‘प्राह’ is a verb which is singular, representing the first person, having past tense and active voice.

TABLE I. MORPHOLOGICAL ANALYSIS OF SANSKRIT TEXT

Sr. No	Document code	Paragraph code	Sentence code	Sentence	Token code	token	lemma	Part of speech	Other information
1	doc1	1	1	ततः असौ प्राह	1	ततः	ततस्	ADV	NA
2	doc1	1	1	ततः असौ प्राह	2	असौ	अदस्	PRON	Gender=Masc
3	doc1	1	1	ततः असौ प्राह	3	प्राह	प्राह	VERB	Number=Sing Person=1 Tense=Past Voice=Act
4	doc1	1	2	क्षत्रियस्य तिस्रः भार्या धर्मम् भवन्	1	क्षत्रियस्य	क्षत्रिय	NOUN	Case=Gen Gender=Masc Number=Sing
5	doc1	1	2	क्षत्रियस्य तिस्रः भार्या धर्मम् भवन्	2	तिस्रः	त्रि	NUM	Gender=Fem Number=Plur
9	doc2	1	1	ि	1	ि	ि	SP	NA

Instead of using any stop word removal technique, only adjectives, adverbs, nouns, and verbs are selected. It resulted in the removal of stop words in the form of discarding numbers, special characters, and pronouns, etc. Also instead of using stemming, lemmatization is used and unique lemmas/terms are identified. The frequency of each term is identified. For example the frequency of 'ध्यानम्' is 30. Group frequency is calculated by adding up the frequency of similar terms 'वृत्त', 'वार्ता' are similar terms and are grouped called as a synset group. Table II shows the frequency of 56198 unique terms in the corpus. Table III shows the frequency of 46, 345 synset groups. In DTMS 'वृत्त' having frequency 4 is not considered as significant terms because its frequency is less than the threshold frequency. But it is an important word. This importance is considered in DSMS as it adds up the frequency and the particular words in synset groups are treated as significant words.

TABLE II. TOKENS AND FREQUENCY FOR DTMS

Sr.No	Token	Frequency
1	ध्यानम्	30
2	वार्ता	26
3	स्तोत्रम्	23
4	फाल्गुन	21
5	वरदः	15
6	सत्यः	4
7	वृत्त	4
		.

TABLE III. TOKENS AND FREQUENCY USING DSMS

Sr. No.	Synset	Frequency
1	वृत्त, वार्ता	30
2	क्षत्रिय, नृप	28
3	ध्यानम्	10
		..
4	स्तोत्रम्	10

TABLE IV. DOCUMENT TERM MATRIX SANSKRIT

Document	क्षत्रिय	वार्या	धर्म	वैश्व	सुत	नृप
doc1	1	2	1	1	1	1
doc 2	0	1	1	0	0	0
doc 3						
doc 4						
doc 5	0	0	1	0	0	0

Table IV shows DTMS for a sample of four documents. The frequent terms are placed as column heads and documents are placed in rows. The entry in the matrix shows the frequency of that term in the document. For example, 'धर्म' has occurred 1 time in document 1. Table V shows the DSMS for the sample of five documents in the corpus. Synset groups are placed in a column, unlike DTMS. It's clear that not only dimensions are reduced but also semantics is being considered that is 'क्षत्रिय, नृप' having the same meaning are grouped while formulating DSMS. In fact, by looking into other significant terms/synset groups context of the corpus is also.

Table VI presents the document similarity matrix using the cosine measure using DSMS. According to the applied 65% threshold, all the documents pairs having a measure of more than 65% are termed as similar. In case of documents, D1 and D2 having measure 0.75 are observed to be similar and D1 and D3 having measure 0.43 is termed to be non-similar.

A confusion matrix is constructed for the DSMS technique and is presented in Table VII. Different evaluation parameters such as F1 score, precision, Matthews Correlation Coefficient, and accuracy are calculated for both the techniques that are DSMS and DTMS and shown in Table VII. All the parameters of the proposed technique (DSMS) produce better results than the existing technique that is DTMS (Table VIII). F1 score, precision, and accuracy are improved by 0.2 measures and Matthews Correlation Coefficient is improved by 0.1 measure.

TABLE V. DOCUMENT SYNSET MATRIX SANSKRIT

Document	क्षत्रिय, नृप	वार्या	धर्म	वैश्व	सुत
doc1	2	2	1	1	1
doc 2	0	1	1	0	0
doc 3					
doc 4					
doc 5	0	0	1	0	0

TABLE VI. COSINE SIMILARITY MATRIX FOR THE CORPUS

Documents	D1	D2	D3		D761
D1	1	0.75	0.45		0.23
D2	0.75	1	0.35		0.67
D3	0.45	0.35	1		0.88
D761	0.23	0.67	0.88		1

TABLE VII. CONFUSION MATRIX

Actual/Predicted	Similar documents	Non-similar documents	Total
Similar documents	431	94	525
Non-similar documents	45	191	236
Total	476	285	761

TABLE VIII. EVALUATION PARAMETERS FOR BOTH TECHNIQUES

Sr. No	Parameter	DSMS	DTMS
1	F1 Score	0.86	0.63
2	Precision	0.82	0.61
3	Matthews Correlation Coefficient	0.60	0.51
4	Accuracy	0.81	0.61

## V. CONCLUSIONS AND FUTURE WORK

DSMS is proposed which identifies the semantics of the term. Synset groups are formed to achieve dimension reduction. Total terms present in the DSMS contribute to identify the context of the corpus. Thus polysemy problem gets solved for the Sanskrit corpus processing which is strongly inflectional language. F1 Score, precision, accuracy and Matthews Correlation Coefficient (MCC) are used to prove the betterment of the technique. F1 score and precision are improved by 0.2 measure and Accuracy and MCC are improved by 0.1 measure. A total of 761 documents are processed including Subhashits and stories. The same approach can be extended for other regional languages.

### REFERENCES

- [1] Bafna P.B., Saini J.R., 2020, "An Application of Zipf's Law for Prose and Verse Corpora Neutrality for Hindi and Marathi Languages", International Journal of Advanced Computer Science and Applications, 11(3).
- [2] Bafna P.B., Saini J.R., 2020, "Marathi Text Analysis using Unsupervised Learning and Word Cloud", International Journal of Engineering and Advanced Technology,9(3).
- [3] Rakholia, R. M., & Saini, J. R. (2015, March). The design and implementation of diacritic extraction technique for Gujarati written script using Unicode Transformation Format. In 2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT) (pp. 1-6). IEEE.
- [4] Kaur, J., & Saini, J. R. (2016). POS Word Class Based Categorization of Gurmukhi Language Stemmed Stop Words. In Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 2 (pp. 3-10). Springer, Cham.
- [5] Rakholia, R. M., & Saini, J. R. (2017). Automatic Language Identification and Content Separation from Indian Multilingual Documents Using Unicode Transformation Format. In Proceedings of the International Conference on Data Engineering and Communication Technology (pp. 369-378). Springer, Singapore.
- [6] Saini, J. R., & Rakholia, R. M. (2016). On continent and script-wise divisions-based statistical measures for stop-words lists of international languages. Procedia Computer Science, 89, 313-319.
- [7] Rakholia, R. M., & Saini, J. R. (2017). A rule-based approach to identify stop words for Gujarati language. In Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications (pp. 797-806). Springer, Singapore.
- [8] Rakholia, R. M., & Saini, J. R. (2016). Lexical classes based stop words categorization for Gujarati language. In 2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA)(Fall) (pp. 1-5). IEEE.
- [9] Raulji, J. K., & Saini, J. R. (2017, January). Generating Stopword List for Sanskrit Language. In 2017 IEEE 7th International Advance Computing Conference (IACC) (pp. 799-802). IEEE.
- [10] Kaur, J., & Saini, J. R. (2020). Designing Punjabi Poetry Classifiers Using Machine Learning and Different Textual Features. International Arab Journal of Information Technology, 17(1), 38-44.
- [11] Rakholia, R. M., & Saini, J. R. (2017). Information Retrieval for Gujarati Language Using Cosine Similarity Based Vector Space Model. In Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications (pp. 1-9). Springer, Singapore.
- [12] Venugopal-Wairagade, G., Saini, J. R., & Pramod, D. (2020). Novel Language Resources for Hindi: An Aesthetics Text Corpus and a Comprehensive Stop Lemma List. arXiv preprint arXiv:2002.00171.
- [13] Bafna P.B., Saini J.R.,2019, "Hindi Multi-document Word Cloud based Summarization through Unsupervised Learning", 9th International Conference on Emerging Trends in Engineering and Technology on Signal and Information Processing (ICETET-SIP-19), Nagpur, India.
- [14] Bafna P.B., Saini J.R., 2019, "Scaled Document Clustering and Word Cloud based Summarization on Hindi Corpus", 4th International Conference on Advanced Computing and Intelligent Engineering, Bhubaneswar, India, in press with Springer.(December).
- [15] Bafna P.B., Saini J.R., 2020, BaSa: A Context based Technique to Identify Common Tokens for Hindi Verses and Proses, in press.
- [16] Murali, N., Ramasree, D. R., & Acharyulu, D. K. (2014). Kridanta Analysis for Sanskrit. Int. Journal on Natural Language Computing, 3(3), 33-49.
- [17] Jha, Deeptanshu, Rashmi Jha, and Varun Varshney. "Natural Language Processing and Sanskrit." International Journal of Computer Engineering & Technology 5, no. 10 (2014): 57-63.
- [18] Bharati, A., Kulkarni, A. P., & Sheeba, V. (2006, April). Building a wide coverage Sanskrit Morphological Analyser: A practical approach. In The First National Symposium on Modelling and Shallow Parsing of Indian Languages, IIT-Bombay.].
- [19] Shevgoor, S. K. (2017, June). A Morphological Approach for Measuring Pair-Wise Semantic Similarity of Sanskrit Sentences. In Natural Language Processing and Information Systems: 22nd International Conference on Applications of Natural Language to Information Systems, NLDB 2017, Liège, Belgium, June 21-23, 2017, Proceedings (Vol. 10260, p. 162). Springer.
- [20] Kulkarni, Malhar, Irawati Kulkarni, Chaitali Dangarikar, and Pushpak Bhattacharyya. "Gloss in sanskrit wordnet." In International Sanskrit Computational Linguistics Symposium, pp. 190-197. Springer, Berlin, Heidelberg, 2010.
- [21] Verma, P., & Verma, A. (2020). Accountability of NLP Tools in Text Summarization for Indian Languages. Journal of Scientific Research, 64(1).
- [22] Nankani, Hitesh, Hritwik Dutta, Harsh Shrivastava, PVNS Rama Krishna, Debanjan Mahata, and Rajiv Ratn Shah. "Multilingual Sentiment Analysis." In Deep Learning-Based Approaches for Sentiment Analysis, pp. 193-236. Springer, Singapore, 2020.
- [23] Sreejith, C., Indu, M., & Raj, P. R. (2013, July). N-gram based algorithm for distinguishing between Hindi and Sanskrit texts. In 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT) (pp. 1-4). IEEE.
- [24] Naik, R. R., & Landge, M. B. (2019). Plagiarism Detection in Marathi Language Using Semantic Analysis. In Scholarly Ethics and Publishing: Breakthroughs in Research and Practice (pp. 473-482). IGI Global.
- [25] www.sanskritebooks.org > category > sanskrit > stories available on 21-04-2020.
- [26] Bafna P.B., Saini J.R., 2020,Marathi Document-Similarity Measurement using Semantics-based Dimension Reduction Technique, International Journal of Advanced Computer Science and Applications 11(4).
- [27] ww.cfil.itb.ac.in/wordnet/webswn/wn.php available on 21-04-2020.
- [28] Raulji J.K., Saini J.R. (2016) "Sanskrit Machine Translation Systems: A Comparative Analysis", International Journal of Computer Application, 136(1):1-4.
- [29] Raulji J.K., Saini J.R. (2016) "Stop-Word Removal Algorithm and its Implementation for Sanskrit Language", International Journal of Computer Application, 150(2):15-17.
- [30] Raulji J.K., Saini J.R. "A Rule Based Architecture for Sanskrit to Gujarati Machine Translation System", Proceedings of International Conference on Emerging Trends in Engineering, Science and Technology (ICRISET-2018), Changa, India, in press.
- [31] Raulji J.K., Saini J.R. (2019) "Sanskrit Lemmatizer for Improvisation of Morphological Analyzer", Journal of Statistics & Management Systems, Taylor and Francis, 22(4):613-625.

- [32] Raulji J.K., Saini J.R. (2020) "Sanskrit Stopword Analysis through Morphological Analyzer and its Gujarati Equivalent for MT System", Proceedings of International Conference on ICT for Sustainable Development (ICT4SD-2019), Panaji, India, 93:427-433.
- [33] Raulji J.K., Saini J.R. (2020) "Bilingual Dictionary for Sanskrit – Gujarati MT Implementation", Proceedings of International Conference on ICT for Sustainable Development (ICT4SD-2019), Panaji, India, 93:463-470.
- [34] Raulji J.K., Saini J.R. (2020) "Sanskrit-Gujarati Constituency Mapper for Machine Translation System", Proceedings of IEEE Bombay Section Signature Conference (IBSSC-2019), Mumbai, India, 1-8.
- [35] Saini J.R., Raulji J.K. (2020) "Peer Analysis of “Sanguj” with Other Sanskrit Morphological Analyzers", Proceedings of 2nd International Conference on Computing Analytics and Networking (ICCAN-2019), Bhubaneswar, India, 1119:65-73.

# An Efficient Model for Mining Outlier Opinions

Neama Hassan<sup>1</sup>, Laila A. Abd-Elmegid<sup>2</sup>, Yehia K. Helmy<sup>3</sup>

Information Systems Department  
Faculty of Commerce and Business Administration  
Helwan University, Helwan, Egypt

**Abstract**—In the internet era, opinion mining became a critical technique used in many applications. The internet offers a featured chance for users to express and share their views and experiences anywhere and at any time through various methods as online reviews, personal blogs, Facebook, Twitter and companies' websites. Such treasure of online data generated by users play an essential role in decision-making process and have the ability to make radical changes in several fields. Although the opinionated text can provide significantly invaluable information for the wide community either are individuals, business, or government, the outlier or anomaly opinions could have the same impact but in opposite manner which harm these fields. Consequently, there is an urge to develop techniques to detect the outlier opinions and avoid their negative impacts on several application domains which rely on opinion mining. In this paper, an efficient model for mining outlier opinions has been proposed. The proposed MOoM model, stands for Mining Outlier Opinion Model, offers for the first time the ability to mine outlier opinions from product's free-text reviews. Accordingly, it can help the decision makers to improve the overall sentiment analysis process and perform further analysis on the outlier opinions to get better understanding for them and avoid their negative impact. The proposed model consists of three modules; Data preprocessing module, Opinion mining module and outlier opinions detection module. The proposed model utilizes the lexicon-based approach to extract sentiment polarity from each review in the dataset. Also, it uses the Distance-based outlier detection algorithm to produce a graded list of review holders with outlier opinions. Experimental study is presented to evaluate the proposed model and the results proved the model's ability to detect outlier opinions in the product reviews effectively. The model is adaptable to be used in other fields rather than product's reviews by customizing its modules' layers.

**Keywords**—Opinion mining; sentiment analysis; anomaly detection; outliers; reviews; text analysis; natural language processing; rapidminer

## I. INTRODUCTION

Data mining targets extracting hidden and implicit information, known as knowledge, from data. To achieve its objective, data mining utilizes various techniques as association rules, clustering, anomaly detection, sequential analysis, and classification. The output of such data mining techniques is used in strategic decision making. A featured technique of data mining known as opinion mining has been issued recently. It can be considered as a special variation of Text Mining where the core content is a set of opinions described through subjective statements [1]. The main objective of opinion mining is to sentimentally analyze opinions, wishes, evaluations, and emotions written by the users in natural language. Accordingly, it utilizes both human

and electronic intelligences. An alternative name of Opinion Mining is Sentiment Analysis [2]. There are five components of the sentiment: 1) The target entity it relates to, 2) The specific target entity aspect to which the opinion refers, 3) The opinion holder, 4) The time of expressed opinion, and 5) The polarity of the opinion which is related to the target entity aspect. For example, a hotel review posted on May 1, 2019 says: "as a business traveller, I found the hotel's location to be great," includes the 1) target entity "hotel", 2) the aspect "location", 3) the opinion holder "I", who travels for business, 4) the opinion posting time "May 1, 2019", and 5) the sentiment "great", which reflects a positive polarity. Not all opinions have these five components, according to [3-5]. There are two main approaches used for sentiment analysis: lexicon-based approach and machine learning approach. Each approach has its own advantages and limitations. A Hybrid technique of both approaches can be used to overcome the shortcomings of the individual techniques. The main difference between Lexicon-based approach and machine learning approach lies on the method of developing the lexicon. Lexicon-based approach works with pre-developed lexicon while as machine learning approach develops its own lexicon dynamically through continuous learning from the data [6]. According to the complexity of the derived knowledge from mining opinions there are three levels of opinion mining; Document level, Sentence level, and Aspect level.

Anomaly Detection (AD) which is also known as Outliers Detection (OD) is an important technique of data mining that is mainly utilized in critical applications like Credit Card fraud detection. In such technique the focus is on up-normal data, which do not conform to normal expected behavior, rather than normal data [7]. Highlighting such up-normal data, known as outliers or anomalies, help the decisions makers to accurately assess the working information regardless to such malicious and vexatious data. However, not all the outliers are harmful attacks as they can just represent data with previously unknown and surprising behavior [8]. The techniques of anomaly detection can be classified in three main categories; Unsupervised, Supervised, and Semi-supervised techniques. The unsupervised outliers detection techniques use unlabeled test data and assume that most of the data instances are normal by searching for such instances seem to leastly fit to the reminder of the data set. On the contrary, supervised outliers detection techniques work with labeled data set and utilizes a classifier to learn from the data. The semi-supervised outliers detection techniques work with a developed model of normal data set that it uses to test other instances of test data set. Anomalies could be one of three types; *Point anomalies*: where a data point exists too far from the other data points;

*Contextual anomalies*: the event is anomalous regarding to specific context and it is common type used in time-series data set; *Collective anomalies*: represent a group of anomalous regarding to the whole data set [9]. This paper proposes a model to sentimentally analyze opinions and extract outlier opinions. The proposed model uses dictionary-based opinion mining approach. Such approach consumes fewer resources and can be adapted easily to different types of opinionated text domains. The proposed model also applies the distance-based outlier detection algorithm for anomaly detection.

The research paper is organized as follows: Section two summarizes the related work in both opinion mining and anomaly detection fields. Section Three describes in details the proposed Mining Outlier opinions Model (MOoM). Section four shows our experimental study. The efficiency of the proposed model is tested and measured in Section five. Finally, the conclusion and future work are highlighted in Section six.

## II. RELATED WORKS

This section reviews the recent presented research work regarding opinion mining and anomaly detection.

### A. Opinion Mining

Opinion mining, or sentiment analysis, has been extensively studied recently with different and novel approaches. The research of such field aims to improve its results especially with the rapid growth of using web technology in different aspects of life. In the tourism field, the classification problem of published tourists' comments about their various experiences have been studied as in [10]. The authors have developed and compared various classifiers of different deep learning techniques. The main objective is to help tourists to get benefit from online published comments in developing their own trip plan efficiently. In the education field, a sentiment analysis lexicon-based approach for automatic analysis of students' comments to predict the expected performance level of teachers have been proposed in [11]. The approach gave an attention to the terms of intensifier words and blind negation words in the analysis process. A hybrid approach on mobile reviews have been applied as in [12]. They proved that the accuracy of the hybrid approach is greater than the accuracy of the individual techniques of Naïve Bayes and KNN. A five Twitter data sets used for applying a convolution algorithm to train the deep neural network in [10]. The results have proved that the proposed GloVe-DCNN model has a better accuracy and faster analysis speed.

### B. Anomaly Detection

Deferent anomaly detection techniques have been successfully applied across several domains. Generally, the anomaly detection techniques can be classified as machine learning-based, distance-based, and density-based techniques. In [13] a hybrid approach of both clustering-based and distance-based techniques is used to mine efficiently the existing outliers. The experimental study proved the efficiency of the proposed approach over distance-based approach by consuming less computational resources. Various outlier

detection algorithms have been used such as K-nearest neighbor in [13] and [14]. Histogram-based outlier score algorithm is used by [15]. an Angle-based anomaly detection method applied for high-dimensional datasets in [16] and many other algorithms which work efficiently on high dimensional datasets.

A framework for the argumentation process to effectively identify outlier opinions of stakeholders have been applied in an argumentation systems in [17]. The proposed framework detects the outlier opinions of stakeholders from both individual and collective viewpoint.

The previous exploration of the most recent research presented for opinion mining and outliers detection highlights the urgent need to propose a model that combines anomaly detection and opinion mining techniques in order to mine outlier opinions. Outlier opinion is the opinion which deviate from the general opinion. Such new knowledge output helps the decision makers to take decisions with minimal effect of outlier opinions.

## III. PROPOSED MODEL (MOoM)

An Efficient Mining Outlier Opinions Model (MOoM) is proposed to mine outlier opinions from free-text product reviews. A rare research work is dedicated for such point. The proposed model integrates both the sentiment analysis domain and the anomaly detection domain to achieve its objective. Fig. 1 clarifies with an example the various modules of the proposed MOoM model starting with manipulating the available product's reviews and ending by a list of review holders which have outlier opinions.

MOoM model consists of three major modules; data pre-processing module, opinion mining module, and outliers' detection module. MOoM architecture appears in Fig. 2.

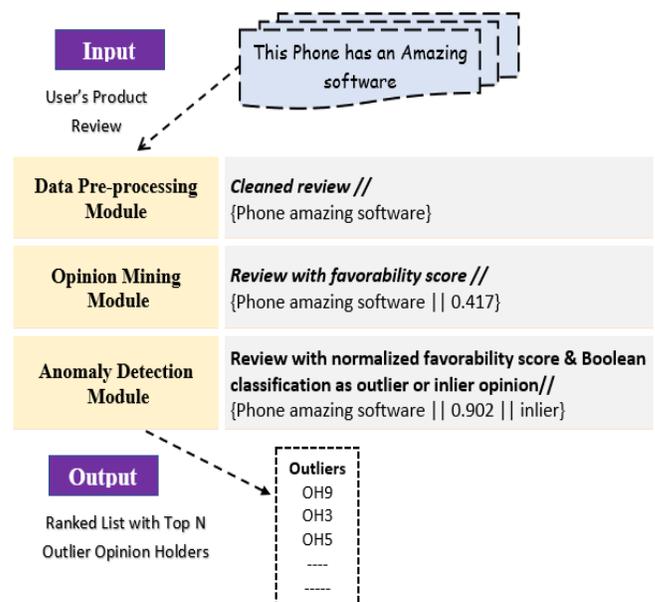


Fig. 1. An Overview of the MOoM Model.

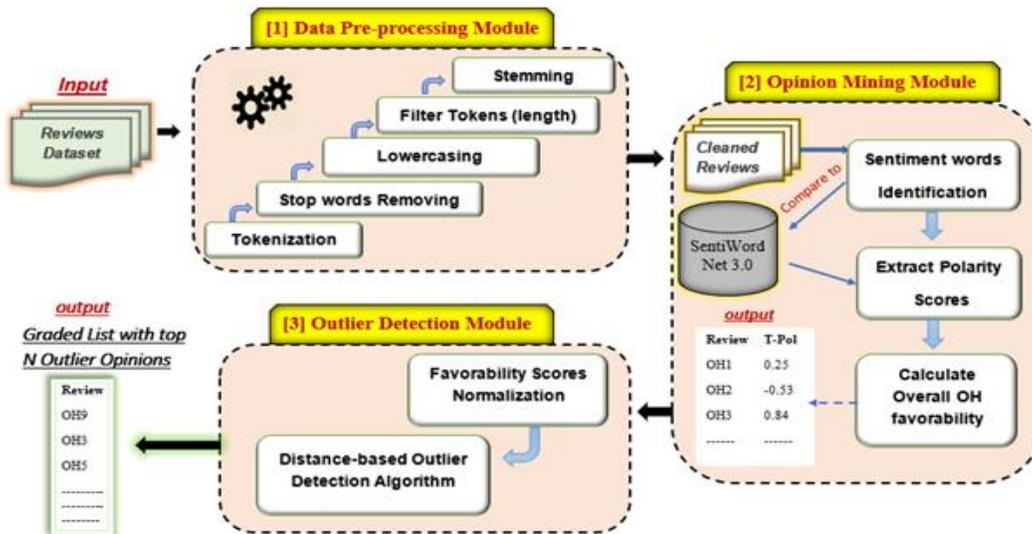


Fig. 2. The Proposed Mining Outlier Opinions Model (MOoM).

### A. Data Pre-Processing Module

In this module, the opinion reviews are transformed to cleaned reviews to be suitable for processing in the next module, opinion mining. To pre-processing of the opinion reviews passes through five NLP tasks as described in the following sections.

1) *Tokenization*: In tokenization step, the large strings of text are divided into tokens which are a small set of words. Bigger chunks of text can be divided into sentences, sentences can be divided into words, etc. Tokenization is done by locating the word boundaries [18]. The sample sentence below shows an example of Tokenization.

This mobile phone is very good.

After tokenization, the output text will be:

This mobile phone is very good advanced processing is commonly conducted after a text reviews has been suitably tokenized.

2) *Lower casing*: Text transformation into the lower case is a simple and effective way to pre-process the text reviews. It is appropriate for almost problems related to text mining and NLP. lower casing is beneficial when the dataset isn't so big and extremely helps to make data consistency [19]. The lower casing is important to make sure that the word matched to respective feature, for example "AMAZING" & "AmAZing" -- should be converted to "amazing".

3) *Stopwords removing*: Stop words are a group of words which are commonly used in a specific language. For example, in English a, the, is, her, are, on, of, with, about, what, when, where, that, this, by, be. and etc. are considered as a stopword. The reason behind removing these stop words is that they are valueless and removing them from reviews enables the model to concentrate on the other words which are most important and consequently, achieving a high accurate

classification [20]. For example, the next review is acceptable if the stop words are removed:

This mobile phone has an amazing software.

4) *Filtering*: Performing more cleansing for data is done by removing non-English words and filtering words by their length, where words with length less than minimum length will be removed [21].

5) *Stemming*: Stemming is the process which used to eliminate the word affixes (circumfixes, prefixes, suffixed, infixes) with the aim of getting root form a word stem. Stemming techniques put word variations like "great", "greatly" and "greatest" to concept of "great" [22].

### B. Opinion Mining Module

The sentiment extraction method presented in this paper is depending on a dictionary-based approach for document level sentiment classification task. The cleaned reviews obtained from the previous Data pre-processing module are passed through three steps; identify sentiment words, extract the polarity scores of sentiment words, and finally calculate the overall favorability for each opinion holder. A detailed explanation of each step is given below.

1) *Identify sentiment words*: The identification of sentiment words is essential to understand the expressed opinions in user reviews. Words which are usually used by the people to express their positive or negative feelings are known as Opinion/sentiment words. Example for positive sentiment words (nice, amazing, and wonderful) and for negative sentiment words (horrible, bad, and terrible). Thus, A part-of-speech (POS) patterns are useful to extract opinionated words. Part of speech tagging is "the process of mapping a word in the text to its corresponding tag [23]. The main aim of doing POS tagging is that adjectives and adverbs would be strong indicators of the opinion of the review, so they help to perform opinion mining because of that the most used opinion words are adjectives and adverbs.

Adjectives are indicated as JJ; Adverbs are indicated as RB.

2) *Extract the polarity scores:* After the identification of sentiment words in each review document, the next step is to the polarity strength of each sentiment word. For this purpose, ‘SentiWordNet’ which is a lexical resource for sentiment analysis has been used. SentiWordNet ‘SWN’ is an opinion lexicon derived from the WordNet database and it is commonly available for the research purposes. In SWN, each word is related to numerical scores which refer to positive and negative opinion information, scores are ranged between -1 and 1. If a word in review document agreed with the word in Wordnet, then its score from SentiWordNet can be used to find its polarity. SentiWordNet is constructed based on a semi-automated process, and simply could be upgraded for incoming versions of WordNet, and for other different languages [24]. In the SentiWordNet database, the terms are categorized based on the parts of speech coming from WordNet, so that be able to apply scores to terms, a part of speech tagging was needed to extract adjectives and adverbs as opinion words then catching their sentiment strengths as the polarity score for each term.

3) *Calculate the overall OH favorability:* After the identification of sentiment words in a review document and extracting their sentiment polarity scores based on WordNet and SentiWordNet lexicon as explained in the two previous sections, now it is needed to calculate the total score of sentiment polarity for each review in the dataset which represents the opinion holder favorability about the product. The overall polarity score of a document is calculated by the Summation of the polarity of all words in a document divided by the total number of words in the document as shown in “equation (1)”.

$$sentscore = \frac{\sum_{i=0}^n(Word\_polarity)}{No.of.words} \quad (1)$$

Word polarity // The polarity value returned by SentiWordNet.

No. of. words // The total number of words in the document.

### C. Anomaly Detection Module

In the anomaly detection module, the list of opinion holders’ favorability scores which obtained from the previous opinion mining module is normalized using appropriate normalization technique to be ready for applying a distance-based outlier detection algorithm which extracts the top K opinion holders with outlier opinions as explained in the following sections.

1) *Favorability scores normalization:* In an unsupervised anomaly detection, the Normalization process has a private significance because of various attributes in the dataset may have distinct units for measurement [25]. The normalized data produces output with higher accuracy and performance in opposite to the unnormalized data which produces output with

lower accuracy and performance according to [12]. Normalization technique is used to scale values in a determined range because all data should lie on similar range to achieve fairly comparison. The proposes MOoM model utilizes the Min-Max normalization technique to normalize the OH favorability scores “equation (2)”. The max represents the largest value, the min represents the smallest value and the other values are ranged between (0,1).

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

X // The original data point.

X<sub>new</sub> // The normalized data point.

X<sub>max</sub>, X<sub>min</sub> // The Maximum and minimum data point respectively.

2) *Distance based outlier detection algorithm:* The Distance based outlier detection algorithm as shown in Algorithm (1) is used to measure dissimilarities of OH favorability cores. This algorithm measures the distance between a data point and its k nearest neighbor. Every data point is ranged based on the distance to its k-th nearest neighbor and the top N data points in this ranked list are defined to be outliers. The K value represents the number of neighbors and the N value represents the number of outliers. Different distance functions can be used to measure the distance between two data points, in our research the Euclidian distance function has been used “equation (3)”. The data point which is numerically differ from the other data is considered as an outlier.

Algorithm 1/Distance based outlier detection algorithm

**Input** Normalized list of OH favorability scores

**Output** Top N OH with outlier options

**Step1** Assign K and N value

K-The number of neighbors

N-The number of outlier parameters

**Step2** Compute the Euclidian distance [Eq3] for each point on the basic of its distance to its k-th nearest neighbor.

**Step3** Each point is ranked on the basic of measured distance and the top N points in this ranking are declared to be outliers.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

The ranked list with top N outlier opinions helps decision-makers to get a better understanding for the outlier opinions and apply further analysis on these anomalies and their OHs, and this would lead to a noticeable improvement in the overall sentiment analysis process and consequently decision making. The value of N is chosen by decision maker to define and filter the top outlier opinions, so it could be varied in different cases.

#### IV. EXPERIMENTAL STUDY

In this section, an experimental study is explained to outline the methodology used for evaluating the proposed model.

1) *Tools and dataset*: A free educational version of RapidMiner studio (Version 9.5.1) is used. Rapid Miner is a powerful software with open source data science platform which provides an integrated environment for data mining, text mining and machine learning processes [26]. The experiment is conducted on a pc with Microsoft Windows 10 operating system with Intel® Core™ i5- 4200U CPU @ 1.60 GHz with 4.00 GB RAM. To validate the proposed model, the experiment is performed on sample dataset about 100 customer reviews, which is taken from amazon mobile phone reviews dataset in (<https://www.kaggle.com/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones>) that is one of the popular dataset sites.

2) *Data pre-processing module*: As illustrated in Fig. 3. *Read Excel* operator is used to input the sample dataset reviews then nominal to text operator used to convert the comment review from nominal to string attribute. To prepare the tested dataset for opinion mining module, the *process document from data* operator which works as a container operator is used. *Tokenize* operator divides the text of a review into a sequence of tokens. The non-letter character mode is used which result in tokens with one single word. *Transform Cases* operator is used to convert all characters in a review to lower case. Then, the noisy words that do not affect the classification task removed from document by using *Filter Stopwords (English)* operator which deletes every token

matches a stopword from the built-in stopwords list. *Filter Token (by length)* operator filters tokens based on the number of characters they contain. For the proposed model Minimum number of characters is chosen to be two.

3) *Opinion mining module*: The experiment has been applied a dictionary-based sentiment analysis approach. *Open WordNet Dictionary* operator is responsible for connecting RapidMiner with WordNet-3.0 dictionary which is stored in a specified directory in pc while defining the wordlist. *Stem (WordNet)* operator used to reduce the length of the words to the minimum length by applying the Porter stemming algorithm and the rule-based replacement for word suffixes. Stem Wordnet uses Wordnet dictionary to define the stem rule. Extract Sentiment operator plays the major role in this module. This operator uses a WordNet 3.0 and a SentiWordNet 3.0.0 database which are connected by Synset IDs to extract sentiment of an input review. This operator allows us to identify the opinionated words by selecting the type of words to be used for calculating sentiment value. In the running experiment, adjectives and adverbs are used as sentiment words. The operator calculates the sentiment of each word to get the total sentiment of a document, where the first meaning of a word has the most influence on a sentiment and each next meaning has less influence on a sentiment. Overall favorability for each opinion holder is then calculated as the average value of all word sentiments as shown in Equation (1). The value of sentiment is in the range (-1.0 to 1.0) where -1.0 means very negative and 1.0 means very positive. Fig. 4 show the output resulting from opinion mining module.

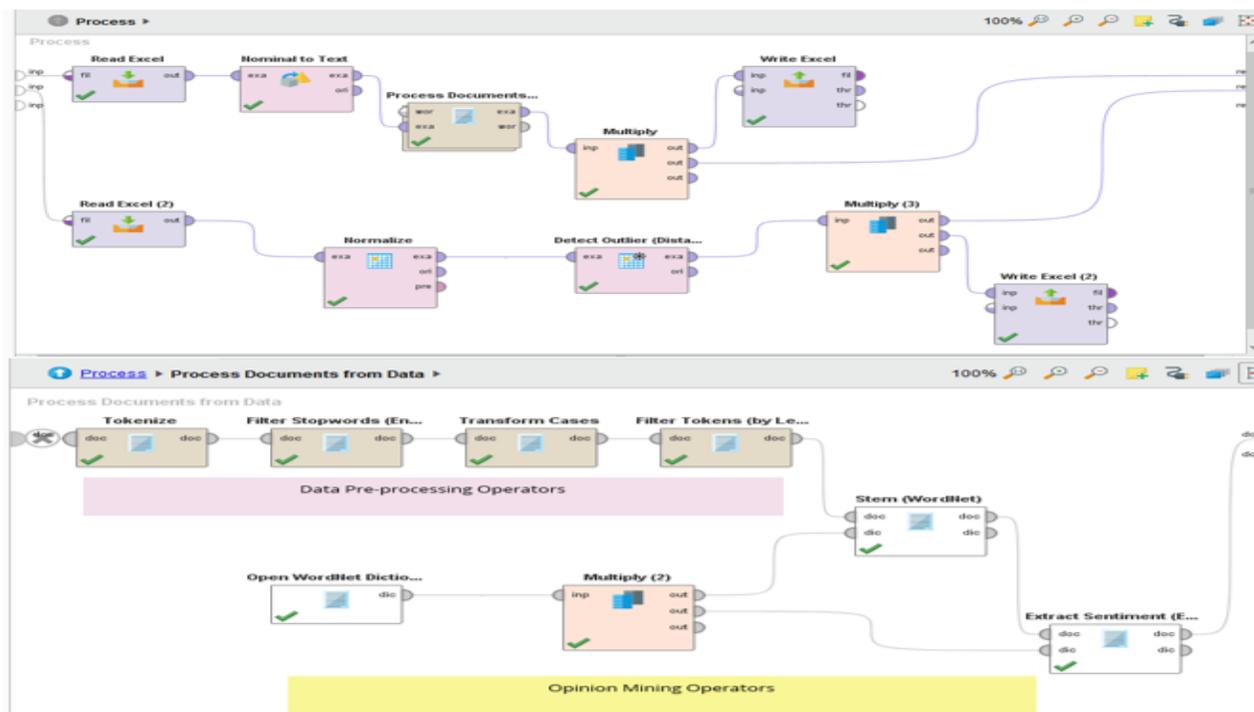


Fig. 3. RapidMiner: Complete Process for MOoM Workflow.

4) *Outlier detection module*: After getting the overall favorability for each opinion holder in the sample dataset, the *Normalize* operator is used to scale these values so they fit in a certain range. Normalization is very significant when working with attributes of different scales for a fair comparison. This Operator performs normalization with four different methods which are Z-normalization, range transformation, proportion transformation and interquartile range. For the running experiment the simplest range transformation method is used to scale all favorability scores between 0 and 1 as illustrated in Equation (2) for Min-Max normalization technique. Then the Detect Outlier (Distances) operator is used to find opinion holders who have an outlier favorability scores in our sample dataset.

As illustrated in algorithm (1) Distance-based outlier detection algorithm, the operator enables us to identify N outliers based on the distance to their k nearest neighbors. Firstly, the variables N and k are specified through parameters as 7 and 5, respectively. Secondly, measure the distance from each point to its 5 nearest neighbors. The operator provides different functions for distance measurement which are Euclidean distance, Squared distance, Cosine distance, Inverted Cosine distance and Angle.

For the running experiment the Euclidean distance is used as shown in Equation (3).

Thirdly, each data point is ranked based on its measured distance to its 5-th nearest neighbor and the top 7 points in this ranking are specified to be outliers. The operator adds a new Boolean attribute called 'outlier'. If the value of this attribute is true that example is an outlier and vice versa.

Fig. 5 shows the output resulting from outlier detection module. In the outlier attribute the true value means that data point is an outlier and the false value means that data point is an inlier.

Row No.	Brand Name	Opinion Hold...	text	sentiment	outlier ↓
37	Samsung	oh37	great	0.486	true
56	Nokia	oh56	excellent	1	true
65	Nokia	oh65	excellent	1	true
93	Nokia	oh93	good	0.613	true
98	Nokia	oh98	friendly said amazingly c...	0.342	true
99	Nokia	oh99	fine easy heavy	0.470	true
100	Nokia	oh100	lots found pretty good	0.475	true
1	Samsung	oh1	lucky found sold liked ol...	0.474	false
2	Samsung	oh2	nice nice clean set easy ...	0.470	false
3	Samsung	oh3	pleased	0.624	false
4	Samsung	oh4	good slow good	0.490	false
5	Samsung	oh5	great lost go eligible	0.393	false
6	Samsung	oh6	stated item cracked side...	0.097	false
7	Samsung	oh7	port loose usable sold	0.367	false
8	Samsung	oh8	good charged charged l...	0.391	false

Fig. 5. The Final Output of Outlier Detection Module.

V. RESULT AND DISCUSSION

This section discusses the results obtained from the previous runs of the conducted experiment. This paper has performed the experimental evaluation on a moderate sized dataset as a sample of Amazon mobile phone review dataset. Fig. 6 depicts a 3D scatter plot after the proposed model has been applied. The data points in green represent the top 7 outlier opinion from ranked list and the data points in blue represent the inlier (normal) opinion.

Table I presents a ranked list of opinion holders who have an outlier favorability score. Oh37 ranked 1 in the outlier ranked list. A decision maker can also change the value of top N outlier opinion based on their domains and objectives.

Evaluation of the proposed model is done using a standard evaluation metrics of Recall, Precision and F-measure. Precision and recall are defined in terms of true positive (TP), false positive (FP) and false negative (FN) as shown in the following equations:

$$\text{Precision}(p) = \frac{TP}{TP+FP} \tag{4}$$

$$\text{Recall}(R) = \frac{TP}{TP+FN} \tag{5}$$

$$F - \text{Measure} = \frac{2 \cdot PR}{P+R} \tag{6}$$

In this paper, we conduct our experiment on sample dataset from Kaggle repository (Amazon-mobile phone reviews dataset). We applied a common data preprocessing steps to obtain a cleaned data ready for further analysis in the opinion mining module. Then we used a dictionary-based approach to perform the sentiment analysis process depending on the SentiWordNet dictionary to obtain the overall favorability score for each opinion holder in our dataset. The results from opinion mining module has been normalized

Row No.	Brand Name	Opinion Holder	text	sentiment
1	Samsung	oh1	lucky found sold lik...	0.242
2	Samsung	oh2	nice nice clean set ...	0.204
3	Samsung	oh3	pleased	0.458
4	Samsung	oh4	good slow good	0.336
5	Samsung	oh5	great lost go eligible	0.125
6	Samsung	oh6	stated item cracke...	-0.302
7	Samsung	oh7	port loose usable ...	0.111
8	Samsung	oh8	good charged char...	0.208
9	Samsung	oh9	originally wanted e...	0.097
10	Samsung	oh10	great responsive bl...	-0.033
11	Samsung	oh11	previously course ...	0.362
12	Samsung	oh12	great side functional	0.258
13	Samsung	oh13	item quickly fixed pl...	0.176
14	Samsung	oh14	disappointed inste...	-0.125
15	Samsung	oh15	ordered model sai...	0.359

Fig. 4. The Final Output of Opinion Mining Process.

using a Min-Max normalization technique in the anomaly detection module. After that, the normalized data used as input for the distance-based outlier detection algorithm to produce a ranked list with the top N opinion holders (OH) which have an outlier opinion. The evaluation metrics like precision, recall, and F-measure have been used to measure our models' performance with 86% for precision, 85.7% for recall and F-measure.

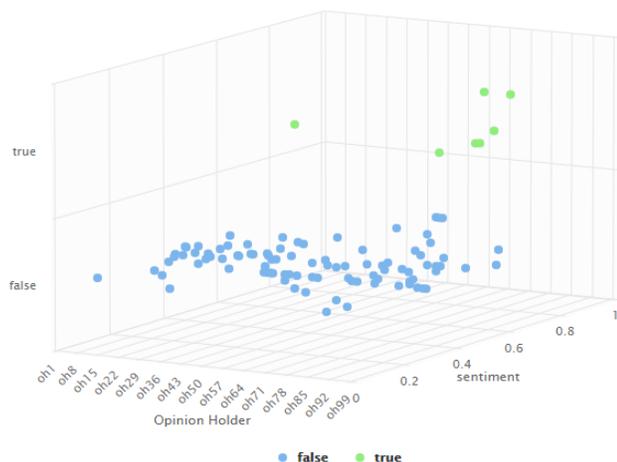


Fig. 6. The 3D Scatter Plot for Outlier Detection Module.

TABLE I. RANKED LIST OF OUTLIER OPINIONS BASED ON A DISTANCE-BASED OUTLIER DETECTION ALGORITHM

<i>Top N opinion holders with outlier opinions</i>
Oh37
Oh56
Oh65
Oh93
Oh98
Oh99
Oh100

TABLE II. SHOWS THE PERFORMANCE'S METRICS OF THE APPLIED MODEL

<i>Performance Indicator</i>	<i>Result</i>
Precision (%)	86%
Recall	0.857
F-Measure	0.857

## VI. CONCLUSION AND FUTURE WORK

The processing of free-text users' opinions are now being giving more attention according to their critical impact. As such opinions have the ability to make radical changes in most if not all fields, especially with the increasing of the people's ability to share and publish their opinions in an easy manner at anytime and anyplace. The importance of outlier detection is coming from the fact that outliers in data are translated into significant (and often critical) actionable information in a wide variety of application domains. As the opinionated comments give individuals, companies and government very useful informative data which is used in sentiment analysis process for decision making and future forecasts of users' behavior.

The anomalies in these opinionated comments could lead to high errors in data analysis and decision-making process. Thus, different application domains which use the opinion mining should be interested with the outlier opinions since they could have a negative impact on their domains. In this paper, we proposed an efficient model for mining outlier opinions (MOoM) to preprocess the opinionated comments and extract the overall favorability score for each opinion holder, then apply a distance-based outlier detection algorithm to generate the top N outlier opinions. To validate the proposed model, a sample dataset from Amazon mobile phone reviews is used. The precision, recall and F-measure of the proposed MOoM model have been evaluated using 100 document reviews. The results of the proposed model can help decision makers to not only analyze the results but also to make more informed decision and decide the best using of this information.

As future work, the following points are considered:

- Various experiments to be conducted using large data sets of different domains.
- Implementation of the hybrid approach for opinion mining process rather than the used dictionary-based approach.
- Exploring different techniques in anomaly detection for mining the outlier opinions.

## REFERENCES

- [1] Wilson, T., J. Wiebe, and P. Hoffmann, Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, 2009. 35(3): p. 399-433.
- [2] Kaur, A. and V. Gupta, A survey on sentiment analysis and opinion mining techniques. *Journal of Emerging Technologies in Web Intelligence*, 2013. 5(4): p. 367-371.
- [3] Liu, B., Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2010. 2: p. 627-666.
- [4] Liu, B., Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 2012. 5(1): p. 1-167.
- [5] Liu, B., Sentiment analysis: Mining opinions, sentiments, and emotions. 2015: Cambridge University Press.
- [6] Paltoglou, G. and M. Thelwall, Sensing social media: A range of approaches for sentiment analysis, in *Cyberemotions*. 2017, Springer. p. 97-117.
- [7] Chandola, V., A. Banerjee, and V. Kumar, Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 2009. 41(3): p. 1-58.
- [8] Agrawal, S. and J. Agrawal, Survey on anomaly detection using data mining techniques. *Procedia Computer Science*, 2015. 60: p. 708-713.
- [9] Ahmed, M., A.N. Mahmood, and J. Hu, A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 2016. 60: p. 19-31.
- [10] Martín, C., et al., Using Deep Learning to Predict Sentiments: Case Study in Tourism. *Complexity*, 2018. 2018.
- [11] Aung, K.Z. and N.N. Myo, Sentiment analysis of students' comment using lexicon based approach. in *Computer and Information Science (ICIS)*, 2017 IEEE/ACIS 16th International Conference on. 2017. IEEE.
- [12] Campos, G.O., et al., On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. *Data Mining and Knowledge Discovery*, 2016. 30(4): p. 891-927.
- [13] Pachgade, M.S. and M.S. Dhande, Outlier detection over data set using cluster-based and distance-based approach. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2012. 2(6).

- [14] Su, M.-Y., Real-time anomaly detection systems for Denial-of-Service attacks by weighted k-nearest-neighbor classifiers. *Expert Systems with Applications*, 2011. 38(4): p. 3492-3498.
- [15] Goldstein, M. and A. Dengel, Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm. *KI-2012: Poster and Demo Track*, 2012: p. 59-63.
- [16] Zhang, L., J. Lin, and R. Karim, An angle-based subspace anomaly detection approach to high-dimensional data: With an application to industrial fault detection. *Reliability Engineering & System Safety*, 2015. 142: p. 482-497.
- [17] Arvapally, R.S., et al., Identifying outlier opinions in an online intelligent argumentation system. *Concurrency and Computation: Practice and Experience*, 2017: p. e4107.
- [18] Mayo, M., A general approach to preprocessing text data. *KDnuggets*, <https://www.kdnuggets.com/2017/12/general-approach-preprocessing-text-data.html>, 2017.
- [19] Vijayarani, S., M.J. Ilamathi, and M. Nithya, Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, 2015. 5(1): p. 7-16.
- [20] Riloff, E. Little words can make a big difference for text classification. in *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*. 1995.
- [21] Baldwin, T., et al. How noisy social media text, how diffrent social media sources? in *Proceedings of the Sixth International Joint Conference on Natural Language Processing*. 2013.
- [22] Asian, J., Effective techniques for Indonesian text retrieval. 2007.
- [23] Saharia, N., et al. Part of speech tagger for Assamese text. in *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*. 2009. Association for Computational Linguistics.
- [24] Ohana, B. and B. Tierney. Sentiment classification of reviews using SentiWordNet. in *9th. it & t conference*. 2009.
- [25] Kandanaarachchi, S., et al., On normalization and algorithm selection for unsupervised outlier detection. *Data Mining and Knowledge Discovery*, 2020. 34(2): p. 309-354.
- [26] Hofmann, M. and R. Klinkenberg, *RapidMiner: Data mining use cases and business analytics applications*. 2016: CRC Press.

# Coronavirus Social Engineering Attacks: Issues and Recommendations

Ahmed Alzahrani

Faculty of Computing and Information Technology  
King Abdulaziz University, Jeddah  
Saudi Arabia

**Abstract**—During the current coronavirus pandemic, cybercriminals are exploiting people's anxieties to steal confidential information, distribute malicious software, perform ransomware attacks and use other social engineering attacks. The number of social engineering attacks is increasing day by day due to people's failure to recognize the attacks. Therefore, there is an urgent need for solutions to help people understand social engineering attacks and techniques. This paper helps individuals and industry by reviewing the most common coronavirus social engineering attacks and provides recommendations for responding to such an attack. The paper also discusses the psychology behind social engineering and introduces security awareness as a solution to reduce the risk of social engineering attacks.

**Keywords**—Social engineering; coronavirus; COVID-19; phishing; vishing; smishing; scams; working remotely; cybersecurity; security awareness; human security behavior

## I. INTRODUCTION

The entire world is suffering from the novel coronavirus or COVID-19, which has damaged the global economy and cybersecurity as well as causing death or damage to millions of people. While the world tries to deal with the coronavirus pandemic, cybercriminals are taking advantage of it to lure Internet users into becoming victims of social engineering attacks. For instance, cybercriminals are sending a huge amount of "phishing" attacks to trick users into clicking on malicious links or attachments so the criminals can steal sensitive data or lock users' devices and force them to pay a ransom to recover their data. Regardless of advanced security controls such as strong firewalls, cryptography methods, intrusion protection systems, and intrusion detection systems, social engineering still the biggest challenge to all organizations [1]. That's because of the human factor the attacker uses to conduct malicious activity when they cannot directly attack a system that has no technical vulnerabilities. Social engineering attacks are now considered as the most powerful security attacks [2]. The attacker uses psychological manipulation to convince users to violate security protocols. To perform a successful social engineering attack, the social engineer (attacker) needs to have a high level of skills, including the ability to fit the proper technique to manipulate a given victim. Relying on technical solutions to prevent such attacks is not practical, human users remain susceptible to manipulation by social engineers seeking to gain sensitive information or unauthorized access [3].

Social engineering attacks fall into four types: physical, technical, social, and socio-technical [3]. In general, there are two methods of social engineering attacks, human-based and computer-based. Human-based social engineering requires interaction with humans to gain the desired information. Impersonation is the most common approach for this type, via a phone call or text message (see Fig. 2), online, or even in person. Computer-based social engineering uses computer software to try to gain the required information. This attack includes sending scam emails asking the user to open an attachment to check the latest statistics about coronavirus or information about coronavirus safety measures (see Fig. 1). Cybercriminals can also create a fake website to trick users into downloading malware to steal users' credentials and online banking information.

The number of social engineering attacks during the coronavirus pandemic is increasing as cybercriminals exploit the situation for their selfish goals. This paper reviews the most common coronavirus social engineering issues and provides some recommendations for both individuals and industry. The paper is organized as follows: Section II provides a review of human security behavior. Section III describes the proper security awareness method to reduce the success of social engineering attacks. Section IV presents the study issues, followed by recommendations in Section V. Cybersecurity lessons from the coronavirus pandemic are discussed in Section VI and Section VII presents the study conclusions and future work.

## II. BACKGROUND

Social engineering attacks manipulate users to reveal sensitive information that the attacker may use against the target organizations. The Verizon [4] data breach investigation's recent report stated that 33% of actions used to attack organizations come through social engineering-based attacks. Therefore, there is a need to analyze the factors affecting human behavior to help security administrators plug gaps in information security.

### A. Human Security Behavior

Most modern organizations depend on information systems. This requires them to manage associated risks such as social engineering attacks. The most challenging risks for an organization arguably center on information security. Many organizations tend to deploy technology-based solutions to mitigate these risks. These are necessary but often insufficient to address a range of potential threats. The human factor is an

essential element of information security. According to the SANS institute [5], there are four categories of behavior that social engineers might use to manipulate human emotions to make a successful attack [5]. The first is a Careless Attack Vector, which helps the attacker exploit a failure to implement proper defensive countermeasures. A common example is password theft when the target has written it down on a piece of paper or a sticky note. Several research studies have tried to improve information-security practices, without significant success [6]. This is, therefore, a fresh examination of the challenges and obstacles related to the way people deal with the need for information security [7]. A number of real-life examples illustrate that complex and expensive security methods are not helpful if users compromise security measures by, for instance, neglecting password protection. Users regularly engage in intentional or non-intentional risk-taking behavior such as careless information handling, surfing on unsecured web-pages, thoughtless use of mobile devices or insecure data practices [8][9]. This type of risk-taking can open avenues through which malicious co-workers or external perpetrators can damage the organization. Second is a Comfort Zone Attack Vector which occurs when users feel comfortable in their environment, such as office, and do not realize the threat. A common technique is shoulder surfing, which means looking over the user's shoulder to get a username or password while the target types on their keyboard. Third: Helpful Attack Vector, which means users generally try to be helpful to everyone, including those they do not know. There are two kinds of direct attack types here.

- Direct Approach - Piggybacking: It's called the "Big Box technique": the attacker carries a big, heavy-looking box as they try to go through security doors [5]. A nearby employee sees this and opens the door to help the person with the box get in. The employee, with good intentions, gives the attacker access to the premises.
- Direct Approach - Impersonation: The attacker calls the organization's IT help desk to gain information such as the target's password. IT support has become aware of this kind of social engineering technique, but the risk should not be disregarded [5]. Another recent example is a scammer calling people and claiming to have a cure for coronavirus or offering a good deal of face masks. People who fall for the scammer provide their credit card information to complete the bogus transaction.

The "Fear Attack Vector" is classified as the most aggressive type of psychological attack [5]. The attacker makes use of the target's fear, anxiety, stress, and pressure to get personal information or gain access to their accounts. For instance, the attacker pretends to be from a government department to trick the user to provide personally identifiable information.

The behavior of some users is probably influenced by the behavior of other users in their group or team [10]. The influencing factors involve both technical and non-technical factors related to protecting sensitive information [11]; both factors are equally important. Users can create, or make possible, many threats to the security of the organization,

broadly divided into two classes [12]. The first kind of threat is intentional, involving malicious users who leak sensitive information. The second kind of risk is non-intentional actions, perhaps as a result of carelessness, resulting in leakage of information.

Consequently, information security is directly associated with the user's security-related behavior. A good understanding of user security behavior can help reduce the success of social engineering attacks. For instance, "autonomy motivator" is a concept of self-determination theory, which focuses on human behavior and the extent to which behavior is self-motivated and self-determined [13]. Autonomy refers to "volition, having the experience of choice, endorsing one's actions at the highest level of reflection" [13]. Since autonomy focusses on "the desire to protect an individual's scope for action and decision-making" [14], users need to be supported in making the right decisions against a number of potential threats, known vulnerabilities of the infrastructure and past and ongoing cyber-attacks. This may help them to make the proper responses to real-world threats such as an e-mail link that might be a phishing attack. Alzahrani and Johnson [13] studied the autonomy motivator by investigating whether the user would fall victim to a phishing attack. The authors used the Decisions and Disruptions (D-D) cybersecurity game, which was developed by Professor Rashid and his team at the University of Bristol cybersecurity group [15]. The authors developed pre- and post-assessment tests on the autonomy motivator to find out whether there is a significant improvement in test scores after participants experience D-D gameplay. Overall results confirmed that the autonomy motivator is positively influenced by the game, which means players became more likely to make the right decision in response to social engineering techniques such as phishing or spear phishing. Users can gain or review valuable information from such awareness methods that may help them to increase their security awareness, which in turn may help them make the right decisions in the real world against social engineering attacks.

#### *B. The Proper Security Awareness Method to Reduce the Success of Social Engineering Attacks*

Many researchers identify the need for information security awareness to promote positive behaviors [16]. Since social engineering relies on the manipulation of human behavior, security experts consider raising awareness and training as the best way to combat social engineering attacks [3]. However, increasing awareness should be designed carefully and effectively by measuring its effect on users' security behavior. Several efforts have been made to improve security awareness and develop methods for spreading awareness of cybersecurity among users. These methods can raise cybersecurity awareness regarding a wide range of social engineering threats. Important methods proposed in previous research include conventional, instructor-led, online, simulation-based, and game-based delivery methods. Each method has advantages and limitations. Organizations or government entities need to adopt the most suitable method to influence security awareness with respect to social engineering attacks.

Conventional delivery methods involve dissemination of cybersecurity awareness using paper and electronic resources. These methods use leaflets and posters for directing users' attention to specific and relevant subjects. This method has the advantage of sharing several messages at one time with the target audience [17]. However, it is challenging to ensure that users have gone through the newsletters and paper-based resources and understood the information that the documents seek to transmit.

Instructor-led delivery methods involve formal presentations, seminars, and classroom-based lectures facilitated by experts in the field to raise the cybersecurity awareness level of users. This method has its limitations. It is an expensive and static solution that usually results in a boring experience that is ineffective for the target audience [18]. The success of this method relies on the capability of the instructor to engage employees in the classroom. Limitations of this method can be addressed by sharing experiences and knowledge among users of an organization and ensuring their participation with interactive activities and group-work-based assignments.

Online delivery methods deliver cybersecurity awareness programs using e-mail broadcasts, online discussion, rich media, and interactive teaching applications. These methods are well suited for teaching over a distributed geographical area. Significant issues in using these methods involve the creation and implementation of an online cybersecurity awareness program as a result of the lack of fully developed plans for delivering it to the target audience and evaluating its effectiveness [18]. These methods require user self-motivation that often hinders the successful delivery of cybersecurity awareness.

Delivering cybersecurity awareness using educational video techniques plays a significant role. This method does not require a classroom trainer and addresses the limitations of conventional delivery methods for holding the trainee's attention for a long enough time to impart knowledge. Online video provides visual and audio learning for the target audience. Users can participate independently in the learning process at any particular time.

Simulation-based delivery methods have been attracting attention as a way to share cybersecurity information with users. [19]. This method could involve sharing simulated phishing emails to evaluate errors in employee response, or a phishing attack followed by training sessions on how to address such attacks.

Game-based delivery methods are another popular method that offer the most effective way to share cybersecurity awareness programs among users. These methods integrate graphics and play, or "gamification", into the training session to create a compelling experience for participants. Game-based

delivery takes the players into a virtual space and simulates a real-world scenario to connect the players to circumstances and their consequences that may arise in the real world [20]. It offers the most effective way to share cybersecurity awareness programs among users. The primary advantage of the game-based delivery method for cybersecurity awareness is that it can challenge, motivate, and engage the target audience. Game-based delivery is a promising technique that can be combined with other approaches in a cybersecurity awareness training context to increase their effectiveness [21]. Game-based learning provides a safe environment in which players can practice their security behavior, and promotes self-learning across different learning styles [22]. Serious games also intrinsically motivate players to deal with game challenges by providing their own conclusions; as a result, they can improve their problem-solving skills [22]. Game-based learning offers an interactive method for training and educating the target audience on a specific subject or topic of social engineering. By playing the game, the target audience can enhance their skills during a fun and engaging experience. Game-based learning offers several advantages over more conventional means [20]. It can involve different case studies for motivating users in learning, according to their specific learning requirements.

Research on game-based learning suggests that it shows significant improvements in the level of players' understanding about setting strong passwords, identifying malware, the need for anti-malware programs, the nature of malware and phishing attacks, and using back-ups as a protective strategy. Game-based learning has achieved higher ratings in the heuristics evaluation process concerning awareness levels, usability, learning content, and fun and enjoyability features. Research found that participants prefer gamified environments over non-gamified environments [21]. Therefore, this may encourage organizations and government entities around the world, such as the computer emergency response team (CERT), to apply serious games to their awareness contexts to keep their users up to date about social engineering attacks and help reduce security breaches.

Also, it is recommended to use the Analytic Hierarchy Process (AHP) to support security decision-making to design effective security awareness programs [23]. AHP is a multiple-criteria decision-making tool used in several applications related to decision-making. One identifies the weights or the most important awareness focus area, to increase users' decision-making ability when they face real-world social engineering attacks. However, these weights may, in turn, be affected by local organizational and cultural factors. Users with the right security awareness method in place can potentially help prevent any kind of social engineering techniques. Table I shows a summary of seminal work in other cybersecurity serious games related to social engineering.

TABLE I. SUMMARY OF RECENT CYBER SECURITY SERIOUS GAMES

Game name	Game type	Goal	Results	Paper & Year
Untitled	Mobile Application.	To increase users' avoidance behavior through motivation to protect themselves against phishing threats.	The game had a positive effect on users' phishing avoidance behaviors.	[24] 2016
Untitled	Mobile Application	To increase participants' mobile threats awareness including phishing and cyber-attack.	The game increased the overall awareness of participants.	[25] 2017
<i>Bird's Life</i>	2D game via multiple platforms: PC, web, and mobile devices.	To improve phishing awareness among college students.	The game had a positive effect on players' understanding of phishing.	[26] 2018
Phishy	Web-based game portal	Phishy is a single-player game that trains enterprise users on phishing awareness.	Users showed a significant improvement in identifying phishing links.	[27] 2018
Pomega	2D Computer Game	To increase users' awareness across five topics: password, phishing, social network, mobile security, and physical security.	Users gained good knowledge about the five security awareness topics.	[28] 2019

### III. ISSUES RELATED TO CORONAVIRUS SOCIAL ENGINEERING ATTACKS

#### A. Coronavirus Malicious Attachments and Malware

Malicious email attachments, posing as PDF files or Microsoft Word documents, are designed to attack the victims' computer once they open them. The main idea behind malicious software (malware) is to cause data loss without the target's consent. According to the World Health Organization (WHO) [29], cybercriminals are taking advantage of the coronavirus outbreaks by sending malicious attachments and links via email or social media applications such as WhatsApp. Hackers and cyber scammers claim to originate from WHO or any public health facilities to send malicious email attachments in the form of advice or information about the coronavirus pandemic. The Emotet is the most popular coronavirus-related malware [30]. It targets banking to obtain financial information. This Trojan malware program is spread through spam emails via attachment files, or malicious links to steal sensitive data.

Coronavirus map is another popular malware designed to trick users who want to keep track of how the virus is spreading around the world. This malware aims to steal users' personal information such as username, password, and credit card details. If the coronavirus continues to spread across globe, more applications will be developed to monitor it which, in turn, motivate cybercriminals to spread malware using the coronavirus theme.

#### B. Coronavirus-Related Phishing, Vishing and Smishing

Phishing is another type of social engineering attack that aims to steal users' sensitive personal information or online banking details, and which can result in identity theft and financial loss. Cybercriminals take advantage of users' anxiety about their health or finances by manipulating their emotions (fear) so they fall victim to social engineering attacks [31]. Cybercriminals are taking advantage of people's need for the latest coronavirus information by spreading phishing websites or using fake coronavirus payment websites to steal their money. For instance, The National Cyber Security Centre (NCSC) has removed more than 2,000 online scams related to coronavirus, including 471 fake online shops, 555 malware

distribution sites, 200 phishing sites, and 832 advance-fee frauds [32]. Attackers also send phishing emails that appear to be from WHO or any authorized health organization to health care employees or users like an email with a PDF attachment that contains information about coronavirus safety measures shown in Fig. 1.

Spear phishing is advanced email phishing that targets well-researched victims such as high-level executives in an organization. The number of coronavirus-related spear-phishing attacks increased by 667% in March, 2020 [33]. According to Barracuda Networks [33] 11% of these attacks were blackmail attacks and conversation hijacking. Also, 77% of these attacks were scams, 22% were brand impersonation, and 1% were compromised business email. The goal of these attacks stealing users' personal information, distributing malware, and financial gain [33]. Attackers also use the Ransomware technique, which is a kind of malicious attack that blocks access to data or systems until a ransom is paid. Barracuda Networks noted that hackers have gained access to some users' information and threatened to infect it with coronavirus unless a ransom was paid.

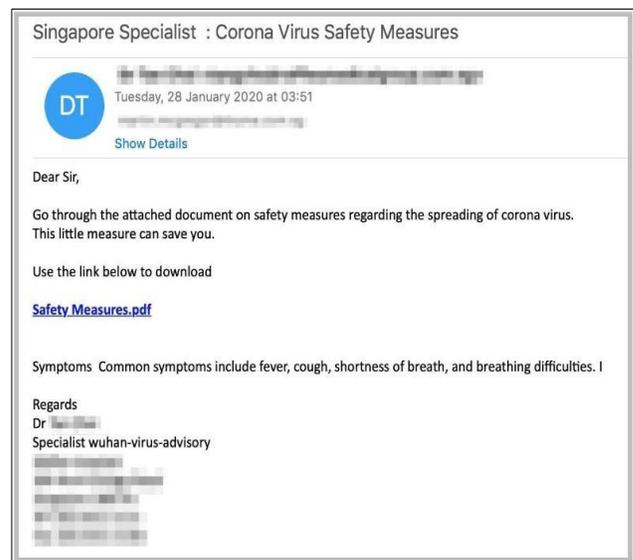


Fig. 1. An Example of a Phishing Attack [31].

Hackers also use SmiShing, or Short Message Service (SMS) phishing attacks, to trick users into downloading any type of malicious code such as viruses or trojans to their mobile devices. The number of smishing attacks increased to 15,9688 in South Korea in Feb, 2020 [34]. The scammers used false information such as providing free face masks or alerts from fake delivery companies. Another common smishing method is using the UK government theme to collect emails, sensitive information, or banking information (see Fig. 2) [35]. The text message tricks people and looks like it is from “COVID” and “UKGOV” including a line to a phishing website.

Smishing is similar to phishing but the hackers use SMS instead of email [3]. Once the victim clicks the link, malicious code is installed on the device.

Vishing or voice phishing is a social engineering technique used to scam users via phone. For instance, the scammer may pretend a WHO employee who is collecting donations or asking for personal information [31]. These three are the most common scams during coronavirus crisis.

The attacks will continue until a cure is found. Meanwhile, security researchers keep investigating these attacks and help users avoid falling victim to any type of social engineering attack. Google blocks more than 100 million phishing emails every day, and 18 million coronavirus-related malware and phishing emails during April 2020 [36]. Similarly, WHO warned people about phishing that appeared to be from WHO and ask for sensitive information such as usernames or passwords, or invite the reader to click a malicious link or open a malicious attachment [29]. Additionally, the NCSC is investigated a huge number of phishing campaigns that had a bad effect on sectors such as transport, engineering, and defense [37]. The campaign includes the same phishing email theme (see Fig. 1) either by malicious links or PDF attachments.

### C. Working Remotely due to the Coronavirus Pandemic

Due to the coronavirus pandemic, many organizations asked their employees to work from home and some staff may have no experience with working remotely. Therefore, this puts these organizations under pressure and raises more security challenges. As mentioned earlier, cybercriminals are exploiting coronavirus-related phishing via email to trick users online by sending suspicious emails contains malicious links or attachments. The scammers manipulate users' emotions by claiming to have a cure for coronavirus or offering financial support [38]. Such an attack may result in malicious code infection and data loss. Organizations need to support staff by providing awareness sessions to avoid such an attack.

A common attack related to working remotely is that hospitals' network gateway devices and Virtual Private Networks (VPN) vulnerabilities are targeted by ransomware campaigns called REvil or Sodinokibi, according to Microsoft [39]. Such an attack leads to stealing credentials and giving access to compromised networks to install ransomware or malware payloads [39]. Microsoft keeps advising organizations and people who use VPNs to work from home to apply security updates [40].



Fig. 2. An Example of Smishing Attack [35].

The following section provides recommendations and guidelines to mitigate the risk of coronavirus-related social engineering attacks among users. These recommendations can help individuals and industries prevent unauthorized access to their confidential information during this health crisis.

## IV. RECOMMENDATIONS

### A. Coronavirus Malicious Attachments and Malware

Cybercriminals are exploiting people's fear of the coronavirus to perform attacks across the globe. They lure the anxious through malicious attachments, fake websites, malicious links, or spam. To avoid falling victim to coronavirus-related malicious attachments and malware, users need to stay aware online and avoid opening email from an unknown address. However, if users' work requires dealing with customers by email, they need to check the attachments carefully, as any attachment might be malicious. Also, it is recommended to be aware of any email with “coronavirus” or “COVID-19” in its title; it should probably be ignored or deleted. People usually look for the latest information about coronavirus from WHO. Thus, it's worth reminding users that WHO never asks for personal information such as a password [29]. Also, WHO never asks to click on any link outside the WHO website and never asks for downloading email attachments that are not requested.

Moreover, users need to make sure to disable macros in their Microsoft Office applications to prevent macro malware [41]. Also, we recommend keeping the computer operating system and security software up to date, because hackers take advantage of security issues that manufacturers may have fixed with a patch, but before a user has installed the patch.

### B. Coronavirus-Related Phishing, Vishing and Smishing

In the current coronavirus situation, cybercriminals use social engineering techniques to trick and convince users to fall victim to their attacks by clicking on a bad link or downloading email attachments (phishing), via phone scam calls (vishing) or by sending SMS containing malicious links to smart devices (smishing). This kind of attack can result in stealing sensitive information from the computer or smart devices, and distributing malware.

The common scam trick is that scammers claim to offer treatment for the coronavirus and use “Buy now, limited supply” to manipulate people's emotions [42]. The user needs

to stay aware of any kind of request online or via phone that asks for personal information. Cybercriminals also use general greetings for phishing emails such as “Dear sir or madam”, without mentioning the receiver’s name (see Fig. 1) [42]. Another indication of a phishing email is including spelling and grammatical mistakes. User should delete such messages. Moreover, avoid visiting any website without checking its authenticity [43]. Websites which emails link to have a high risk of being malicious and including phishing links.

In general, NCSC listed some signs that may help people to avoid cyber scams [44].

- Authority: Scammers claim to be an official employee from an organization like a bank, healthcare sector, or government.
- Urgency: Scammers ask you to respond to their request immediately or within 24 hours.
- Emotion: Criminals manipulate people's emotions either by spreading fear or hope to trick them into cyber scams.
- Scarcity: Cybercriminals offer something with limited quantity such as medicine, face masks, or hand sanitizer. So people who fear missing a good deal may respond quickly.
- Current events: Scammers make use of current news stories or events such as tax refunds or donations for help to fight coronavirus to make their scam look real.

### C. Working Remotely

Due to the coronavirus pandemic, most organizations across globe have their employees work from home, which may raise security challenges. NCSC provides some general recommendations to mitigate the risk of working remotely [45].

- The organization needs to provide its employees with clear guidelines on how to use their software, regardless of their experience level.
- Employees are already under stress due to the coronavirus pandemic, so organizations need to support them all the time not only by providing technical guides.
- Increase awareness among employees to avoid cybersecurity attacks such as coronavirus-themed emails, phishing, and scams.
- Educate employees on how to report a problem, especially security incidents.
- Encrypt data to protect it if devices are stolen, and keep the VPN fully patched.
- Disable removable media to prevent coronavirus malware.

Also, things users can do [46]:

- Configure Wi-Fi encryption, especially in older installations.

- The anti-virus software and operating system must be fully updated.
- Security tools such as privacy tools and add-ons for browsers must be up to date.
- Perform regular backups to protect against damaged or stolen data.
- Set up a screen lock for when a device must be left unattended.
- Turn on two-factor authentication for all accounts.

Furthermore, users need to stay vigilant by understanding the coronavirus-related phishing and malware in (Section V(A) and V(B) to support secure remote working.

## V. CYBERSECURITY LESSONS FROM THE CORONAVIRUS PANDEMIC

Here are brief recommendations based on lessons learned during the coronavirus pandemic. Organizations and government entities need to rethink planning for a worst-case scenario which will be useful in unexpected crises such as a pandemic. This could be done by studying the risks and finding ways to mitigate them. To build skilled and educated users, proper security awareness methods should be implemented, as discussed in Section III. This could be done by enterprises and CERT across the world. However, awareness is temporal in nature and it must be refreshed or renewed regularly to keep users motivated and updated about changing attack methods and trends. New technologies are emerging in addition to quick and challenging security risks to organizations or individuals who must be updated to remain aware. Finally, enhance the cybersecurity infrastructure such as firewalls and anti-malware software, and keep monitoring all activity within the network to detect any kind of compromises.

## VI. CONCLUSION AND FUTURE WORK

In conclusion, social engineering attacks can not be stopped by using only technology or security systems. Social engineers manipulate the users to access sensitive information. During the coronavirus pandemic, cybercriminals use psychological and analytical techniques to manipulate users. This study attempts to help individuals and industries by reviewing the most common coronavirus social engineering attacks, and provides recommendations to respond to such an attack. Cybercriminals make use of people's fear and anxiety about coronavirus by spreading malicious attachments purporting to provide covid-19 information and malware via email across the globe. Coronavirus-related phishing, vishing, and smishing are growing in frequency and intensity. Such attacks increase the chances of distributing malicious code, stealing sensitive information, and ransomware attacks. The paper discussed cybersecurity issues related to working remotely and provides recommendations to avoid security breaches.

Future work may analyze these coronavirus-related security issues to implement detection and countermeasure techniques to reduce the success of social engineering attacks. Similarly, attention should be paid to implementing proper social engineering awareness to reinforce good security behavior.

REFERENCES

- [1] N. Pokrovskaja, S. S.- Management, T. And, and U. 2017, "Social engineering and digital technologies for the security of the social capital development," *ieeexplore.ieee.org*, pp. 16–18, 2017.
- [2] A. M. Aroyo, F. Rea, G. Sandini, and A. Scutti, "Trust and social engineering in human robot interaction: Will a robot make you disclose sensitive information, conform to its recommendations or gamble?," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3701–3708, 2018.
- [3] A. Yasin, R. Fatima, L. Liu, A. Yasin, and J. Wang, "Contemplating social engineering studies and attack scenarios: A review study," *Secur. Priv.*, vol. 2, no. 4, Jul. 2019, doi: 10.1002/spy2.73.
- [4] Verizon, "2019 Data Breach Investigations Report- Executive Summary," 2019. Accessed: May 12, 2019. [Online]. Available: <https://enterprise.verizon.com/resources/reports/2019-data-breach-investigations-report.pdf>.
- [5] C. Lively, "Psychological based social engineering," SANS Inst., 2003.
- [6] M. Bada, A. M. Sasse, and J. R. C. Nurse, "Cyber security awareness campaigns: Why do they fail to change behaviour?," *arXiv Prepr. arXiv1901.02672*, 2019.
- [7] I. Kirlappos, S. Parkin, and M. S. Society, "Shadow security as a tool for the learning organization," *ACM SIGCAS Comput. Soc.*, vol. 45, pp. 29–37, 2015.
- [8] M. Siponen and A. Vance, "Neutralization: new insights into the problem of employee information systems security policy violations," *MIS Q.*, pp. 487–502, 2010.
- [9] J. Stanton, K. Stam, P. Mastrangelo, and J. Jolton, "Analysis of end user security behaviors," *Comput. Secur.*, vol. 24, no. 2, pp. 124–133, 2005.
- [10] R. Cialdini and M. Trost, "Social influence: Social norms, conformity and compliance," McGraw-Hill, 1998.
- [11] M. T. Siponen, "An analysis of the traditional IS security approaches: implications for research and practice," *Eur. J. Inf. Syst.*, vol. 14, no. 3, pp. 303–315, Sep. 2005, doi: 10.1057/palgrave.ejis.3000537.
- [12] J. Leach, "Improving user security behaviour," *Comput. Secur.*, vol. 22, no. 8, pp. 685–692, 2003.
- [13] A. Alzahrani and C. Johnson, "Autonomy Motivators, Serious Games, and Intention Toward ISP Compliance," *Int. J. Serious Games*, vol. 6, no. 4, pp. 67–85, 2019.
- [14] A. Alzahrani, C. Johnson, and S. Altamimi, "Information security policy compliance: Investigating the role of intrinsic motivation towards policy compliance in the organisation," in 2018 4th International Conference on Information Management (ICIM), May 2018, pp. 125–132, doi: 10.1109/INFOMAN.2018.8392822.
- [15] R. Awais and B. Shreeve, "Decisions & Disruptions," Lancaster University, CC-BY-NC, 2017. <https://sites.google.com/view/decisions-disruptions/> (accessed Dec. 14, 2017).
- [16] J. D'Arcy, A. Hovav, and D. Galletta, "User awareness of security countermeasures and its impact on information systems misuse: a deterrence approach," *Inf. Syst.*, vol. 20, no. 1, pp. 79–98, 2009.
- [17] M. Wilson, J. H.-N. S. Publication, and U. 2003, "Building an information technology security awareness and training program," *citadel-information.com*, vol. 800, no. 50, pp. 1–39.
- [18] J. Valentine, "Enhancing the employee security awareness model," *Elsevier*, vol. 2006, no. 6, pp. 17–19, 2006.
- [19] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," *Commun. ACM*, vol. 50, no. 10, pp. 94–100, 2007.
- [20] S. Boyle, "An Introduction to Games based learning, sl: UCD Dublin," 2011.
- [21] I. Rieff, "Systematically Applying Gamification to Cyber Security Awareness Trainings," 2018.
- [22] A. Cook, R. Smith, L. Maglaras, and H. Janicke, "Using gamification to raise awareness of cyber threats to critical national infrastructure," 2016.
- [23] A. Alzahrani and C. Johnson, "AHP-based Security Decision Making: How Intention and Intrinsic Motivation Affect Policy Compliance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 6, 2019, doi: 10.14569/IJACSA.2019.0100601.
- [24] N. A. G. Arachchilage, S. Love, and K. Beznosov, "Phishing threat avoidance behaviour: An empirical investigation," *Comput. Human Behav.*, vol. 60, pp. 185–197, 2016, doi: 10.1016/j.chb.2016.02.065.
- [25] N. Micallef and N. A. G. Arachchilage, "Involving users in the design of a serious game for security questions education," *arXiv Prepr. arXiv1710.03888*, 2017.
- [26] P. Weanquoi, J. Johnson, and J. Zhang, "Using a Game to Improve Phishing Awareness," *J. Cybersecurity Educ. Res. Pract.*, vol. 2018, no. 2, p. 2, 2019.
- [27] G. C. J. S. Pandit, S. Vaddepalli, H. Tupsamudre, V. Banahatti, and S. Lodha, "Phishy-a serious game to train enterprise users on phishing awareness," in Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, 2018, pp. 169–181.
- [28] V. Visoottiviset, R. Sainont, T. Boonnak, and V. Thammakulkrajang, "POMEGA: Security game for building security awareness," in Proceeding of 2018 7th ICT International Student Project Conference, ICT-ISPC 2018, 2018, pp. 1–6, doi: 10.1109/ICT-ISPC.2018.8523965.
- [29] World Health Organization, "Beware of criminals pretending to be WHO," 2020. <https://www.who.int/about/communications/cyber-security>.
- [30] K. Okerefor and O. Adebola, "Tackling The Cybersecurity Impacts of the Coronavirus Outbreak as a Challenge to Internet Safety," *J. Homepage http://ijmr.net.*, vol. 8, no. 2, 2020.
- [31] C. Conley, "Coronavirus Cyber Attacks - The Secret Threat | SANS Security Awareness," 2020. <https://www.sans.org/security-awareness-training/blog/coronavirus-cyber-attacks-secret-threat> (accessed Apr. 24, 2020).
- [32] National Cyber Security Centre, "Public urged to flag coronavirus related email scams as... - NCSC.GOV.UK," 2020. <https://www.ncsc.gov.uk/news/public-urged-to-flag-covid-19-threats-new-campaign> (accessed Apr. 24, 2020).
- [33] S. Magazine, "Coronavirus-Related Spear Phishing Attacks See 667% Increase in March 2020 | 2020-04-16 | Security Magazine," 2020. <https://www.securitymagazine.com/articles/92157-coronavirus-related-spear-phishing-attacks-see-667-increase-in-march-2020> (accessed Apr. 24, 2020).
- [34] C. Mu-Hyun, "South Korea sees rise in smishing with coronavirus misinformation," *ZD Net*, 2020. <https://www.zdnet.com/article/south-korea-sees-rise-in-smishing-with-coronavirus-misinformation/#ftag=RSSbaffb68> (accessed Apr. 24, 2020).
- [35] National Cyber Awareness System, "COVID-19 Exploited by Malicious Cyber Actors," 2020. <https://www.us-cert.gov/ncas/alerts/aa20-099a> (accessed Apr. 24, 2020).
- [36] N. Kumaran and S. Lugani, "Protecting against cyber threats during COVID-19 and beyond," 2020. <https://cloud.google.com/blog/products/identity-security/protecting-against-cyber-threats-during-covid-19-and-beyond> (accessed Apr. 24, 2020).
- [37] National Cyber Security Centre, "Phishing campaign," 2018. <https://www.ncsc.gov.uk/news/phishing-campaign> (accessed Apr. 24, 2020).
- [38] National Cyber Security Centre, "NCSC issues guidance as home working increases in response," 2020. <https://www.ncsc.gov.uk/news/home-working-increases-in-response-to-covid-19> (accessed Apr. 25, 2020).
- [39] Microsoft Security, "Microsoft works with healthcare organizations to protect from popular ransomware during COVID-19 crisis: Here's what to do," 2020. <https://www.microsoft.com/security/blog/2020/04/01/microsoft-works-with-healthcare-organizations-to-protect-from-popular-ransomware-during-covid-19-crisis-heres-what-to-do/> (accessed Apr. 24, 2020).
- [40] National Cyber Security Centre, "Weekly Threat Report 3rd April 2020 - NCSC.GOV.UK," 2020. Accessed: Apr. 24, 2020. [Online]. Available: <https://www.ncsc.gov.uk/report/weekly-threat-report-3rd-april-2020>.
- [41] Norton, "What Are Malicious Websites?," 2020. <https://us.norton.com/internetsecurity-malware-what-are-malicious-websites.html> (accessed Apr. 25, 2020).

- [42] S. Symanovich, "Coronavirus phishing emails: How to protect against COVID-19 scams," Norton, 2020. <https://us.norton.com/internet-security-online-scams-coronavirus-phishing-scams.html> (accessed Apr. 26, 2020).
- [43] T. Ahmad, "Corona Virus (COVID-19) Pandemic and Work from Home: Challenges of Cybercrimes and Cybersecurity," Available SSRN 3568830, 2020.
- [44] National Cyber Security Centre, "Phishing attacks: dealing with suspicious emails and messages," 2018. <https://www.ncsc.gov.uk/guidance/suspicious-email-actions> (accessed Apr. 26, 2020).
- [45] National Cyber Security Centre, "Home working: preparing your organisation and staff," 2020. <https://www.ncsc.gov.uk/guidance/home-working> (accessed Apr. 26, 2020).
- [46] ENISA, "Top Tips for Cybersecurity when Working Remotely," 2020. <https://www.enisa.europa.eu/news/executive-news/top-tips-for-cybersecurity-when-working-remotely> (accessed Apr. 26, 2020)

# Hybrid based Energy Efficient Cluster Head Selection using Camel Series Elephant Herding Optimization Algorithm in WSN

N. Lavanya<sup>1</sup>, T. Shankar<sup>2</sup>

Department of Communication Engineering  
School of Electronics Engineering  
Vellore Institute of Technology, Vellore, Tamil Nadu, India

**Abstract**—The rapid growth in wireless technology is enabling a variety of advances in wireless sensor networks (WSNs). By providing the sensing capabilities and efficient wireless communication, WSNs are becoming important factor in day to day life. WSNs have many commercial, industrial and telecommunication applications. Maximizing network lifespan is a primary objective in wireless sensor networks as the sensor nodes are powered by a non-rechargeable battery. The main challenges in wireless sensor networks (WSNs) are area of coverage, network's lifetime and aggregating. Balanced node establishment (clustering) is the foremost strategy for extending the entire network's lifetime by aggregating the sensed information at the head of the cluster. The recent research trend suggests Meta-heuristic algorithms for the intelligent selection of ideal Cluster Heads (CHs). The existing Cluster Head Selection (CHS) algorithm suffers from the inconsistent trade-offs between exploration – exploitation and global best examine constraints. In this research, a novel Camel series Elephant Herding Optimization (CSEHO) algorithm is proposed to enhance the random occurrences of Camel algorithm by the Elephant Herding Optimization algorithm for optimal CHS. The Camel algorithm imitates the itinerant actions of a camel in the desert for the scavenging procedure. The visibility monitoring condition of the camel algorithm improves the efficiency of exploitation, whereas the exploration inefficiency of a Camel algorithm is compensated optimally by the Elephant Herding Optimization operator (Clan and separator). The superior performance of the proposed CSEHO algorithm is validated by comparing its performance with various other existing CHS algorithms. The overall attainment of the offered CSEHO algorithm is 21.01%, 31.21%, 44.08%, 67.51%, and 85.66%, better than that of EHO, CA, PSO, LEACH, and DT, respectively.

**Keywords**—Camel algorithm; Cluster Head Selection; Elephant Herding Optimization; meta-heuristic algorithm; network lifespan; wireless sensor network

## I. INTRODUCTION

A sensor network incorporates multiple minimal prices and low-battery sensor nodes. Every sensor nodes are situated in a specific zone and typically install a remote system by the method of self-sorting out. The sensor nodes can investigate typically at some of the extraordinary and critical circumscribe that individuals cannot handle [1]. Due to various complex factors, information communication between nodes in an effective manner is almost impossible. An energetic clustering

is a well-known system to increase the effectiveness of sensed data transmission. The fundamental concept of clustering is to boundaries around the nodes of the active sensor into separate groups and chooses leaders for all groups. The groups are known as clusters while group leaders are known as Cluster Heads (CHs) of the groups to perform data communication [2]. In such condition, electing the CH under several constraints such as less amount of energy consumption, distance, delay and so on is the vital character of every clustering algorithm. In the real-world, CH renewed through different iterations contributes towards the finest attainment. The individual cluster comprises a CH with finite cluster associates. The responsibility of the CH is that it should organize each and every node existing in the cluster.

Information communicated to the sink node (BS) from every sensor node over an energetic CH is an important experiment exhibited for the routing technique [3]. Best CH election structure results in reduction of transmission energy, transmission time, transmission distance and so on. This paper besides deliberates some energy-efficient clustering algorithm in a wireless sensor network and introduces an optimal meta-heuristic algorithm for efficient CH election in WSN. Among the several meta-heuristic optimization techniques, maximum of them were implemented to achieve the best CH election in WSN. Meta-heuristics are generally inspired by physical phenomena, animal behaviour, or evolutionary concepts [4]. The ease enables a researcher to simulate various natural ideas, suggest fresh algorithms, integrate many meta-heuristics, or enhance present meta-heuristics. This further, helps other researchers to rapidly know and apply meta-heuristics to their issues. In general, meta-heuristics may be categorized into two foremost groups: single-value established and a set solution (population) established. In the first case, the search method begins with one candidate solution. In the course of iterations, this single candidate solution is then enhanced. However, population-based meta-heuristics use a set of solutions (population) to conduct optimization. In this technique, the exploration procedure begins with an arbitrary primary population (set solutions), and during the progression of rounds, this population is improved. Based on the shared information, population-based meta-heuristics can suddenly jump towards a better optimal location and avoid local optimal location and superior exploration than the single value-established algorithms [4]. The existing meta-heuristic method

frequently suffers from numerous problems like an exploration - exploitation tradeoff, moderate convergence speed and in addition it is not as much of ability in considering the multi-objective. This inspires us to obtain a global search with fast convergence; a Camel series Elephant Herding Optimization (CSEHO) algorithm is proposed to select an energy efficient cluster head selection in WSN.

Camel algorithm (CA) is the latest renowned optimizer unit that works on the itinerant performance of a camel in the desert for scavenging procedures. Camels have the ability to identify the optimal food source location present in the desert and also share the identified optimal food location to other camels that are present in the visibility region [5]. Hence for exploitation, a camel algorithm is employed. For a further broader search, Elephant Herding Optimization (EHO) is employed in the proposed work. EHO algorithm is constructed on the assembling attitude of elephant groups under the supervision of an elephant queen (matriarch). The main aim of the CSEHO is to prolong the lifespan of the modelled WSN by employing EHO for exploration and CA for exploitation. This proposed CSEHO technique combines the advantages of both CA and EHO, bringing the equilibrium among the exploitation and exploration stages of optimization, resulting in better performance of the network.

The remaining portions of the research article are arranged in a manner listed as Section 2 deliberates the associated works proposed in the literature review; Section 3 elaborates on the method offered; Section 4 presents the simulation results and discussion; Section 5 presents the conclusion for the work implemented and the scope to extend further.

## II. LITERATURE REVIEW

Heinzelman, W.B. et al. (2002) [3,6] proposed a centralized protocol called LEACH-C, in this the responsibility of the CH election and the delivery of info into the sensor network are maintained by the Sink node. Because of the fact that the steady-state level is entirely implemented at the Sink node, sensor nodes are not affected by overheads at the period of cluster establishment. The drawback of the LEACH protocol is that it performs an arbitrary election of CHs, which election inefficient CHs and thus makes to extremely unproductive lifespan and energy absorbent by the network. Younis, O., & Fahmy, S. (2004) [7] proposed a Hybrid Energy Efficient Distributed (HEED) clustering algorithm. The residual energy of nodes and intra communication costs are the two main parameters that have been used in this approach for the selection of CH among sensor nodes. HEED provides even distribution of CHs, and the chances for two nodes within the same communication range can be selected as CHs is avoided. The major drawback found in this method is the overhead caused by energy dissipation. To extend the life span of a WSN, nodes are determined to sleep by the Sleep Scheduling methodology. Basically, Sensor nodes are very lucky to sleep consuming Sleep Scheduling mechanisms structured by the base station to increase energy efficient management [8].

In recent years several investigators have established natural inspiration based optimization techniques that imitate certain genetic actions or physical occurrences. Hussain et al. investigated smart techniques using a genetic algorithm (GA)

for cluster construction and control. The optimal cluster head selection process is carried out at the base station using GA [9]. The GA is sensitive to the primary population and it may converge at local minima. The authors in [10,11] have recommended a PSO algorithm that is founded based on the natural activities of a group of birds. The PSO uses the cumulative effect of personal best and global best to generate a new direction for each particle in the search region. The gravitational search algorithm for the node localization in WSN was presented in [12]. Hence it needs an improvement in energy consumption and stability of localization method. In [13], enhanced PSO based clustering energy optimization algorithm (EPSO-CEO) in WSN for CH selection is done through PSO algorithm and concerning the less power consumption in WSN with high delay.

In [14], an optimal based clustering technique using artificial bee colony (ABC) algorithm and fractional calculus were presented for the purpose of maximizing the network energy and lifetime. The cluster heads are selected optimally using the ABC algorithm. In [15], Dynamic clustering based routing protocol with the generalized ACO algorithm was presented to increase the lifespan of SNs by energy constraints. The protocol consumes around 25% of the whole energy consumption of the network for data transmission. The overhead of this approach is relatively high and so the network lifetime is relatively short. In [16], Harmony search algorithm (HSA) based CH selection algorithm was offered to solve the wide range of NP hard problem with effective mapping and fitness function by means of energy, distance and node degree. This technique doesn't consider fault tolerance and delay.

Yang [17] proposed a flower pollination algorithm initiated on the development of flower pollination of blossoming flowers seeing the biological reproduction cycle. The researcher focuses on the specialty of sunflowers' motion in the exploration for the best orientation in the direction of the sun. The random fertilization is considered between the marginal distances of sunflower  $i$  and  $i+1$ . In general, millions of pollen gametes are frequently released by every flower patch. For ease, we consider that every sunflower generates only one game of pollen and reproduces individually. The authors in [18] suggested a non-dominated fashion of cluster head selection using a multi-objective evolutionary algorithm (MOEA) called NSGA-II for extending the lifespan of the WSN. The energy consumption function is analysed to improve the life period of the WSN. The network loss the node at the prime stage is a major limitation of this method. The authors in [19,20] suggested a routing method to incorporate cluster establishment, and multipath routing through reduced energy consumption and routing overheads.

Several research undertakings the hybridization of two optimization algorithms to overcome the disadvantages of the algorithms working independently. Shankar, T. et al [21] suggested a hybrid HSA-PSO bring exploration-exploitation trade-off in the optimization problem of cluster head selection in WSNs. This method combines extraordinary examine efficiency of HSA and dynamic nature of PSO, to produce an improved performance. The performance in terms of the first dead node is good. However, the last dead node round number needs improvement. Lavanya, N. et al. [22] suggested a hybrid

squirrel harmony search algorithm for extending the lifespan of the WSN. The Squirrel search algorithm is established by engaging the method of improving the positions of the squirrels with respect to predator presence probability and seasonal monitoring condition. The stability among the exploitation and the exploration is achieved by the gliding constant. The performance in terms of the last dead node is good. However, the residual energy needs improvement.

From the literature, the single optimization methods either excel in the exploration phase of optimization or the exploitation phase, but lags in balance among the exploration and exploitation phases. Hence the proposed work presents the hybridization of EHO algorithm for replacing the random search of a camel algorithm to balancing exploration and exploitation, thereby providing better performance in the optimal selection of CHs in WSN.

### III. RESEARCH METHODOLOGY

#### A. Wireless Sensor Network Model

In this research, the overall transmission energy of the SN is considered by free space network model as given in equation (1-2). It is assumed that the separation distance between the receiver unit and transmitter unit is  $x$ , the information length is  $I$ , and a group of sensor nodes are scattered in a rectangle fashion [18].

$$E_{Tx}(I, x) = \begin{cases} IE_{ec} + IE_{fs}x^2, & x < x_0 \\ IE_{ec} + IE_{amp}x^4, & x \geq x_0 \end{cases} \quad (1)$$

$$E_{Rx}(I) = IE_{ec} \quad (2)$$

where  $E_{ec}$  is the electronics energy;  $E_{Rx}$  signifies the energy used up by the receiver for the communication of bit length  $I$ , and  $E_{Tx}$  signifies the energy used up by the transmitter for the same information. In the molded WSN,  $n$  number of sensors is arbitrarily settled in a field of area 'M x N' m<sup>2</sup>. For computing the optimal 'k' Cluster head, the following fitness function is employed [13]:

$$f_{obj} = \varepsilon \times f_1 + (1 - \varepsilon) \times f_2 \quad (3)$$

where  $f_1 = \max_k \left\{ \sum_{\forall node, i \in C_k} \frac{d(node_i, CH_k)}{|Cluster_k|} \right\}$  and  $f_2 = \frac{\sum_{i=1}^N E(node_i)}{\sum_{j=1}^k E(CH_j)}$  are distance and energy optimization fitness function;  $\varepsilon \in [0, 1]$ ;

#### B. Camel Algorithm

Mohammed et al. (2016) proposed a meta-heuristic optimization method which imitates the itinerant performance of a camel in the desert for scavenging process. The Camel familiarizes itself with the desert to tolerate scarcity of water and high temperature for extended periods of time so that it can survive with limited or insufficient of existing food resources. Over the snowy and cold times, the camel can manage themselves without intake for more than a few months. The camels spread through an influenced area in search of the food spot. When a certain camel enters a food resource, it interacts with the other members of the caravan to follow that commercial food resource. The caravan members can adjust their route until they reaches the optimal food resource.

Because of the presence of sand dunes, the camel may fail to follow the optimal path to the food resource and start searching alternate food sources in a random location. Other camels may identify improved food locations during this trip, so other camels can change their route based on the received information about the new location. The cycle goes on until the camels grasp an oasis.

#### C. Camel Algorithm Implementation

Camel algorithm starts when random position of camels begins scavenging. The sand dunes and distance affects the camel's visibility. The following steps are implemented in the camel algorithm:

Step 1: Random initialization: In a desert, it is considered that there is  $N$  number of camels scattered randomly to identify the optimal food source and the initial position of all the camels is given in the following equation:

$$x_i = x_{min} + rd \times (x_{max} - x_{min}) \quad (4)$$

where  $x_{min}$  and  $x_{max}$  are minimum and maximum of search bounds correspondingly; whereas  $rd$  refers to arbitrarily created value well-defined in the period  $[0, 1]$ .

Step 2: Fitness assessment: Based on the initial energy and position the fitness standards of every camel are obtained, consequently they are arranged in an ascending fashion. The lowest fitness is declared as global optimal solution (global\_best).

Step 3: The camel endurance ( $E$ ): The travelling distance particularly impacts the camels' energy and it reduces the camels' endurance directly. The multiple camels travel along different paths toward the optimal solution and meeting the different values of distance resultant in dissimilar endurance value for each individual. The distance  $D$  of camel  $i$  at the iteration  $t$  is given as follows:

$$D_i(t) = D_{min} + rd \times (D_{max} - D_{min}) \quad (5)$$

where  $D_{min}$  and  $D_{max}$  are minimum and maximum of travelling distance; whereas  $rd$  refers to arbitrarily created value well-defined in the period  $[0, 1]$ .

It is suggested that the impact of distance on camel endurance  $E$  as given in the following equation:

$$E_i(t) = 1 - \frac{(D_i(t) - D_{min})}{(D_{max} - D_{min})} \quad (6)$$

where  $D_i(t)$  refers to initial distance of the camel. The equation (6) shows that, the distance is inversely related to the camel endurance.

Step 4: New location generation: The sand dunes may fail to update the camel location towards an optimal food resource identified by other camels. The searching movement of the camels can be modelled accurately from the following two Scenarios:

Scenario 1: The visibility of the camel is greater than the visibility threshold. The new location updates function as given in the following equation:

$$x_i(t+1) = x_i(t) + E_i(t) \times (x_{global\_best} - x_i(t)) \quad (7)$$

where  $x_{global\_best}$  and  $x_i(t)$  are refers to global best and initial position,  $E_i(t)$  is the endurance.

Scenario 2: The visibility of the camel is lesser than the visibility threshold. On bearing this condition, the random updating of the camels is accomplished in the following equation (8):

$$x_i(t + 1) = x_{min} + rd \times (x_{max} - x_{min}) \quad (8)$$

where  $x_{min}$  and  $x_{max}$  are minimum and maximum of search bounds correspondingly; whereas rd refers to arbitrarily created value well-defined in the period [0, 1].

Step 5: New locations fitness assessment: Evaluate the fitness for all the new locations and assign the new best location as the global best if it is improved than the existing global best.

Step 6: Stopping Criterion: The procedure is recurrently executed for a defined number of search iterations. The pseudocode of camel algorithm for header node selection as follows:

Pseudocode of Camel Algorithm:

```

Initializing the WSN and CA parameters
 $r_{max} \leftarrow$  Number of iterations for Data Transmission
 $NI \leftarrow$  Internal iteration for CHS using camel search
 $N_{CA} \leftarrow$  camel search matrix size
for  $count \leftarrow 1$  to  $r_{max}$  do
  To create an arbitrary determination of normal nodes
  for  $i \leftarrow 1$  to  $NI$  do
    for  $j \leftarrow 1$  to  $N_{CA}$  do
       $CA(j,:) \leftarrow$  Arbitrarily chosen cluster head for CA
       $f_{obj}(j) \leftarrow$  Fitness standards for  $CA(j,:)$  using equation (3)
    end
    Arrange the Fitness standards of CA related to  $f_{obj}$ 
    Declare  $global\_best$  from the fitness values
    for  $j \leftarrow 1$  to  $N_{CA}$  do
      if  $v >$  visibility threshold then
        Update the camel position using equation (7)
      else
        Update the camel position using equation (8)
      end
    end
    if updated  $global\_best >$  old  $global\_best$ 
       $global\_best \leftarrow$  new  $global\_best$ 
    end
    Update the camel visibility
  end
  Selecting the  $global\_best$  as the Cluster heads for Data Transmission
end

```

#### D. Elephant Herding Optimization (EHO) Algorithm

G.G.Wang et al. [23] proposed a new meta-heuristic optimization method which imitates the assembling attitude of elephant groups represented as the Elephant Herding Optimization (EHO) algorithm. The elephant population is divided into a small number of subgroups called as a clan

under the supervision of an elephant queen (matriarch). When the elephant king (male elephant) maturing up, he would like to stay away from the family and contact with the clan at low frequency. The supervision and stay away nature of the elephant builds the EHO algorithm into two operatives: clan updating (supervising) operative and separating (staying away) operative. The elephants are updated using their present location and matriarch through clan updating operative, and the separating operative is then executed.

#### E. Mathematical Model of Elephant Herding Optimization (EHO) Algorithm

The process of EHO is initialized by the leadership activities of elephant group. In woodland, it is considered that there is P number of elephants scattered randomly and this elephant population (P) is separated into N number of clans. Each clan may update the position under two operatives [24]. The elephant clan updating behaviour is represented as follows.

Clan operative: The elephants survive in an organized manner under the headship of an elephant queen (matriarch) in each set. As a result, for all the elephants in clan  $c_{ni}$ , its update location is determined by the elephant queen (matriarch)  $c_i$ . The updated locations of elephant  $j$  in clan  $c_{ni}$  as given in the subsequent equation:

$$x_{new,C_{ni},j} = x_{C_{ni},j} + \varepsilon \times rd \times (x_{best,C_{ni}} - x_{C_{ni},j}) \quad (9)$$

where  $x_{new,C_{ni},j}$ ,  $x_{C_{ni},j}$  and  $x_{best,C_{ni}}$  are recently updated location, past location and fittest location of elephant  $j$  in clan  $C_{ni}$ , whereas scaling parameter ( $\varepsilon$ ) and rd refer to arbitrarily created values well-defined in the period [0, 1].

Updating the fittest elephant: Based on the clan members, the elephant queen (fittest elephant) can update its own location in following equation:

$$x_{new,C_i,j} = \delta \times (x_{center,C_{ni}}) \quad (10)$$

where  $\delta$  refer to arbitrarily created values well-defined in the period [0, 1], whereas  $x_{center,C_{ni}}$  represents the midpoint of clan  $c_{ni}$ . It is obtained by equation (11):

$$x_{center,C_{ni},d} = \frac{1}{n_{C_{ni}}} \times \sum_{j=1}^{n_{C_{ni}}} x_{C_{ni},j}, d \quad (11)$$

Separating operative: In elephant clusters, the male elephant will live separately from the clan when they reach maturity. This separating procedure offering the separating operator presenting each generation as the elephant individuals with the poorest fitness, which is realized as follows:

$$x_{worst,C_{ni}} = x_{min} + (x_{max} - x_{min} + 1) \times rd \quad (12)$$

where  $rd \in [0,1]$ ;  $x_{min}$  and  $x_{max}$  are minimum and maximum of search bounds; whereas  $x_{worst,C_{ni}}$  refers to the lowest fitness in clan  $c_{ni}$ .

New location fitness assessment: Evaluate the fitness for all the new locations and assign the new best location as the global best if it is improved than the existing global best.

Stopping Criterion: The procedure is recurrently executed for a defined number of search iterations. The pseudocode of EHO algorithm for header node selection as follows:

Pseudocode of EHO Optimization

```

Initializing the WSN and EHO parameters
rmax ← Number of iterations for Data Transmission
NI ← Internal iteration for CHS using EHO
NEP ← Number of CHS
ni ← Number of nodes in each CH
for count ← 1 to rmax do
  To create an arbitrary determination of normal nodes
  for l ← 1 to NI do
    for j ← 1 to NEP do
      Arrange all the elephant in Cni related to fObj
      xbest,Cni = Best fitness elephant
      xworst,Ci = lowest fitness elephant
      clan updating operative:
      for j = 1 to NEP do
        for k = 1 to ni do
          Update all the elephant in Cni using equation 9 and 10
        end
      end
      Separating operative:
      for j = 1 to NEP do
        Substitute xworst,Ci using equation 12
      end
      Evaluate recently updated locations
    end
    Choosing the global_best as the Cluster heads for Data Transmission
  end

```

F. Proposed HCSEHO Algorithm for Optimal Cluster Head Selection

The proposed camel series elephant herding optimization algorithm integrates the general attributes of camel algorithm and elephant herding optimization algorithm. The camel visibility is directly related to distance and sand dunes. The designed objective of camel algorithm is to identify the optimal location for all the nodes based on the camel visibility. The visibility monitoring condition is established in camel algorithm to exclude the offered technique from being bound to the results of local best and provide an accurate approach towards reaching the optimal result. Even though the visibility monitoring condition is established in camel algorithm, still it faces the high dimensionality limitation. In order to achieve the exploration and exploitation with high search efficiency in high dimensionality, the randomly searching process of the camel algorithms is switched by EHO algorithm. The flow figure of the offered HCSEHO scheme is presented in the Fig. 1.

In HCSEHO algorithm, employing camel algorithm allows the camels to travel around from one region to other region by

modifying the positions and visibility at the end of every round. The energetic performance of Camel algorithm decides to find optimal cluster heads (CHs) present in the n number of clusters with high search efficiency of EHO algorithm. The frame work of the proposed CSEHO algorithm is to perform energy proficient transmission from every member nodes present in the different cluster to the base station through optimal CHs. The following steps are implemented in HCSEHO algorithm for optimal cluster head selection:

Step 1: Initialization of WSN: Initialize the sensor network as per the constraints mention in parameter (Table I).

Step 2: Fitness assessment and global best declaration: The fitness assessment for each node is obtained with the help of objective function given in equation (3). Subsequently it is arranged in an ascending manner and the least solution of fitness is assigned to global best.

Step 3: Visibility check: The foraging of camels is impacted due to present of sand dunes. The visibility monitoring condition offers an accurate approach in the direction of the optimal result. If the condition for visibility checking is satisfied, the camel position is updated using camel algorithm. Else, it follows EHO algorithm for optimal cluster heads selection.

Step 4: Stop criterion: The procedure is recurrently executed for a defined number of search iterations.

Step 5: Data Transmission: Data transmission is initiated among the optimal CHs and the Sink node. This procedure is repeated for each iteration of information communication and continued until the iteration reaches a maximum number of rounds.

TABLE I. SIMULATION PARAMETERS OF HCSEHO ALGORITHM

Parameter	Value
Sensor search section (m <sup>2</sup> )	(200*200)
Number of nodes (n)	100
Initial energy of a node(E <sub>o</sub> ) (J)	0.5
Information length (I) (bits)	4096
E <sub>ec</sub> (nJ/bit)	70
E <sub>amp</sub> (pJ/bit/m <sup>2</sup> )	120
Energy data aggregation(nJ)	5
Number of Number of iterations for Data Transmission (r <sub>max</sub> )	3000
Number of iterations for cluster head selection	5
Total optimal Cluster Heads (k)	5
Visibility threshold	0.1
Scale factor(α,β)	α ∈ [0, 1] δ ∈ [0, 1]
Particle location [x <sub>max</sub> , x <sub>min</sub> ]	[0,200]

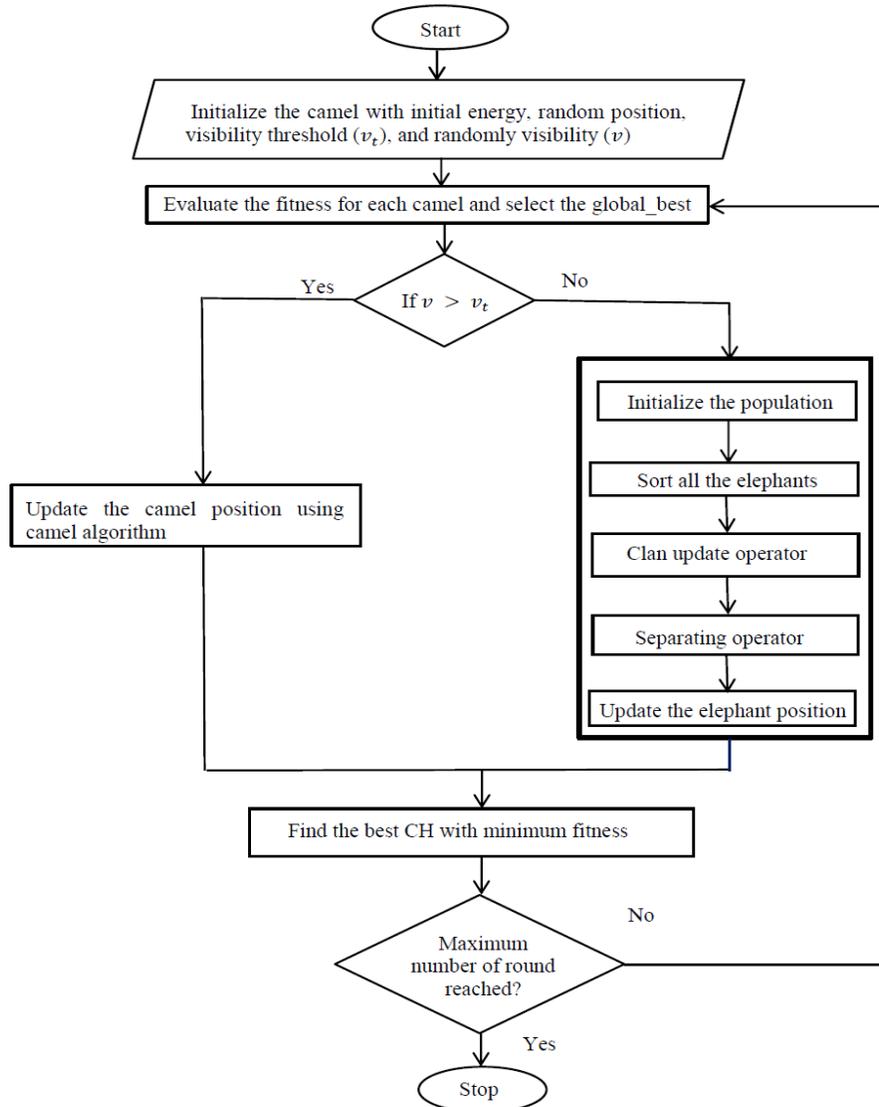


Fig. 1. Flowchart of the Proposed HCSEHO.

#### IV. RESULTS AND DISCUSSION

The simulation of the recommended HCSEHO method of CHS is executed in MATLAB R2018a plat form. The necessary parameters of the simulation taken into consideration for the offered technique are specified in Table I. The recommended HCSEHO is analyzed by comparing with the surviving CHS methods explicitly, Direct Transmission, LEACH [6], PSO [13], CA [5], and EHO [23].

Fig. 2 shows the comparative illustration of the performance of the different algorithms in terms of count of the alive nodes for an increasing number of rounds of data transmission. The first node in the WSN for the proposed HCSEHO is alive till the round number 2038, whereas, for the algorithms: Camel, EHO, PSO, LEACH, and DT, the nodes die at 1298, 1602, 1201, 270, and 44, respectively. Half of the total number of nodes in the WSN for the proposed HGWSFO is alive till the round number 2147, whereas, for the algorithms: Camel, EHO, PSO, LEACH, and DT, the nodes die 1445,

1710, 1222, 370, and 74, respectively. All the nodes in the WSN for the proposed HGWSFO stay alive till the round number 2198, whereas, for the algorithms: Camel, EHO, PSO, LEACH, and DT, the nodes stay alive till 1512, 1736, 1229, 714, and 315 rounds, respectively. The lifetime of the nodes for the proposed HCSEHO is 31.21% more than the Camel algorithm, 21.01% more than the EHO algorithm, 44.08% more than the PSO algorithm, 67.51% more than the LEACH protocol, and 85.66% more than the DT.

Fig. 3 shows the comparative illustration of the performance of the different algorithms in terms of count of the dead nodes for an increasing number of rounds of data transmission. The first node in the WSN for the proposed HCSEHO dies at the round number 2038, whereas, for the algorithms: Camel, EHO, PSO, LEACH, and DT, the nodes die at 1298, 1602, 1201, 270, and 44, respectively. Half of the total number of nodes in the WSN for the proposed HCSEHO die at the round number 2147, whereas, for the algorithms: Camel, EHO, PSO, LEACH, and DT, the nodes die at 1445, 1710,

1222, 370, and 74, respectively. All the nodes in the WSN for the proposed HCSEHO die at the round number 2198, whereas, for the algorithms: Camel, EHO, PSO, LEACH, and DT, the nodes die at 1512, 1736, 1229, 714, and 315, respectively. The lifetime of the nodes for the proposed HCSEHO is 31.21 % more than the Camel algorithm, 21.01% more than the EHO algorithm, 44.08% more than the PSO algorithm, 67.51% more than the LEACH protocol, and 85.66% more than the DT.

Fig. 4 shows the comparative illustration of the performance of the different algorithms in terms of residual energy in J for an increasing number of iterations of information transmission. The residual energy in the WSN for all the algorithms at the initial round is 50 J. When 1500 rounds are reached, the residual energy of the proposed HCSEHO is 15.49 J, whereas, for the algorithms: Camel and EHO, the residual energies are 0.2251 J and 2.585 J, respectively. The residual energies of PSO, LEACH, and DT become zero at the round number 1500. The residual energy of the proposed HCSEHO at 1700th round is 10.94 J. The residual energy of the proposed HCSEHO declines to zero by the round number 2198, while for the algorithms: Camel, EHO, PSO, LEACH, and DT, the residual energy declines at 1512, 1736, 1229, 714, and 315, respectively. The residual energy of proposed HCSEHO is 52.65% more than the Camel algorithm, 39% more than the EHO algorithm, 67% more than the PSO algorithm, 87.10% more than the LEACH protocol, and 95.10% more than the DT. The following table.2 shows the comparative analysis of the performance of various CHS methods.

Fig. 5 shows the comparative illustration of the performance of the different algorithms in terms of throughput obtained in bps for an increasing number of rounds of data transmission. The throughput in the WSN for all the algorithms at the initial round is 409600 bps. When 1700 rounds are reached, the throughput of the proposed HCSEHO and EHO are 409600 bps and 282653 bps, whereas, for the algorithms: Camel, PSO, LEACH, and DT the throughput is 0 bps. The throughput of the proposed HCSEHO declines to zero at the round number 2198, while the algorithms: Camel, EHO, PSO, LEACH, and DT, declines at 1512, 1736, 1229, 714, and 315, respectively. The throughput lifetime of the proposed HCSEHO is 32.69% more than the Camel algorithm, 20.35% more than the EHO algorithm, 43% more than the PSO algorithm, 83% more than the LEACH protocol, and 95% more than the DT.

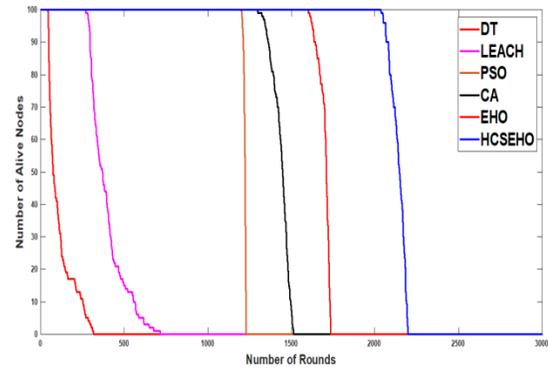


Fig. 2. Comparison of Alive Nodes Obtained for different CHS Algorithms.

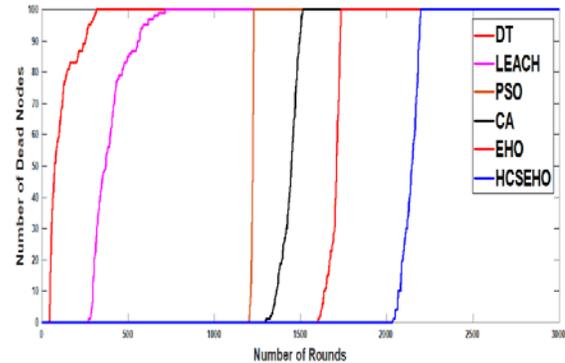


Fig. 3. Comparison of Dead Nodes Obtained for different CHS Algorithms.

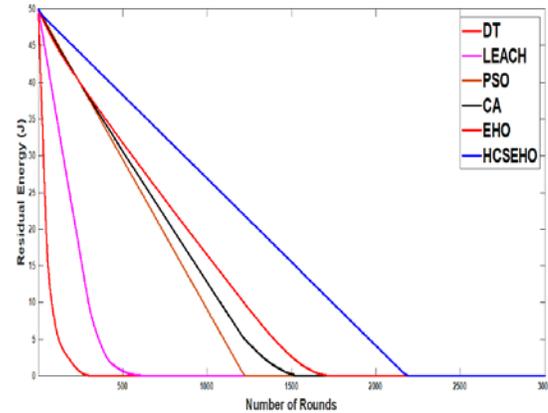


Fig. 4. Comparison of Residual Energy Obtained for different CHS Algorithms.

TABLE II. TABLE OF COMPARISON

Algorithm	Dead Nodes (Rounds)			Residual Energy (J)			Throughput (bits/round)		
	FND	HND	LND	After 1000 rounds	After 1500 rounds	After 1700 rounds	After 1000 rounds	After 1500 rounds	After 1700 rounds
DT	44	74	315	0	0	0	0	0	0
LEACH	270	370	714	0	0	0	0	0	0
PSO	1201	1222	1229	9.008	0	0	409600	0	0
CA	1298	1445	1512	12.73	0.2251	0	409600	28675	0
EHO	1602	1710	1736	16.45	2.585	0.0829	409600	409600	282653
Proposed HCSEHO	2038	2147	2198	26.88	15.49	10.94	409600	409600	409600

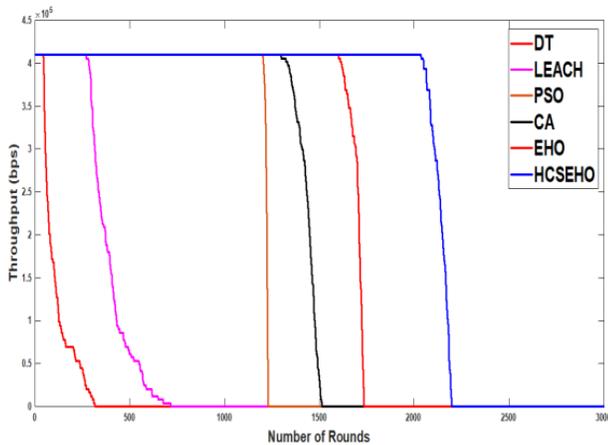


Fig. 5. Comparison of Throughputs Obtained for different CHS Algorithms.

## V. CONCLUSIONS

This research methodology suggests HCSEHO algorithm for energy efficient cluster head selection in WSN by means of integrating two meta-heuristic algorithms explicitly, Camel and Elephant Herding Optimization. The energy consumption and separation distance are considered as the fitness function for selecting optimal CHs. The visibility monitoring condition of the camel algorithm enhances the efficiency of exploitation, whereas the exploration inefficiency of Camel algorithm is compensated in an optimal way by the EHO algorithm. The superior performance of the EHO replaces the random occurrence of the camel algorithm under the clan and separator operator. However, there is still an inadequacy in the EHO algorithm with respect to shortfall of exploitation towards an optimal convergence. The proposed HCSEHO is developed by integrating EHO algorithm with a high search efficient optimization algorithm called camel. The proposed HCSEHO is validated by comparing its performance with various other existing CHS algorithms in terms of throughput, residual energy, alive nodes, and dead nodes. It is found that the lifetime of the WSN guided by the proposed HCSEHO CHS shows 21.01%, 31.21%, 44.08%, 67.51%, and 85.66%, enhancement when compared to EHO, CA, PSO, LEACH, and DT methods, respectively.

## VI. FUTURE WORK

In the future, this work can be extended by conducting experiments for various types of sensor nodes.

### REFERENCES

[1] I. F. Akyildiz, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks", *IEEE Communications magazine*, 2002, 40(8), pp. 102-114.

[2] A. Lindgren, A. Doria, and O. Schelén, Probabilistic routing in intermittently connected networks. In *ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc 2003: 01/06/2003-03/06/2003*.

[3] Z. Jiang, and A. Wei. "An energy balanced algorithm of LEACH protocol in WSN". *International Journal of Computer Science Issues (IJCSI)*, 2013, 10(1), pp. 354.

[4] S. Mirjalili, S.M Mirjalili, and Andrew Lewis, "Grey Wolf Optimizer", *Advances in Engineering Software*, 2014, (69), pp.46-61.

[5] S. A. Ramzy, F. M. Alnahwi and S. A. Abdulkareem, "A Modified Camel Traveling Behaviour Algorithm for Engineering Applications", *Australian Journal of Electrical and Electronics Engineering*, 2019.

[6] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks". *IEEE Transactions on wireless communications*, 2002, 1(4), pp. 660-670.

[7] O. Younis, and S. Fahmy, "HEED a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks", *IEEE Transactions on mobile computing*, 2004, (4), pp.366-379.

[8] N. Lavanya, and T. Shankar, "A Review on Energy-Efficient Scheduling Mechanisms in Wireless Sensor Networks". *Indian Journal of Science and Technology*, 2016, 9(32).

[9] S. Hussain, A. W. Matin, and O. Islam, "Genetic algorithm for hierarchical wireless sensor networks". *Journal of Networks*, 2007, 2(5), pp. 87-97.

[10] J. Rejina Parvin, and C. Vasanthanayaki, "Particle swarm optimization-based clustering by preventing residual nodes in wireless sensor networks". *IEEE sensors journal*, 2015, 15(8), pp.4264-4274.

[11] P. S. Rao, P. K. Jana, and H. Banka, "A particle swarm optimization based energy efficient cluster head selection algorithm for wireless sensor networks". *Wireless networks*, 2017, 23(7), pp.2005-2020.

[12] Z. Wei-Guo, Y. Shao-Pu, L. Kui, and W. Li-Ying, "Gravitational search algorithm for node localization in wireless sensor network", *Information Technology Journal*, 2016, 12(20), pp.5806-5811.

[13] C. Vimalarani, R. Subramanian, and S. N. Sivanandam, "An enhanced PSO-based clustering energy optimization algorithm for wireless sensor network", *The Scientific World Journal*, 2016.

[14] R. Kumar, and D. Kumar, "Multi-objective fractional artificial bee colony algorithm to energy aware routing protocol in wireless sensor network. *Wireless Networks*", 2016, 22(5), pp.1461-1474.

[15] Z. Ye, and H. Mohamadian, "Adaptive clustering based dynamic routing of wireless sensor networks via generalized ant colony optimization". *Ieri Procedia*, 2014,10, pp.2-10.

[16] P. Lalwani, S. Das, H. Banka, and C. Kumar, "CRHS: clustering and routing in wireless sensor networks using harmony search algorithm", *Neural Computing and Applications*, 2018, 30(2), pp.639-659.

[17] X-S Yang, Flower pollination algorithm for global optimization. In: *International conference on unconventional computing and natural computation*. Springer, Berlin, 2012, pp 240-249.

[18] N. Lavanya, and T. Shankar, "Energy optimization in wireless sensor network using NSGA-IF", *ARPN Journal of Engineering and Applied Sciences*, 2006, 12(23).

[19] A. Genta, D. K. Lobiya, and J. H. Abawajy, "Energy Efficient Multipath Routing Algorithm for Wireless Multimedia Sensor Network". *Sensors*, 2019, 19(17), pp. 3642.

[20] Y.H. Niharika, and C. Agrawal, "A Review of Enhancing Energy Efficiency of Wireless Sensor Network", *International Journal of Electrical, Electronics and Computer Engineering*, 2019, pp.46-49.

[21] T. Shankar, S. Shanmugavel, and A. Rajesh, "Hybrid HSA and PSO algorithm for energy efficient cluster head selection in wireless sensor networks", *Swarm and Evolutionary Computation*, 2016, 30, pp.1-10.

[22] N. Lavanya, and T. Shankar, "Energy Efficient Cluster Head Selection Using Hybrid Squirrel Harmony Search Algorithm in WSN", *International Journal of Advanced Computer Science and Applications*, 2019,10(12).

[23] Gai-Ge Wang, Suash Deb, and Leandro dos S. Coelho, "Elephant Herding Optimization", 2015, 3rd International Symposium on Computational and Business Intelligence.

[24] A. E. Mostafa, Ragab A. El Schiemy, Yasser I. Rashwan, X.Z. Gao, "On the performance improvement of elephant herding optimization algorithm", *Knowledge-Based Systems*, 2019,166, pp.58-70.

# Indexed Metrics for Link Prediction in Graph Analytics

Marcus Lim<sup>1</sup>, Azween Abdullah<sup>2</sup>, NZ Jhanjhi<sup>3</sup>  
School of Computer Science & Engineering (SCE)  
Taylor's University  
Selangor, Malaysia

Mahadevan Supramaniam<sup>4</sup>  
Research and Innovation Management Centre  
SEGI University, Malaysia

**Abstract**—With the explosive growth of the Internet and the desire to harness the value of the information it contains, the prediction of possible links (relationships) between key players in social networks based on graph-theory principles has garnered great attention in recent years. Consequently, many fields of scientific research have converged in the development of graph analysis techniques to examine the structure of social networks with a very large number of users. However, the relationship between persons within the social network may not be evident when the data-capture process is incomplete or a relationship may have not yet developed between participants who will establish some form of actual interaction in the future. As such, the link-prediction metrics for certain social networks such as criminal networks, which tend to have highly inaccurate data records, may need to incorporate additional circumstantial factors (metadata) to improve their predictive accuracy. One of the key difficulties in link-prediction methods is extracting the structural attributes necessary for the classification of links. In this research, we analysed a few key structural attributes of a network-oriented dataset based on proposed social network analysis (SNA) metrics for the development of link-prediction models. By combining structural features and metadata, the objective of this research was to develop a prediction model that leverages the deep reinforcement learning (DRL) classification technique to predict links/edges even on relatively small-scale datasets, which can constrain the ability to train supervised machine-learning models that have adequate predictive accuracy.

**Keywords**—Link prediction; social network analysis; criminal network; deep reinforcement learning

## I. INTRODUCTION

The rapid accessibility of the internet and social media platforms has resulted in the exponential growth of social networks. Thus, providing a medium for the gathering of internet users with common interests would increase the number of possible associations and facilitate the establishment of new communities.

The discovery of new links is a valuable attribute of the friend recommender systems employed by social network platforms. A number of algorithms employed by IBM within its own internal social network were investigated by Chen et al. [1], who found them to facilitate the establishment of connections among its employees. The technique of predicting

the presence of hidden/missing links or the formation of new structural connections is usually described as a link-prediction problem in social networks.

In social network analysis (SNA), the problem of predicting links poses an ongoing challenge and represents a key research topic [2]. There are two main approaches for predicting links between nodes within a network. One approach is based on the features of the nodes and the other is based on the topological properties of the connected nodes within the network. In the context of social network platforms, users are represented as nodes or vertices and user-related information or profile attributes may not be easily accessible. The choice to use topological properties for link prediction is preferable as it is mainly based on models derived from graph-theory analysis. Even though the topological properties of real-world social networks may not always be consistent, classical link prediction metrics based on topological properties, for example, common neighbours [3] and Adamic/Adar [4], seldom factor metadata as weights when formulating SNA metrics.

The ability to predict links accurately has many valuable applications in a range of domains that can be modelled using a network-oriented structure. In the field of bio-informatics, link prediction is used to identify the structure of connecting proteins [5]. Link prediction is also applied in e-commerce to develop recommendation systems [6]. In the domain of criminal-network analysis (CNA), link prediction is critical for the swift identification of key terrorist or criminal groups [7]. As the problem of predicting links is pertinent to a wide range of domains, many algorithms have been explored in recent years to address this issue. Many of these algorithms have relied mainly on classical machine-learning techniques that require training on the relevant features of a large dataset to achieve adequate predictive accuracy.

In SNA, the topological properties of network-oriented domains are considered along with environmental factors that can have an impact on changes in the links or relationships among users over time [8]. These environmental factors, commonly referred to as metadata, such as judicial convictions, arrest records and community crime rates in the context of criminal networks, furnish supporting information that may subsequently be used to shape the structural configurations of the networks [9].

---

This research received the funding under Taylor's Flagship Project under project # (TUFR/2017/004/06) from the Centre for Research Management, Taylor's University, Malaysia.

Deep reinforcement learning (DRL) is a machine-learning (ML) algorithm that utilises the reinforcement-learning (RL) model and incorporates a deep-learning (DL) algorithm to serve as a function approximator. In the latest developments, a DRL model has been successfully shown to be capable of self-learning across multiple domains by assimilating layers of feature learning via self-simulation based only on the provision of basic rules applicable to a particular domain [10, 11].

A DL algorithm that formulates an ML model processes multiple layers of feature learning that are typically progressively extracted from a sufficiently large dataset to acquire a precise abstraction of the domain properties. DL, which is also referred to as a deep neural network (DNN), functions as an artificial neural network (ANN) by imitating the learning process of brain neurons [12-14]. DL automates representation learning by the abstraction of domain features via layers of an ANN from the input to the output layer. The use of DNN in the formulation of ML models therefore minimises its reliance on human input to programme feature abstraction rules for specific domains.

RL is an ML framework that involves the development of programmes that function as agents that use trial and error to learn to navigate within an environment to achieve pre-defined goals [15]. These agents are guided towards achieving a goal by performing a series of tasks, which when completed are assessed as either a success or failure by a system of domain-related rules. Successes are usually given positive marks as rewards and negative marks as punishments.

Our research is expected to contribute to the development of a new set of enhanced indexed link-prediction metrics that can better predict missing links than classical link-prediction metrics. An evaluation experiment was performed on a time-series criminal-network dataset. This model was developed by indexing SNA metrics with metadata to enhance the capability of law enforcement agencies to more accurately identify critical unknown relationships in criminal networks. The model proposed in this research may have some limitations in that it is constructed based on relatively small dataset which are characteristics of criminal or terrorist networks compared to social networks such as Facebook. The relatively small dataset may have an impact on the predictive performance of certain machine learning models being trained.

In the rest of this paper, we our research work is presented as follows: In Section II, we review relevant research work involving ML models that incorporate weighted metrics. In Section III, we describe our development of the proposed and baseline models and the training methodology used. In Section IV, we describe the properties of the dataset and the experimental setup and discuss the experimental results. We present our research conclusions in Section V and in Section VI we consider the trajectory of subsequent research work.

## II. RELATED WORK

Supervised ML algorithms are usually the preferred techniques for solving link-prediction problems. ML was first reported in 2003 by Liben-Nowell and Kleinberg [16] based

on their research on the value of the structural attributes of graphs and their development of models trained on bibliographic datasets. In 2006, Hasan et al. conducted research [7] based on the technique developed by Liben-Nowell and Kleinberg. Subsequently, many other researchers have developed models using the same technique. A majority of the models proposed by these researchers were trained and evaluated on co-authored or bibliographic datasets [16], [17], [18]. Song et al. developed link-prediction models trained on a feature matrix based on node similarity and proximity measures extracted from large-scale real-world datasets such as MySpace and Facebook for use in matrix factorisation [19].

In 2011, Zaki and Al Hasan conducted a survey and provided a review of other link-prediction approaches based on linear algebra, Bayesian probabilistic models and Bayesian networks [20].

Cukierski, et al. [21] constructed a model for predicting links based on the random-forest classification technique based on an extraction of 94 graph features. The results from their model trained on Flickr datasets were found to achieve a high level of predictive accuracy.

In the development of a highly accurate link-prediction algorithm, it is critical to compute set feature metrics derived from the structural properties of graph datasets. As the accuracy of such models also depends on the use of large-scale datasets, the generation of a feature matrix requires significant computer resources. Social networks such as Facebook, which had some 700 million users in 2011 and a monthly average incremental of 20 million, poses a considerable challenge in terms of computer resources [22]. Furthermore, the structural configurations of these networks also exhibit certain attributes, for example, a power-law degree distribution [23] and small-world properties [24]. These properties must be considered when local structural features are being computed as nodes, which requires huge computer resources. As such, a subgraph is used.

Silver et al. made a major contribution to the advancement of DRL research with their development of AlphaGo, a programme that plays the ancient strategic board game Go. AlphaGo, which incorporates the Monte Carlo tree search (MCTS) technique, achieves accurate super-intuitive judgement by identifying the scope of analyses that have the highest likelihoods of success [25]. The game of Go, which is considered by artificial intelligence experts to be the holy grail in the field of computer science, has more possible variations in board positions than all the atoms in the universe that are visible to man. In 2017, AlphaGo demonstrated its extreme intelligence by beating the world's best professional Go player by a clear margin.

Silver et al. achieved another milestone in artificial general intelligence with the development of AlphaGo Zero. With only the basic rules of the games provided, AlphaGo Zero succeeded in mastering different two-player games with complete information such as Go, Shogi and Chess, using its self-learning algorithm to compete with different versions of itself [26].

Using self-simulated dataset generated within three days for training purposes, AlphaGo Zero defeated AlphaGo with a score of 3–0.

Unlike classical supervised ML techniques, such as the random forest, DRL has the capability of being trained on a self-generated dataset via self-play. As a result, the DRL technique is relevant to the modelling of network-oriented domains with datasets that are comparatively small, for example criminal syndicates, using a self-simulated dataset.

The research journals we reviewed offer scant evidence of any examination of ML models for predicting links that integrate SNA metrics indexed with metadata measurements and the DRL technique with respect to dynamic criminal networks. In this research, we conducted experiments to fill this research gap in the construction of models trained on a time-series criminal-network dataset.

### III. MODELS AND METHODOLOGY

In this research, we constructed two DRL models for link prediction: a baseline model for predicting links using just classical SNA metrics (BSNA-DRL) and another that uses classical SNA metrics indexed with metadata weights (ISNA-DRL).

#### A. Baseline Model SNA (BSNA-DRL) using Classical Link-Prediction Metrics

##### 1) Common Neighbours (CN)

$$CN_{xy} = |\Gamma(x) \cap \Gamma(y)| \quad (1)$$

The CN score of the node pair  $x$  and  $y$ ,  $CN_{xy}$ , denotes the number of directly connected nodes that are common to  $x$  and  $y$ .  $\Gamma(x)$  and  $\Gamma(y)$  denote the set of directly connected nodes of both  $x$  and  $y$  [27].

##### 2) Jaccard Coefficient (JC)

$$JC_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \quad (2)$$

The JC score of the node pair  $x$  and  $y$ ,  $JC_{xy}$ , denotes the number of neighbours common to  $x$  and  $y$  as a ratio of the total number of directly connected nodes of both  $x$  and  $y$  [27].

##### 3) Adamic/Adar measure (AA)

$$AA_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log k_z} \quad (3)$$

The AA score of the node pair  $x$  and  $y$ ,  $AA_{xy}$ , denotes the summation of the inverse value of the degree  $k$  of node  $z$ , which is the neighbour common to all directly connected nodes of both  $x$  and  $y$  [27],[28].

##### 4) Preferential Attachment (PA)

$$PA_{xy} = \Gamma_x \times \Gamma_y \quad (4)$$

The PA score of the node pair  $x$  and  $y$ ,  $PA_{xy}$ , indicates the probability that two nodes will be connected, which is proportional to the degree of the nodes [27].

#### B. Indexed Model SNA (ISNA-DRL) using Classical Link-Prediction Metrics Factored with Metadata Weights

In this research, classical link-prediction metrics were factored with two types of metadata to measure their impact on the precision of the link-prediction model compared with that of the baseline model. The metadata used included the number of criminal records and education level. Persons associated with criminals with lengthy criminal records are expected to form a relationship in the future, and those with a low education level who are associated with criminal networks are expected to have a higher likelihood of forming a relationship in the future.

##### 1) Indexed Common Neighbours (iCN)

$$iCN_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{\log(md1_z) + \log(md2_z)}{2} \quad (5)$$

The indexed CN score of the node pair  $x$  and  $y$ ,  $iCN_{xy}$ , denotes the common neighbour node  $z$  that is factored by the weighted average of the metadata, i.e., the  $md1$  and  $md2$  value attributes of node  $z$ . The  $iCN_{xy}$  value increases with the likelihood of a link forming between nodes  $x$  and  $y$ .

##### 2) Indexed Jaccard Coefficient (iJC)

$$iJC_{xy} = \frac{\sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{\log(md1_z) + \log(md2_z)}{2}}{|\Gamma(x) \cup \Gamma(y)|} \quad (6)$$

The indexed JC score of the node pair  $x$  and  $y$ ,  $iJC_{xy}$ , denotes the  $iCN_{xy}$  value as a ratio of the total number of directly connected nodes of both  $x$  and  $y$ . The  $iJC_{xy}$  value increases with the likelihood of a link forming between nodes  $x$  and  $y$ .

##### 3) Indexed Adamic/Adar measure (iAA)

$$iAA_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{\log(md1_z) + \log(md2_z)}{2} \times \frac{1}{\log k_z} \quad (7)$$

The indexed AA score of the node pair  $x$  and  $y$ ,  $iAA_{xy}$ , denotes the summation of the weighted average of the metadata ( $md1$ ,  $md2$ ) value attributes of the common neighbour node  $z$ , factored with the inverse value of the degree  $k$  of node  $z$ .

##### 4) Indexed Preferential Attachment (iPA)

$$iPA_{xy} = \sum_{x' \in \Gamma_x} \frac{\log(md1_{x'}) + \log(md2_{x'})}{2} \times \sum_{y' \in \Gamma_y} \frac{\log(md1_{y'}) + \log(md2_{y'})}{2} \quad (8)$$

The indexed PA score of the node pair  $x$  and  $y$ ,  $iPA_{xy}$ , indicates that the probability that two nodes will be connected

is proportional to the degree of nodes  $x$  and  $y$  after factoring the weighted average of the metadata (md1, md2) value attributes of the neighbouring nodes  $x'$  and  $y'$  of node  $x$  and  $y$ , respectively.

C. Proposed ISNA-DRL Model

The proposed metadata-indexed SNA link-prediction CNA model (ISNA-DRL) (Fig. 1), which represents an extension of the research on link prediction in the criminal network domain, applies the MCTS method in network searches [29-31]. To assess the predictive precision of the ISNA-DRL model, we used the area under the curve (AUC) score [32].

The ISNA-DRL model leverages a value network (indexed-SNA-metrics neural net) (Fig. 1), which is a DNN trained with features extracted from indexed SNA metrics. During the training of the model, the indexed SNA feature matrix extracted from the number of criminal records and education levels were factored as a score for the indexed-SNA-metrics neural net. The indexed-SNA-metrics neural net is a function approximator that generates output values used to rank each pair of nodes based on the likelihood of links forming or disappearing. The MCTS commences its tree search from the pair of nodes with the maximum combined indexed-SNA-metrics score estimated by the value network. The cumulative scores obtained by the RL agent from all the completed simulated network instances are then fed back to the neural net for re-calibrating the ISNA-DRL model's hyper-parameters to improve its predictive accuracy in subsequent iterations (Fig. 1).

Notes (Fig. 1):

- a) The topological features of the criminal-network dataset are used to compute the indexed SNA values.
- b) Features are extracted from the metadata for computation of the indexed SNA metrics.
- c) The indexed SNA metrics of the criminal-network dataset are used to formulate the features matrix.
- d) The indexed SNA feature matrix derived from the metadata features, e.g., the number of arrest records and the education levels, are processed by the indexed-SNA-metrics value network.
- e) The indexed-SNA-metrics value network approximates the node pairs that will most likely to form a link or have their link disappear.
- f) The MCTS commences to traverse the network by initiating a network-instance simulation by identifying links/edges, which are then ranked in accordance with their probability scores ( $P_0, P_1$ ) as approximated by the indexed-SNA-metrics value network.
- g) The simulation of the network states,  $S_0$  to  $S_N$ , occurs by the roll-out of the link-prediction process by the MCTS policy network. These simulated network states are assessed regarding their prediction accuracy by comparing their results with the test datasets ( $T_1$  to  $T_{10}$ ).
- h) The predictive accuracy scores of the predicted instances of the network are processed by the RL agent and fed back to the indexed-SNA-metrics neural net to adjust the related hyper-parameters and improve the accuracy of subsequent simulations.

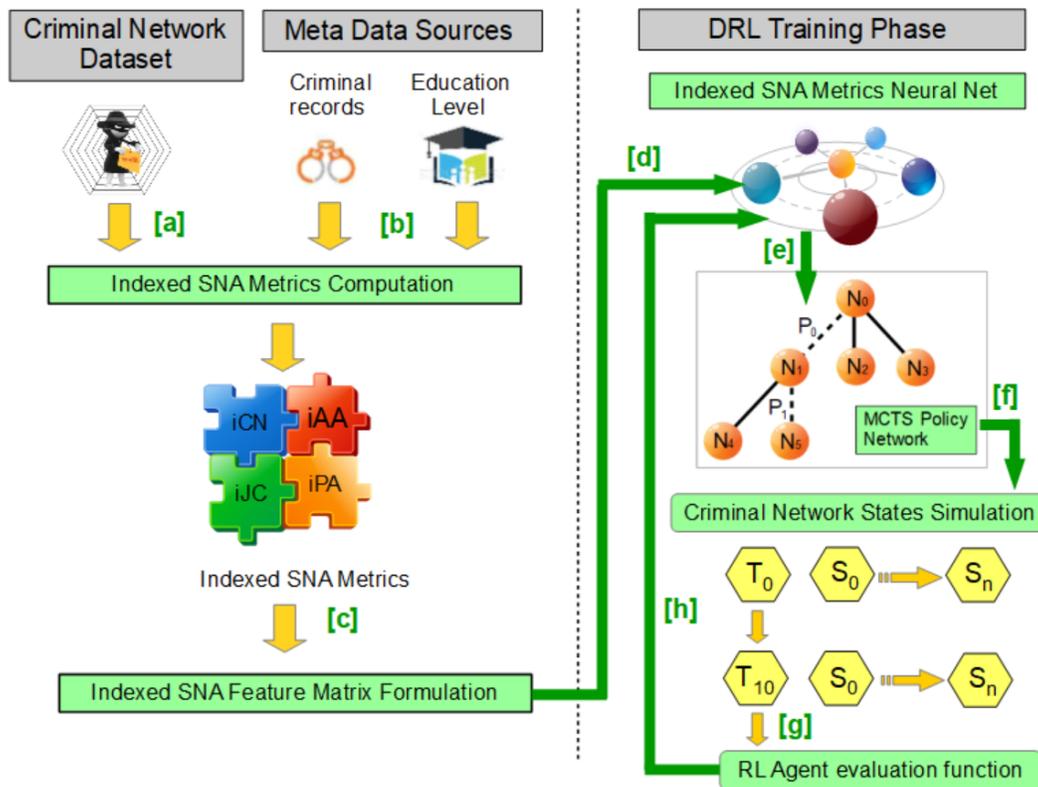


Fig 1. Proposed ISNA-DRL Link-Prediction Model based on SNA and Metadata Metrics.

The SNA measurements selected for the formulation of the indexed SNA feature matrix included the CN, AA, JC and PA metrics [33]. The indexed SNA metrics derived ( $iCN_{xy}$ ,  $iJC_{xy}$ ,  $iAA_{xy}$ ,  $iPA_{xy}$ ) (5-8) from the structural properties of the network and metadata were extracted for representation learning by the value network to approximate the ranking score of the links. These scores were then processed by the MCTS function to generate the most probable network instances. The indexed SNA scores are formulated as feature vectors, which are stored as data records. This array of features is then used to train the indexed SNA value network to perform a binary classification of the links/edges predicted to have either positive or negative labels. The link/edge that is predicted to be most likely to form is given a positive label and that predicted to be most likely to disappear is given a negative label.

The training of the ISNA-DRL model involves the representation learning of arrays of features, which comprise scores of indexed SNA metrics that denote the probability of the formation or disappearance of links/edges in the future. The cumulated value of the indexed SNA prediction metrics are calculated for every link in each feature array. The ISNA-DRL model was evaluated using the test dataset to ascertain whether the trained model predicted network instances with the required accuracy (Fig. 1).

The MCTS network tree search process creates a network instance for each link that is predicted by the policy network to be most likely to change in the following iteration. Each iteration of the MCTS network traversal, which starts from the root node, creates an initial simulated instance,  $S_0$ , from which the search process continues to the next node, navigating in accordance with the score of a positive or negative link most likely to form.

The generation of a probable network instance,  $S_2$ , as a result of a new link being predicted from the current state,  $S_1$ , is the outcome of the navigation of the RL agent to  $S_2$  from  $S_1$ , based on the rules of the default policy network. The prediction made regarding a new network instance based on the likely formation of a new link is determined by the values of the SNA prediction metrics used to rank the links. When a simulation has been completed, every simulated network state is assessed against the original network dataset to determine the accuracy of the prediction. Any variances found from this assessment are evaluated by a cost function to adjust the hyper-parameters of the indexed-SNA-metrics neural net and the MCTS function to achieve a better prediction. These hyper-parameters are then incorporated into the subsequent network-instance simulation in accordance with the link-prediction rules (Fig. 1).

The link prediction accuracies of the BSNA-DRL and ISNA-DRL models constructed on classical SNA metrics ( $CN_{xy}$ ,  $JC_{xy}$ ,  $AA_{xy}$ ,  $PA_{xy}$ ) (1-4) and indexed SNA metrics ( $iCN_{xy}$ ,  $iJC_{xy}$ ,  $iAA_{xy}$ ,  $iPA_{xy}$ ) (5-8), respectively, were evaluated based on their AUC indices. The AUC index of an ML model indicates the precision of the modelling process in identifying the underlying domain patterns, with the score ranging from 0 to 1. The higher is the AUC index achieved by the model, the more accurate are its predictions likely to be.

#### D. Metadata Indexing

Metadata indexing refers to the process of factoring the measurements of various pieces information obtained from the environment into SNA measurements for link prediction, which can influence the precision of the links predicted by the models. With reference to the criminal-networks domain, metadata may include the number of criminal records, education level and age, which can shape the structural configurations of a dynamic network and affect the underlying metrics on which link predictions are based [34]. In the proposed ISNA-DRL model (Fig. 1), the number of criminal records and the education levels of the members of the criminal network were factored into the feature matrix formulated from the indexed SNA metrics (5-8), which were then used to train the indexed SNA value network. The output of the indexed SNA value network is an approximation of a set of ranked scores that identify the node pairs with the highest likelihood of changing over time.

#### E. Time-Series Dataset

The graph algorithm is used in modelling network-oriented domains that evolve over time, such as online social or criminal groups, whose topological configurations may vary with time [34]. Participants in a network, which are denoted as nodes, may enter or exit the group as time passes. The structural configurations of the network may also change, for example, when the strength of the relationships or links among the participants change over time. The dynamic nature of such real-world networks are reflected in a time-series dataset.

### IV. EXPERIMENTS AND RESULTS

In this research we used a time-series dataset of the Caviar drug import syndicate [35]. This dataset contains a series of eleven time-series snapshots of arrest raids conducted to seize drugs from the criminal network over a 2-year period. We evaluated both the proposed ISNA-DRL and baseline BSNA-DRL models based on their AUC indexes, as these values are not skewed by the presence of imbalanced classes and this method is typically employed to evaluate the accuracy of ML classifier models.

#### A. Experiment Setup

The BSNA-DRL and ISNA-DRL prediction models were trained using a multidimensional feature matrix which was computed based on classical SNA metrics ( $CN_{xy}$ ,  $JC_{xy}$ ,  $AA_{xy}$ ,  $PA_{xy}$ ) (1-4) and indexed SNA metrics ( $iCN_{xy}$ ,  $iJC_{xy}$ ,  $iAA_{xy}$ ,  $iPA_{xy}$ ) (5-8), respectively, by factoring the metadata features derived from the Caviar dataset. This computation combines classical SNA link-prediction metrics with metadata indices to derive values that represent the probability of the formation of positive or negative links at each of the time-series snapshots of the Caviar dataset (Fig. 1).

We randomly divided the time-series dataset into training and test datasets with an 80:20 proportion, respectively. Of the eleven time-series snapshots of arrest raids, ten were used as the training set from which a random selection of positive edges was made. Then, a random selection of negative edges was made until the numbers of negative and positive edges

were equal. Each of the snapshots from the time-series test dataset was then processed by the model to predict the likely network topology from the test dataset.

The tenth and eleventh time-series snapshot of the Caviar dataset (Fig. 2 and 3) depict the actual evolution of the network during the interval. The eleventh time-series snapshot was used for testing to determine the precision of the BSNA-DRL and ISNA-DRL models in making predictions.

The experiment was conducted in two stages. In the first stage, we used the feature matrix computed for each classical SNA measurement to train both the models. The objective of this step was to determine the impact of factoring metadata on the prediction properties of each SNA measurement made independently from the others. Fig. 4 shows the predictive accuracies of the BSNA-DRL and ISNA-DRL models based on the use of individual SNA metrics.

In the next stage of the experiment, we combined all four classical metrics to formulate a feature matrix, which was used to train both models.

**B. Results and Discussion**

The ISNA-DRL model was able to identify more links/edges (Fig. 6) than the BSNA-DRL model (Fig. 5), and these links/edges were expected to change in the eleventh time-series snapshot when compared with the actual time-series snapshot at T<sub>11</sub> (Fig. 2) and T<sub>10</sub> (Fig. 3).

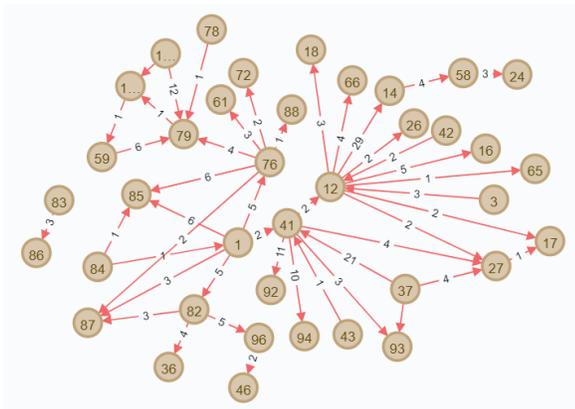


Fig 2. Actual Criminal Network at Time-Stamp T<sub>11</sub>.

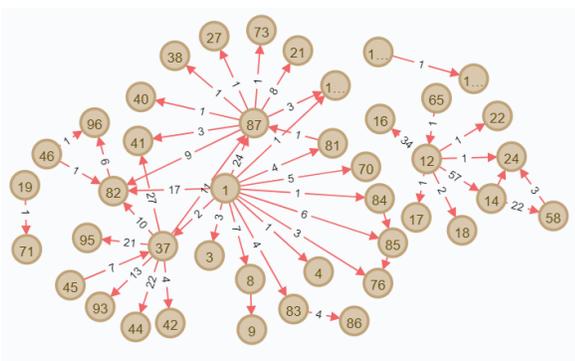


Fig 3. Actual Criminal Network at Time-Stamp T<sub>10</sub>.

**AUC Score of Link Prediction Models**

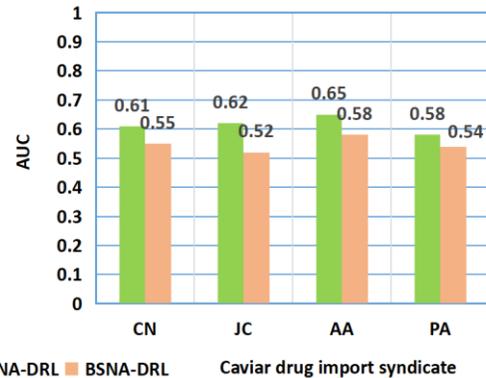


Fig 4. AUC Scores of Models Built with Individual SNA Metrics.

Although the predictions of the ISNA-DRL model for four links/edges were incorrect, i.e., node pairs (76,87), (1,81), (14,24), (12,14) (Fig. 6), the BSNA-DRL model incorrectly predicted five more links/edges, i.e., node pairs (41,93), (37,82), (41,87), (1,83) and (46,82) (Fig. 5).

A comparison of the predicted structural configurations of the Caviar network at the eleventh time step, T<sub>11</sub>, with the experimental results indicates that the indexed SNA metrics (*iCN<sub>xy</sub>*, *iJC<sub>xy</sub>*, *iAA<sub>xy</sub>*, *iPA<sub>xy</sub>*) (5-8) of the ISNA-DRL model (Fig. 6) achieved better prediction accuracy than the BSNA-DRL model (Fig. 5).

The AUC scores of the ISNA-DRL model that uses individual classical SNA metrics such as CN, JC, AA and PA (Fig. 4) to factor metadata scores are better by 0.06, 0.10, 0.07 and 0.04, respectively, than the AUC scores obtained by the BSNA-DRL model (Fig. 4), which did not incorporate metadata indexing. This indicates that metadata indexing does not cause inconsistent results when all the indexed SNA metrics are combined.

The better performance of the ISNA-DRL model than the BSNA-DRL model seemed to be related to the use of the SNA metrics indexed with metadata (*iCN<sub>xy</sub>*, *iJC<sub>xy</sub>*, *iAA<sub>xy</sub>*, *iPA<sub>xy</sub>*) (5-8), which reflect the real-life characteristics of criminal activity.

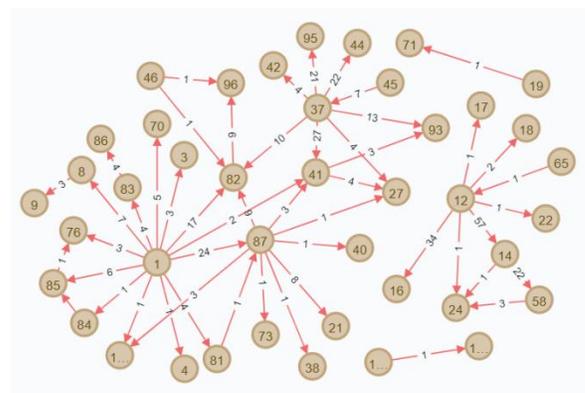


Fig 5. Network Predicted by BSNA-DRL Model.

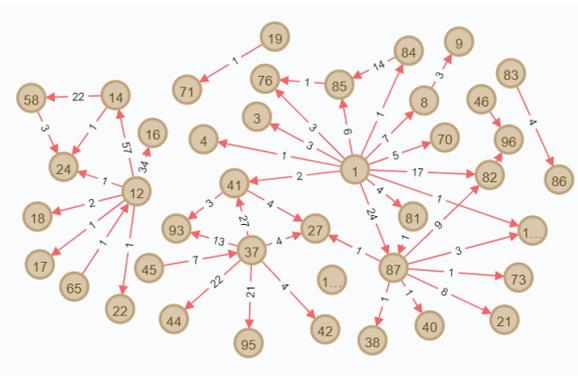


Fig 6. Predicted Network by ISNA-DRL Model.

To determine the overall prediction precisions of the models, we also conducted an experiment in which we combined all four classical SNA metrics to train both the proposed ISNA-DRL and BSNA-DRL models (Fig. 7 and 8).

The results indicate that the use of metadata scores also improved the precision of the predictions made by the ISNA-DRL trained with indexed SNA feature matrix. The overall improvement of the ISNA-DRL model using the combined indexed SNA metrics ( $iCN_{xy}$ ,  $iJC_{xy}$ ,  $iAA_{xy}$ ,  $iPA_{xy}$ ) (5-8) could be due to the fact that metadata scores provide further information related to the nodes, which influences the selection of node pairs towards those with a higher likelihood of forming positive or negative links in the future.

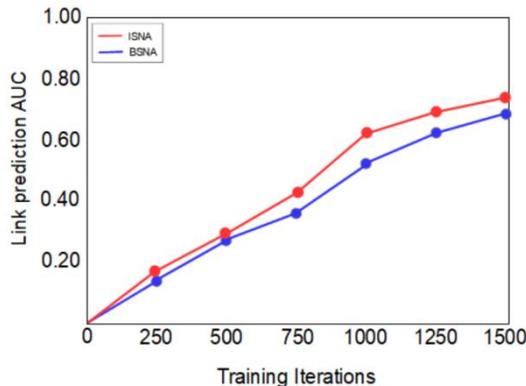


Fig 7. AUC Metrics of Link-Prediction Models with Combined SNA Metrics.

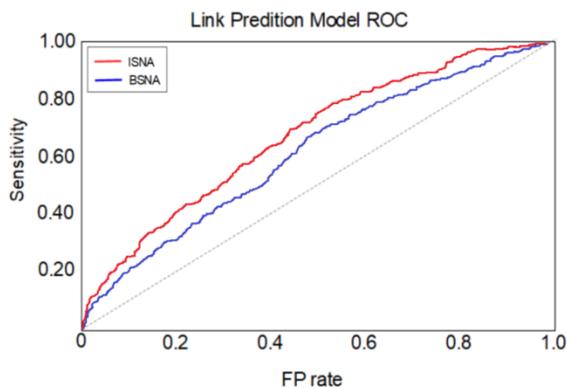


Fig 8. ROC Curve of Link-Prediction Model with Combined SNA Metrics.

In general, the experiment conducted to evaluate the link-prediction performance associated with metadata indexing indicated that the ISNA-DRL model, which factors metadata scores with SNA link-prediction metrics, performed better than the baseline BSNA-DRL model (Fig. 7 and 8), which was trained with a feature matrix formulated without metadata indexing and simulated self-generated datasets using the DRL technique.

The experimental results for the proposed DRL link-prediction model, which was trained on a time-series criminal-network dataset, are consistent with those obtained by Lim, Marcus et al. [30], [31].

In both [30] and [31] (Table II), the experiments were also conducted with link prediction models constructed using relatively small time-series dataset and leveraging on DRL.

In [30], the link prediction model, TDRL-CNA, only use features formulated from classical SNA metrics to train the model. However, in the TDRL-CNA model, additional SNA features metrics, i.e. Hub Index and Preferential Attachment index were formulated as weights in the hidden layers of the SNA metrics neural network. This model which uses breath first search (BFS) ranking algorithm, performed with approximately with the same level of predictive accuracy (AUC score of 0.78) compared to our ISNA-DRL model (AUC score of 0.74) (Table II). This result seem to indicate that the ISNA-DRL model despite being a more simplified model compared to the TDRL-CNA model, was able to achieve a comparable level of performance by using SNA metrics which were indexed by metadata.

TABLE I. AUC SCORES OF BSNA-DRL LINK-PREDICTION MODEL AND ISNA-DRL MODELS

Model	AUC	Time-score(Hr)	Iterations
BSNA-DRL	0.69	1.12	1500
ISNA-DRL	0.74	1.85	1500

TABLE II. COMPARISON OF DRL LINK PREDICTION MODELS FROM RELATED RESEARCH WORKS

Model	ISNA-DRL	TDRL-CNA	MDRL-CNA
<b>ML technique</b>	DRL	DRL	DRL with metadata fusion
<b>Tree search ranking algorithm</b>	MCTS	BFS	MCTS
<b>SNA metrics</b>	metadata indexed	classical	classical
<b>Dataset</b>	11 time-periods	11 time-periods	20 time-periods
<b>Maximum nodes</b>	42	27	55
<b>Training time-score (hour)</b>	1.85	Not available	4.3
<b>Training iterations</b>	1500	1500	2500
<b>AUC Score</b>	0.74	0.78	0.79
<b>Authors</b>	Current work	[30]	[31]

In [31], the link prediction model, MDRL-CNA, incorporated meta data features in the formulation of the feature matrix used to train the neural networks of the model instead of factoring into the SNA metrics computation. This model was able to achieve a performance which is better than the ISNA-DRL model with an AUC score of 0.79. However, due to additional complexity of fusing metadata in the weight formulation by leveraging on DL, additional computing resources of 4.3 hours were used to train the MDRL-CNA model compared to the 1.85 hours required to train the ISNA-DRL model. Therefore, considering the resources required to train the models, the factoring of meta data into the formulation of SNA metrics may prove a viable option in constructing a link prediction model where time is a constraint.

## V. CONCLUSION

In the experiments conducted in this study, the link-prediction model constructed from combined indexed SNA metrics that factored metadata scores (ISNA-DRL model) performed consistently better than the BSNA-DRL model that did not factor metadata scores. This result is supported by the respective AUC scores of 0.74 and 0.69 achieved by the ISNA-DRL and BSNA-DRL models (Table I). The experimental results also indicate that models constructed by leveraging the DRL technique can be successfully trained on smaller and self-generated datasets.

The incorporation of metadata, i.e., criminal records and education level, with classical SNA metrics enhanced the predictive precision of the ISNA-DRL algorithm, which is likely due to the incorporation into the model of real-life factors that may shape criminal-network behaviour. The improved predictive accuracy of the proposed model can contribute significantly to disrupting the activities of criminal syndicates.

## VI. FUTURE WORK

Future research should focus on the investigation of the results obtained when more than two metadata scores are factored with SNA metrics in the construction of a link-prediction algorithm, which is expected to increase the precision of the ISNA-DRL model. However, the factoring of SNA metrics with an increased number of metadata scores must be explored regarding its ability to either further improve the accuracy of the ISNA-DRL model or diminish its predictive performance due to over-fitting.

## REFERENCE

- [1] J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy, "Make new friends, but keep the old: recommending people on social networking sites," in Proceedings of the 27th international conference on Human factors in computing systems, ser. CHI '09. New York, NY, USA: ACM, 2009, pp. 201–210. [Online]. Available: <http://doi.acm.org/10.1145/1518701.1518735>
- [2] Wang, Peng, BaoWen Xu, YuRong Wu, and XiaoYu Zhou. "Link prediction in social networks: the state-of-the-art." Science China Information Sciences 58, no. 1: 1-38, 2015.
- [3] Aich, Suman, and Anita Mehta. "The clash of the Titans: How preferential attachment helps the survival of the smallest." The European Physical Journal Special Topics 223, no. 13: 2745-2758, 2014.
- [4] Güneş, İsmail, Şule Gündüz-Öğüdücü, and Zehra Çataltepe. "Link prediction using time series of neighborhood-based node similarity scores." Data Mining and Knowledge Discovery 30, no. 1: 147-180, 2016.
- [5] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic block models for relational data with application to protein-protein interactions," Proceedings of International Biometric Society-ENAR Annual Meetings, 2006.
- [6] Ricci, Francesco, Lior Rokach, and Bracha Shapira. "Recommender systems: introduction and challenges." In Recommender systems handbook, pp. 1-34. Springer, Boston, MA, 2015.
- [7] M. A. Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," SDM Workshop of Link Analysis, Counterterrorism and Security, 2006.
- [8] Hajibagheri, Alireza, Gita Sukthankar, and Kiran Lakkaraju. "Leveraging network dynamics for improved link prediction." In International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, pp. 142-151. Springer, Cham, 2016.
- [9] Arnoux, Thibaud, Lionel Tabourier, and Matthieu Latapy. "Combining structural and dynamic information to predict activity in link streams." In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, pp. 935-942. 2017.
- [10] Li, H., Kumar, N., Chen, R., & Georgiou, P. "A Deep Reinforcement Learning Framework for Identifying Funny Scenes in Movies". 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 3116-3120, 2018.
- [11] Bahdanau, Dzmitry, Jan Chorowski, Dmitriy Serdyuk, Philemon Brakel, and Yoshua Bengio. "End-to-end attention-based large vocabulary speech recognition." In 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 4945-4949. IEEE, 2016.
- [12] Zeiler, M.D., & Fergus, R. "Visualizing and Understanding Convolutional Networks". ECCV, Vol.1, 2014.
- [13] Chen, Xi-liang, Lei Cao, Chen-xi Li, Zhi-xiong Xu, and Jun Lai. "Ensemble network architecture for deep reinforcement learning." Mathematical Problems in Engineering 2018 (2018).
- [14] Duan, Y., Chen, X., Houthoofd, R., Schulman, J., & Abbeel, P. "Benchmarking Deep Reinforcement Learning for Continuous Control". ICML.Vol.1.2016.
- [15] Shi, Yangyang et al. "Contextual spoken language understanding using recurrent neural networks." In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5271-5275. IEEE, 2015.
- [16] D. Liben-Nowell and J. Kleinber, "The link-prediction problem for social networks," Journal of the American Society for Information Science and Technology, 58, no. 7, 2007.
- [17] Anitha, M. V., and Linda Sara Mathew. "Dynamic Link Prediction in Biomedical Domain." In International Conference On Computational Vision and Bio Inspired Computing, pp. 219-225. Springer, Cham, 2019.
- [18] Gupta, Anshul, Shalki Sharma, and Hirdesh Shivhare. "Supervised Link Prediction Using Forecasting Models on Weighted Online Social Network." In Proceedings of International Conference on ICT for Sustainable Development, pp. 249-261. Springer, Singapore, 2016.
- [19] H. H. Song, T. W. Cho, V. Dave, Y. Zhang, and L. Qiu. "Scalable proximity estimation and link prediction in online social networks," in Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, ser. IMC '09. New York, NY, USA: ACM, 2009, pp.322–335. [Online].Available: <http://doi.acm.org/10.1145/1644893.1644932>.
- [20] M. A. Hasan and M. J. Zaki, Social Network Data Analytics, C. C. Aggarwal, Ed. Springer, 2011.
- [21] W. J. Cukierski, B. Hamner, and B. Yang, "Graph-based features for supervised link prediction," International Joint Conference on Neural Networks, 2011.
- [22] Insidefacebook"<http://www.insidefacebook.com/2011/06/12/facebooksees-big-traffic-drops-in-us-and-canada-as-it-nears-700-million-usersworldwide/>.

- [23] Stai, Eleni, Vasileios Karyotis, and Symeon Papavassiliou. "A hyperbolic space analytics framework for big network data and their applications." *IEEE Network* 30, no. 1: 11-17, 2016.
- [24] Grabow, Carsten, Stefan Grosskinsky, Jürgen Kurths, and Marc Timme. "Collective relaxation dynamics of small-world networks." *Physical Review E* 91, no. 5: 052815, 2015.
- [25] D.Silver et al. "Mastering the Game of Go without Human Knowledge". *Nature*, 550, no. 7676: 354–359, 2017.
- [26] D. Silver et al., "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play", *Science*, 362, no. 6419, 1140-1144, DOI. 10.1126. science. aar6404, 2018.
- [27] Mallek, Sabrine, Imen Boukhris, Zied Elouedi, and Eric Lefèvre. "Evidential link prediction in social networks based on structural and social information." *Journal of computational science* 30: 98-107, 2019.
- [28] L.A. Adamic, E. Adar, Friends and neighbors on the web, *Social Networks* 25, 211–230, 2003.
- [29] Marcus Lim, Azween Abdullah, NZ Jhanjhi. "Performance Optimization of Criminal Network Hidden Link Prediction Model with Deep Reinforcement Learning", *Journal of King Saud University - Computer and Information Sciences*, 2019.
- [30] Lim, Marcus, Azween Abdullah, Nz Jhanjhi, Muhammad Khurram Khan, and Mahadevan Supramaniam. "Link Prediction in Time-Evolving Criminal Network With Deep Reinforcement Learning Technique." *IEEE Access* 7 (2019): 184797-184807.
- [31] Lim, Marcus, Azween Abdullah, N. Z. Jhanjhi, and Muhammad Khurram Khan. "Situation-Aware Deep Reinforcement Learning Link Prediction Model for Evolving Criminal Networks." *IEEE Access*, 2019.
- [32] Lim, Marcus, Azween Abdullah, NZ Jhanjhi, and Mahadevan Supramaniam. "Hidden Link Prediction in Criminal Networks Using the Deep Reinforcement Learning Technique." *Computers* 8, no. 1: 8, 2019.
- [33] Özcan, Alper, and Şule Gündüz Öğüdücü. "Multivariate temporal link prediction in evolving social networks." In 2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS), pp. 185-190. IEEE, 2015.
- [34] Holme, Petter. "Modern temporal network theory: a colloquium." *The European Physical Journal B* 88, no. 9: 234, 2015.
- [35] Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies. 1, 2002.

# Radar GPR Application to Explore and Study Archaeological Sites: Case Study

Ahmed Faize<sup>1</sup>, Gamil Alsharahi<sup>2</sup>, Mohammed Hamdaoui<sup>3</sup>

Department of Physics

Polydisciplinary Faculty of Nador Mohammed First University, Morocco

**Abstract**—The issue of exploring and searching for archaeological sites is very important for a greater knowledge of the history of ancient nations and peoples. Recently, Ground Penetrating Radar GPR technology appeared to detect objects buried and study as deep as tens of meters. This work aims to apply this technique in studying and exploring some archaeological sites using the 500 MHz antenna. This study has proven its effectiveness and success. Also, one of the important programs used in the processing of the obtained data is called Reflexw.

**Keywords**—Archaeological sites; exploring; ground penetrating radar; processing data

## I. INTRODUCTION

Ground Penetrating Radar (GPR) is used to image the subsurface. It is a geophysical method based on the propagation, reflection and scattering of high frequency (from 10 MHz to 2.5 GHz) electromagnetic (EM) waves in the subsurface. The investigation depth depends on the EM wave attenuation, which grows as the conductivity of the subsoil materials increases, and on the frequencies used. Generally, the penetration depth is higher at lower frequency and varies from about 1m to some tens of meters [1], [9].

We evaluate the applicability and the effectiveness of the GPR attribute analysis for archaeological purposes and we test the attribute analysis on GPR data obtained in the river harbor area of the Aquileia Archaeological Park, NE Italy, where the cultural heritage of the Roman imperial period is buried at different depths beneath a silty loam layer at an average depth not greater than 3–4 m [2]. During its development years, geophysical survey has served field archaeology by defining possible sites underground, before excavation or preservation. Now we can see the art taking off as a research method in its own right [3].

A ground-penetrating radar (GPR) survey, using mostly a 500 MHz antenna, was carried out in an urban area (Lecce, Italy) to obtain a detailed characterization of the most superficial layers, where presumably archaeological structures are buried, and to quickly identify anomalous zones for excavation [4].

This technique has been successfully applied in GPR surveys along different archaeological sites such as: Roman buildings and other historical sites [5], [7]. They are commonly wide-open areas with large targets that often have a well-known geometry. So usually these “Pseudo 3D” GPR surveys are based on the use of mid-low frequency antennas with a

space between profiles of half a meter that involves a vast interpolation among data. Such methodology is generally considered sufficient to obtain images with enough resolution to show the profile of those archaeological targets [6], [8].

## II. SIGNAL PROCESSING

The processing of the recorded results is done by a two-stage signal processing process. First, a first processing is done on the radargrams, which allows the visualization of the underground data obtained directly by the GPR, thanks to the Reflexw software which uses standard filtering (Dewow, bandpass, gain improvement). Then, the data recovered after the passage of the Reflexw software, are processed by the GPR-Slice software, which is a software for processing images of slices in time [10]-[11]. Indeed, although the radargrams given by the Reflexw software, inform, in a satisfactory way, on the existence and the localization of the objects sought in-depth, a meticulous and advanced 2D and 3D treatment proves to be extremely useful to access better results and interpretations.

## III. RESULTS AND DISCUSSION

In what follows, we describe the most notable results of the geophysical study conducted in:

### A. Zone 1

The results of the B-Scans in Fig. 1(a) and (b) reveal the presence of strong horizontal secondary reflectors (red circles). We also note, in Fig. 1, an anomaly, green circle, characterized by the hyperbolic reflector with propagation speed equal to 0.1 m / ns and vertex located at position 15.5 m, from the origin of the radargram, and 0.6 m deep. Fig. 2(a) shows four time-slice representations (Time-Slice).

The area of the area is identified by x and y, and the depth by time ranges from 0 ns to 20 ns (0 m to 1.8 m deep). The amplitudes, in absolute value, of the reflected signal, are represented by different colors, ranging from blue to orange. Gray is reserved for low amplitudes while yellow and orange are reserved for high amplitudes of signals returned by materials. The latter could be of natural or cultural origin.

At least four different rectangular structures seem to be present: the first is represented by a clear anomaly of 1 m in length and 1 m in width which is visible from 10 ns (0.5 m deep) up to 20 ns (1.8 m deep). The second structure is represented by an anomaly of 1 m in length and 1.4 m in width, visible from the same temporal depth as that indicated above.

The third is a large anomaly 4 m long and 5 m wide. The last structure is 1 m long and 1.8 m wide.

Fig. 2(b) gives the general diagram of zone 1. This diagram represents the geometric structures suspected during the interpretation of the results given by the GPR. Analysis by 3D imagery (Fig. 3) confirms the presence of the anomalies observed in the radargrams and time slicer presentations.

**B. Zone 2**

The radargrams in Fig. 4 show a strong secondary-horizontal reflector located at a depth of 0.9 m and located between 3.5 m and 8.2 m, in the direction X. We also observe hyperbolas (speed of 0, 06 at 0.1 m / ns), which may indicate other secondary reflectors.

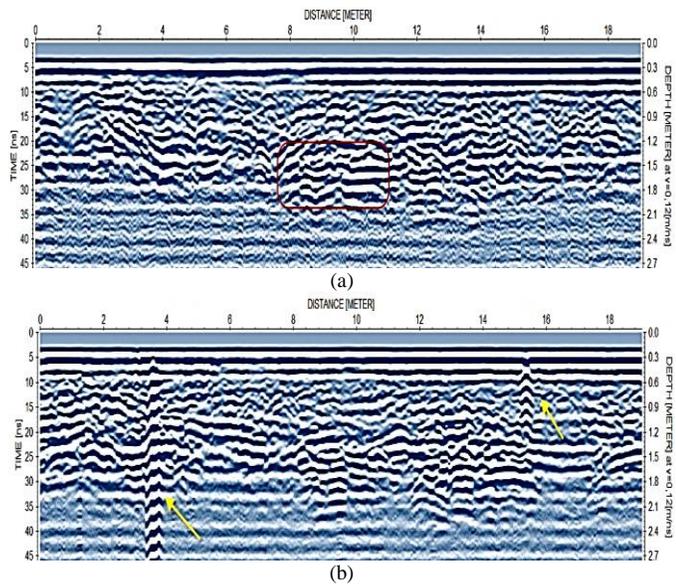


Fig. 1. Radargrams or B-Scans of Zone 1 Obtained using the 500 MHz Antenna (a): The Red Square Encircles the Reflection on a Flat Surface, (b): The Yellow Arrows Indicate the Response of Shallow Metal Objects.

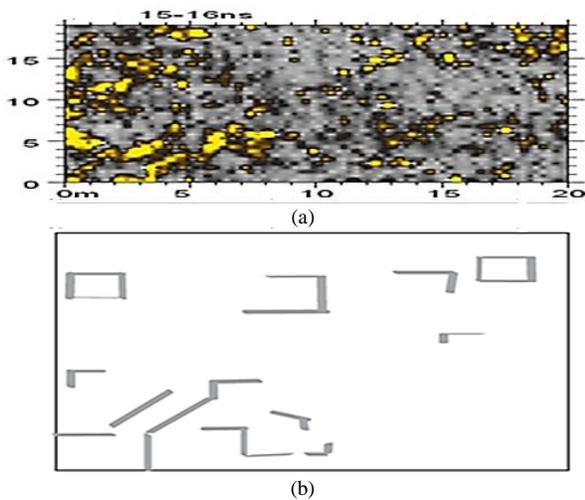


Fig. 2. (a) Time Slice or C-scan between 15 and 16 ns Obtained by the 500 MHz Antenna, (b) General Interpretation of Zone 1. We can Observe Rectangular Structures as well as Several Linear Anomalies.

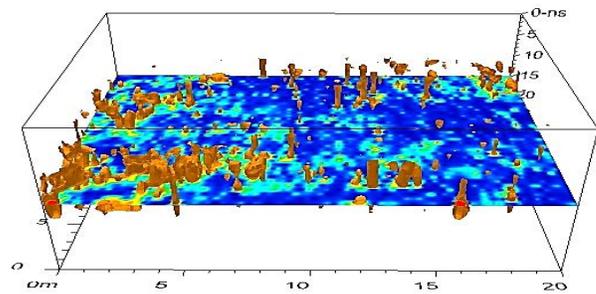


Fig. 3. GPR iso-Surface Slice Obtained using the 500 MHz Antenna. We can Observe Four Small Rooms Around 15-17 ns.

The figure radargram (Fig. 4) shows a strong attenuation of the signal, between 15 m and 17 m, which could be due to the presence of a mound at this level (red circle).

Fig. 5(a), in time slices, shows the presence of rectangular structures. The longer one, which can be associated with anomaly 1, becomes wider and deeper. This change in dimensions is probably due to a collapse somewhere in this structure. A wetter area (strongly attenuated signal) represented by a red circle in Fig. 4, corresponds to the mound already observed in the radargram given by the Reflexw software.

Fig. 5(b) gives the general diagram of zone 2. This diagram represents the geometric structures suspected during the interpretation of the results given by the GPR.

**C. Zone 3**

The radargrams obtained by the 500 MHz antenna, illustrated in Fig. 6, reveal the presence of accentuated hyperbolas which indicate the presence of a long structure in the X direction (probably a wall).

On the radargram of Fig. 6, one can note, the presence of multiple reflections, which testify to the existence of several metallic fragments dispersed on the field of investigation, at a depth of 30 cm. This complicates the task of exploring in this area. This "metallic" noise is observable in the 2D image.

Fig. 7 illustrates a slice in horizontal time going from 0 to 18 ns. This time slot was constructed with the data collected by the 500 MHz antenna. Strong radar reflection, coming from obstacles close to the surface (0.4 m), is observable on the 7 - 18 ns bands. This signal would correspond to a modern utility gap hiding much of the potential characteristics.

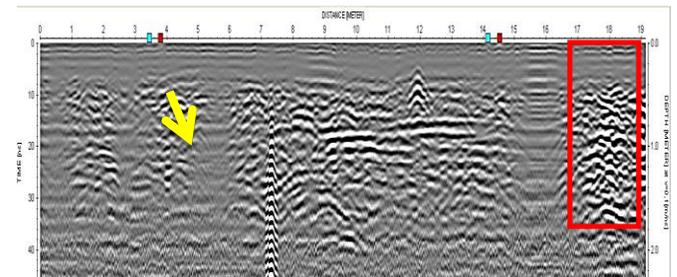


Fig. 4. Radargram or B-Scan of Zone 2 Processed by the 500 MHz Antenna. between X = 17 m and X = 19 we can Observe different Responses of the GPR Signal of a Structure (Probably a Wall). The Yellow Arrow Marks the Response of a Metallic Reflector and the Red Square Marks the Area where the Signal is Attenuated.

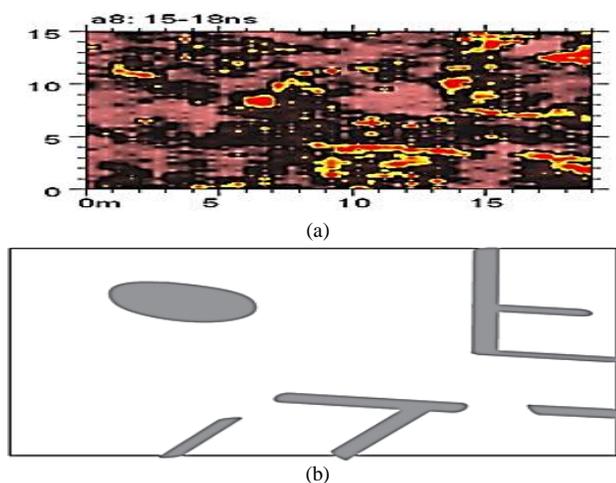


Fig. 5. (a) Time Section between 15 and 18 ns Obtained by the 500 MHz Antenna in Zone 2, this Section Presents some Interesting Devices, such as a Long Linear Anomaly. In Addition, there is a Square Attached to this Linear Function which is Centered on (X = 11 m, Y = 9 m). (b) General Interpretation of Zone 2.

Analysis by 3D imagery (Fig. 8) confirms the presence of the anomalies observed in the radargrams and the time slice representations discussed above.

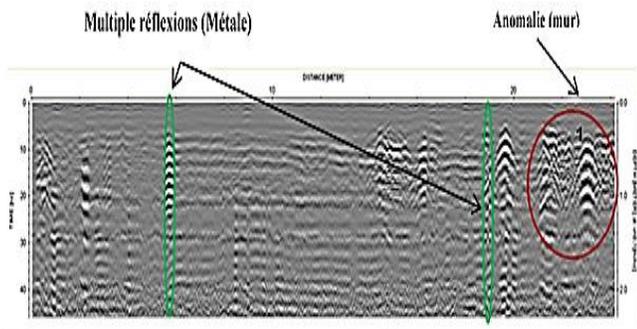


Fig. 6. Radargram in Zone 3 (a) of Line Number 13 Obtained with the 500 MHz Antenna. We can Observe a Linear Anomaly around 7 ns- 18 ns (Red Circle) and also Multiple Reflections (Green Circle).

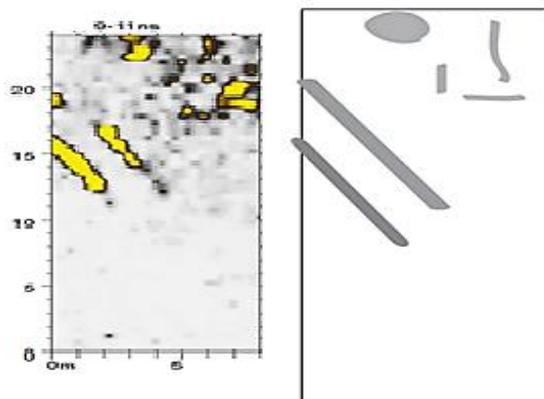


Fig. 7. Time Section of Area 16 at 7ns-16ns Obtained with the 500 MHz Antenna in the Y Direction, (b) Illustration of the Time Section at 9-11 ns and (c) General Interpretation of the Result in Zone 3. Two Parallel Anomalies Appeared at a Depth of 7 ns - 16 ns.

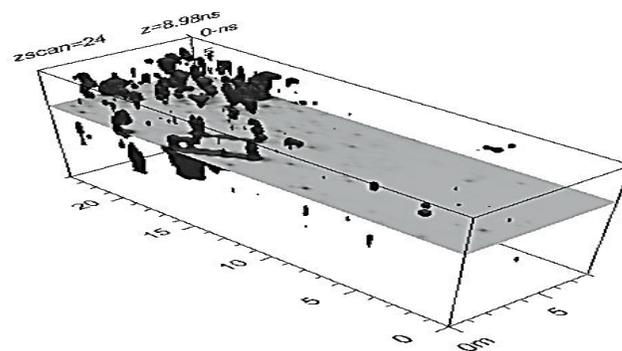


Fig. 8. GPR iso-Surface Slice Obtained in Time Windows 0-45ns with the 500 MHz Antenna.

#### IV. CONCLUSION

The sites studied are highly potential places from an archaeological point of view. They are indicated and recommended by specialists in the field of archeology. Analysis of the data collected by the GPR system, for the areas studied, reveals the presence of a significant amount of rectangular structures, long coasts, which can be interpreted, in this kind of sites, as archaeological remains, probably walls, chambers or remains of foundations.

However, it should be noted that certain areas of the sites studied were, in the past, places where modern activities were carried out, such as agriculture for example. The construction of greenhouses and warehouses has left scrap debris, which is buried at a shallow depth, which makes the interpretation of results (radargrams confused by parasitic signals caused by this debris extremely difficult.

#### REFERENCES

- [1] BANNING, "Edward Bruce. Archaeological survey". Springer Science & Business Media, 2002.
- [2] W.Zhao, E.Forte, M.Pipan, G.Tian. "Ground penetrating radar (GPR) attribute analysis for archaeological prospection". Journal of Applied Geophysics, 2013, vol. 97, p. 107-117.
- [3] CONYERS, Lawrence B. "Ground-penetrating radar for anthropological research". Antiquity, 2010, vol. 84, no 323, p. 175-184.
- [4] V. Basile, M.T. Carrozzo, S. Negri, L. Nuzzo, T. Quarta, A.V. Villani. "A ground-penetrating radar survey for archaeological investigations in an urban area (Lecce, Italy)". Journal of Applied Geophysics, 2000, vol. 44, no 1, p. 15-32.
- [5] ALSHARAHI, Gamil, Ahmed Faize, Carmen Maftai, and Abdellah Driouach. "GPR Application for Risks Detection in Subsurface Engineering Construction Projects". Ovidius University Annals of Constanta-Series Civil Engineering 21, no. 1 (2019): 51-58.
- [6] Nikos Papadopoulos, Apostolos Sarris, Myeong-Jong Yil, Jung-Ho Kim. "Urban archaeological investigations using surface 3D ground penetrating radar and electrical resistivity tomography methods". Exploration Geophysics, 2009, vol. 40, no 1, p. 56-68.
- [7] ALSHARAHI, G., A. Faize, M. Louzazni, A. M. M. Mostapha, M. Bayjja, and A. Driouach. "Detection of cavities and fragile areas by numerical methods and GPR application". Journal of Applied Geophysics 164 (2019): 225-236.
- [8] HOLCOMB, Derrold W. et SHINGIRAY, Irina Lita. "Imaging radar in archaeological investigations: An image processing perspective. In: Remote Sensing in Archaeology". Springer, New York, NY, 2006. p. 11-45.
- [9] Alsharahi, Gamil, Ahmed Faize, Carmen Maftai, Mohamed Bayjja, Mohamed Louzazni, Abdellah Driouach, and Abdellatif Khamlichi. Analysis and "Modeling of GPR Signals to Detect Cavities: Case

- Studies in Morocco". *Journal of Electromagnetic Engineering and Science* 19, no. 3 (2019): 177-187.
- [10] G. Alsharahi, A. Mint Mohamed Mostapha, A. Faize, A. Driouach. "Modelling and simulation resolution of ground-penetrating radar antennas". *Journal of electromagnetic engineering and science*, 2016, vol. 16, no 3, p. 182-190.
- [11] ALSHARAHI, Gamil, DRIOUACH, Abdellah, and FAIZE, Ahmed. "Performance of GPR influenced by electrical conductivity and dielectric constant". *Procedia technology*, 2016, vol. 22, p. 570-575.

# Generalized Approach to Analysis of Multifractal Properties from Short Time Series

Lyudmyla Kirichenko<sup>1</sup>

Department of Applied Mathematics  
Kharkiv National University of  
Radio Electronics  
Kharkiv, Ukraine

Abed Saif Ahmed Alghawli<sup>2\*</sup>

Department of Computer Science  
College of Sciences and Humanities  
Prince Sattam Bin Abdulaziz  
University Aflaj  
Kingdom of Saudi Arabia

Tamara Radivilova<sup>3</sup>

V.V. Popovskyy Department of  
Infocommunication Engineering  
Kharkiv National University of  
Radio Electronics  
Kharkiv, Ukraine

**Abstract**—The paper considers a generalized approach to the time series multifractal analysis. The focus of research is on the correct estimation of multifractal characteristics from short time series. Based on numerical modeling and estimating, the main disadvantages and advantages of the sample fractal characteristics obtained by three methods: the multifractal fluctuation detrended analysis, wavelet transform modulus maxima and multifractal analysis using discrete wavelet transform are studied. The generalized Hurst exponent was chosen as the basic characteristic for comparing the accuracy of the methods. A test statistic for determining the monofractal properties of a time series using the multifractal fluctuation detrended analysis is proposed. A generalized approach to estimating the multifractal characteristics of short time series is developed and practical recommendations for its implementation are proposed. A significant part of the study is devoted to practical applications of fractal analysis. The proposed approach is illustrated by the examples of multifractal analysis of various real fractal time series.

**Keywords**—Fractal time series; multifractal analysis; estimation of multifractal characteristics; generalized Hurst exponent; practical applications of fractal analysis

## I. INTRODUCTION

In the last years, there has been a growing interest in complex systems that have a fractal structure: informational, biological, physical, technological, financial and other. The dynamics of such systems generate time series with fractal (self-similar) properties. Time series fractal analysis is used to simulate, analyze and control complex systems in various fields [1]-[5]. Processes with fractal properties can be divided into two groups: monofractal and multifractal. Monofractal processes are homogeneous in the sense of fractal properties and have single scaling exponent. Multifractal processes have heterogeneous scale properties and are characterized by a set of scaling exponents.

There are a large number of methods for estimating the parameters of self-similar and multifractal processes from time series [6]-[11]. Methods based on wavelet transforms are of particular importance among the research methods of fractal nonstationary processes [2], [12], [13]. Currently, the two most popular tools of time series multifractal analysis are

the method of multifractal detrended fluctuation analysis (MFDFA) and method of wavelet transform modulus maxima (WTMM). The MFDFA method is focused on time series with trend components and has been widely used due to [6], [14]. The WTMM method is based on continuous wavelet transform and was originally proposed in [15], [16]. Both methods are a powerful tool for the statistical processing of nonstationary processes. A significantly smaller number of studies are carried out using the method of multifractal analysis based on discrete wavelet transform [2], [17], [18].

In recent years, many studies were focused on the statistical properties of fractal characteristics estimates obtained by the above methods. However, most of the studies done have certain limitations. The main attention was paid to the statistics of estimates of self-similarity degree obtained by different methods [7], [8], [12], [18]. When studying the properties of multifractal characteristics, time series of great length were usually considered. So, for example, in [14], [19], when conducting a comparative analysis between MFDFA and WTMM, the length of the realizations was over 60 thousand values. In this case, the sample fractal characteristics are quite close to the theoretical ones. At the same time, the time series obtained in practice have a much smaller range of values. In [9], [20], aimed at studying the accuracy of estimating short time series, the minimum series length was over 1000 values, which is also quite large.

Many studies also compare the estimates of multifractal characteristics calculated using the same method [9], [14], [19]. The most common in practice is the MFDFA method. Quite a few researchers use several methods to evaluate the multifractal properties of time series. Despite numerous publications related to the practical application of fractal analysis, still, there is no universal approach to the estimation of fractal characteristics by time series. Besides, some practical questions of estimation from short non-stationary time series remain unsolved.

This work aims to study the application of multifractal analysis methods to short time series and to propose a generalized approach to the analysis of the multifractal properties of time series.

---

\*Corresponding Author  
Submission Date: April 20, 2020  
Acceptance Date: May 12, 2020

## II. SELF-SIMILARITY PROPERTY AND ITS ESTIMATION METHODS

### A. Definitions and Properties of Self-Similar Stochastic Processes

A process  $X(t)$  is called self-similar if it is invariant in distribution when changing the time-scale:

$$\text{Law}\{X(at)\} = \text{Law}\{a^H X(t)\}, \quad \forall a > 0, t > 0 \quad (1)$$

The parameter  $H$  is called Hurst exponent. It is a measure of self-similarity and together with that a measure of long-range dependence of process. In this case, the correlation function  $K(\tau)$  of the process  $X(t)$  decreases hyperbolically:  $K(\tau) \sim \tau^{-\beta}$ , where  $H = 1 - (\beta/2)$ ,  $0 < \beta < 1$ .

$$X_t^{(m)} = \frac{1}{m}(X_{m-m+1} + \dots + X_m), \quad m, t \in N. \quad (2)$$

A process with discrete time  $X$  is called self-similar with the parameter  $H$  if the expression is true.

$$\text{Law}\{m^{1-H} X^{(m)}\} = \text{Law}\{X\}, \quad (3)$$

that is, the process does not change the distribution laws after averaging over blocks of length  $m$ .

The Hurst parameter  $H$  is a measure of the long-term dependence duration. The case  $0.5 < H < 1$  means that the process has persistent behavior or long memory. In other words, if positive (or negative) process increments have been observed for some time in the past, then with the probability close to the value  $H$ , this trend will continue.

The case  $0 < H < 0.5$  means an antipersistent process. Here, high process values follow low ones, and vice versa. In other words, the probability that the process will change its trend in the opposite direction is as great as the parameter  $H$  closer to 0.

When  $H = 0.5$  the deviations of the process from the mean are indeed random and do not depend on the previous values, that corresponds to the case of the ordinary Brownian motion.

The  $q$ -moments of the self-similar random process  $X(t)$  can be expressed as:

$$E\left[|X(t)|^q\right] \propto t^{qH} \quad (1)$$

In particular, for the value  $q = 2$ , we have a scaling relation for the variance.

$$\text{Var}[X(t)] \propto t^{2H} \quad (2)$$

Most methods for estimating the Hurst exponent are based on scaling change of some series characteristics close to variance. For example, in the method of R/S analysis change of the normalized range of the cumulative series  $\frac{R}{S}(\tau) \propto \tau^H$  is

investigated, in the method of detrended fluctuation analysis the fluctuation function  $F(\tau) \propto \tau^H$  is investigated, etc.

The property of self-similarity corresponds to the linear dependence of the logarithm of the considered value, on the logarithm of the time  $\log \tau$ . For example, in cases of detrended fluctuation analysis, we have linear regression  $\log F(\tau) = H \log(\tau) + \text{const}$  which is constructed by the least squares method. The self-similarity parameter  $H$  can be estimated by the tangent of the line inclination angle.

### B. Methods for Estimating the Hurst Exponent

To estimate the Hurst exponent of a time series, many methods have been proposed in [2], [6]-[8]. Consider the most popular.

Rescaled range method. This method was proposed by H. Hurst and is currently the most well-known and popular method of fractal analysis [7], [8]. It is widely used in telecommunication technologies, in the study of the self-similar properties of information traffic, the study of the dynamics of financial markets, the study of biomedical signals, etc.

In this method in the study of a time series  $x(t)$  of length  $\tau$ , the following relation is determined.

$$\frac{R(\tau)}{S(\tau)} = R/S = \frac{\max(x^{cum}(t, \tau)) - \min(x^{cum}(t, \tau))}{\sqrt{\frac{1}{\tau-1} \sum_{i=1}^{\tau} (x(i) - \bar{x})^2}}, \quad t = 1, \tau,$$

where  $R(\tau)$  is a range of the cumulative series  $x^{cum}(t, \tau)$ ;  $S(\tau)$  is the standard deviation of the original series;

$$\bar{x}(\tau) = \frac{1}{\tau} \sum_{i=1}^{\tau} x(i); \quad x^{cum}(t, \tau) = \sum_{i=1}^t x(i) - \bar{x}(\tau)$$

For a self-similar process, this ratio for large values  $\tau$  has the following scaling:

$$E\left[\frac{R}{S}\right] \propto \tau^H \quad (3)$$

The dependence of  $\frac{R(\tau)}{S(\tau)}$  on  $\tau$  in double logarithmic scale for a self-similar time series, have to be a straight line, approximated by the least squares method. The value  $H$  is calculated as the tangent of line angle dependencies  $\log \frac{R(\tau)}{S(\tau)}$  on  $\log(\tau)$ , as it is shown in Fig. 1.

The method of detrended fluctuation analysis (DFA). This method was originally proposed for the analysis of long-term correlations in the structure of the heart rhythm. Currently, it is one of the most widely used methods for studying various non-stationary time series [6], [14], [19]-[23].

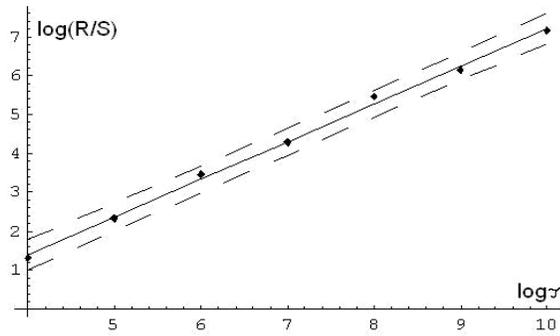


Fig. 1. The Dependence of  $\log(R/S)$  on  $\log(\tau)$ , Obtained by Rescaled Range Method.

In the DFA method, for the initial time series  $x(t)$ , the cumulative series is constructed  $y(t) = \sum_{i=1}^t x(i)$ , which is divided into  $N$  segments with a length  $\tau$ , and for each segment  $y(t)$ , the fluctuation function is calculated:

$$F^2(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} (y(t) - Y_m(t))^2, \quad (4)$$

where  $Y_m(t)$  is a local  $m$ -polynomial trend that limited by this segment.

The function  $F(\tau)$  averaged throughout the entire series  $y(t)$  has a scaling dependence on the length of a series segment:

$$F(\tau) \propto \tau^H \quad (5)$$

The plot of dependence  $\log F(\tau)$  on  $\log \tau$  in a certain range of values is straight line approximated by the least squares method. The value  $H$  is calculated as the tangent of straight line angle of dependence  $\log F(\tau)$  on  $\log \tau$ , as it is shown in Fig. 2.

Discrete Wavelet Transform Method. The main development of the method was in the papers [2], [12]. Currently, it is used in the analysis of both stationary and non-stationary time series in various fields of research.

The method of wavelet estimation of self-similarity degree  $H$  is based on the properties of the detail wavelet coefficients of time series decomposition by Discrete Wavelet Transform. The basis of this method is a statement that the wavelet energy value  $E_j$  at the wavelet decomposition level  $j$  satisfies the scaling relation:

$$E_j \propto 2^{(2H+1)j} \quad (6)$$

The plot of dependence  $\log_2(E_j)$  on  $j$  is a straight line approximated by the least squares method. The value of exponent  $H$  can be found by evaluating tangent of straight line angle. The wavelet energy spectrum of a self-similar process and dependence  $\log_2(E_j)$  on  $j$  are shown in Fig. 3.

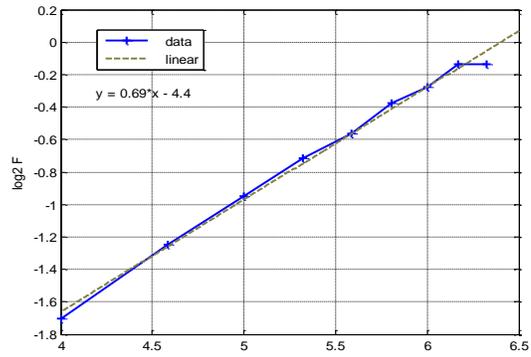


Fig. 2. Dependence  $\log F(\tau)$  on  $\log(\tau)$ , Obtained by the DFA Method.

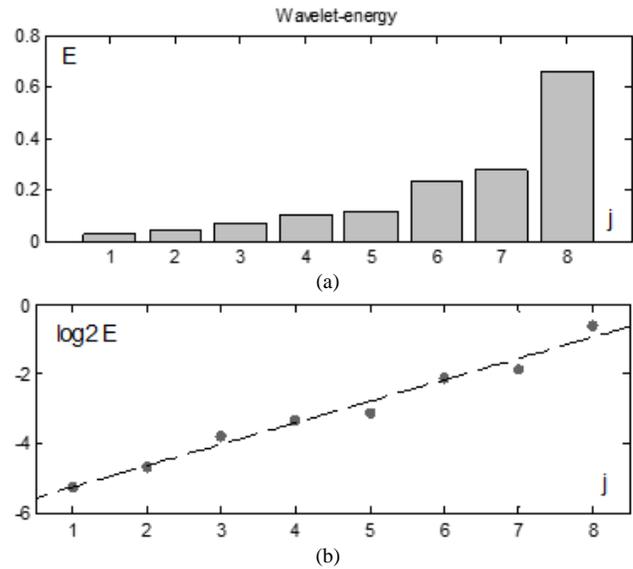


Fig. 3. The (a) Wavelet Energy Spectrum of a Self-Similar Process; (b) Dependence  $\log_2(E_j)$ .

### III. ESTIMATION METHODS OF MULTIFRACTAL CHARACTERISTICS

#### A. Definitions and Properties of Multifractal Stochastic Processes

Multifractal stochastic processes also are invariant in distribution, but in this case, the change of quantitative characteristics of the process depends on the magnitude of the stretching in time.

$$\text{Law}\{X(at)\} = \text{Law}\{M(a) \cdot X(t)\}, \quad \forall a > 0, t > 0 \quad (7)$$

where  $M(a)$  is a random function which independent of  $X(t)$ . In the case of a self-similar (monofractal) process  $M(a) = a^H$ .

For multifractal processes, the following relationship of  $q$ -moments holds:

$$E\left[|X(t)|^q\right] \propto t^{q^H(q)}, \quad (8)$$

where  $h(q)$  is a function of generalized Hurst exponent. Value  $h(q)$  at  $q=2$  is equal to the degree of self-similarity  $H$ . Generalized Hurst exponent of the monofractal process does not depend on the parameter  $q$ :  $h(q)=H$ . For the case  $q=2$ , it again gets at the formula (5).

Thus, before investigating the multifractal properties of the time series, it is necessary to determine the existence of self-similar properties. A more detailed analysis of the self-similarity of time series is presented in [21]. Consider some popular methods of time series estimation of multifractal properties.

### B. The Method of Multifractal Detrended Fluctuation Analysis

When performing multifractal fluctuation detrended analysis (MFDFA), the dependence of the fluctuation function  $F_q(s)$  on the parameter  $q$  is investigated:

$$F_q(s) = \left\{ \frac{1}{N} \sum_{i=1}^N [F^2(s)]^{\frac{q}{2}} \right\}^{\frac{1}{q}} \quad (9)$$

$F_q(s)$  is obtained by raising the expression (7) to the power  $q$  and then averaging over all segments. Since at  $q=0$  (12) contains uncertainty, the following expression is usually used instead:

$$F_q(\tau) = \text{Exp} \left\{ \frac{1}{N} \sum_{i=1}^N \ln[F^2(\tau)] \right\}$$

Changing the length of the segment  $s$  at a fixed  $q$ , we find the dependence  $F_q(s)$ . If the time series has multifractal properties, then the fluctuation function is represented by a power dependence:

$$F_q(s) \propto s^{h(q)}, \quad (10)$$

where  $h(q)$  is a function of generalized Hurst exponent.

For the monofractal time series, the fluctuation function  $F_q(s)$  is the same for all segments and the generalized Hurst exponent does not depend on the parameter  $q$ :  $h(q)=H$ . For multifractal series,  $h(q)$  is a non-linear function: with positive  $q$ , the main contribution to the function  $F_q(s)$  is provided by segments that exhibit large deviations  $F^2(s)$ , while for negative  $q$  the segments with small  $F^2(s)$  dominate.

Fig. 4 shows the fluctuation functions  $F_q(s)$  for the monofractal (a) and multifractal (b) processes of the parameter values  $q = \{-5, -2, 0, 2, 5\}$ .

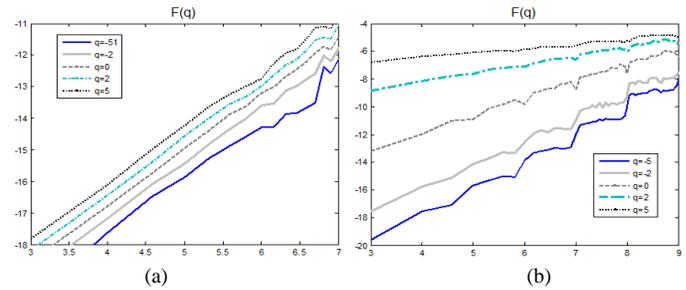


Fig. 4. Functions  $F_q(s)$  for (a) the Monofractal and (b) Multifractal Time Realizations.

We can offer the following step-by-step algorithm for the estimation of the multifractal characteristics of time series  $X(t)$ ,  $t = 1, 2, \dots, n$ .

1. For the calculation, it is necessary to convert the original series to cumulative  $X^{cum}(k) = \sum_{t=1}^k X(t)$ ,  $k = 1, 2, \dots, n$ . If the original time series is cumulative, this step is skipped.
2. The range  $q$  is specified, in which it is required to determine the generalized Hurst exponent  $h(q)$ .

The series  $X^{cum}(t)$  is divided into  $N$  non-overlapping segments of length  $\tau$  and fluctuation function is calculated for each segment where  $Y_m(t)$  is a local  $m$ -polynomial trend within a given segment.

3. The function  $F(\tau)$  is averaged over the whole series  $X^{cum}(t)$ :

$$F(\tau) = \sqrt{\frac{1}{N} \sum_{i=1}^N F_i^2(\tau)}$$

4. The fluctuation function is calculated for value  $q$ :

$$F^2(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} (X^{cum}(t) - Y_m(t))^2$$

5. The value  $\tau$  increases and steps 3–5 are repeated.
6. Linear regression  $\log F_q(\tau) = k \log(\tau) + b$  is constructed by the least squares method. The value  $k$  is equal to the value of  $h(q)$  for a given value  $q$ .
7. Performing steps 2-6 for all given values of the parameter  $q$ , the function  $h(q)$  is obtained.

### C. Wavelet Transform Modulus Maxima Method

The method of wavelet transform modulus maxima (WTMM) is based on the mathematical apparatus of wavelet analysis. The continuous wavelet transform of the function  $X(t)$  is described by

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} X(t)\psi_{ab}(t)dt, \quad (11)$$

where  $W(a,b)$  is the wavelet spectrum,  $\psi_{ab}(t)$  is the wavelet function with scale  $a$  and shift  $b$ .

The WTMM algorithm involves the study of the behavior of a function  $X(t)$  in two stages. At the first stage, the wavelet transform (14) is performed. The result of the wavelet transform is the wavelet spectrum  $W(a,b)$ .  $W(a,b)$  can be represented as the surface of the wavelet coefficients in three-dimensional space. An example of multifractal time series is shown in Fig. 5(a). The surface of wavelet coefficients for this time series is shown in Fig. 5(b). The most important information is contained in local extremum lines of function  $W(a,b)$  (it is shown in Fig. 5(c)), which are searched at every scale  $a$ .

The choice of a suitable wavelet function is determined by several aspects. This function must have  $m$  zero moments [2]. On the one hand, large  $m$  allows ignoring large-scale polynomial trends and analyzing small-scale variations of the function  $X(t)$ . On the other hand, an increase in the number of zero moments leads to an increase in the number of local extrema lines and the appearance of a large number of additional lines terminating on small scales [24]. Such lines are too short for estimating power dependences and become hindrances in the numerical analysis of singularities.

Usually, wavelets of the Gauss family are used as wavelet functions [15], [19], [24]. If high order derivatives of the Gauss function are taken, then the number of additional short lines of local maxima increases, which is due to oscillating "tails" of wavelet functions. Therefore, when conducting a multifractal analysis, it is advisable to limit the value  $m = 2$ .

The basic information about possible local features of  $X(t)$  at a point  $t_0$  lies in the asymptotic behavior of the coefficients  $W(a,t_0)$  at small scales  $a$ . If the coefficients on a small scale diverge,  $X(t)$  has a feature in  $t_0$ . If the coefficients  $W(a,t_0)$  are close to zero in a neighborhood  $t_0$  on a small scale, then  $X(t)$  is regular at this point. The first stage of the WTMM algorithm is completed by construction of the "skeleton" of the coefficients  $W(a,b)$ .

The second stage of the WTMM algorithm is to calculate the partition function:

$$Z(q,a) = \sum_{l \in L(a)} \left( \sup_{a' \leq a} |W(a',x_l(a'))| \right)^q, \quad (12)$$

where  $L(a)$  is the set of all lines maxima  $l$  of the wavelet coefficients modules of the scale  $a$ ;  $x_l(a)$  is the location of the maximum at this scale.

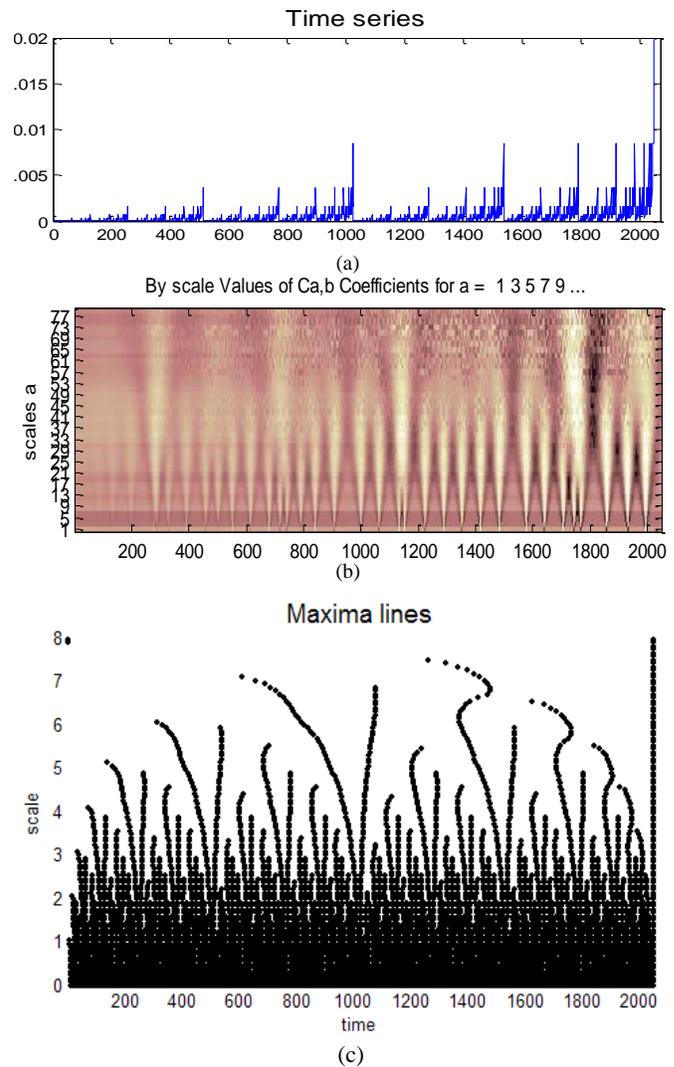


Fig. 5. (a) The Time Series, (b) The Surface of the Wavelet Coefficients, (c) The Lines of Local Maxima.

To calculate  $Z(q,a)$  the maximum value of the wavelet coefficients modulus along each line on scales smaller than the specified scale  $a$  is chosen. In this case, the following relation holds:

$$Z(q,a) \propto a^{\tau(q)}, \quad (13)$$

where  $\tau(q)$  is scaling exponent that is related to the generalized Hurst exponent  $h(q)$  by the ratio [14]:

$$\tau(q) = qh(q) - 1, \quad (17)$$

The multifractal analysis based on the wavelet transform allows studying singularities with negative values of  $q$  [15,16]. Partial functions  $Z(q,a)$  at  $q < 0$  characterizing scaling features for weak singularities (small fluctuations), and at  $q > 0$  for strong singularities (large fluctuations).

The statistical sums  $Z(q, a)$  of a realization of the multifractal process for the values  $q = \{-5, -2, 0, 2, 5\}$  are shown in Fig. 6(a). The scaling exponents  $\tau(q)$  of multifractal and monofractal realizations are shown in Fig. 6(b).

The next step-by-step algorithm can be offered to estimate the multifractal characteristics of time series  $X(t)$ ,  $t = 1, 2, \dots, n$  by WTMM.

1. It is necessary to convert the original series to cumulative  $X^{cum}(k) = \sum_{t=1}^k X(t)$ ,  $k = 1, 2, \dots, n$  for the calculation. If the original time series is cumulative, this step is skipped.
2. The wavelet transform by (14) is performed and the spectrum of wavelet coefficients  $W(a, b)$  is found for this series.
3. A set of local maximum lines is determined for the surface  $W(a, b)$ .
4. The range  $q$  is specified, where it is required to determine the scaling exponent  $\tau(q)$  and generalized Hurst exponent  $h(q)$ .
5. The partial function  $Z(q, a)$  is calculated by (15) for given  $q$  and  $a$ .
6. The linear regression  $\log Z(q, a) = k \log(a) + b$  is constructed by the least squares method. The value  $k$  is equal the value of  $\tau(q)$  for a given value  $q$ .
7. By performing steps 5-6 for all given values of the parameter  $q$ , we obtain the function  $\tau(q)$ .
8. The values of the generalized Hurst exponent  $h(q)$  are determined by (17).

#### D. Multifractal Analysis based on Discrete Wavelet Transform

In subsection 2.2, a method of determining the degree of self-similarity using DWT is considered. It is based on the properties of the detail coefficients obtained by the time series decomposition.

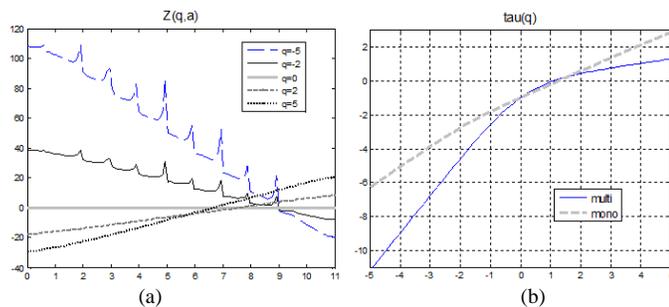


Fig. 6. (a) The Functions  $Z(q, a)$  of Multifractal Realization with different  $q$ ; (b) The Functions  $\tau(q)$  of Multifractal and Monofractal Realizations.

Similarly, in [17,18], a method of estimating multifractal characteristics based on DWT was proposed (MFDWT).

If for a multifractal process  $X(t)$  there are  $q$ -th moments, then for the detail wavelet coefficients obtained by DWT, the following relation holds:

$$E|\det(j, k)|^q = E|\det(0, k)|^q 2^{-j(\xi(q) - \frac{q}{2})}, \quad (14)$$

where  $\det(j, k)$  is  $k$ -th detail wavelet coefficient of level  $j$ ,  $\xi(q)$  is a nonlinear function of a parameter  $q$  associated with the generalized Hurst exponent  $h(q)$  considering (17) by the expression:

$$h(q) = \frac{\xi(q) + 1}{q} - \frac{1}{2}. \quad (15)$$

Based on the expressions (18-19) and taking into account the properties of the detail wavelet coefficients, the following step-by-step algorithm can be offered to estimate the multifractal characteristics of time series  $X(t)$ ,  $t = 1, 2, \dots, n$  by MFDWT.

1. The original series has to be converted to cumulative series  $X^{cum}(k) = \sum_{t=1}^k X(t)$ ,  $k = 1, 2, \dots, n$ . If the original time series is cumulative, this step is skipped.
2. The range  $q$  is specified, in which it is required to determine the multifractal characteristics  $\xi(q)$  and  $h(q)$ .
3. The range of decomposition levels of DWT is selected. The following value is calculated for each decomposition level  $j$ :

$$E_j^{(q)} = \frac{1}{N_j} \sum_{k=1}^{N_j} |\det(j, k)|^q$$

4. Each value  $j$  is assigned a logarithm of  $E_j^{(q)}$ . The tangent of the inclination angle of the approximating straight line is equal to value  $\xi(q)$ .
5. By performing steps 3-4 for all given values of the parameter  $q$ , the function  $\xi(q)$  is obtained.
6. By (19) the corresponding values of the generalized Hurst exponent  $h(q)$  are calculated.

The functions  $\log_2 E_j^{(q)}$  of the multifractal process realizations for the values  $q = \{-5, -2, 0, 2, 5\}$  are presented in Fig. 7(a). The functions  $\xi(q)$  of different multifractal realizations are shown in Fig. 7(b).

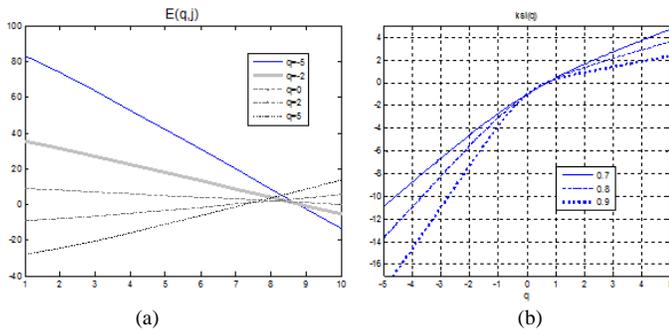


Fig. 7. (a) The Functions  $\log_2 E_j^{(q)}$  of Multifractal Realizations for  $q = \{-5, -2, 0, 2, 5\}$ ; b) the Functions  $\xi(q)$  of Multifractal Realizations.

#### IV. ESTIMATION OF MULTIFRACTAL PROPERTIES OF TIME SERIES

##### A. Experiment Description

The section presents the results of a numerical experiment, during which the realizations of different fractal stochastic processes were modeled: fractal Brownian motion (monofractal process),  $\alpha$ -stable process (bi-fractal process) and binomial multiplicative stochastic cascade (multifractal process). The length of the realizations were chosen to be 250, 500, 1000 and 2000 values. For each generated time series, using the methods of MFDFA, WTMM and MFDWT the functions of the generalized Hurst exponent  $h(q)$  were calculated. Then functions  $h(q)$  were averaged over a set of realizations. The parameter  $q$  had values in the range  $-5 \leq q \leq 5$ .

The following characteristic was chosen as a measure of proximity to theoretical values:

$$\Delta h = \frac{1}{n_q} \sum_q \left| \hat{h}(q) - h_t(q) \right|, \quad (16)$$

where  $h_t(q)$  is the theoretical function of the generalized Hurst exponent;  $\hat{h}(q)$  is sample function;  $n_q$  is the number of values  $q$ ;  $\Delta h$  is the average deviation of the sample value from the theoretical. Since the methods can have different errors for positive and negative values of parameter values  $q$ , it makes sense instead of  $\Delta h$  to consider separately  $\Delta h(q < 0)$  and  $\Delta h(q > 0)$ .

Model time series for estimation.

Fractal Brownian motion. FBM is the most famous and simple model of the self-similar process [25]. On the one hand, the FBM is a self-similar process with the Hurst exponent  $H$ . On the other hand, FBM can be considered as a monofractal process, in which the generalized Hurst exponent is a constant:  $h(q) = H$ .

Stable process with independent increments. A random process is a  $\alpha$ -stable process if its finite-dimensional distributions are stable random variables [26]. The parameter

$\alpha$  is called the stability index and determines the heavy tails of the distribution. For a  $\alpha$ -stable process with independent increments, equality holds.

$$\text{Law}\{X(at)\} = \text{Law}\{a^{1/\alpha} X(t)\}, \quad \forall a > 0, t > 0$$

It is shown that such processes are bi-fractal. The corresponding generalized Hurst exponent is described by:

$$h(q) = \begin{cases} 1/\alpha & q \leq \alpha \\ 1/q & q > \alpha \end{cases}$$

Stochastic binomial cascade. The simplest model of a multifractal process is the deterministic binomial cascade [25]. When constructing it, the initial unit segment is divided into two equal intervals, weight coefficients  $p_1$  and  $p_2 = 1 - p_1$  are assigned to each of them. As a result, in the second step, there are four intervals with weights  $p_1^2$ ,  $p_1 p_2$ ,  $p_2 p_1$  and  $p_2^2$ . With an increase in the number of iterations, we obtain an ordered set of weights, which has multifractal properties. When constructing stochastic cascades, the weight coefficients are values of some random variable [27]. If a random variable with a beta distribution  $Beta(a, b)$  is used as a random variable, then in the case of  $a = b$ , the scaling exponent  $\tau(q)$  can analytically determine over the interval of values  $q > -1$ :

$$\tau(q) = -\log_2 \frac{Beta(\alpha+q, \alpha)}{Beta(\alpha, \alpha)} - 1 \quad (17)$$

The theoretical value of the function  $h(q)$  is determined in accordance with (17).

##### B. The Results of the Generalized Hurst Estimation Exponent by Time Series

Estimation of multifractal characteristics for monofractal realizations.

A typical FBM realization with  $H=0.8$  is shown in Fig. 8(a) and the corresponding realization of increments (fractal Gaussian noise) is shown in Fig. 8(b). The length of realizations is 1000 values.

The sample values of the generalized Hurst exponent calculated by the methods MFDFA (a), WTMM (b) and MFDWT (c) are presented in Fig. 9. A straight line is the theoretical values  $h(q)$ . It is shown that with an increase in the length of the realization, sample characteristics tend to theoretical values. However, at small lengths, the obtained estimates demonstrate false multifractal properties. Therefore, in the case of the realization of a small length, it is necessary to conduct additional studies to confirm the monofractal properties.

Table I represents the values of the deviations  $\Delta h$ , calculated in accordance with (20). The estimates of the generalized Hurst exponent obtained by the MFDFA method are much closer to their theoretical values than the estimates calculated using wavelet transforms. Also, when using the WTMM method, determining the wavelet spectrum, local maxima lines and choosing the appropriate scale range is a

more complicated task than determining the polynomial degree and the segment lengths in MFDFA. The MFDWT method demonstrates incorrect estimation of  $h(q)$  with negative  $q$ , however, for positive  $q$  the estimates by MFDWT method are close to the estimates by WTMM.

Estimation of multifractal characteristics for bi-fractal realizations.

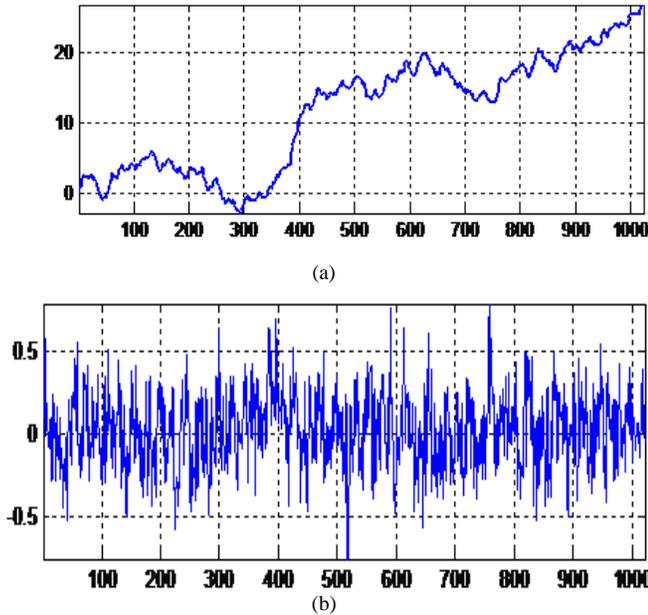


Fig. 8. (a) FBM Realization, (b) Realization of Increments.

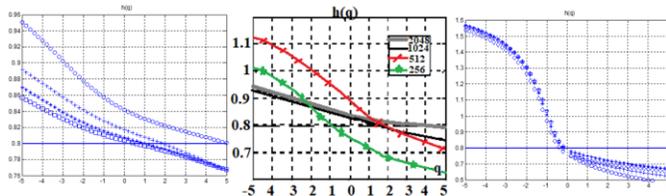


Fig. 9. Sample  $h(q)$  of Monofractal Realization, Calculated by the Methods (a) MFDFA, (b) WTMM and (c) MFDWT.

TABLE I. DEVIATIONS OF SAMPLE  $h(q)$  FOR MONOFRACTAL REALIZATIONS

Length	Method	$\Delta h (q < 0)$	$\Delta h (q > 0)$
250	MFDFA	0.094	0.018
	WTMM	0.18	0.13
	MFDWT	0.54	0.15
500	MFDFA	0.06	0.015
	WTMM	0.3	0.06
	MFDWT	0.51	0.12
1000	MFDFA	0.035	0.013
	WTMM	0.27	0.023
	MFDWT	0.50	0.1
2000	MFDFA	0.025	0.014
	WTMM	0.15	0.018
	MFDWT	0.46	0.08

Let us consider the estimation of  $h(q)$  for the realizations of a  $\alpha$ -stable process. A typical realization of a  $\alpha$ -stable process with  $\alpha=1.5$  is shown in Fig. 10(a) and the corresponding realization of increments is shown in Fig. 10(b). The length of realizations is 1000 values.

The sample functions of the generalized Hurst exponent obtained using the MPDFA, WTMM and MFDWT methods are shown in Fig. 11(a), (b) and (c), respectively.

The obtained estimates of  $h(q)$  show false multifractal properties. The estimates obtained by the MFDFA method are closer to their theoretical values than the estimates calculated using wavelet transforms, as evidenced by Table II. The MFDWT method, as well as in the monofractal case, demonstrates incorrect estimation of multifractal properties. However, for positive  $q$  the estimates of  $h(q)$  by the MFDWT method have the smallest difference with the theoretical values.

Estimation of multifractal characteristics for realizations of a stochastic binomial cascade.

Let consider the estimation of multifractal properties for realizations of stochastic multiplicative binomial cascades whose weights have a beta distribution. For comparison with analytical characteristics (21), calculations were performed in the range  $-1 < q < 5$ .

The typical realization of a binomial stochastic cascade with a length of 1000 values is shown in Fig. 12(a). In this case the weights have a uniform distribution (i.e.  $a = b = 1$ ). The corresponding sample values of the generalized Hurst exponents are shown in Fig. 12(b) by MFDFA, (c) WTMM and (d) MFDWT. The theoretical function  $h(q)$  is shown by the thin line.

The numerical values of the deviations are given in Table III. The WTMM method shows the worst estimation of the stochastic multifractal cascades. The MFDWT method shows good results and it is not inferior to the MFDFA when evaluating multifractal realizations.

Table IV shows the deviations values for the positive  $q$  of all types of fractal time series. This is quite convenient and clear for a general results comparison. The estimation analysis showed that the estimates of the generalized Hurst exponent obtained from short time series have shifts that decrease with the increasing length of the time series. The method MFDFA has significant advantages in accuracy when estimating from short realizations, especially in the case of monofractal ones.

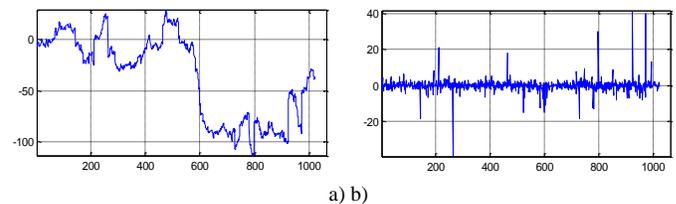


Fig. 10. (a) Realization of a  $\alpha$ -Stable Process, (b) Realization of Increments.

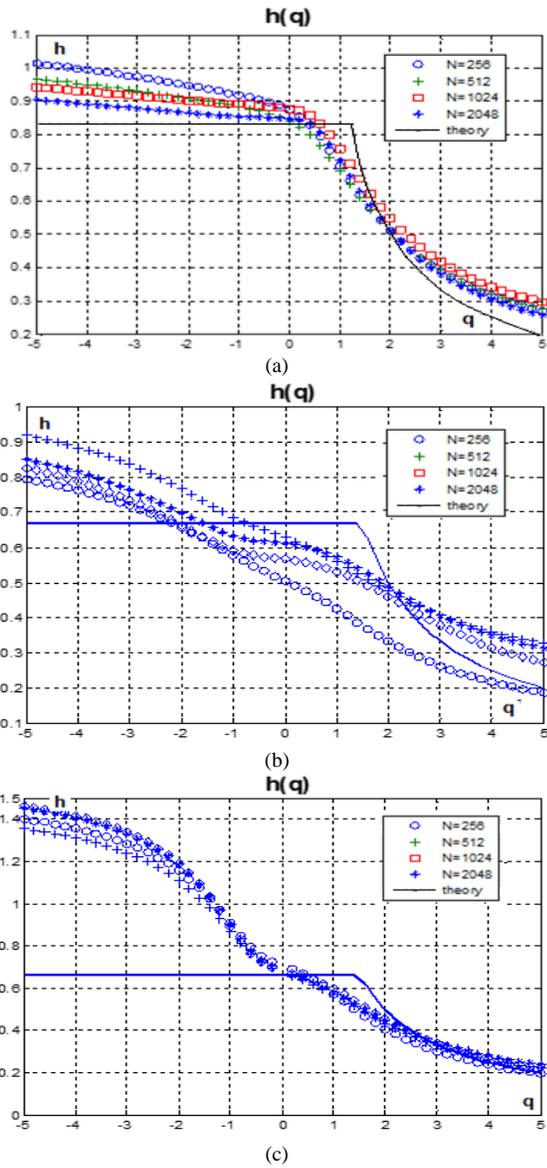


Fig. 11. Sample  $h(q)$  of bi-fractal Realizations, Calculated by the Methods (a) MFDFA, (b) WTMM, (c) MFDWT.

TABLE II. DEVIATIONS OF SAMPLE  $h(q)$  FOR BI-FRACTAL REALIZATIONS

Length	Method	$\Delta h (q < 0)$	$\Delta h (q > 0)$
200	MFDFA	0.096	0.056
	WTMM	0.076	0.12
	MFDWT	0.47	0.047
500	MFDFA	0.064	0.053
	WTMM	0.13	0.086
	MFDWT	0.45	0.042
1000	MFDFA	0.057	0.05
	WTMM	0.08	0.075
	MFDWT	0.52	0.038
2000	MFDFA	0.054	0.041
	WTMM	0.08	0.07
	MFDWT	0.54	0.034

TABLE III. DEVIATIONS OF SAMPLE  $h(q)$  FOR REALIZATIONS OF STOCHASTIC MULTIFRACTAL CASCADE

Length	Method	$\Delta h(q < 0)$	$\Delta h(q > 0)$
250	MFDFA	0.18	0.023
	WTMM	0.8	0.11
	MFDWT	0.17	0.016
500	MFDFA	0.14	0.016
	WTMM	0.5	0.09
	MFDWT	0.17	0.016
1000	MFDFA	0.16	0.012
	WTMM	0.56	0.077
	MFDWT	0.18	0.011
2000	MFDFA	0.18	0.010
	WTMM	0.4	0.061
	MFDWT	0.15	0.01

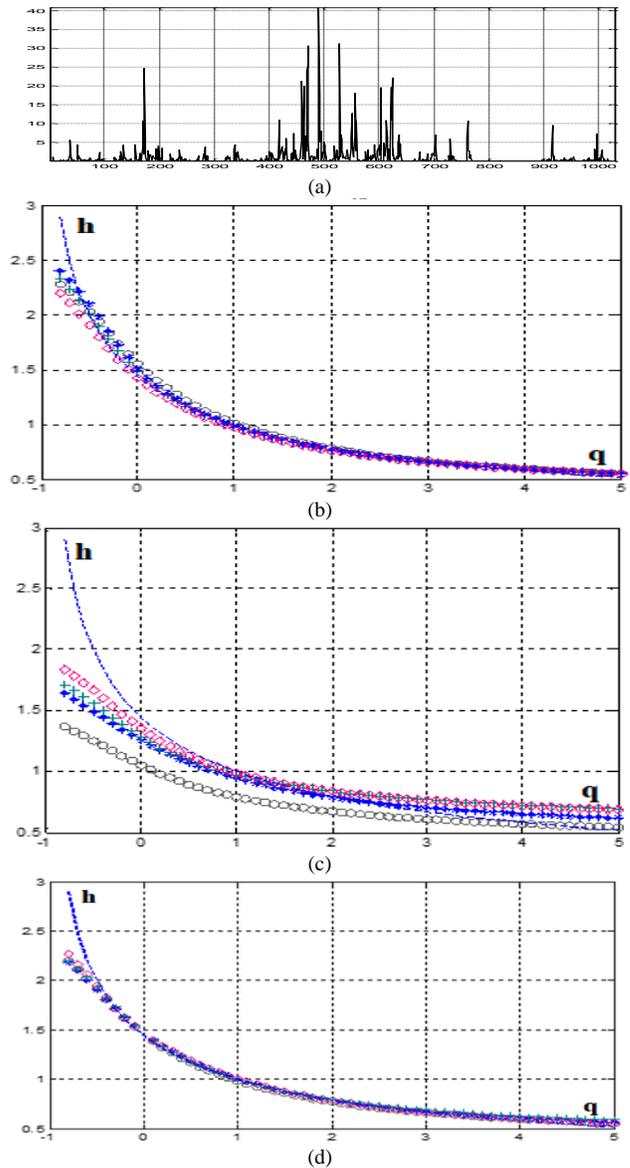


Fig. 12. (a) The Stochastic Cascade Realization;  $h(q)$  Obtained by (b) MFDFA, (c) WTMM, (d) MFDWT.

TABLE IV. DEVIATIONS  $\Delta h(q > 0)$  FOR MONOFRACTAL, BI-FRACTAL, AND MULTIFRACTAL REALIZATIONS

Length	Method	Monofractal	Bifractal	Multifractal
200	MF DFA	0.018	0.056	0.023
	WTMM	0.13	0.12	0.11
	MFDWT	0.15	0.047	0.024
500	MF DFA	0.015	0.053	0.016
	WTMM	0.06	0.086	0.09
	MFDWT	0.12	0.042	0.016
1000	MF DFA	0.013	0.05	0.012
	WTMM	0.023	0.075	0.077
	MFDWT	0.10	0.038	0.011
2000	MF DFA	0.010	0.041	0.010
	WTMM	0.018	0.07	0.061
	MFDWT	0.08	0.034	0.01

### C. Testing the Hypothesis about Monofractality of Time Series

The results of the multifractal analysis of model time series demonstrate the need to develop a tool to distinguish between mono- and multifractal time series. Since the MF DFA method showed the best results in evaluating monofractal realizations, a method based on estimates obtained by MF DFA is proposed. It allows to accept or reject the hypothesis of the presence of time series monofractal properties.

For self-similar processes, the question of the estimates Hurst exponent  $H$  distribution was considered in several publications [2], [8], [19], where it was shown, that estimates are normal random variables. The analysis of the sample distribution of the generalized Hurst exponent  $h(q)$  showed that the estimates at  $q > 0$  have a normal distribution, the parameters of which depend on the value  $q$ . For  $q < 0$  the sample values  $h(q)$  in the general case are not normal random variables.

The multifractal time series have a much greater difference in values  $\Delta h = h(q_1) - h(q_2)$  than monofractal ones. A random variable  $\Delta h$  has a normal distribution  $N(m_h, s_h)$ , the parameters of which depend on the length of the time series and the values  $q$ . The criterion of the monofractal degree was proposed to be considered the value  $\Delta h = h(0.1) - h(5)$ . By numerical simulation of monofractal processes with varying degrees of self-similarity, the sample values  $m_h$  and  $s_h$  were obtained. Table V presented these values for the series of different lengths  $N$ .

The same characteristics calculated from cascade processes with different degrees of heterogeneity, given by the beta distribution parameter  $a$ , are also presented in Table V.

The table shows that even with very weak multifractal properties of the process, mono- and multifractal time series can be distinguished. Thus, the value of a random variable  $\Delta h$  can be used as test statistics for acceptance of the monofractal properties hypothesis. In this case, the null hypothesis is the assumption that the series is monofractal. Having obtained the

estimate  $\hat{h}(q)$  using the MF DFA method, the observable value  $\Delta \hat{h} = \hat{h}(0.1) - \hat{h}(5)$  is calculated. The hypothesis is accepted with a significance level  $\alpha$  if the observable value is within the range of acceptable values:

$$\Delta \hat{h} < m_h(N) + t_\alpha s_h(N),$$

where  $N$  is the length of time series;  $m_h$  and  $s_h$  are values calculated for the monofractal process;  $t_\alpha$  is the corresponding quantile of standard normal distribution.

TABLE V. PARAMETERS OF THE RANDOM VARIABLE  $\Delta h$

	N=250		N=500		N=1000		N=2000	
	$m_h$	$s_h$	$m_h$	$s_h$	$m_h$	$s_h$	$m_h$	$s_h$
Monofractal	0.07	0.05	0.0	0.03	0.03	0.02	0.0	0.02
1	5	2	5	8	5	6	3	2
Multifractal								
a=1	0.87	0.23	0.8 6	0.18	0.85	0.16	0.8 2	0.12
a=5	0.32	0.13	0.2 9	0.09 5	0.28	0.08	0.2 7	0.07
a=10	0.23	0.1	0.2	0.07	0.18	0.06	0.1 7	0.05

### D. A Generalized Approach to Estimating the Multifractal Properties of Small Length Time Series

Summing up the research results, the following scheme for multifractal analysis of some random process represented by a time series  $X(t)$  of length  $N$  can be proposed. It should be noted that before proceeding to fractal analysis, it is necessary to find out from a priori known information whether the series is cumulative (for example, the exchange rate) or increments (for example, information traffic).

The main stages of fractal analysis can involve various methods of estimation. Since the application of the wavelet transform apparatus requires appropriate software and experience, the description of the scheme is structured in such way that the use of wavelet estimation methods is a desirable but not necessary element. However, the application of the MF DFA method is necessary for two reasons: this method has sufficient accuracy and is designed for non-stationary time series. Consider the step by step implementation of the generalized approach to the estimation of the multifractal properties of self-similar time series.

1) A preliminary study of the time series structure is a necessary stage of multifractal analysis, its full description is given in [21] This stage includes determining intervals of various scaling, detecting and removing short-term autoregressive dependence, estimation of the Hurst exponent  $H$ .

2) All three considered methods MF DFA, WTMM, and MFDWT are intended for analysis non-stationary series. But before proceeding to an estimation of multifractal properties, it is necessary to investigate time series structure by correlation function, Fourier spectrum or spectrum of wavelet energy, which allow identifying the trend and cyclical

components of the series. The results of the preliminary study allow selecting the input parameters correctly for the multifractal analysis method.

3) When estimating the generalized Hurst exponent using the MFDFA method, it is necessary to first carry out a rough estimation of the Hurst exponent  $H = h(2)$  using local polynomials of increasing degree and determine the degree to estimation. After that, evaluate the function  $h(q)$ , removing the local polynomial trend of this degree.

4) When evaluating the scaling exponent  $\tau(q)$  by the WTMM method, it needs to select the appropriate scale range  $a_1 \leq a \leq a_2$  on which the local extremums of the function  $W(a, b)$  are determined. The right choice of the scale range allows ignoring the influence of the trend components of the time series. Then it is necessary to carry out a rough estimation of the partial function  $Z(q, a)$  to make sure that there are intervals of linear dependence  $\log Z(q, a)$  on  $\log a$ .

5) When estimating the scaling exponent  $\xi(q)$  using the MFDWT method, it needs to first select the appropriate range of values of levels  $j_1 \leq j \leq j_2$  at which there is a linear dependence  $\log_2 E_j$  on the level number  $j$ .

6) With weakly expressed multifractal properties, it is necessary to test the hypothesis of the monofractal properties of time series.

7) If estimates of multifractal characteristics were obtained by different methods, then to increase accuracy, it is desirable to conduct comparative analysis taking into account a priori information and the results of numerical studies.

## V. FRACTAL ANALYSIS OF REAL TIME SERIES

### A. Fractal Analysis of Temperature Series

Well-known time series with the property of self-similarity are temperature series. The specialized site [28] presents daily temperature series (maximum, minimum and average values) for different cities over 50 years from 1942 to 1992. Fig. 13 shows a fragment of the temperature series for 1952-1956 (a) and the corresponding series of daily increments (b).

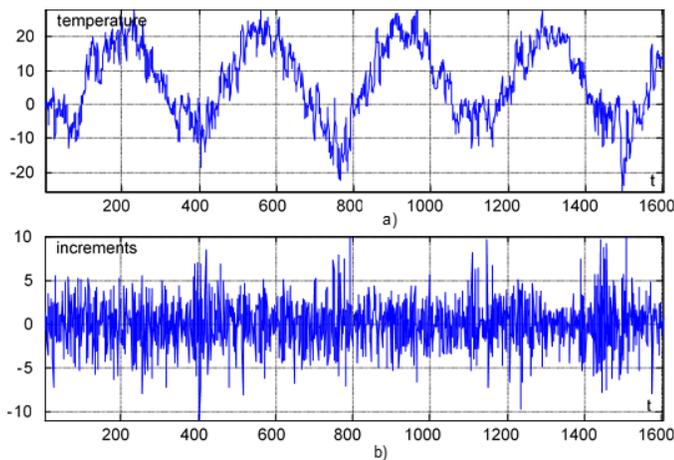


Fig. 13. A Series of Daily Temperatures (a) and a Series of Daily Increments (b).

Before the multifractal analysis, the fluctuation function  $F(\tau)$  was investigated. The presence of a section with a linear dependence of  $F(\tau)$  corresponds to the self-similar behavior of the time series. If the function  $F(\tau)$  has several linear sections, this implies the several time scaling for different time intervals. Fig. 14(a) shows a graph  $F(\tau)$  plotted over a time interval from  $2^5$  (month) to  $2^{9.5}$  (two years) days.

It is obvious that the graph has two sections close to linear with different tilt angles. It is worth noting the high expulsion in the second section when the value of the argument is equal to  $2^{8.5}$  which is almost equal to one year and corresponds to the annual seasonal component of the time series. Multifractal analysis by the MFDFA method was carried out separately for each section. For the first section, the Hurst exponent  $H = 0.21$ , for the second  $H = 0.92$ . Fig. 14(b) shows the multifractal spectrum functions for both sections.

Thus, it can be concluded that at intervals from two months to six months (Section 1), the series of temperature dependence is antipersistent, and at time values from six months to one and a half years (Section 2), the series has a strong long-term dependence.

The research of the monofractal property, carried out by the method presented in subsection 4.3, has shown that in both cases the series has almost monofractal properties. Table VI shows the results of hypothesis testing for the generalized Hurst exponent sample values  $\hat{h}(q)$  of average daily temperatures series. Critical values  $\Delta h$  were obtained based on calculated data for the significance level  $\alpha = 0.05$  and the corresponding series length.

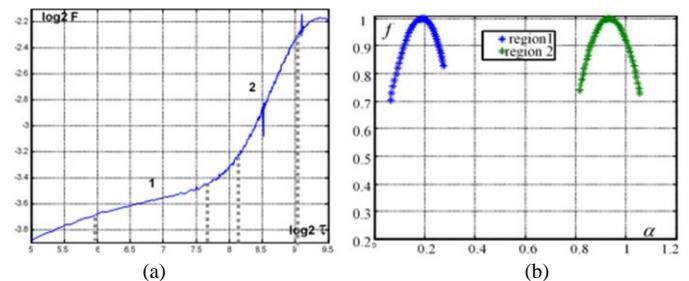


Fig. 14. The Fluctuation Function  $F(\tau)$  for the Temperature Series and the Functions  $f(\alpha)$  for the Sections of Antipersistent and Persistent Dependence.

TABLE VI. HYPOTHESIS TESTING ABOUT THE MONOFRACTAL PROPERTY OF TEMPERATURE SERIES

Series	Length	Critical value $\Delta h$	Observed value $\hat{\Delta h}$	Fractal properties
Temperature series (section 1)	2000	0.0731	0.0264	mono
Temperature series (section 2)			0.0213	mono

### B. Fractal Analysis of Atmospheric Layer Return Echo

One of the methods for studying processes occurring in the atmospheric boundary layer is acoustic sounding and subsequent analysis of both echo signals and ensembles of sequentially obtained echo signal profiles – echograms. Let's conduct a fractal analysis of some of the experimental data obtained at the existing acoustic sounding station of the Kharkiv National University of Radio Electronics.

In this case, the maximum depth of the sounding atmospheric layer  $Z = 372\text{ m}$ ; the value of the resolution in depth  $\Delta Z = 5\text{ m}$ ; the carrier frequency of the sounding pulse  $f = 1800\text{ Hz}$ , sounding pulse repetition period  $T = 2\text{ sec}$ . Fig. 15 shows graphs of signal values  $X(t)$  reflected at different depth of the sounding layer – 50 m and 200 m.

Signals multifractal analysis was carried out by the method of the discrete wavelet transform (MFDWT). Fig. 16 shows the dependence  $\log_2(E_j)$  on the decomposition level  $j$ . Estimates of the Hurst exponent obtained for signals at various depths are respectively equal to  $H_{50} = 0.73$  and  $H_{200} = 0.55$ .

Fig. 17 shows the values of the Hurst exponent  $H$  as a function of the reflective layer depth  $Z$ , calculated for the numerical data of one of the echograms.

As seen from the figure, at low depth, where the proximity of buildings, structures, tall trees and other objects of the metropolis are affecting, the values of the Hurst exponent do not exceed 0.6. Then, due to the influx of heat in the lower layers of the atmosphere, the values of the Hurst exponent increase to 0.75, which indicates the presence of long-term dependence in atmospheric processes at low depth. As the depth and distance from the megapolis increase, the value of Hurst exponent decreases and tends to 0.5, which corresponds to weakly correlated random processes.

A multifractal analysis carried out by MFDWT confirms that with increasing depth of the reflecting atmosphere layer, the correlation structure of the processes changes, which is reflected in a decrease in the values of the generalized Hurst exponent  $h(q)$ . Fig. 18 shows the values of the generalized Hurst exponent for signals reflected at 50, 150, and 200 m depth. In this case, the values of the parameter  $q$  changed in the range  $1 \leq q \leq 10$ .

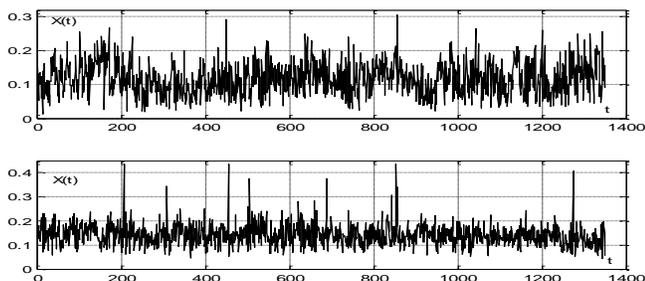


Fig. 15. Reflected Signals at 50 m (Top) and 200 m (Bottom) Depth of the Sounding Layer.

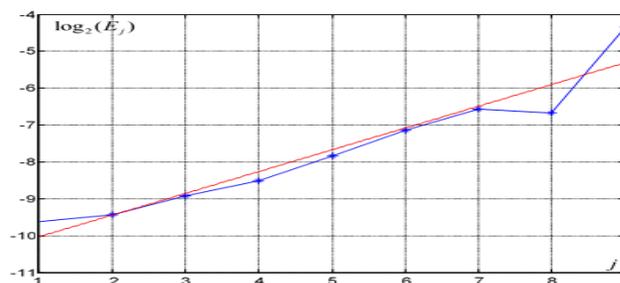


Fig. 16. Dependence  $\log_2(E_j)$  on the Level  $j$  for Signal Reflected at Depth 50 m.

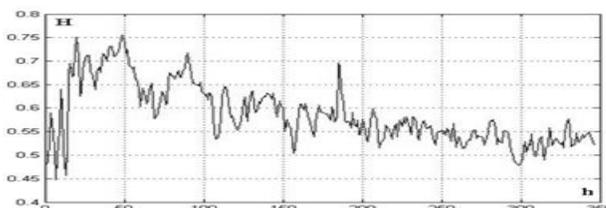


Fig. 17. Hurst Exponent  $H$  as a Function of the Reflective Layer Depth.

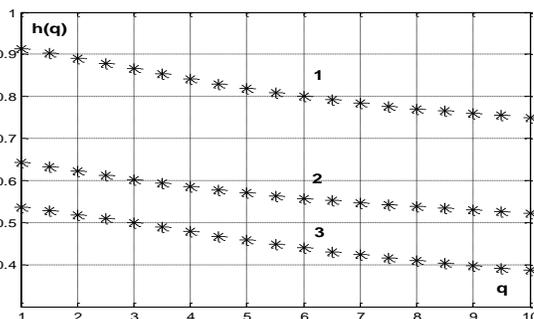


Fig. 18. Generalized Hurst Exponent for Signals Reflected at Depth 50 m (Line 1), 150 m (Line 2) and 200 m (Line 3).

Monofractal property testing presented in Subsection 4.3 has shown that signals reflected at different depth have weak multifractal properties. Table VII has the results of the sample generalized Hurst exponent analysis for signals of various depths. The critical value  $\Delta h$  was obtained using calculated data for significance level  $\alpha = 0.05$  and corresponding series length.

### C. Electroencephalogram Signals Research

Fractal geometry has been used in biology for over a quarter of a century. The use of fractal methods opens up new opportunities in studying the functional organization of living systems. Numerous experimental and clinical data lead to the conclusion that research on the fractal properties of different biological systems will lay the foundation for fractal diagnostics [29], [30].

Multifractal characteristics of electroencephalogram records of different activity.

In [31], the study of how the multifractal characteristics of the electroencephalogram records (EEG) change when a person performs a physical action, and when he imagines that he is doing it, was carried out. Experimental data were taken

from the [32], where are the EEG of persons, who, with one signal clasped their hand into a fist, and with another signal, they only imagined that they were doing it. Fig. 19 shows typical EEG realizations in cases where the respondent squeezed his fist (left) and when the respondent imagines it (right).

Multifractal characteristics of EEG records for the above-mentioned cases were studied using the method of wavelet transform modulus maxima (WTMM). Fig. 20 shows the multifractal spectrum functions  $f(\alpha)$  corresponding to considered EEG records. Spectrum  $f(\alpha)$  for real fist squeezed states are represented by line 1, and line 2 corresponds to an imaginary squeezed case.

Therefore, the functions of the multifractal spectrum obtained from the EEG realization make it possible to differentiate the real and imaginary actions of the respondent.

Multifractal characteristics of EEG for different phases of wakefulness and sleep.

The EEG realizations of laboratory animals, in different phases of wakefulness and sleep, have been researched. Fig. 21 shows typical EEG for various phases: wakefulness (awake), slow-wave sleep (sws) and rapid eye movement sleep (rem).

TABLE VII. HYPOTHESIS TESTING ABOUT THE MONOFRACTAL PROPERTY OF REFLECTED SIGNALS

Series	Length	Critical value $\Delta h$	Observed value $\hat{\Delta h}$	Fractal properties
Depth 50 m	1300	0.075	0.105	multi
Depth 150 m			0.079	multi
Depth 200 m			0.092	multi

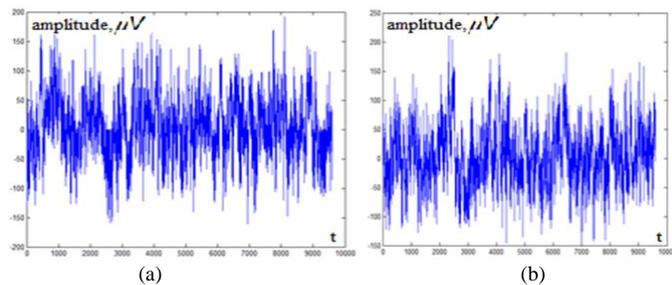


Fig. 19. EEG Records: Respondent Clenches Fist (a) and Respondent Imagines it (b).

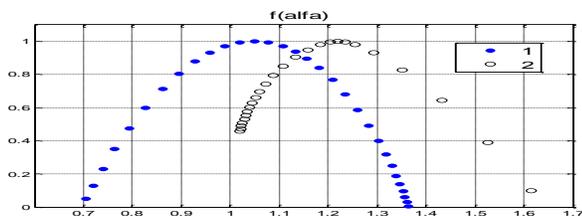


Fig. 20. Functions of  $f(\alpha)$  for Really Squeezed Fist (Line 1) and Imaginary Squeezed Fist (Line 2).

Before the multifractal analysis, the research of the self-similar properties of the EEG realization was carried out, which identified the long-term dependence for the awake, and persistence for the slow sleep phase and weak autocorrelation dependence for the fast sleep phase. Fig. 22 shows a fluctuation function  $F(\tau)$  typical for the EEG realizations in the wakefulness.

Multifractal analysis conducted by the MFDFA method detected significant differences in fractal properties of EEG for wakefulness and sleep. Fig. 23 shows the functions of the generalized Hurst exponent  $h(q)$  for the EEG realizations presented above.

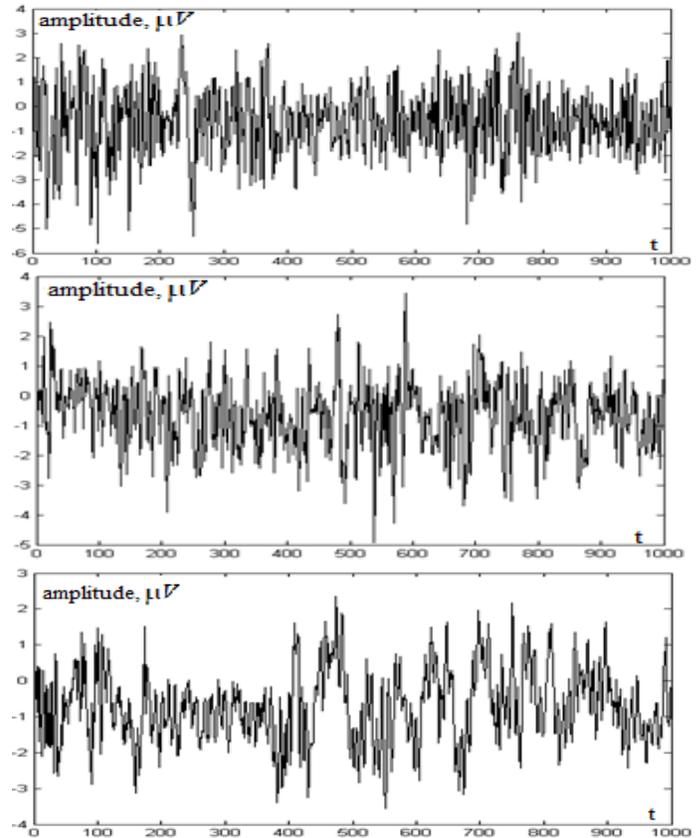


Fig. 21. EEG Realizations: awake (Top), rem (Middle) and sws (Bottom).

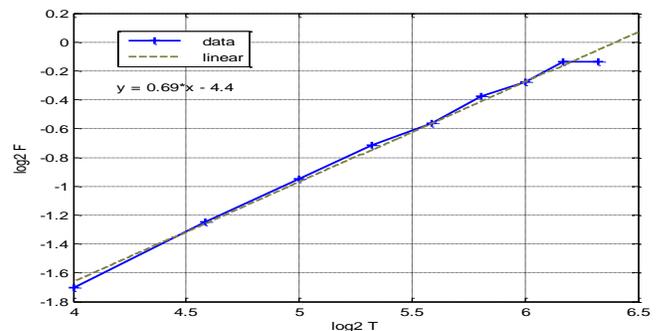


Fig. 22. Fluctuation Function  $F(\tau)$  Typical for Wakefulness.

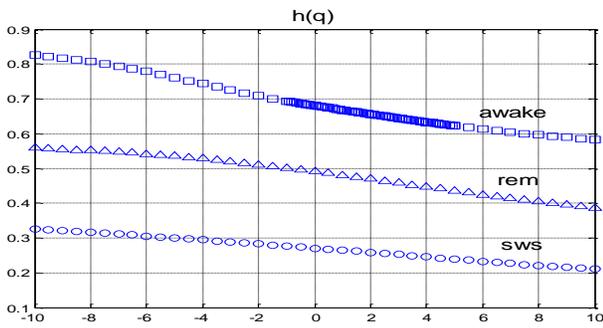


Fig. 23. Generalized Hurst Exponent for AWAKE (Squares), SWS (Circles) and REM (Triangles).

The analysis also showed that there is an undoubted long-term dependence for the EEG realizations in the awake phase: in this case, Hurst exponent  $H$  is significantly exceeded 0.5. The slow sleep phase is characterized by ant persistence, Hurst exponent takes values in the range of less than 0.5. For REM sleep, the Hurst exponent  $H$  is close to 0.5, it takes values greater or less than 0.5. In this case, the EEG realizations are characterized by very weak autocorrelation dependence.

Table VIII presents the testing results of the sample generalized Hurst exponent  $\hat{h}(q)$  for the studied realizations of EEG. The critical values  $\Delta h$  are based on the calculated data for significance level  $\alpha = 0.5$  and corresponding series length.

D. Multifractal Analysis of Seismic Wave

One of the known examples of time series with clearly expressed fractal properties is seismic waves. Nowadays real-time seismic monitoring is not only important but also is the most complicated task of seismological practice. There are various seismic signal detection methods based on amplitude ratio analysis, spectral characteristics, wavelet analysis, etc. Some methods based on the estimation of fractal characteristics of the seismic process such as Hurst exponent and fractal dimension [33], [34].

Numerical studies show that seismic waves have multifractal properties and different stages of their development have different functions of generalized Hurst exponent  $h(q)$ . Fig. 24 (top) shows a seismic time series where two windows are highlighted: before the activity burst and after. Obviously, in these windows, the function of the generalized Hurst exponent  $h(q)$ , shown in Fig. 24 (bottom) has significant differences.

Table IX contains the results of the analysis of generalized Hurst exponent  $\hat{h}(q)$  for studied seismic wave realizations in various stages of development. Critical values  $\Delta h$  are given based on calculated data for the significance level  $\alpha = 0.05$  and corresponding series length.

The results of the analysis allow asserting that seismic wave realizations have strong multifractal properties that allow to distinguish them from white noise which is monofractal. These properties can be used when detecting seismic signals.

E. Analysis Fractal Properties of Community Activities Time Series in Social Networks

In recent years, there have been studies of community dynamics in social groups, which show that the corresponding time series have self-similarity properties [35], [36]. The comparative fractal analysis for two groups on the social network Facebook found on keywords related to cyber threats was carried out. Two groups ThreatPost and ThreatSignal were selected to research each with 14000 and 84000 users respectively. For each of these social groups, data on the number of likes, comments, and involvement levels over the past five years has been collected.

Table X presents the results of the sample generalized Hurst exponent  $\hat{h}(q)$  for the studied series. The critical values  $\Delta h$  are given based on calculated data for the significance level  $\alpha = 0.05$  and the corresponding series length.

TABLE VIII. HYPOTHESIS TESTING ABOUT MONOFRACAL PROPERTY OF EEG

Series	Length	Critical value $\Delta h$	Observed value $\hat{h}$	Fractal properties
EEG (really squeezed fist)	2000	0.0721	0.15	multi
EEG (imaginary really squeezed fist)			0.12	multi
awake EEG	2500	0.0603	0.081	multi
rem EEG			0.072	multi
sws EEG			0.058	mono

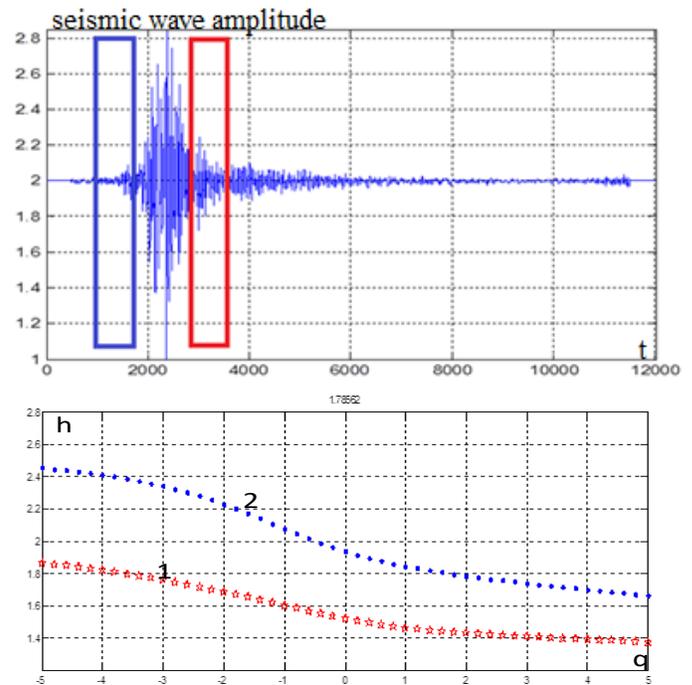


Fig. 24. Seismic Series and  $h(q)$  before Burst (Line 1) and after (Line 2).

TABLE IX. HYPOTHESIS TESTING ABOUT THE MONOFRACTAL PROPERTY OF SEISMIC WAVES

Series	Length	Critical value $\Delta h$	Observed value $\Delta \hat{h}$	Fractal properties
Seismic wave in the initial stage	1000	0.0861	0.1344	multi
Seismic wave in the final stage			0.3212	multi

TABLE X. HYPOTHESIS TESTING ABOUT MONOFRACTAL PROPERTY OF COMMUNITY ACTIVITIES SERIES

Series	Length	Critical value $\Delta h$	Observed value $\Delta \hat{h}$	Fractal properties
ThreatPost	1000	0.0861	0.9522	multi
ThreatSignal			0.3518	multi

Fig. 25 (top) shows a time series of a daily number of likes for both groups. The fractal analysis showed that these time series have strong multifractal properties. Fig. 25 (bottom) shows the values of the generalized Hurst exponent for these series. Both series have persistence ( $H_1, H_2 > 0.5$ ), but the series of likes for the group ThreatSignal has much greater fractal heterogeneity, which is expressed in a much greater range of values  $\Delta h(q)$ .

Research has shown that a series of likes are quite strongly correlated with the series of involvement levels, so their multifractal characteristics are very close. The series of comments have a rather close to each other multifractal structure.

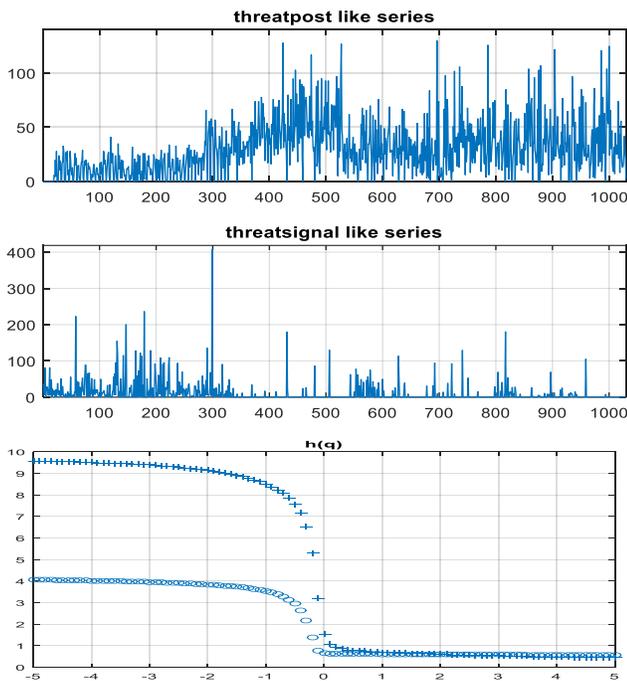


Fig. 25. Time Series of Likes for Groups and Corresponding Values  $h(q)$   
(+ - Group ThreatPost, o - ThreatSignal).

Thus, the research confirmed that many of the time series of social network activity indicators have fractal properties and the application of fractal analysis allows to detect differences and to reveal characteristic features of different social groups' dynamics.

## VI. CONCLUSION

The work has considered the features of the numerical implementation of multifractal analysis methods: multifractal detrended fluctuation analysis, wavelet transform modulus maxima, multifractal analysis using discrete wavelet transform. The properties of generalized Hurst exponent estimates obtained by these methods from short time series have been investigated. For this testing, each method on realizations of fractal processes of different types has been carried out. The advantages and disadvantages of each method were considered.

It is shown that the estimates of the generalized Hurst exponent, obtained from short time series, have a bias that decreases with increasing series length. It is shown that the method of multifractal detrended fluctuation analysis has significant advantages in accuracy when evaluating the characteristics of monofractal processes. A method has been developed that allows to accept or reject the hypothesis of the presence of monofractal properties in a time series. It is based on the study of sample values of the generalized Hurst exponent, obtained by the method of multifractal detrended fluctuation analysis.

Summing up the results of research, a generalized approach to the estimation of multifractal properties of time series and practical recommendations for the implementation of this approach are proposed. Examples of the practical use of the considered approach and methods for multifractal analysis of biomedical signals, natural phenomena and social networks are presented.

It should be noted certain limitations of the proposed approach. Non-stationary time series are considered in the work, but multifractal properties are assumed to be practically constant during the time realization. This requirement is consistent with the short length of the studied time series. In multifractal analysis of sufficiently long time series, it makes sense to conduct additional studies of fractal characteristics using the sliding window method.

## ACKNOWLEDGMENTS

This research was supported by the Deanship of Scientific Research at Prince Sattam bin Abdulaziz University.

## REFERENCES

- [1] Brambila, F.: Fractal Analysis - Applications in Physics, Engineering and Technology, <https://www.intechopen.com/books/fractal-analysis-applications-in-physics-engineering-and-technology>, last accessed 2020/03/18.
- [2] Abry, P., Goncalves P., Vehl, J. L.: Scaling, Fractals and Wavelets. John Wiley & Sons, London (2009).
- [3] Daradkeh, Y.I., Kirichenko, L., Radivilova, T.: Development of QoS Methods in the Information Networks with Fractal Traffic. International Journal of Electronics and Telecommunications 64(1), 27-32 (2018). doi: 10.24425/118142.

- [4] Ageyev, D., Kirichenko, L., Radivilova, T., Tawalbeh, M., Baranovskyi, O.: Method of self-similar load balancing in network intrusion detection system. In: 2018 28th International Conference Radioelektronika (RADIOELEKTRONIKA), Prague, Czech Republic, pp. 1-4 (2018). doi: 10.1109/RADIOELEK.2018.8376406.
- [5] Kirichenko, L., Bulakh, V., Radivilova, T.: Fractal time series analysis of social network activities. 2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kharkov, Ukraine, pp. 456-459 (2017). doi: 10.1109/INFOCOMMST.2017.8246438.
- [6] Kantelhardt, J.W.: Fractal and Multifractal Time Series, 2008, <https://arxiv.org/abs/0804.0747>, last accessed 2020/03/18.
- [7] Clegg, R. G.: A practical guide to measuring the Hurst parameter. International Journal of Simulation. Systems, Science & Technology 7(2), 3-14 (2006).
- [8] Kirichenko, L., Radivilova, T., Deineko, Z.: Comparative Analysis for Estimating of the Hurst Exponent for Stationary and Nonstationary Time Series. Information Technologies and Knowledge 5, 371-388 (2011).
- [9] Xike Zhang, Gui Zhang, Luo Qiu, Bo Zhang, Yurong Sun, Zifan Gui, Qiuwen Zhang A Modified Multifractal Detrended Fluctuation Analysis (MFDFA) Approach for Multifractal Analysis of Precipitation in Dongting Lake Basin, China Water 2019, 11, 891; doi:10.3390/w11050891.
- [10] Kirichenko, L., Radivilova, T., Bulakh V.: Machine Learning in Classification Time Series with Fractal Properties. Data 4(1) 5, 1-13 (2019). doi:10.3390/data4010005.
- [11] L. Kirichenko, T. Radivilova and V. Bulakh, "Machine Learning in Classification Time Series with Fractal Properties." Data, vol.4(1) 5, pp.1-13, 2019. doi:10.3390/data4010005.
- [12] Abry, P., Veitch, D.: Wavelet analysis of long-range dependent traffic. IEEE/ACM Transactions Information Theory 1(44), 2-15 (1998).
- [13] Flandrin, P.: Wavelet Tools for Scaling Processes. Turbulence: measurements and signals. Lecture given at the Summer School, Lyon, pp.123-128 (2002).
- [14] Kantelhardt, J.W., Zschiegner, S.A., Bunde, A., Havlin, S., Koscielny-Bunde, E., Stanley, H.E.: Multifractal detrended fluctuation analysis of non-stationary time series. Physica A. 316, 87-114 (2002).
- [15] Muzy, J.F., Bacry, E., Arneodo, A.: Wavelets and Multifractal Formalism for Singular Signals: Application to Turbulence Data. Physical review letters 67, 3515-3518 (1991).
- [16] Muzy, J.F., Bacry, E., Arneodo, A.: Multifractal formalism for fractal signals: the structure-function approach versus the wavelet-transform modulus-maxima method. Physical Review E, American Physical Society (APS) 47 (2), 875-884 (1993).
- [17] Veitch, D., Abry, P., Flandrin, P., Chainais, P.: Infinitely Divisible Cascade Analysis of Network Traffic Data. In: Proceedings of 2000 IEEE International Conference: Acoustics, Speech, and Signal Processing, ICASSP'00, vol.1 (2000). doi: 10.1109/ICASSP.2000.861931.
- [18] Abry, P., Flandrin, P., Taqqu, M.S., Veitch, D.: Self-similarity and long-range dependence through the wavelet lens. In: Doukhan P., Oppenheim G., Taqqu M.S. (Eds.), Long Range Dependence: Theory and Applications, pp. 125-160 (2002).
- [19] Oswiecimka, P., Kwapin, J., Drozd, S.: Wavelet versus detrended fluctuation analysis of multifractal structures. Physical Review E: Statistical, Nonlinear, and Soft Matter Physics 74, 161-203 (2006).
- [20] Lopez, J.L., Contreras, J.G.: Performance of multifractal detrended fluctuation analysis on short time series, <https://arxiv.org/pdf/1311.2278.pdf>, last accessed 2020/03/18.
- [21] Kirichenko, L., Radivilova, T., Bulakh, V. Generalized approach to Hurst exponent estimating by time series. Inform. Autom. Pomiar Gospod. Ochr. Srodowiska 8, 28-31 (2018). doi:10.5604/01.3001.0010.8639.
- [22] Bulakh, V., Kirichenko, L., Radivilova, T.: Time Series Classification Based on Fractal Properties. In: 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, pp. 198-201 (2018). doi: 10.1109/DSMP.2018.847853.
- [23] Ivanisenko, I., Kirichenko, L., Radivilova, T.: Investigation of self-similar properties of additive data traffic. 2015 Xth International Scientific and Technical Conference "Computer Sciences and Information Technologies" (CSIT), Lviv, pp. 169-171 (2015). doi: 10.1109/STC-CSIT.2015.7325459.
- [24] Pavlov, A.N., Anisshchenko, V.S.: Multifractal analysis of complex signals Physics - Uspekhi 50 (8), 819 - 834 (2007). doi: 10.1070/PU2007v050n08ABEH006116.
- [25] Feder, J.: Fractals. Plenum, New York, 1988.
- [26] Nakao, H.: Multi-scaling properties of truncated Levy flights. Physics Letters A 266(4-6), 282-289 (200). doi: [https://doi.org/10.1016/S0375-9601\(00\)00059-1](https://doi.org/10.1016/S0375-9601(00)00059-1).
- [27] Riedi, R.H.: Multifractal Processes. Available online: [https://www.researchgate.net/publication/2839202\\_Multifractal\\_Processes](https://www.researchgate.net/publication/2839202_Multifractal_Processes), last accessed 2020/03/18.
- [28] Weather archive for the cities of the CIS (19th and 20th centuries), [http://thermo.karelia.ru/weather/w\\_history.php](http://thermo.karelia.ru/weather/w_history.php), last accessed 2020/03/16.
- [29] Ching, E. S., Tsang, Y. K.: Multifractality and scale invariance in human heartbeat dynamics. Physical Review E 76(4), 041910 (2007).
- [30] Harikrishnan, K. P., Misra, R., Ambika, G.: Can the multifractal spectrum be used as a diagnostic tool. Chaotic Modeling and Simulation 1, 51-57 (2013).
- [31] Alghawli, A., Kirichenko, L.: Multifractal Properties of Bioelectric Signals under Various Physiological States. Information Content & Processing International Journal 2(2), 138-163 (2015).
- [32] PhysioNet: the research resource for complex physiologic signals, [www.physionet.org](http://www.physionet.org), last accessed 2020/03/18.
- [33] Boschetti, F., Dentith, M. D., List, R. D.: A fractal-based algorithm for detecting first arrivals on seismic traces. Geophysics 61(4), 1095-1102 (1996).
- [34] Cao, M. S., Ren, Q. W., Wang, H. H., Gong, T.: A method of detecting seismic singularities using combined wavelet with fractal. Chinese journal of Geophysics 48(3), 740-749 (2005).
- [35] Yang, C. C., Sageman, M.: Analysis of terrorist social networks with fractal views. Journal of Information Science 35(3), 299-320 (2009).
- [36] Liu, Q., Zhao, X., Willinger, W., Wang, X., Zhao, B. Y., Zheng, H.: Self-similarity in social network dynamics. ACM Transactions on Modeling and Performance Evaluation of Computing Systems (TOMPECS) 2(1), 1-26 (2016).

# Play-Centric Designing of a Serious Game Prototype for Low Vision Children

Nurul Izzah Othman<sup>1</sup>, Nor Azan Mat Zin<sup>2\*</sup>, Hazura Mohamed<sup>3</sup>

Faculty of Information Science and Technology  
The National University of Malaysia  
43600 Bangi, Malaysia

**Abstract**—Currently, with the advancement of Information and Communications Technology (ICT), gaming industry becomes one of the fastest growing industries. This trend leads to development of serious games as alternative tool for creating effective learning experience. However, most educational applications such as serious games mainly used visuals such graphics and animations that pose challenges for low vision children. Visually impaired users especially children with low vision would face difficulty using the applications. They have problems to see highly visual elements of the games. Accessibility refers to how a certain software or application is accessible to disabled users. Several accessibility aspects should be considered when designing user interfaces for children with low vision. Thus, a game designed to fulfill their needs is needed. However, the challenge of serious game design is not only to consider users' accessibility needs, but also the playability aspects as well so that the visually impaired children can enjoy playing regardless of their disabilities. This paper presents a study on designing a low fidelity serious game prototype for low vision children using play-centric design approach, focusing on playability to obtain feedback from low vision children. Then, based on users' feedbacks, the game prototype will be refined to improve the game design.

**Keywords**—*Serious game; play-centric design; accessibility; low vision; low fidelity prototype*

## I. INTRODUCTION

The development of Information and Communications Technology (ICT) leads to the gaming industry becoming one of the fastest growing industries. Games become the choice of younger generation. Today, games are important for leisure activities of most children. The game features will make learning become interesting and rewarding for children [1]. This trend leads to the development of serious games around the world. Serious games act as an alternative tool for creating learning experiences [2]. Serious games are digital games that have been designed with goals beyond entertainment such as for communication, defense, training, healthcare, and marketing [3]. The purpose of a serious game is basically to educate users while enabling an enjoyable experience [4]. As an example, serious games have been designed for training purposes. InfoSecure is a serious game designed to create awareness on information security [5]. Serious games are also used for health-related problems including rehabilitation, neurological disabilities, psychomotor coordination, speech problems, and obesity [6], as in the case of Kinect game for autistic rehabilitation. Besides, serious games have also been

implemented in learning [7]–[9]. In learning, serious games provide benefits because they are fun and motivate students to participate in the learning activities [5]. However, how far the serious game is accessible to the disabled user is not widely studied. Accessibility refers to how a certain software or application is accessible to disabled users.

According to [10], accessibility refers to how disabled user is enabled to use ICT systems, applications, and websites. Accessibility refers to ICT systems or applications that are designed and developed so that user with disabilities can also use them. They could perceive, understand, navigate, and interact with the ICT systems. Accessibility also enables products to be more beneficial for all users and makes products more user friendly to everyone [11]. Accessibility encompasses auditory, cognitive, neurological, physical, speech, and visual components [12]. Different types of disability may become a barrier in accessing ICT applications. For example, a user with hearing impairment has difficulties to access audio while a visually impaired user has difficulties to access visual information. They have different accessibility requirements due to their disabilities. Therefore, ICT applications should be designed with the consideration of accessibility aspect so that the application can be used by users with disabilities. The user interface of ICT application is designed based on certain disabilities such as physical, cognitive, audio, and visual. For users with physical disabilities, alternative environment should be designed instead of using mouse or input to computers because they would encounter difficulties using the tools. Users with cognitive disabilities require suitable user interface to support their cognitive disabilities. Meanwhile, hearing impaired users need graphical elements on the user interface as an alternative to audio due to their disability to access audio elements. On the other hand, visual impairment may cause difficulties to access graphical elements on the user interface. Thus, visually impaired user needs audio as a visual substitute on the user interface.

Most ICT applications use graphical based interfaces. Therefore, visually impaired users especially children with low vision would face difficulty using the applications. Low vision is one of the visual impairments that causes a person to encounter difficulties in performing daily activities. Low vision occurs when the visual acuity decreases and can not be corrected by glasses or contact lenses. The level sight of a person who has low vision does not exceed 6/60 or 20/200 [13]. As a result, they have problems to see visual elements

\*Corresponding Author

and subsequently may encounter challenges in learning. The choice of efficient tools is important to enable children with low vision to access learning materials. Many educational applications such as serious games are based on visual and animation that pose challenges for children with low vision. According to [14],[15], the interfaces usually cause problems for them due to small text, graphic, and crowded backgrounds. They need high contrast and bright colours. Moving objects also present problems to them. Due to these issues, they need an accessible game designed to fulfill their play experience.

Despite the accessibility barriers, children with low vision should have the opportunity to play serious games like other normal children and have fun along the way. Therefore, serious game designs for them should not only consider the accessibility needs, but the playability aspects as well so that the visually impaired children can enjoy playing regardless of their disability.

Design phase is important in game development. A serious game design for children with low vision also needs to be balanced between serious purposes and gaming experience in order for them to have a chance to experience games like normal children. Thus, the prototyping stage plays an important role in designing a game. The game prototype that has been designed can be tested and redesigned iteratively based on player or user feedback. This paper presented a study to design a low fidelity serious game prototype for children with low vision using play-centric design approach. The prototype was tested for the gaming experience.

The next section of this paper presents the Literature review, followed by Methodology in Section III, Results and discussion in Section IV and finally Conclusion in Section V.

## II. LITERATURE REVIEW

### A. Low Vision

According to [13], about 1.3 billion people live with visual impairment. Visual impairment becomes a challenge for visual accessibility. Visual accessibility is not limited to blindness only. Low vision is a type of visual impairment that affects people around the world. Low vision is vision loss that affects daily activities which can differ among individuals. Eye care professionals use the term “low vision” to describe permanently reduced vision that cannot be improved with regular glasses, contact lenses or surgery.

Children with low vision have sight balance that can be assisted by using low vision aids. The aids could help someone in doing daily activities. Several low vision aids such as monoculars, magnifying lenses, and non-optic devices are used to improve the ability to see visual elements [16]. Apart from that, children with low vision use non-optical devices such as reading lights, bold line papers, and large print books. Those with low vision can also use computers and software if the interfaces are designed based on their needs. However, children with low vision has several issues that become a difficulty for them to access ICT systems or applications. They have difficulty to access visual elements on user interfaces. The contents of ICT applications such as text, graphics, and animations may be inaccessible to them. Instead, audio elements can be used to substitute visual elements [12].

Several aspects should be considered when designing user interfaces for children with low vision such as screen design, graphic presentation, text readability, and graphic colours [16]. These considerations are important to ensure they benefit from the development of technology especially serious games because everyone should have the opportunity to play serious games including children with low vision.

1) *Designing serious game for low vision children:* The serious game concept was introduced by Abt in 1970. He presented the use of board and card games or pen and paper based games. This concept was also presented in the form of war games during a project conducted by Atari in the 1980s for military training. Designs of a serious game are different based on target users since they are different [17]. Despite the potential of serious games, research on serious games for visually impaired users are limited because most existing games for visually impaired users are for entertainment purposes. Audio acts as a replacement for visual information in games. Thus, several audio games have been developed for visually impaired users [18]. Original games that have been modified to enable the visually impaired users to play them include Audio Quake, Rock Vibe, Terraformers, Em Busca do Santo Grau, and many others [19]–[23]. Apart from audio, haptic interactions are also used to replace visual information in games [24]. As an example, Game2senses is one of digital games which use haptic to replace visual element [25]. However, for visually impaired users such as children with low vision, they still have vision balance which enables them to play serious games like normal children if the games are designed based on their needs. The user interface should be accessible and easy to use so they can play the game rather than be stressed by the user interface. According to [15], an accessible game will provide equal opportunity for playing experience to its players. However, the gaming industry has been facing a lack of accessible games [26].

Generally, people with visual impairment could not play mainstream video games because the important game content is in the form of graphics. Therefore, several accessibility elements have been identified from literature and are described as in Table I. These elements should be considered when designing serious games for low vision children.

A serious game should be balanced between game enjoyment and serious purpose. According to [33], learning is usually considered as a serious activity and not for the purpose of fun. Nevertheless, research on serious games have shown that learning through games provide positive effects because children can learn while playing. Game becomes an effective pedagogy. Children will be bored if games are not entertaining. Children could learn and improve their skills during playing and get motivated by fun elements in games. Therefore, specific attention should be given to the game design so that children can focus and enjoy playing the game. These highlight the importance of play-centric approach in game design process.

TABLE I. ACCESSIBILITY ELEMENTS FOR LOW VISION CHILDREN

Accessibility Elements	Descriptions
Multimedia (Audio, Text, Graphics, and Animation) [15], [27]–[29]	<ol style="list-style-type: none"> <li>1. Sound is used as accessibility element and important for those with vision problems. Audio instruction have to be clear. Sound recordings are incorporated in dialogues, menus, and tasks. One sound should be played at a time because low vision children only focus on one sound at a time.</li> <li>2. Text-to-speech instructions should include audio.</li> <li>3. If the low vision children does not understand a task, the task can be listened again.</li> <li>4. Format and style for text should use at least size 18 and up.</li> <li>5. The colour combinations for fonts and backgrounds are very contrasting.</li> <li>6. The best type of text for visually impaired children is Serif.</li> <li>7. Information in the form of graphics need to be bigger</li> <li>8. Use of bright colour</li> </ol>
Navigation [27], [30]–[32]	<ol style="list-style-type: none"> <li>9. Menu layout should be simpler in design.</li> <li>10. Start button and menu should include audio to assist navigation.</li> </ol>
Language [28], [31]	<ol style="list-style-type: none"> <li>11. Use language that user can understand.</li> </ol>
Object Speed [31]	<ol style="list-style-type: none"> <li>12. Animation movements should be slower for low vision children</li> <li>13. Animation movements should be in the same direction and not the opposite.</li> </ol>
Screen Design [28] [30]	<ol style="list-style-type: none"> <li>14. Simple screen design that reduces children's cognitive load. The user interface is easy to be adapted for low vision children.</li> </ol>

### B. Play-Centric Design

According to [34], play-centric design is an approach that puts player experience as a priority in the design process that creates audience involvement with meaningful interactions and complex storylines. It is an iterative method that depends on feedback from players at the early stage of design process. Ideas on satisfying gaming experience are generated during the initial process. In play-centric design, ideas are prototyped and playtested during the design phase. Prototyping is important in the early stages of game design and becomes a tool for playtesting. Prototypes allow designers to obtain feedback from users and get ideas. Prototyping is an important part in design process in order to understand player experience. Different approaches include paper prototyping, physical prototyping, scenario writing, and interactive prototyping. Prototype also acts as a basis for playtesting. Playtesting is conducted during design process to identify whether the game achieves player expectations and obtain feedback regarding the play experience. There are several ways to conduct playtesting such as informal and qualitative. Table II shows several methods for playtesting.

By prototyping and playtesting, tester will know whether the prototype is playable or not. Therefore, play-centric design is important in game design and development. Based on the approach, the game prototype that has been designed will be tested and redesigned based on tester feedback. Next section discusses the methodology in this study.

TABLE II. PLAYTESTING METHODS

Playtesting Methods	Descriptions
One on one testing	Testing is conducted with individuals, where tester watches them play the game.
Group testing	Testing is conducted with a group of people which allows them to play together and ask questions
Feedback Forms	Testing is conducted by giving each person who play the game a set of questions to answer after playing.
Interview	Tester sits down face to face with playtesters and give an interview after the testing session.
Open discussion	One on one discussion or group discussion after playtesting session where tester guides the conversation and prepares specific questions.

### III. METHODOLOGY

The accessibility elements were applied to the game design. so that the storyline could be easily understood by children. Language is an important element in order to enhance accessibility. In this game, Malay language was used as a medium to convey the game content. The multimedia elements were presented in the storyline with bigger character sizes and using bright colours. The size of text used to convey the storyline should be bigger and the colour for the character and text should be in high contrast with the background colour. The storyline should also be accompanied with background audio. Furthermore, in the game design, the multimedia elements should be adapted with the fantasy elements where the game environment should use bright colours and have high contrast between the graphics and background colour. The graphics for fantasy elements should also be minimal. In order to design rewards for the game, the text size should be bigger and accompanied with background audio. The colours used for graphics should be bright and have high contrast. The challenges should be adapted to the children's ability. Meanwhile, graphics, animations, and audio were incorporated into the challenges with bigger graphic size, bright colours, and audio instructions for the tasks. Audio can also be used for guidance in game obstacles. The game's objective should be clear and included in the background audio so that children with low vision can understand it.

Apart from the accessibility aspect, gaming experience is important in order to ensure the children have fun and improve the playability of the game. Based on the play-centric design approach, there are four phases involved: conception phase, design and prototyping, playtesting, and implementation.

#### A. Serious Game Design and Prototyping

The design of serious game starts with conception phase. There are several game aspects to be identified in this phase such as objective, genre, content, character, storyline, gameplay, theme, and target player. In designing a serious game for children with low vision, the intention is to avoid them from getting stressed during playing. The children should feel entertained and comfortable while learning. We chose an adventure game and set the storyline in the animal kingdom by generating simple conflicts between the animals. This serious game is a math game that targets children with low vision aged seven years old. Friendship and kindness among friends are the theme for this game. This game was

developed for Android tablets. The storyline of this game is about a rabbit named “Bunny” who tries to save his friends who are kidnapped by a tiger. Bunny needs to complete the tasks provided for each level in order to get clues and tools to be used in the next challenges. The clues and tools will assist Bunny to save his imprisoned friends from the tiger. The game environment revolves around the animal kingdom which consists of a jungle, rivers, and an old palace. Other characters in the game are monkey, bear, tiger, and turtle who help Bunny to save his friends. Animals are selected because children can identify better with animal characters [35], [36].

The game content involves mathematical subject for Primary Levels 1 (7 years old) and 2 (8 years old) which consists of three levels. The player will control the main character with the aim to save his friends who were kidnapped by the tiger. The results of the player’s actions will culminate to the completion of each level. Fig. 1 shows the serious game flowchart.

The game starts with a brief introduction scene where Bunny is seen playing with monkey and bear in an area near the forest when a tiger comes and kidnaps the monkey and bear. Bunny is sad and meets a turtle. According to the turtle, Bunny needs to finish mathematical tasks if he wants to find the tiger’s place and release his friends. Level 1 starts thereafter. The player’s first challenge is to choose the way to the tiger’s place. The player is required to complete the task of recognising numbers. Upon completion of the task, guidance is given to the player to go to the place of the tiger. Then, he faces the next obstacle, the lion. The player needs to complete the subsequent tasks in order to get to the second level. The player will get marks and stars for every task that has been successfully completed.

At the second level, the first challenge starts when the player wants to cross the lake. To cross the lake, he/she needs to complete the task of recognising geometric shapes. Upon completion of the task, the player will cross the lake with a given boat. The player will face the next obstacle — a big rock that blocks the journey. The player needs to complete another set of geometric shapes task. After successfully completing the task, the big rock is moved to the side and the player continues the journey.

The player proceeds to the third level. At the third level, the player goes to an old castle. He/she needs to choose between two doors to enter the castle. He/she has to complete mathematical operations task. After successfully completing the task, direction is provided and the player goes to the correct door. The player meets his friends which are monkeys and bears imprisoned in a cage. To free his friends from the cage, the player needs to answer another set of mathematical operations. Upon completion of the task, the player gets the key to unlock the cage.

Therefore, based on gameplay ideas, designs, and storyline, a low fidelity prototype was produced. Prototype is used to test the ideas and game playability in order to assess its usability and needs for improvement. The low fidelity prototype was a storyboard. Fig. 2 shows samples of the low fidelity prototype.

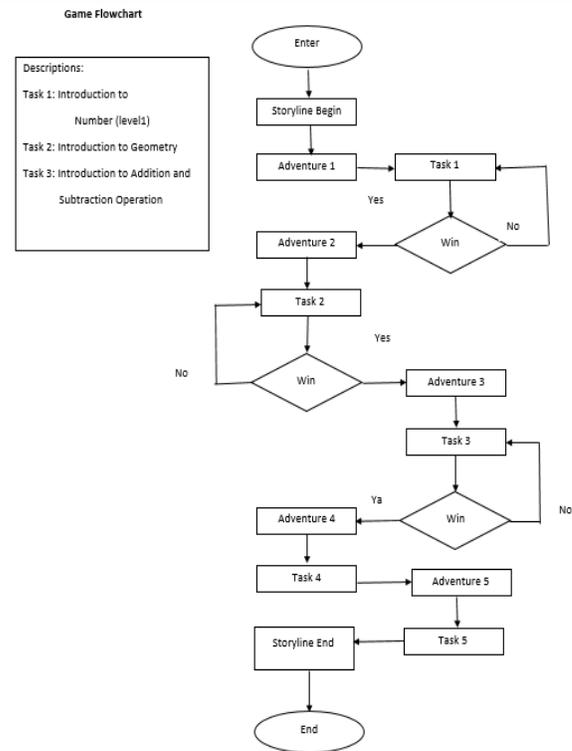


Fig. 1. Serious Game Flowchart.



Fig. 2. Sample Screens of Low Fidelity Prototype.

**B. Playtesting**

Playtesting method is conducted after the design and prototyping phase to elicit feedback from users regarding their play experience. In this study, playtesting was conducted informally and qualitatively. This method was used to enable users to play more comfortably. The goal of this play test was to get feedback on their play experience including problems affecting the experience. Normal children were selected for playtesting because they can give feedback on the game design and playability based on their gaming experience. Normal children could be selected when it is difficult to engage with real end-users [37]. The playtesting procedures for the study are as follows:

- Procedures were started with the researchers introducing themselves and explaining the purpose of the playtesting.
- Researchers introduced game prototypes to testers.
- Observation began with researchers asking the children to play the low fidelity prototype.
- This playtesting involved six children with digital gaming experience.
- Players played in pairs where they can talk while playing the game. They were asked to think aloud during playing.
- The play session was recorded.
- After the session, a focus group was conducted with the testers to get their feedback on the game.
- The feedback from the playtesting were used to refine the prototype.

Fig. 3 shows the playtesting process with children.

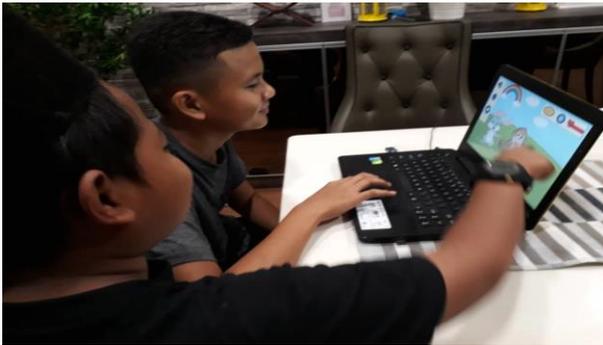


Fig. 3. Playtesting Process with Children.

Fig. 4 shows the summary of game design process.

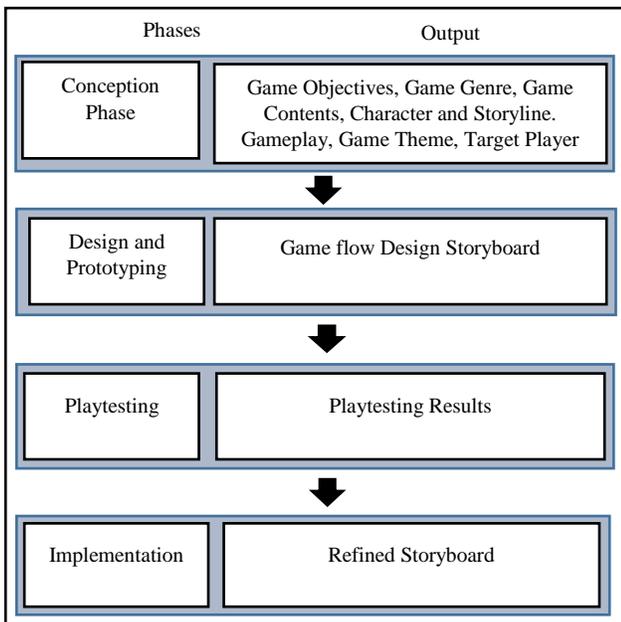


Fig. 4. Game Design Process.

#### IV. RESULTS AND DISCUSSION

Based on the observation, the players took about 25 minutes to complete all three levels of the game. According to the players, the game was fun as they can play while learning and thinking. Table III presents results from the playtesting.

TABLE III. PLAYTESTING RESULTS

Game Aspects	Playtesting Results
Gaming Experiences	The respondents took around 25 minutes to complete all three levels of the game. The respondents indicated that the game was fun to play. They can also play while learning and thinking.
Storyline	Easy to understand.
Challenges	The tasks need to include time to complete a task. At Level 3, a guard must be placed at the castle.
Rewards	Rewards should also be added in a more attractive form. When doing a task, praise should be given to players.
Game Interface	Graphics and colours are suitable for children. The game interface is easy to navigate. The 'next' button on the game interface needs to be changed to make it clearer to the player.

According to the players, the game storyline was easy to understand. In terms of the challenges, each task should include time limit to complete it. At level 3, a guard must be placed at the castle. Rewards should also be presented in a more attractive form. Upon finishing a task, compliments should be given to players. On the aspect of game interface, according to the testers, the graphics and colours were suitable for children. The game interface was also easy to navigate. However, the 'next' button on the game interface should be changed to make it clearer to players. The following figures show the aspects of the prototype that had been refined based on the players' feedback.

- Each task should have time limit to complete it (Challenge) as shown in Fig. 5.
- A guard should be placed at the door to improve the element of challenge in the game as shown in Fig. 6. If the task fails to be completed, Bunny will be captured by the guard.
- Positive feedback is given each time a task is completed as shown in Fig. 7.



Fig. 5. The Hourglass shows the Amount of Time it Takes to Complete the Task.



Fig. 6. A Guard at the Door.



Fig. 7. Positive Feedback - Praise for Encouragement.

- The existing 'star' badge can be used to assist future tasks as shown in Fig. 8. The number on the 'star' badge increases upon successful completion of the previous task.
- Rewards should be presented in a more attractive form. Fig. 9 shows the rewards after completing tasks.

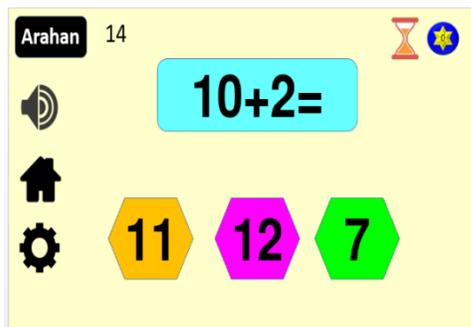


Fig. 8. The Existing 'Star' Badge can be used to Assist Future Tasks.



Fig. 9. Rewards after Completing Tasks.

## V. CONCLUSION

This paper presents a study on designing a low fidelity serious game prototype for low vision children using play-centric design approach in order to obtain feedback on game playability from children. In this study, ideas are prototyped and playtested. Children are involved at the early stage of the design process. They contributed game ideas to improve the playability of the game. Thus, play-centric design approach is important to ensure playability and good gaming experience. A serious game should be balanced between playability and serious purpose to ensure the children will have fun and motivated when playing game. Based on testers feedbacks, there are several game aspects that should be improved such as gaming experiences, storylines, challenges, rewards, and game interface. Future work will involve refining the game prototype which will be verified using heuristic evaluation and finally the development of final high fidelity game prototype.

## ACKNOWLEDGMENT

The authors would like to acknowledge UKM for the publication sponsorship under grant PP-FTSM-2020 and the Ministry of Higher Education for sponsorship under MyBrain15 programme.

## REFERENCES

- [1] N. Azli, M. Masrop, H. Ishak, G. Zainuddin, and S. R. Ramlan, "Digital Games Based Language Learning for Arabic Literacy Remedial," *Creat. Educ.*, vol. 10, pp. 3213–3222, 2019.
- [2] S. Cano, U. Cauca, J. M. Arteaga, U. Cauca, and V. B. Amador, "Model for Analysis of Serious Games for Literacy in Deaf Children from a User Experience Approach," in *Interacción '15 Proceedings of the XVI International Conference on Human Computer Interaction*, 2015.
- [3] R. Dörner and U. Spierling, "Serious Games Development as a Vehicle for Teaching Entertainment Technology and Interdisciplinary Teamwork: Perspectives and Pitfalls," in *SeriousGames '14 Proceedings of the 2014 ACM International Workshop on Serious Games*, 2014, pp. 3–8.
- [4] L. He, X. Hu, and D. Wei, "The Case Analysis of Serious Game in Community Vocational Education," *Int. Conf. Comput. Sci. Netw. Technol.*, 2011, pp. 1863–1866.
- [5] A. Ghazvini and Z. Shukur, "A Serious Game for Healthcare Industry: Information Security Awareness Training Program for Hospital Universiti Kebangsaan Malaysia," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 9, pp. 236–245, 2018.
- [6] A. Shapi'i, N. A. A. Rahman, M. S. Baharuddin, and M. R. Yaakub, "Interactive Games using h=Hand-e=Eye Coordination Method for Autistic Children Therapy," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4–2, pp. 1381–1386, 2018.
- [7] N. A. M. Zin, A. Jaafar, and S. Y. Wong, "Digital Game-Based Learning (DGBL) Model and Development Methodology for Teaching History," *WSEAS Trans. Comput.*, vol. 8, no. 2, pp. 322–333, 2009.
- [8] N. M. Diah, N. A. M. Zin, "Interactive Writing Tool for Jawi Literacy," vol. 7, no. October, pp. 48–57, 2013.
- [9] N. Zainuddin, A. Ab. Rahman, M. S. Sahrir, and H. A. Khafiz, "A Study on the Effectiveness of Global Zakat Game (GZG) as a Zakat Teaching and Learning Tool," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 9, no. 3, pp. 1488–1500, 2019.
- [10] Katie Cunningham, *Accessibility Handbook*. 2012.
- [11] S. Z. Mohid and N. A. M. Zin, "Courseware Accessibility for Hearing Impaired," *Information Technology (ITSim)*, in *2010 International Symposium*, 2010, pp. 26–30.
- [12] The World Wide Web Consortium, "World Wide Web Consortium," 2018. [Online]. Available: <https://www.w3.org/>. [Accessed: 15-Feb-2018].

- [13] World Health Organization, "World Health Organization," 2017. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs282/en/>. [Accessed: 01-Mar-2017].
- [14] N. I. Othman, N. A.M. Zin, and H. Mohamed, "Accessibility Requirements In Serious," in 2019 International Conference on Electrical Engineering and Informatics (ICEEI), 2019.
- [15] I. Dela Torre and I. Khaliq, "A Study on Accessibility in Games for The Visually Impaired," in 2019 IEEE Games, Entertain. Media Conf. GEM 2019, 2019, pp. 1–7.
- [16] S. R. Vaughn, C. S. Bos, and J. S. Schumm, Teaching Students Who Are Exceptional, Diverse, and at Risk. 2003.
- [17] S. R. Z. Abidin, S. F. M. Noor, and N. S. Ashaari, "Low-Fidelity Prototype Design for Serious Game for Slow-reading Students," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 3, pp. 270–276, 2019.
- [18] T. Westin, K. Bierre, D. Gramenos, and M. Hinn, "Advances in Game Accessibility from 2005 to 2010," 6th International Conference, UAHCI, 2010, pp. 400–409.
- [19] L. V. Neto, P. H. Fontoura Junior, R. A. Bordini, and J. L. Otsuka, "Design and implementation of an educational game considering issues for visual impaired people inclusion," Proc. - IEEE 19th Int. Conf. Adv. Learn. Technol. ICALT 2019, no. Dc, pp. 298–302, 2019.
- [20] M. T. Atkinson, C. H. C. Machin, and A. E. Lawrence, "Making the Mainstream Accessible: Redefining the Game," ACM SIGGRAPH Symp. Videogames, vol. 1, no. 212, pp. 21–28, 2006.
- [21] D. Archambault, R. Ossmann, T. Gaudy, K. Miesenberger, I. Inova, and S. Bernard, "Computer Games and Visually Impaired People," pp. 1–21, 2008.
- [22] G. R. White and G. Mcallister, "Toward Accessible 3D Virtual Environments for the Blind and Visually Impaired," in DIMEA '08: Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts 2008, pp.134-141.
- [23] D. Grammenos, A. Savidis, and C. Stephanidis, "Designing universally accessible games," Comput. Entertain., vol. 7, no. 1, p. 1, 2009.
- [24] M. De Pascale, S. Mulatto, and D. Prattichizzo, "Bringing Haptics to Second Life for Visually Impaired People," in International Conference on Human Haptic Sensing and Touch Enabled Computer Applications, pp 896-905, 2008.
- [25] P. Escudeiro, N. Escudeiro, P. Oliveira, and M. Barbosa, "Blind's Inclusion in Mobile Games," in 2017 27th EAEEIE Annual Conference (EAEEIE), 2017, pp. 1–5.
- [26] D. Mcpheron, "Video Gaming Accessibility," in The 20th International Conference on Computer Games, 2015, pp. 1–5.
- [27] R. Baguma and J. T. Lubega, "A Web Design Framework for Improved Accessibility for People with Disabilities ( WDFAD )," in W4A '08: Proceedings of the 2008 international cross-disciplinary conference on Web accessibility, 2008, pp. 134–140.
- [28] Z. Sun and J. Wen, "Exploration of Chinese Website Accessibility Evaluation Model," in International Conference on Computer Science and Software Engineering, 2008, pp. 1357–1360.
- [29] N. Aziz, A. A. Mutalib, and S. M. Sarif, "The design principles of assistive courseware for low vision (AC4LV) learners," ARPN J. Eng. Appl. Sci., vol. 10, no. 3, pp. 1447–1456, 2015.
- [30] C. Perrenoud and K. Phan, "Emergence of Web Technology: An Implementation of Web Accessibility Design in Organizations," in Proceedings of PICMET'12: Technology Management for Emerging Technologies, 2012, pp. 633–645.
- [31] Z. Sun and J. Wen, "Exploration of Chinese Website Accessibility Evaluation Model," in International Conference on Computer Science and Software Engineering, 2008, pp. 1357–1360.
- [32] M. S. Hassouna, N. Sahari, and A. Ismail, "University website accessibility for totally blind users," J. Inf. Commun. Technol., vol. 16, no. 1, pp. 63–80, 2017.
- [33] A. F. S. Barbosa and F. G. M. Silva, "Serious Games - Design and Development of OxyBlood," in Proceedings of the 8th International Conference on Advances in Computer Entertainment, 2011, pp. 1-8.
- [34] T. Fullerton, GAME DESIGN WORKSHOP A Playcentric Approach to Second Edition, 2nd ed. Elsevier Inc., 2008.
- [35] R. Raisamo, S. Patomäki, M. Hasu, and V. Pasto, "Design and evaluation of a tactile memory game for visually impaired children," Interact. Comput., vol. 19, no. 2, pp. 196–205, 2007.
- [36] E. Sudarmilah, A. Susanto, R. Ferdiana, and N. Ramdhani, "Developing a game for preschoolers: What character, emotion and reward will tend to hack preschoolers?," in Proc. 2015 Int. Conf. Data Softw. Eng. ICODSE 2015, 2016, pp. 89–92.
- [37] M. A. Lievesley and J. S. R. Yee, "Surrogate users: A Pragmatic Approach to Defining User Needs," in Conf. Hum. Factors Comput. Syst. - Proc, 2007, pp. 1789–1794.

# Principal Component Analysis on Morphological Variability of Critical Success Factors for Enterprise Resource Planning

Ayogebob Epizitone<sup>1</sup>, Oludayo. O. Olugbara<sup>2</sup>

ICT and Society Research Group, South Africa Luban Workshop  
Durban University of Technology  
Durban, South Africa

**Abstract**—The concept of critical success factors (CSFs) has been widely used as a measure to tackle the hurdles associated with numerous implementations of enterprise resource planning (ERP) systems. This study evaluates the morphological variability of CSFs using the analytical principal component analysis technique to identify principal components (PCs) that can be adopted for a successful ERP system implementation. The dataset of 205 CSFs from 127 different studies was evaluated for the morphological variability in those studies. According to the results, 66 PCs were identified and ranked accordingly. The first 49 PCs with eigenvalues greater than 1 accounted for 89.67 % of the variability recorded. The first 6 PCs respectively accounted for 13.67%, 19.37%, 24.67%, 29.41%, 33.52% and 36.94% cumulative variations. In general, the graphical illustration of the study results show the palpable division between the taxonomic groups for 3 PCs.

**Keywords**—Enterprise resources; morphological variability; principal component; resource planning; success factor

## I. INTRODUCTION

Critical success factors (CSFs) have been identified to be an essential precept for a successful ERP system implementation [1]. A critical success factor is a variable that has a significant impact on delivering a measurable improvement to project success [2]. The relevance of CSFs classification in ERP systems has been emphasised in various studies using different methods [1,3]. Certain authors, Hentschel, Leyh and Baumhauer [4] as well as Denolf, Trienekens, Wognum, van der Vorst and Omta [5], have postulated that despite the strong focus on avoiding failure in system implementation using CSFs approach, CSFs remain rarely researched. On the other hand, Saxena and McDonagh [6], have contended that CSFs remain the most-researched areas over the past years within the domain of enterprise systems.

However, despite this contention, there exists a consensus among researchers that CSF is a highly significant concept that can help address the inherent challenges associated with ERP system implementation [7]. Moreover, this has led to the identification of diverse CSFs in the literature. Consequently, the overarching objective of this study is to apply principal component analysis (PCA) to analyse morphological variability of CSFs for successful implementation of ERP systems. The realisation of the objective of this study affords

the following distinctive contributions. An enhanced understanding of the concept of CSFs that acknowledges their morphological variability for an efficacious implementation of ERP systems. The application of a robust analytical method to provide valued acumen to the CSFs phenomena of ERP system implementation. The remainder of this paper is succinctly summarized as follows. The next section provides the background discussion with respect to the related literature. This is followed by the description of the material and methods of the study. Next is the presentation of results and discussion and the paper is briefly concluded.

## II. BACKGROUND

The nature of CSFs has been reported in the literature to be inconsistent and repetitive, yielding the need for more analytical scientific methods [8-10]. Epizitone and Olugbara [11] highlighted this need by emphasising on the holistic nature of CSFs in different application settings. This view is further supported by the adoption of a mixed method research approach to tackle the complex phenomena of CSFs [12]. The determination of morphological variability of CSFs is a significant part of a successful implementation of an ERP project. The significance of CSFs classification has been emphasised in various related studies with a lot of attentions paid to the importance of CSFs and the success of ERP system implementations [9]. Consequently, the application of PCA to extract relevant information regarding CSFs from a large dimensional dataset is considered to enhance a deeper understanding of the intrinsic characteristics of CSFs [13].

PCA is a useful mathematical technique for emphasising variations and exposing hidden patterns in a dataset. It is predominantly applied for dimensionality reduction in application domains such as computer vision and pattern discovery in data mining [14]. It has been successfully used to specify principal components in varieties of datasets in many other areas of data science [14-18]. The technique has the potential to reveal essential characteristics while capturing the main structures of CSFs variability [19]. It is useful for discovering, reducing and identifying meaningful variables in a dataset. Hanci and Cebeci [15] have reported PCA to be a multivariate statistical technique with the capability of converting a lot of likely correlated factors into a set of smaller factors called principal components (PCs). The direction of the first PC is the same with the largest eigenvalue allied with its

eigenvector. While the direction of the second PC is determined by the eigenvector, which is related to the second largest eigenvalue.

The PCA technique involves a mathematical procedure that is based on the eigen analysis, which computes eigenvalues and corresponding eigenvectors of a square symmetric matrix with sums of squares and cross products [20]. The paramount objective of this study as earlier stated is to apply the PCA technique to analyse morphological variability of CSFs for successful implementation of ERP systems. The analysis technique would help to identify different PCs for promoting ERP system adoption [21]. It is assumed that the results of this study will provide the knowledge of CSFs that is appropriate for use in a successful implementation of an ERP system.

### III. MATERIAL AND METHODS

In this study, a total of 205 CSFs identified from 127 studies [22-27] was compiled and represented in a binary format displaying the feature of the identified variables for further analysis. The study dataset shown in Table S1 describes each factor as well as provides 205 qualitative CSFs and 127 quantitative instances that are suitable for PCA. The dataset was subjected to PCA to characterise the CSFs and identify the weight of each factor. The PCA technique was applied to a transformed dataset that was standardised into units of classes and attributes to determine the morphological variability. The number of PCs was determined using the minimal eigenvalue of unity called Kaiser criterion [28]. The dataset consisted of attributes 1 to 205 coded numerically as @ ATTRIBUTE F1-F205, while the related papers investigated for the extraction of factors were coded as @ATTRIBUTE class (P1-P127). All statistical procedures for the evaluation of morphological variability were obtained using the IBM SPSS statistics version 25 and WEKA 3.8.3. These statistical tools mutually afford an added validation advantage in identifying variations among the CSFs for ERP system implementation. The focus was on their morphological variation as it influences implementation success whilst providing the chance to analyse more than one factors in association.

### IV. RESULTS AND DISCUSSION

Table S2 illustrates the 66 PCs and 49 PCs identified by WEKA and SPSS with their corresponding eigenvalue, variance, and cumulative percentage. The WEKA statistical software identified 66 PCs with eigenvalue 28.015 to 0.518, variance 13.667 to 0.253 and cumulative variance 13.666 to 95.205 %. Each component of the 49 PCs identified by the IBM SPSS accounted for the following percentage range: eigenvalue 28.015 to 1.007, variance 13.667 to 0.491 and cumulative variance 13.666 to 89.199. The 49 PCs identified by the SPSS tool also featured in the 66 PCs identified by the WEKA software and they constitute the first 49 components in the WEKA result. These 49 components only consider the eigenvalues that are greater than unity as in other studies that applied PCA to different practical problems [28-30]. However, the 49 PCs had eigenvalue ranging from 28.015 for the first component to 1.007 for the last component. In addition, the 17 subsequent PCs had eigenvalues less than unity falling in the range of 0.962 to 0.517 for the PC 50 and PC 66, respectively.

Fig. 1 shows the ranking of the 66 PCs with the first six components respectively having the following scores: 86.33%, 80.00%, 75.36%, 70.59%, 66.45% and 63.06%.

The Vendor (F1) extraction value for PC represents the lowest value of 0.631, while the maximum extracted values are for F31-Professional training services, F32-Setting realistic deadlines, F37-User participation in defying new processes, F58-Deep understanding strategy, F60-Former major change experience, F81-Business change is first to be considered, F85-Level of implementation acceleration, F139- Opportunities for growth and F146-Data model is compatible with data requirements. It can be noted from the first component that these factors loadings were integrated to account for the high eigenvalue.

The first 6 PCs cumulative variations are 13.67%, 19.37%, 24.67%, 29.41%, 33.52% and 36.94% respectively as shown in Table I. These PCs can be seen to be distinctively illustrated by screen plot in Fig. 2, while Fig. 3 represents the component transformation matrix and Fig. 4 is the component matrix. These results illustratively provide the appropriate visualisation of the CSFs morphological variability that justifies the significance of these factors and their interpretations. The communalities shown in Table II, present each factor loading used for extraction that can be seen within a range of 0.631 minimum to 0.995 maximum for the component extracted. Table II further shows the result of the analysis presented for the communality showing the contribution of each factor.

The PC one (PC1) has an eigenvalue of 28.015, which explains 13.667 as the total variance with the same value for the cumulative variance. Taking into composition the contributions of individual weighted factor values for the PC occurring from the factors in Table II. The contribution of 10 factors can be seen in the table identifying different groups for the 6 PCs (Table I). Fig. 5 shows the first six components in rotated space. The contribution can further be seen in Table S3. The first group for PC 1 includes CSF such as Business change is first to be considered with eigenvector of 0.166 variation that reflects environment to the level of implementation acceleration and using ERP to fulfil cross-functional areas with 0.15 variation. This component presents the largest variability in the dataset as compared to the subsequent components [15, 29].

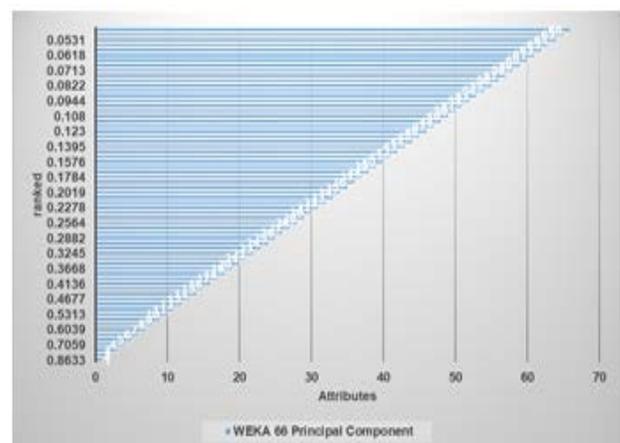


Fig. 1. Ranking of 66 PCs.

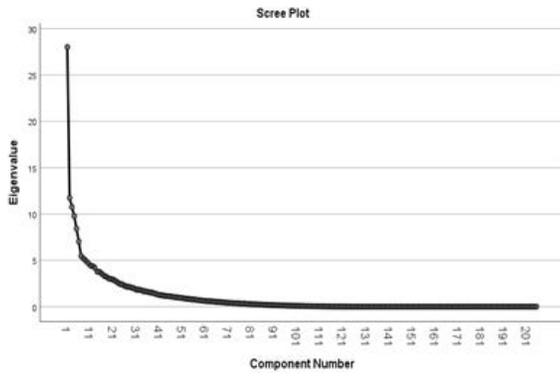


Fig. 2. Scree Plot of the PCs.

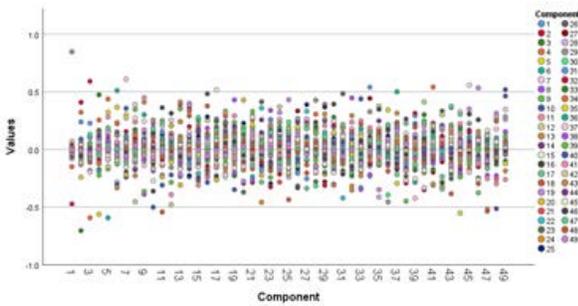


Fig. 3. Component Transformation Matrix for the Identified Components.

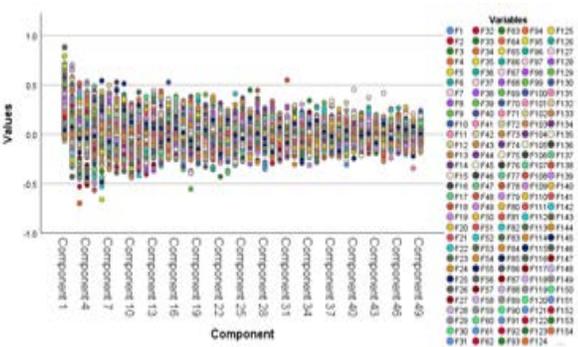


Fig. 4. Component Matrix of 49 PCs.

TABLE I. FIRST SIX COMPONENTS 10 FACTOR LOADINGS

PC1	0.166F81+0.166F37+0.166F146+0.166F32+0.166F60 +0.166F31+0.166F85+0.166F58+0.166F139+0.15 F117...
PC2	0.205F118+0.193F133+0.189F178+0.169F13+0.162F132 +0.148F173+0.145F66+0.143F14+0.14 F41+0.136F131...
PC3	-0.213F128-0.213F181-0.213F184-0.162F182+0.157F17 +0.157F171+0.148F14-0.137F114+0.134F18-0.134F121...
PC4	-0.169F149+0.165F177-0.161F88+0.159F125+0.159F20 +0.159F117+0.149F94+0.149F92+0.148F87+0.144F93...
PC5	-0.195F17-0.195F171-0.172F18-0.172F109+0.168F137 +0.167F134-0.161F15+0.151F74-0.15F172+0.149F21...
PC6	0.25 F95-0.205F77-0.204F175+0.2 F82+0.184F97+0.183F145 +0.171F91+0.164F107-0.164F170-0.146F28...

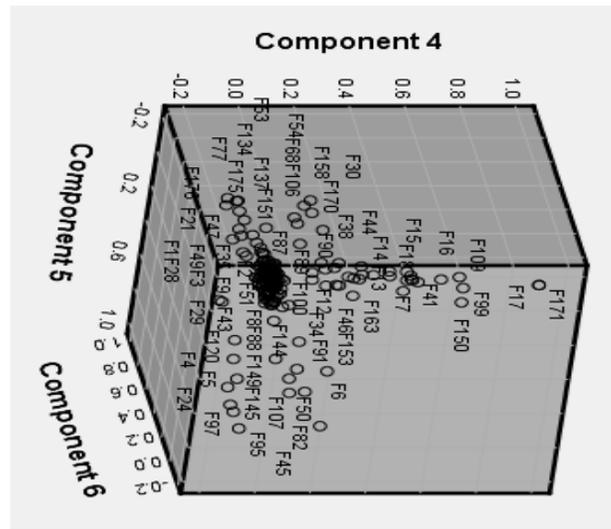
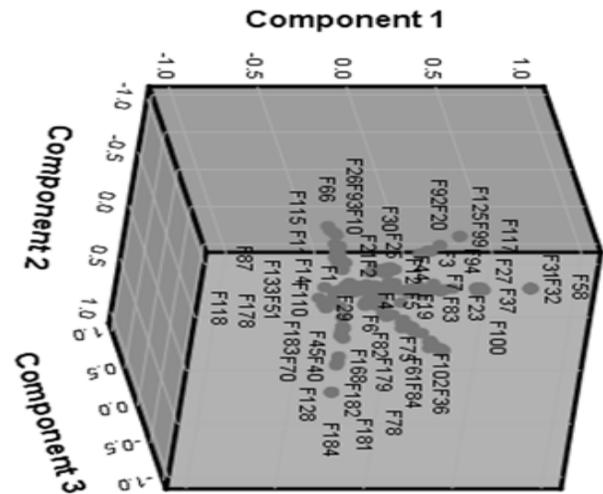


Fig. 5. Principal Component Plot in Rotated Space Graph for the First Six-Component (PC1 – PC6).

Several other groups were identified that consisted of different variational contributions from CSFs. In the second PC, the only factor that reached the eigenvector value of 0.205 is End users' attitudes (F118). The second component also had high eigenvector for the ERP easy to learn (Learnability/awareness (F133): 0.193, Availability of reliable data networks (F178): 0.189, Assign responsibility/Clear roles and responsibilities (F13), 0.169 and ERP usefulness (F132) 0.162. The highest eigenvector for PC three was achieved by Defined project miles' stones (F17): 0.157, Coordinate project activities (F171): 0.157 and subsequently Social influence (F177): 0.165. For PC four, software configuration (F137): 0.168, Vanilla ERP (F134): 0.167 for PC five and Education on new business processes (F95): 0.25, Architecture choices (F82): 0.2 for PC six.

TABLE II. COMMUNALITY

Factors	Initial(I)	Extraction									
F1	1.000	0.631	F53	1.000	0.874	F105	1.000	0.881	F157	1.000	0.847
F2	1.000	0.833	F54	1.000	0.869	F106	1.000	0.840	F158	1.000	0.860
F3	1.000	0.949	F55	1.000	0.969	F107	1.000	0.821	F159	1.000	0.856
F4	1.000	0.898	F56	1.000	0.870	F108	1.000	0.929	F160	1.000	0.963
F5	1.000	0.873	F57	1.000	0.838	F109	1.000	0.983	F161	1.000	0.887
F6	1.000	0.926	F58	1.000	0.995	F110	1.000	0.814	F162	1.000	0.903
F7	1.000	0.887	F59	1.000	0.791	F111	1.000	0.862	F163	1.000	0.874
F8	1.000	0.816	F60	1.000	0.995	F112	1.000	0.814	F164	1.000	0.886
F9	1.000	0.911	F61	1.000	0.881	F113	1.000	0.863	F165	1.000	0.855
F10	1.000	0.881	F62	1.000	0.928	F114	1.000	0.914	F166	1.000	0.958
F11	1.000	0.950	F63	1.000	0.778	F115	1.000	0.816	F167	1.000	0.776
F12	1.000	0.862	F64	1.000	0.960	F116	1.000	0.902	F168	1.000	0.817
F13	1.000	0.860	F65	1.000	0.895	F117	1.000	0.985	F169	1.000	0.886
F14	1.000	0.962	F66	1.000	0.869	F118	1.000	0.983	F170	1.000	0.851
F15	1.000	0.751	F67	1.000	0.827	F119	1.000	0.885	F171	1.000	0.974
F16	1.000	0.910	F68	1.000	0.829	F120	1.000	0.847	F172	1.000	0.871
F17	1.000	0.974	F69	1.000	0.982	F121	1.000	0.970	F173	1.000	0.985
F18	1.000	0.969	F70	1.000	0.987	F122	1.000	0.862	F174	1.000	0.817
F19	1.000	0.793	F71	1.000	0.841	F123	1.000	0.908	F175	1.000	0.875
F20	1.000	0.985	F72	1.000	0.884	F124	1.000	0.930	F176	1.000	0.872
F21	1.000	0.866	F73	1.000	0.930	F125	1.000	0.985	F177	1.000	0.950
F22	1.000	0.902	F74	1.000	0.854	F126	1.000	0.888	F178	1.000	0.947
F23	1.000	0.890	F75	1.000	0.928	F127	1.000	0.838	F179	1.000	0.958
F24	1.000	0.887	F76	1.000	0.823	F128	1.000	0.984	F180	1.000	0.963
F25	1.000	0.809	F77	1.000	0.890	F129	1.000	0.960	F181	1.000	0.984
F26	1.000	0.889	F78	1.000	0.861	F130	1.000	0.947	F182	1.000	0.951
F27	1.000	0.896	F79	1.000	0.880	F131	1.000	0.907	F183	1.000	0.862
F28	1.000	0.912	F80	1.000	0.855	F132	1.000	0.960	F184	1.000	0.984
F29	1.000	0.824	F81	1.000	0.995	F133	1.000	0.918	F185	1.000	0.877
F30	1.000	0.895	F82	1.000	0.923	F134	1.000	0.886	F186	1.000	0.719
F31	1.000	0.995	F83	1.000	0.902	F135	1.000	0.797	F187	1.000	0.830
F32	1.000	0.995	F84	1.000	0.942	F136	1.000	0.831	F188	1.000	0.981
F33	1.000	0.885	F85	1.000	0.995	F137	1.000	0.892	F189	1.000	0.934
F34	1.000	0.835	F86	1.000	0.960	F138	1.000	0.854	F190	1.000	0.869
F35	1.000	0.829	F87	1.000	0.876	F139	1.000	0.995	F191	1.000	0.890
F36	1.000	0.874	F88	1.000	0.863	F140	1.000	0.785	F192	1.000	0.843
F37	1.000	0.995	F89	1.000	0.809	F141	1.000	0.980	F193	1.000	0.932
F38	1.000	0.846	F90	1.000	0.896	F142	1.000	0.972	F194	1.000	0.885
F39	1.000	0.868	F91	1.000	0.829	F143	1.000	0.809	F195	1.000	0.796
F40	1.000	0.914	F92	1.000	0.979	F144	1.000	0.823	F196	1.000	0.947
F41	1.000	0.905	F93	1.000	0.921	F145	1.000	0.886	F197	1.000	0.925
F42	1.000	0.839	F94	1.000	0.979	F146	1.000	0.995	F198	1.000	0.970
F43	1.000	0.806	F95	1.000	0.926	F147	1.000	0.887	F199	1.000	0.790
F44	1.000	0.903	F96	1.000	0.844	F148	1.000	0.841	F200	1.000	0.878
F45	1.000	0.819	F97	1.000	0.909	F149	1.000	0.928	F201	1.000	0.832
F46	1.000	0.835	F98	1.000	0.856	F150	1.000	0.969	F202	1.000	0.972
F47	1.000	0.867	F99	1.000	0.984	F151	1.000	0.878	F203	1.000	0.972
F48	1.000	0.902	F100	1.000	0.941	F152	1.000	0.822	F204	1.000	0.904
F49	1.000	0.867	F101	1.000	0.772	F153	1.000	0.812	F205	1.000	0.883
F50	1.000	0.883	F102	1.000	0.977	F154	1.000	0.775			
F51	1.000	0.938	F103	1.000	0.817	F155	1.000	0.972			
F52	1.000	0.839	F104	1.000	0.880	F156	1.000	0.936			

These results report the presence of great morphological variability for some of the CSFs presenting specification of the CSFs diversification of ERP system implementation based on the taxonomy of the groups possibly identified by the selection of these CSFs. In this paper, we have explained the morphological variability and tried to model the CSFs to diverse components that are relevant to ERP system implementation. It can be seen from these results that taxonomic groups were conceivably attained by selecting these features. Azadeh, Afshari-Mofrad and Khalojini [30] and García, Rivera and Iniesta [31] applied PCA to their studies to characterised CSFs. The current study explicates on the diversity of CSFs variability based on different identity groups. Many studies undertaken on CSFs have selected certain CSFs to contextualise their results. However, results of the current study are attained from the inclusion of all the identified CSFs to provide a holistic nature of CSFs with different morphology. A similar approach to Ahmad, Haleem and Syed [22], study where all CSFs identified were retained for further analysis [3], characterised CSFs using a hybrid approached of PCA and impact factor analysis to identify, validate, rank and classify factors as critical, active, inert and reactive. Bhatti [32] applied PCA on a smaller dataset consisting of data from 53 inputs, using the reliability and validity scale to explain and characterise 11 CSFs with eigenvalue greater than 1 that only assimilation factor loads greater than 0.5. Madapusi and Ortiz [33] report findings on ERP, discussing two factors that account for 50.315 of the variability following a lesser Cronbach alpha statistic of 0.60 as compared to Bhatt [32] who used 0.75.

The projection of the 205 CSFs morphology in the two-dimensional graph of the component plot is shown in Fig. 3 and Fig. 4. The first, second and third PC coordinates of the PCA is realised using the morphological data accounted for 24.67% of the diversity observed (Fig. 3). While the subsequent three PCs four, five and six in Fig. 4 accounted for 12.30%. Overall, these displays denote an obvious division between taxonomic groups of CSFs relevant for the success of ERP system implementation.

## V. CONCLUSION

Employing different markers of the CSFs, diversification was estimated by exploring the morphological attributes that provide essential preliminary method for gauging different CSFs while concurrently elucidating their performance under successful implementation. The substantial knowledge presented by the results of this study is the CSFs variability applicable to various implementations of ERP systems. In this study, 205 different CSFs were analysed by using data obtained from 127 studies presenting different morphological findings of CSFs. The low variability of the first six principal components demonstrates that the diversity of the pool was significantly with the highest CSFs having eigenvectors not limited to values such as 0.25, 0.205, 0.193, 0.169 and 0.157.

The results of this study provide an important contribution to the ERP CSFs body of knowledge with a special attention paid to the morphological features of a disparate model from several morphological taxonomies of the identified CSFs using a robust analytical method. The study results can help

practitioners not to neglect any CSF, rather they should attach significant consideration to their roles in ensuring a successful implementations of ERP systems.

## ACKNOWLEDGMENT

The first author would like to thank her fellow researchers for their inspirations and the Durban University of Technology for the support provided during the study.

## REFERENCES

- [1] N. Ahmed, A. A. Shaikh, and M. Sarim, "Critical success factors plays a vital role in ERP implementation in developing countries: An exploratory study in Pakistan," *International Journal of Advanced Computer Science And Applications*, vol. 8, no. 10, pp. 21-29, 2017.
- [2] Z. Alias, E. Zawawi, K. Yusof, and N. Aris, "Determining critical success factors of project management practice: A conceptual framework," *Procedia-Social and Behavioral Sciences*, vol. 153, pp. 61-69, 2014.
- [3] B. M. Kalema, O. O. Olugbara, and R. M. Kekwaletswe, "Identifying critical success factors: the case of ERP systems in higher education," *The African Journal of Information Systems*, 6(3), p.1, 2014.
- [4] R. Hentschel, C. Leyh, and T. Baumhauer, "Critical success factors for the implementation and adoption of cloud services in SMEs." In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- [5] J. M. Denolf, J. H. Trienekens, P. N. Wognum, J. G. van der Vorst, and S. O. Omta, "Towards a framework of critical success factors for implementing supply chain information systems," *Computers in industry*, vol. 68, pp. 16-26, 2015.
- [6] D. Saxena, and J. McDonagh, "Yet another 'list' of critical success 'factors' for enterprise systems: Review of empirical evidence and suggested research directions." In *The UK Academy for Information Systems*, UK Academy of Information Systems Conference.
- [7] C. Leyh, "Critical success factors for ERP projects in small and medium-sized enterprises-The perspective of selected German SMEs." In *2014 Federated Conference on Computer Science and Information Systems* (pp. 1181-1190). IEEE, pp. 1181-1190.
- [8] R. C. Thompson, O. O. Olugbara, and A. Singh, "Deriving critical success factors for implementation of enterprise resource planning systems in higher education institution," *African Journal of Information Systems*, vol. 10, no. 1, 2018.
- [9] S. Dezdar, and A. Sulaiman, "Successful enterprise resource planning implementation: taxonomy of critical factors," *Industrial Management & Data Systems*, vol. 109, no. 8, pp. 1037-1052, 2009.
- [10] S. Finney, and M. Corbett, "ERP implementation: a compilation and analysis of critical success factors," *Business Process Management Journal*, vol. 13, no. 3, pp. 329-347, 2007.
- [11] A. Epizitone, and O. O. Olugbara, "Critical success factors for ERP system implementation to support financial functions," *Academy of Accounting and Financial Studies Journal*, vol. 23, no. 6, pp. 1-11, 2019.
- [12] A. Epizitone, and O. O. Olugbara, "Mixed method approach to determination critical success factors for successful financial ERP system implementation," *Academy of Accounting and Financial Studies Journal*, vol. 24, no. 2, pp. 1-10, 2020.
- [13] F. Tang, and H. Tao, "Binary principal component analysis." In *BMVC*, pp. 377-386.
- [14] H. Li, "Multivariate time series clustering based on common principal component analysis," *Neurocomputing*, vol. 349, pp. 239-247, 2019.
- [15] F. Hanci, and E. Cebeci, "Determination of morphological variability of different pisum genotypes using principal component analysis," *Legume Research-An International Journal*, vol. 42, no. 2, pp. 162-167, 2019.
- [16] H. Li, "Accurate and efficient classification based on common principal components analysis for multivariate time series," *Neurocomputing*, vol. 171, pp. 744-753, 2016.
- [17] Z.-X. Li, J.-S. Guo, X.-B. Hui, and F.-F. Song, "Dimension reduction method for multivariate time series based on common principal component," *Control and Decision*, vol. 28, no. 4, pp. 531-536, 2013.

[18] R. Karayel, and H. Bozoglu, "Determination of morphological variability of local pea genotypes," *Ekin Journal of Crop Breeding and Genetics*, vol. 1, no. 2, pp. 56-64, 2015.

[19] C. R. Bern, K. Walton-Day, and D. L. Naftz, "Improved enrichment factor calculations through principal component analysis: Examples from soils near breccia pipe uranium mines, Arizona, USA," *Environmental Pollution*, vol. 248, pp. 90-100, 2019.

[20] M. A. Espósito, L. A. Milanesi, E. Martin, V. Cravero, A. Lopez, and E. Cointry, "Principal component analysis based on morphological characters in pea (*Pisum sativum* L.)," *Int. J. Plant Breed*, vol. 1, no. 2, pp. 135-137, 2007.

[21] A. P. Chan, D. C. Ho, and C. Tam, "Design and build project success factors: multivariate analysis," *Journal of construction engineering and management*, vol. 127, no. 2, pp. 93-100, 2001.

[22] N. Ahmad, A. Haleem, and A. A. Syed, "Compilation of critical success factors in implementation of enterprise systems: a study on Indian organisations," *Global Journal of Flexible Systems Management*, vol. 13, no. 4, pp. 217-232, 2012.

[23] F. Fui-Hoon Nah, J. Lee-Shang Lau, and J. Kuang, "Critical factors for successful implementation of enterprise systems," *Business process management journal*, vol. 7, no. 3, pp. 285-296, 2001.

[24] F. F.-H. Nah, and S. Delgado, "Critical success factors for enterprise resource planning implementation and upgrade," *Journal of Computer Information Systems*, vol. 46, no. 5, pp. 99-113, 2006.

[25] C. Leyh, "Critical success factors for ERP projects in small and medium-sized enterprises—the perspective of selected ERP system vendors," *Multidimensional Views on Enterprise Information Systems*, pp. 7-22: Springer, 2016.

[26] L. Shaul, and D. Tauber, "CSFs along ERP life-cycle in SMEs: a field study," *Industrial Management & Data Systems*, vol. 112, no. 3, pp. 360-384, 2012.

[27] T. M. Somers, and K. G. Nelson, "A taxonomy of players and activities across the ERP project life cycle," *Information & Management*, vol. 41, no. 3, pp. 257-278, 2004.

[28] J.B. Adegbenro, M.T. Gumbo, and O.O. Olugbara. "Exploring technological knowledge of office data processing teachers: Using factor analytic methods." In *TPACK: Breakthroughs in Research and Practice*, pp. 155-184. IGI Global, 2019.

[29] A. Leilah, and S. Al-Khateeb, "Statistical analysis of wheat yield under drought conditions," *Journal of Arid Environments*, vol. 61, no. 3, pp. 483-496, 2005.

[30] A. Azadeh, M. Afshari-Mofrad, and M. Khalojini, "The role of organisational infrastructure in successful ERP implementation: an empirical study by hierarchical regression and PCA," *International Journal of Business Information Systems*, vol. 10, no. 1, pp. 40-67, 2012.

[31] J. L. García, D. G. Rivera, and A. A. Iniesta, "Critical success factors for Kaizen implementation in manufacturing industries in Mexico," *The International Journal of Advanced Manufacturing Technology*, vol. 68, no. 1-4, pp. 537-545, 2013.

[32] T. Bhatti, "Critical success factors for the implementation of enterprise resource planning (ERP): empirical validation." pp. 1-10, 2005.

[33] A. Madapusi, and D. A. C. Ortiz, "The influence of technical competence factors in ERP system implementations," *American Journal of Business*, vol. 10, no. 2, 2019.

SUPPLEMENTARY DOCUMENTS

PRINCIPAL COMPONENT ANALYSIS ON MORPHOLOGICAL VARIABILITY OF CRITICAL SUCCESS FACTORS FOR ENTERPRISE RESOURCE PLANNING

TABLE SI. QUALITATIVE AND QUANTITATIVE ERP CSFs FOR PCA

CSF Factors	Description	Occurrences
F1	Vendor	4
F2	Selection of appropriate vendor	5
F3	ERP vendor characteristics /reputation	4
F4	Partnership with vendor	26
F5	Vendor support	33
F6	Use of vendors' tools	17
F7	Keeping suppliers and customers informed	5
F8	Project Management	71
F9	Project leader	9
F10	Appointment & availability of competent project manager	4
F11	project manager /Full time	8
F12	Scope creep Management (Detail schedule)	19
F13	Assign responsibility/ Clear roles & responsibilities	7
F14	control project scope	6
F15	Evaluate any propose change	3
F16	Control and assess scope expansion requests/ assessment	3
F17	Define project miles stones	1
F18	Set realistic milestone and end dates	3
F19	Knowledge transfer management	9
F20	Management of conflicts	2
F21	Management of legacy systems	18
F22	Clear and defined project plan	4
F23	Planning required upgrades	2
F24	Management of expectations	21
F25	Management of risks	14
F26	Effective project management methodology	7
F27	Project tracking	1

F28	Total quality management approach	9
F29	Interdepartmental communication and cooperation	81
F30	Open and honest communication (Targeted and effective communication, among stakeholders, expectations communicated at all level and progress communication)	15
F31	Professional training services	1
F32	Setting realistic deadlines	1
F33	Project Management to implement project plan	3
F34	Change management	70
F35	Change management program	18
F36	Understanding the political structure (Political influence)	4
F37	User participation in defying new processes	1
F38	Understanding the organizational culture / (norms, values & beliefs)	36
F39	Developing a culture of continuous improvement	5
F40	National culture	3
F41	Recognizing the need for change	4
F42	Commitment to change -perseverance and determination	4
F43	Project team competence (formulation, composition and involvement)	67
F44	Team finest cross functional knowledge /small internal team	11
F45	Trust between various shareholders	5
F46	Good relations between project team and users (Partnership, trust, risk-sharing and incentives)	14
F47	Team morale and motivation	18
F48	Full time team members	8
F49	Balanced and cross functional project team	16
F50	Allocating valuable resources/Dedicating resources	25
F51	Culture of resistance/ enabling constraints	8
F52	Counselling to staff to minimize resistance to change	6
F53	Staff retention	6
F54	Empowered decision makers	18
F55	Work time schedule	4
F56	Performance tied to compensation	7
F57	Availability of qualified implementation team	4
F58	Deep understanding strategy	1
F59	Organizational characteristics	5
F60	Former major change experience	1
F61	Having in place advanced technology	4
F62	Former major IT change experience/Previous organization's experience with complex IS	3
F63	Interdepartmental coordination/ company wide	9
F64	Organization transformation and software migration	1
F65	Clear organizational strategy	7
F66	Organization encouragement of continuous learning	2
F67	Organization structure	16
F68	Implementation strategy	16
F69	Project definition and organization	1
F70	Implementation promotion	2
F71	Consultant's domain knowledge & experience	7
F72	Appointment of consultant/ external consultant involve in implementation (third party)	12
F73	Managing consultants	3
F74	Use of consultants (Consultant selection and relationship)	42
F75	Decision making process style/Strategic Decision making	7
F76	Focused performance measures plan	5
F77	Planning the cost of ERP implementation-Project cost planning and management	11
F78	Regard as a technological, business, and organizational project	8
F79	Alignment between business strategy and IT strategy	17
F80	Ensuring fair time to fulfil the implementation	6
F81	Business change is first to be considered	1
F82	Architecture choices	10

F83	Functional requirements are clearly defined before deciding on ERP adaptation/country related/carefully defined information and system requirement	9
F84	Continues focus on organizational resistance	4
F85	Level of implementation acceleration	1
F86	Implementation approach	1
F87	Implementer's domain knowledge & experience	3
F88	Project champion	49
F89	Education and training	58
F90	Education and training to technical staff /IT workforce re-skilling	21
F91	Education and training to end users	48
F92	Education on future business processes	3
F93	Adequate training to the implementation team	2
F94	Developing a clear education and training plan	3
F95	Education on new business processes	11
F96	Top management support	105
F97	Management and project steering committees	17
F98	management leadership	10
F99	Willingness to become involved	2
F100	Developing an understanding of the needs, capabilities & IT limitations	2
F101	Exhibiting strong commitment	4
F102	Resolving political conflicts (Political influence)	3
F103	Willingness to adopt modern technologies/Adaption Mechanism	5
F104	dedicated staff of vendor and institute for implementation	4
F105	Financial budget /funding Model	16
F106	Business vision	37
F107	Project mission /goals (Clear Goals and Objective)	42
F108	beliefs on ERP (management, users, teams and managers)/ Perception	5
F109	Justification for investment in ERP (investment plan)	2
F110	BPR	79
F111	User involvement	38
F112	User participation in the overall process approach	11
F113	User uses the system according to guidance	6
F114	Users' trust	2
F115	Key users' business knowledge	2
F116	Appointment & availability of competent key users	3
F117	Using ERP to fulfil cross functional areas	2
F118	End users' attitudes	3
F119	ERP System	4
F120	Level of Customization	43
F121	System flexibility to changing conditions	3
F122	System integration	16
F123	Systems reliability	5
F124	System interoperability	1
F125	System cross functionality	2
F126	System testing	13
F127	System quality	7
F128	Systems Changes and Upgrade	1
F129	System support	2
F130	ERP Version	1
F131	ERP ease of use/complexity	7
F132	ERP usefulness	5
F133	ERP easy to learn (Learnability/ awareness)	3
F134	Vanilla ERP	10
F135	Suitability of software and hardware considerations	9
F136	IT Infrastructure	32
F137	software configuration	15

F138	Environment	7
F139	Opportunities for growth	1
F140	Competition in industry/trend	7
F141	External/stakeholder pressure	3
F142	Competitors' adoption of ERP	2
F143	Uncertainty about environment	4
F144	Data Management	24
F145	Data analysis Plan	14
F146	Data model is compatible with data requirements	1
F147	Data quality control	7
F148	Developing a plan for migrating and cleaning up data	5
F149	Data conversion Plan	20
F150	Selection of data to be converted	2
F151	Data accuracy and integrity	14
F152	Package selection	39
F153	Careful and professional package selection process/modules	15
F154	Fit between ERP and business process, information needs and strategic goals/multi-site issues	15
F155	Planning the package selection process	2
F156	Software development	4
F157	Developing a plan for testing interfaces with integrated legacy systems	8
F158	Developing proper troubleshooting tools /Troubleshooting/crises management	18
F159	Robustness and Error Prevention (Working closely with vendors and consultants to resolve software problems and troubleshooting errors)	6
F160	Developing proper troubleshooting skills and techniques for the IT workers	4
F161	Planning and Establishing Software development, testing and troubleshooting architecture	17
F162	Appropriate modelling methods and Techniques (pre-implementation analysis) / Necessary preconditions	12
F163	Configuration of overall ERP architecture	7
F164	Monitoring management	28
F165	Monitoring and evaluation of performance metrics (fast effects)	31
F166	Monitoring progress against clear milestones	4
F167	User support organization and involvement	8
F168	User friendliness, Help, and Documentation/Document ERP success	5
F169	User acceptance feedback management/Analysis of user feedback (user satisfaction/satisfaction and system satisfaction)	21
F170	Enforce project timeliness /Timeframe	12
F171	Coordinate project activities	1
F172	Track milestones and targets	3
F173	Implementation experience/ with ERP implementation in similar scope	3
F174	Appropriate business and legacy systems including building a business case	23
F175	Post-implementation evaluation/audit	10
F176	Client consultation	8
F177	Social influence	1
F178	Availability of reliable data networks	3
F179	standardization and process measurement	5
F180	Follow the PDCA cycle	1
F181	System's Response Time to Users' Requests	1
F182	Interest/users groups	2
F183	Policies and Standards/ Government policies/Model	4
F184	Availability of applications (as result of Obsolescence of Hardware and Software)	1
F185	Discipline/Base point analysis; Process discipline; benchmarking	7
F186	Contingency plans (Co-ordinated analysis; contingency plans)	3
F187	Effective management techniques	2
F188	Controlled ROI on ERP implementation	2
F189	Operational Efficiency	1
F190	Internal readiness	2
F191	security of interface	4

F192	Integrated department and solve the problem of human resources management/Allocation of Best Internal Business Personnel	9
F193	Cost of update/upgrade/maintenance and integration	3
F194	Confidentiality	1
F195	Feasibility /evaluation of ERP project	1
F196	Strategic initiatives	2
F197	stimuli (environmental and customer needs)	3
F198	ERP treated as a program not a project	1
F199	Technical task and tools/Factors	4
F200	Reporting structure (project manager reporting to mgmt.	2
F201	Required Organizational Buy-In and Project Ownership	2
F202	Value Chain Connectivity	2
F203	IT provider and Integrator Push	2
F204	Globalization	1
F205	Procurement Management	2

TABLE SII. COMPARATIVE RESULTS PCS OF THE CSF WITH RESPECT TO WEKA AND SPSS

WEKA 66 Principal Components								SPSS 49 Principal Components			
Initial Eigenvalue								Component	Initial Eigenvalues		
Principal Component	eigenvalue	proportion	cumulative	Principal Component	eigenvalue	proportion	cumulative		Eigenvalue	% of Variance	Cumulative %
PC1	28.01467	0.13666	0.13666	PC50	0.96216	0.00469	0.89668	PC1	28.015	13.666	13.666
PC2	11.74692	0.05730	0.19396	PC51	0.95278	0.00465	0.90133	PC2	11.747	5.730	19.396
PC3	10.75159	0.05245	0.24641	PC52	0.88444	0.00431	0.90564	PC3	10.752	5.245	24.641
PC4	9.78366	0.04773	0.29413	PC53	0.86610	0.00422	0.90987	PC4	9.784	4.773	29.413
PC5	8.41872	0.04107	0.33520	PC54	0.83310	0.00406	0.91393	PC5	8.419	4.107	33.520
PC6	7.02066	0.03425	0.36944	PC55	0.79243	0.00387	0.91780	PC6	7.021	3.425	36.944
PC7	5.46663	0.02667	0.39611	PC56	0.77909	0.00380	0.92160	PC7	5.467	2.667	39.611
PC8	5.19389	0.02534	0.42145	PC57	0.74995	0.00366	0.92526	PC8	5.194	2.534	42.145
PC9	4.97628	0.02427	0.44572	PC58	0.70508	0.00344	0.92870	PC9	4.976	2.427	44.572
PC10	4.71714	0.02301	0.46873	PC59	0.67867	0.00331	0.93201	PC10	4.717	2.301	46.873
PC11	4.44978	0.02171	0.49044	PC60	0.65077	0.00317	0.93518	PC11	4.450	2.171	49.044
PC12	4.38653	0.02140	0.51184	PC61	0.62464	0.00305	0.93823	PC12	4.387	2.140	51.184
PC13	4.18899	0.02043	0.53227	PC62	0.60563	0.00295	0.94118	PC13	4.189	2.043	53.227
PC14	3.77951	0.01844	0.55071	PC63	0.59025	0.00288	0.94406	PC14	3.780	1.844	55.071
PC15	3.76175	0.01835	0.56906	PC64	0.57522	0.00281	0.94687	PC15	3.762	1.835	56.906
PC16	3.55745	0.01735	0.58641	PC65	0.54523	0.00266	0.94953	PC16	3.557	1.735	58.641
PC17	3.32842	0.01624	0.60265	PC66	0.51769	0.00253	0.95205	PC17	3.328	1.624	60.265
PC18	3.22817	0.01575	0.61839					PC18	3.228	1.575	61.839
PC19	3.03787	0.01482	0.63321					PC19	3.038	1.482	63.321
PC20	3.00068	0.01464	0.64785					PC20	3.001	1.464	64.785
PC21	2.90889	0.01419	0.66204					PC21	2.909	1.419	66.204
PC22	2.75999	0.01346	0.67550					PC22	2.760	1.346	67.550
PC23	2.59826	0.01267	0.68818					PC23	2.598	1.267	68.818
PC24	2.44134	0.01191	0.70009					PC24	2.441	1.191	70.009
PC25	2.40218	0.01172	0.71180					PC25	2.402	1.172	71.180
PC26	2.23643	0.01091	0.72271					PC26	2.236	1.091	72.271
PC27	2.15058	0.01049	0.73320					PC27	2.151	1.049	73.320
PC28	2.13079	0.01039	0.74360					PC28	2.131	1.039	74.360
PC29	2.05022	0.01000	0.75360					PC29	2.050	1.000	75.360
PC30	1.98248	0.00967	0.76327					PC30	1.982	0.967	76.327
PC31	1.84063	0.00898	0.77225					PC31	1.841	0.898	77.225
PC32	1.83486	0.00895	0.78120					PC32	1.835	0.895	78.120
PC33	1.79143	0.00874	0.78994					PC33	1.791	0.874	78.994

PC34	1.68255	0.00821	0.79815					PC34	1.683	0.821	79.815
PC35	1.65753	0.00809	0.80623					PC35	1.658	0.809	80.623
PC36	1.61361	0.00787	0.81410					PC36	1.614	0.787	81.410
PC37	1.53884	0.00751	0.82161					PC37	1.539	0.751	82.161
PC38	1.51210	0.00738	0.82899					PC38	1.512	0.738	82.899
PC39	1.40574	0.00686	0.83584					PC39	1.406	0.686	83.584
PC40	1.34829	0.00658	0.84242					PC40	1.348	0.658	84.242
PC41	1.26481	0.00617	0.84859					PC41	1.265	0.617	84.859
PC42	1.25485	0.00612	0.85471					PC42	1.255	0.612	85.471
PC43	1.17802	0.00575	0.86046					PC43	1.178	0.575	86.046
PC44	1.14210	0.00557	0.86603					PC44	1.142	0.557	86.603
PC45	1.13026	0.00551	0.87154					PC45	1.130	0.551	87.154
PC46	1.10956	0.00541	0.87695					PC46	1.110	0.541	87.695
PC47	1.05925	0.00517	0.88212					PC47	1.059	0.517	88.212
PC48	1.01597	0.00496	0.88708					PC48	1.016	0.496	88.708
PC49	1.00652	0.00491	0.89199					PC49	1.007	0.491	89.199

TABLE III. RANK ATTRIBUTES WITH 5 FACTOR LOADINGS

Ranked	Attributes	Contribution
0.8633	1	0.166F81+0.166F37+0.166F146+0.166F32+0.166F60...
0.806	2	0.205F118+0.193F133+0.189F178+0.169F13+0.162F132...
0.7536	3	-0.213F128-0.213F181-0.213F184-0.162F182+0.157F17...
0.7059	4	-0.169F149+0.165F177-0.161F88+0.159F125+0.159F20...
0.6648	5	-0.195F17-0.195F171-0.172F18-0.172F109+0.168F137...
0.6306	6	0.25 F95-0.205F77-0.204F175+0.2F82+0.184F97...
0.6039	7	-0.212F72-0.18F65-0.178F9-0.178F67-0.177F105...
0.5786	8	0.232F196+0.203F98+0.196F205+0.191F63-0.17F203...
0.5543	9	-0.23F55-0.181F112-0.175F123+0.168F203+0.168F202...
0.5313	10	0.202F11+0.191F73+0.187F55+0.184F26+0.176F83...
0.5096	11	-0.202F203-0.202F202+0.198F22+0.176F104-0.169F154...
0.4882	12	0.194F2-0.182F180+0.18 F1-0.178F126+0.176F104...
0.4677	13	0.216F127+0.199F197+0.188F112+0.185F113+0.184F86...
0.4493	14	-0.225F201-0.213F160-0.203F105-0.198F200-0.194F34...
0.4309	15	-0.272F130-0.184F63-0.182F135-0.179F16-0.173F22...
0.4136	16	0.203F173+0.193F180+0.186F69+0.166F191+0.166F162...
0.3974	17	-0.195F90+0.186F4-0.18F162-0.177F194+0.17 F80...
0.3816	18	0.308F69+0.219F157+0.218F70-0.212F64-0.212F86...
0.3668	19	0.225F179+0.223F180+0.197F186+0.189F122-0.172F111...
0.3521	20	0.232F189-0.189F64-0.189F86+0.173F39+0.172F28...
0.338	21	0.225F170-0.209F168+0.208F204+0.201F46-0.2F186...
0.3245	22	0.258F33-0.19F194-0.179F75+0.173F71+0.163F26...
0.3118	23	0.242F197+0.227F189+0.173F190+0.171F174+0.161F12...
0.2999	24	-0.263F52+0.198F89+0.193F167-0.192F204+0.188F129...
0.2882	25	-0.236F194-0.21F204-0.208F131-0.206F193+0.176F38...
0.2773	26	0.3 F33-0.2F8-0.178F110+0.167F175+0.166F187...
0.2668	27	-0.303F198-0.247F53-0.221F18+0.195F204-0.195F188...
0.2564	28	-0.241F151-0.221F45-0.211F96-0.192F8+0.184F161...
0.2464	29	0.235F191+0.22 F198-0.207F10-0.202F62+0.192F124...
0.2367	30	-0.193F131-0.178F136-0.172F45-0.151F124-0.151F10...
0.2278	31	-0.404F198-0.208F188-0.195F18+0.172F193+0.162F194...
0.2188	32	0.224F204-0.216F192-0.203F174-0.184F164+0.164F63...
0.2101	33	-0.226F88+0.217F25+0.193F205+0.177F111+0.174F136...
0.2019	34	-0.263F183+0.241F79+0.179F66-0.175F62-0.168F177...
0.1938	35	0.219F23-0.216F62-0.21F126+0.2 F11+0.18 F51...
0.1859	36	0.252F9+0.212F27+0.211F52-0.186F198+0.18 F172...

0.1784	37	0.274F195+0.239F192+0.229F8-0.175F35-0.161F108...
0.171	38	-0.208F110+0.188F72+0.181F43+0.172F198-0.168F89...
0.1642	39	0.268F192-0.259F151-0.221F135+0.203F29-0.18F63...
0.1576	40	0.391F195-0.21F205+0.2 F71+0.182F10+0.18 F105...
0.1514	41	-0.192F195+0.191F176-0.189F15+0.185F108+0.167F29...
0.1453	42	0.329F7+0.243F96+0.207F158+0.204F172+0.188F38...
0.1395	43	-0.182F29+0.175F183+0.173F28+0.167F135-0.156F80...
0.134	44	0.391F195-0.256F144-0.215F201-0.2F80-0.19F154...
0.1285	45	-0.242F108-0.228F27-0.227F172+0.225F191+0.167F59...
0.123	46	-0.251F96-0.194F161-0.189F168+0.181F164-0.178F91...
0.1179	47	0.265F200+0.221F163+0.214F27+0.201F195-0.196F66...
0.1129	48	-0.34F101-0.219F115+0.204F7-0.192F78-0.176F163...
0.108	49	-0.239F91-0.227F190-0.219F16-0.208F108-0.185F195...
0.1033	50	-0.289F186-0.219F103+0.202F66+0.201F19+0.183F45...
0.0987	51	-0.383F1+0.228F195-0.211F199+0.172F201-0.172F98...
0.0944	52	-0.203F52+0.202F110+0.177F205-0.176F34-0.171F41...
0.0901	53	-0.242F101+0.242F1+0.225F115-0.169F23-0.16F195...
0.0861	54	0.246F42-0.225F101+0.215F176+0.21 F57+0.172F103...
0.0822	55	0.232F45+0.217F143-0.206F168+0.201F1+0.163F7...
0.0784	56	-0.218F140+0.186F127+0.18 F169-0.18F168+0.171F119...
0.0747	57	0.202F68+0.202F154-0.178F1+0.166F2-0.156F195...
0.0713	58	-0.253F112+0.221F107-0.219F113-0.191F144+0.189F8...
0.068	59	0.202F152+0.2 F83+0.188F38-0.178F71-0.178F25...
0.0648	60	0.187F74+0.175F105-0.171F157-0.17F96-0.161F136...
0.0618	61	-0.241F107-0.231F89+0.225F8+0.202F122+0.159F172...
0.0588	62	-0.203F43+0.187F2+0.183F106+0.177F71+0.17 F82...
0.0559	63	0.21 F46+0.202F29-0.194F115-0.18F25+0.172F91...
0.0531	64	-0.308F167+0.207F194-0.183F144-0.178F23-0.177F135...
0.0505	65	0.271F76+0.253F43+0.199F144+0.197F151-0.166F35...
0.0479	66	-0.209F122-0.187F21-0.184F63-0.182F91-0.178F101...

# A Comprehensive Science Mapping Analysis of Textual Emotion Mining in Online Social Networks

Shivangi Chawla<sup>1</sup>, Monica Mehrotra<sup>2</sup>

Department of Computer Science  
Jamia Millia Islamia  
New Delhi, India

**Abstract**—Textual Emotion Mining (TEM) tackles the problem of analyzing the text in terms of the emotions, it expresses or evokes. It focuses on a series of approaches, methods, and tools to help understand human emotions. The understanding would play a pivotal role in developing relevant systems to meet human needs. This work has drawn significant interest from researchers worldwide. This article carries out a science mapping analysis of TEM literature indexed in the Web of Science (WoS), to provide quantitative and qualitative insight into the TEM research. To explain the evolution of mainstream contents, various bibliometric indicators and metrics are used which identify annual publication counts, authorship patterns, performance of countries/regions, and institutes. To further supplement this study, various types of network analysis are also performed like co-citation analysis, co-occurrence analysis, bibliographic coupling, and co-authorship pattern analysis. Additionally, a fairly comprehensive manual analysis of top-cited and most-used journal and proceeding papers is also conducted to understand the growth and evolution of this domain. As per the authors' knowledge, this manuscript provides the first thorough investigation of TEM's research status through a bibliometric examination of scientific publications. Expedient results are recorded that will allow TEM researchers to uncover the growth pattern, seek collaborations, enhance the selection of research topics, and gain a holistic view of the aggregate progress in the domain. The presented facts and analysis of TEM will help the researchers' fraternity to carry out the future study.

**Keywords**—*Emotion mining; emotion models; bibliometric analysis; science mapping analysis; co-citation analysis; network analysis*

## I. INTRODUCTION

With the frenzied profusion of social media services in recent years, the amount of data stored in electronic media is exponentially increasing. In this era of digitization, most people have an online life too apart from their daily routine activities where an insatiable desire is seen among them for sharing their opinions, thoughts, ideas, and feelings. This has created a lot of User Generated Content (UGC) to which researchers are paying active interest [1]. This user data is a topic of paramount importance among computer science researchers as it is a key to unlock the great potential of computing where machines can understand the highly emotional human being and respond and assist accordingly. A great deal of online social media communication is textual and hence, Text Analysis, Opinion Mining, Sentiment analysis, and Emotion Mining take their role. All the above areas are enough mature except Emotion Mining [2].

Emotions are affective-cognitive states that are fundamental to the human experience that show their existence in every single communication and mining of these emotional states is indeed an interesting topic with wide theoretical and practical applications. In the neurosciences, emotion mining can assist a deeper understanding of the mental health of a patient, detection of stress, anxiety and depression levels, mental health disorders which can help to adapt medications and prevent suicides in extreme cases. In the field of customer service, customer satisfaction is the utmost priority for a company selling its product and services. Emotion mining can't only help to gauge customer satisfaction, but also it can help employ improvement measures and study its impact on users as well. A successful attempt at mining user emotions can lead to the smart user interface of computers that can understand and respond according to human emotions.

According to psychological studies, every human action has one or more emotions(s) attached to it, for example, writing, reading, facial expressions, speech, music, body movements, and gestures, etc. Emotion Mining can be done from each of these media and is a separate field of study with its research challenges. Research efforts in this domain, date back to the early '90s, however, limited to data having audio and video aids captured using various sensors for the study [3]. With the advent of web 2.0, most of the user data is in the form of text and its great potential in affective computing has kicked forward the growth of Textual Emotion Mining (TEM). Emotion Mining from text is of the utmost challenge than from any other media because of the absence of any kind of aid which is implicit in audio and video data. This paper only takes into account the problem of TEM.

Research in the field of TEM started getting the attention of affective computing researchers with the work of Alm et al. [4] in 2005. They targeted narrative text of children's fairy tales for automatic emotional classification of fairy tale sentences into one of Ekman's six emotional categories. Their work is followed by a massive amount of literature targeting the classification of textual emotions from a variety of data domains including news headlines, news articles, web blogs, novels, chat messages, microblog texts, and suicide notes, etc. Due to this gargantuan growth, the existing literature in the concerned domain opens up many research avenues along with information overload making it difficult to obtain a clear picture of the process of TEM. Taking into account the substantial accomplishments of TEM research and the

supremacy of bibliometric and scientometric techniques [5]–[8], this paper aims to chart a landscape of the TEM domain visually and to scrupulously check the evolution of research in this sector. Specifically, the present study is an applied scientific method that intends to carry out a systematic bibliometric analysis of the TEM-related academic publications over the past 15 + years (Jan 2005 ~ Apr 2020). The results will enable concerned scholars to understand the knowledge structure as well as the recent trends in the TEM research and to decide or alter further study.

Currently, there is no scientific and comprehensive analysis of TEM research based on quantitative and statistical perspective. Therefore, this article employs different bibliometric methods [7], [9] to comprehensively map the landscape of TEM by responding to the following research questions:

RQ1: What is the period in which TEM work originated and how it grew over time? What is the count of papers in TEM since the inception of this domain?

RQ2: What are the most influential countries and institutes publishing TEM work?

RQ3: Who are the influential, productive, and top-cited authors of TEM domain who gave directions to researchers fraternity, during the period under study?

RQ4: What are the most cited and hot publications in TEM?

RQ5: Can Network analysis help visualize the evolution of a domain?

Besides these research questions, this manuscript also presents a manual analysis of top-cited papers of this domain to discuss the major approaches, emotion models, data sources used in their studies. It also reports the level at which emotion analysis was done listing the dataset and lexicon utilized. The motivation and major contributions of the proposed work are as follows:

The article attempts to satisfy the above-mentioned questions. The answers to these queries may prove to be of significant importance in deriving an understanding of the emergence and development of the field of TEM. It will provide a nice visualization of the evolution trends of the domain and grab an understanding of various aspects of TEM research. The readers of this manuscript will be able to trace the panorama of the TEM research field.

The contribution of this paper is four-overlay. Firstly, it attempts to make readers understand the concept and terminology of emotion. For many years, the term “emotion” was not properly understood or synonymically used with terms like sentiment, mood, etc. Second, it demonstrates the progress of the TEM domain in various demi-decades since 2005. Third, the use of various bibliometric indicators in the study, shed light on TEM literature from various angles by documenting most popular authors, publication venues, top institutions, etc leaving newcomers with an indication of venues that welcome the topic. Fourth, by reviewing the top-

cited papers according to WoS it tries to show the hallmarks of TEM research.

This paper is sorted out in six sections. Section 1 starts with the introduction of the field giving insights into the basic definitions and discusses the motivation behind this study. In Section 2, we discuss the preliminary background enlisting fundamental concepts that ground the TEM literature. Section 3 explains the methodology used to collect data and analyze it. Section 4 describes the empirical findings from the science mapping of the TEM field. A comprehensive manual analysis of the TEM field is provided in Section 5. Section 6 presents the conclusion, with a discussion of the limitations and highlights future work.

## II. RELATED WORK AND BACKGROUND

This section intends to present the preliminary concepts that describe the origin and significance of this domain. It also presents the related work describing the previous survey articles on TEM published so far.

### A. The Concept of Emotion

Before recognizing emotions in the text we should seek an answer to a very important question “*what we understand by emotion*”. This is considered to be the first step towards developing any effective emotion mining system. Kleinginna and Kleinginna [10] reviewed 92 different definitions of emotions and suggested this broad formal definition of emotion: “Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labelling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behaviour that is often, but not always, expressive, goal-directed, and adaptive.”

### B. Emotion-Related Terms

Socrates [11] wrote ‘*The beginning of wisdom is the definition of terms*’. Research in the area of emotion mining revolves around a lot of words that look synonymous with each other but carry a lot of difference in their meanings. These include subjectivity terms like opinions, sentiments, feelings, emotions, and affect which are commonly used interchangeably in most literature. However, a proper understanding of these terms and a clear differentiation among these terms is crucial. Scherer [12] also stated that inconsistencies in the definitions of emotion-related terms lead to failure in their proper apprehension and usage. It is also noted that blurred definition boundaries often lead to the introduction of unwanted noise into the scientific investigation and hence, lower the performance of automatic emotion detectors. Hence, after understanding the fungible aspect of the above mentioned emotion-related terms, this part of the section explores these terms with an attempt to distinguish between them. Table I presents a comparison of these terms (affect, opinion, sentiment, emotion, and mood) for better understanding and proper apprehension.

TABLE I. COMPARATIVE ANALYSIS OF EMOTION-RELATED TERMS

Subjectivity Term	Affect	Opinion	Sentiment	Emotion	Mood
Definition	The conscious subjective part of an emotion considered apart from bodily changes; also a set of observable manifestations of a subjectively experienced emotion	A view, judgment, or appraisal formed in the mind about a particular matter	An attitude, thought, or judgment prompted by feeling; a specific view or notion	The affective aspect of consciousness; a state of feeling; a conscious mental reaction (such as anger or fear) subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body	A conscious state of mind or predominant emotion
State of Mind	Non-Conscious	Conscious	Conscious	Pre-conscious	Conscious
Target-Oriented	Not Target Oriented	Always targeted towards an object	Always targeted towards an object	May or may not be target-oriented	May or may not be target-oriented
Persistence	Persist for Long-term duration	Longer episodes of human behavior	Longer episodes of human behavior	Short episodes of human behavior	Longest Duration
Synonyms	Feeling	Feeling, Sentiment	Emotion, Opinion	Sentiment	Feeling
Field of Study	Affective Computing	Opinion Mining	Sentiment Analysis	Emotion Mining	Mood Detection

### C. Related Work

This study intends to present an exploratory analysis, investigating the field of textual emotion recognition by pulling together most of the existing literature of this domain. Although there exist some surveys devoted to the topic of TEM, these lack the perspective of bibliometric inspection of literature.

One of the earliest surveys on TEM is the contribution of Kao et al. [13]. They presented a classification of emotion mining works into three categories namely keyword-based, learning-based, and hybrid methods. Another work by Binali and Potdar [14] discussed all the current emotion theories and techniques that lay the ground for textual emotion recognition. They also designed an evaluation framework for the meticulous evaluation of existing approaches. Jain and Kulkarni [15] presented a review of TEM literature enlisting some information retrieval methods utilized for research in text mining and then, suggested a system TextEmo. Tripathi et al. [16] reported the different approaches, datasets, and lexicons that have been used by TEM researchers to bring about a collective understanding of this domain. Another detailed survey article dedicated to the current domain is given by Yadollahi et al. [2] where they presented the current state of text sentiment analysis starting from opinion mining to emotion mining. Their study documented the sentiment analysis literature from a new and different perspective i.e., with an emphasis on emotion mining. The paper begins with the taxonomy of sentiment analysis through which they shed light on different tasks under opinion mining and emotion mining and then presented a thorough survey of publications discussing popular computational resources i.e. datasets and lexicons. A somewhat recent yet comprehensive review article on emotion mining is the contribution of Sailunaz et al. [17]. They focused on reviewing emotion mining research efforts based on text and speech and hence presented a very detailed survey covering various models, datasets, techniques, their features, and possible extensions for a better outcome.

Yet another addition to TEM surveys by Apte and Khetwat [18] covered various aspects of emotion detection like feature extraction/reduction techniques, approaches utilized for emotion analysis including the challenges encountered in the studied domain. The most recent and widest review article by Nourah and Mohamed [19] studied the implicit and explicit approaches to emotion detection: Keyword-based, Rule-based, Machine Learning based, Deep Learning based, as well as hybrid approaches. They also report best performing feature sets and point some open challenges.

### III. METHODOLOGY

The current study uses the method of science mapping to examine the TEM research domain. Science mapping-“a general process of domain analysis and visualization”-aims at detecting the intellectual structure of a scientific domain [6], [7] This method typically applies several bibliometric analysis techniques for visualizing significant patterns and trends within a large body of literature. This section is documented to cover the following phases in our study- study setup and data collection, data pre-processing, science mapping tools selection; and the procedure used for further analysis.

#### A. Data Collection

The current study uses the bibliographic data obtained from the Clarivate Analytics Web of Science(WoS) database [20], [21]. More specifically, the WoS Core selection is used in this analysis. This is because compared to other databases like Google Scholar, Scopus, and Research Gate, WoS is internationally recognized among the research community for accommodating the highest quality articles [22]. Bibliometric analysts find the WoS to be a valuable database for both finding and assessing various types of publications since it offers a collection of essential metadata including abstracts, references, citations count, authors, institutions, and countries.

To search for articles in the WoS database, keyword selection was done with the aim of search optimization to locate every related article. We use the "Topic" filter to get the

maximum number of appropriate TEM related documents. "Topic" in WoS tells that the record will be shown based on the presence of supplied search terms in Title, Abstract, Author Keywords, or Keywords Plus. The search data range was fixed to 2005-2020 and only articles published during these 15+ years were taken into account. We used several search strings to collect the published literature in WoS. Table II depicts the search queries used and the statistics of the data downloaded.

Although the present article focuses on research literature covering the domain of emotion mining, we can see that the topic is getting attention in the general public as well. For obtaining a clear and luminous picture of public interest, searches were made with different search strings (refer to TS in Table II) in Google Search Engine. Fig. 1 illustrates the year-wise increase in searches on Google.

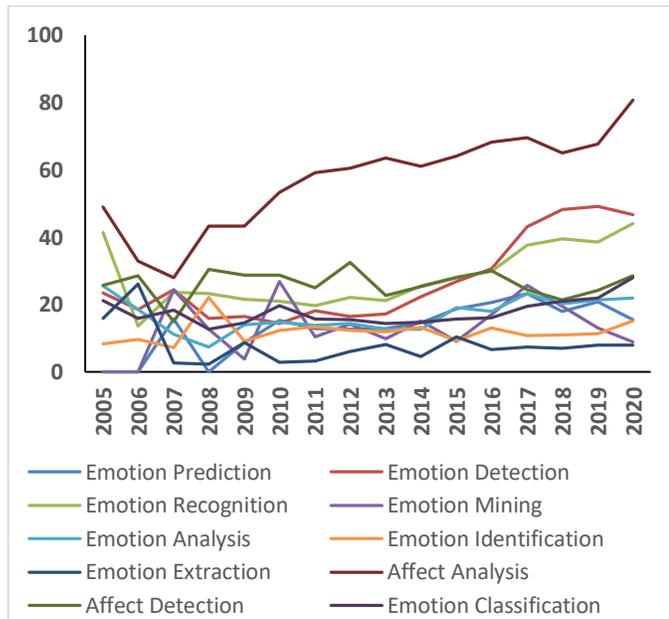


Fig. 1. Google Trends Related to different Search Terms.

**B. Data Preprocessing**

EM\_DS is punctiliously preprocessed to detect and fix viable typographical mistakes that may be present in the title of the publications, names of the authors, and date of publications. After this, the complete content of the paper including the title, abstract, and the author-supplied keywords is manually verified to check whether the search term is

effectively present or not. Papers giving negative results are excluded. Once the preprocessing phase has been completed, only 280 articles remain in the dataset, and these documents are used to mine the knowledge required to perform the bibliometric analysis. This new preprocessed dataset is named as TEM\_DS.state the units for each quantity that you use in an equation.

TEM-DS includes journal articles (~87%), proceeding papers (~5%), reviews, editorial materials, and book chapters (~8%). Each article in the WoS is assigned to one or more subject categories. As TEM is a subfield of ‘computer science’, these statistics are in line with the main venue of publications in the computer science subject category (~93%). Other major subjects include engineering, telecommunications, linguistics, management science, information science, library science, business, and economics.

**C. Selection of Tools and Metrics**

Analysis of TEM\_DS is done in the following manner. The exported “Plaintext” files are first converted to CSV format and then imported into mongoDB (version 4.0) database. Then, we merge the data into a single collection through mongo shell scripts, followed by the execution of various aggregation and find queries. The results are then fed to mongo shell scripts to obtain the desired outputs. Further analysis was done through Microsoft Excel.

As for the visualization tools, this study opts for the popular information visualization software VOSviewer. VOSviewer 1.6.15 [23], [24] is employed to handle WoS data, which is then used to perform network analysis based on the information related to the co-citation of references and journals, co-authorship, co-occurrence of keywords and the bibliographic coupling of cited references. The visualizations presented in figure ... Are created through this software. Microsoft Word and Excel were used for the manual investigation of content along with python scripts.

Additionally, we use Google Trends, a web facility powered by Google which provides the data related to the frequency of usage of a search term.

Apart from this we also employ various standard bibliometric indicators, as described below-

TP: Total Papers

TC: Total Citations

ACPP: Average Citations per paper (TC/TP)

TABLE II. DATA STATISTICS

Source	Category	Period	Search Query Used	#fields in each record	Date of download
Web of Science Core Collection	Articles, Reviews, Proceeding papers, Editorial Material and Book Chapters	01.01.2005 to 15.04.2020	TS=(("Emotion Mining") OR ("Emotion Detection") OR ("Emotion Recognition") OR ("Emotion Prediction") OR ("Emotion Identification") OR ("Emotion Analysis") OR ("Emotion Classification") OR ("Affect Analysis") OR ("Affect Detection"))	60	15.04.2020

#### D. Analytical Procedure

The analysis procedure involves both computational as well as manual investigation of publications. As depicted in Fig. 2, computational analysis of TEM\_DS is done using three different techniques present in WoS publication records, viz occurrence-based, content-based, and network-based analysis. All three types of analysis uncover a different aspect of the concerned domain leaving scholars with a wealth of information necessary to grasp the perspicuous evolution footprints of TEM authors. For example, relevant researchers may obtain valuable information about the authors, countries, and the affiliating institutes that are influential and productive.

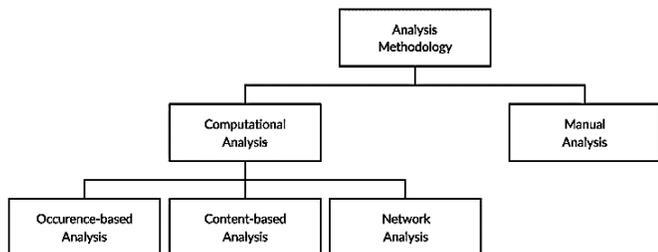


Fig. 2. Hierarchy Depicting Methods of Analysis.

#### IV. SCIENCE MAPPING ANALYSIS

In this section, the task of computational analysis of the TEM\_DS dataset is described, along with the various bibliometric indicators used. The subsections below present details of various types of analytical methods used along with the tables and figures illustrating the results.

##### A. Occurrence-Based Analysis

The Computational analysis using occurrence-based metadata aims at observing year-wise research publications trends as well as predominant institutions, countries, and authors.

1) *Annual publication distribution*: Firstly, we have measured the total number of published articles on TEM for each of the years from 2005 to 2020 (till 15.4.2020). Fig. 3 shows the total publication count in TEM on a year-wise plot. The increase in the number of publications can be observed since 2005. The lesser count of articles in 2020 is justifiable since it is the ongoing period and also some of the published works from 2020 are yet to be incorporated in WoS.

2) *Country-wise distribution*: Table III presents the 10 most productive countries/ regions in terms of the total publications (TP). China has emerged as a leading contributor to TEM research and is far ahead of other countries. The USA (42), Japan (33), and India (18) stand at the second, third, and fourth positions respectively.

3) *Institute-wise distribution*: Predominant institutions which contributed remarkably to the field and the study during the time frame of 2005-2020, are considered important for visualizing the development dynamics at the institution-level.

Table IV lists the most influential institutions in the decreasing order of the publications count (TP). Tokushima University, for example, contributes the largest number of research publications. Three of the top-performing institutions are located in China which again depicts the country's dominant rank in this research domain. National Institute of Informatics from Japan observes the highest citation count (TC) and the highest ACPP is recorded by National Research Council, Canada.

4) *Most influential authors*: The authors who are responsible for a significant count of published literature over the studied period are referred to as highly productive. Similarly, authors whose published articles got cited the most, are named to be the top-cited authors of the domain. We have also analyzed the TEM-DS dataset to recognize the most productive and cited authors (refer to Table V). We can observe that during the study period, Ren Fuji is the most active author of TEM and Saif M. Mohammad is the most cited author, in terms of total citations followed by Yanghui Rao and Qing Li.

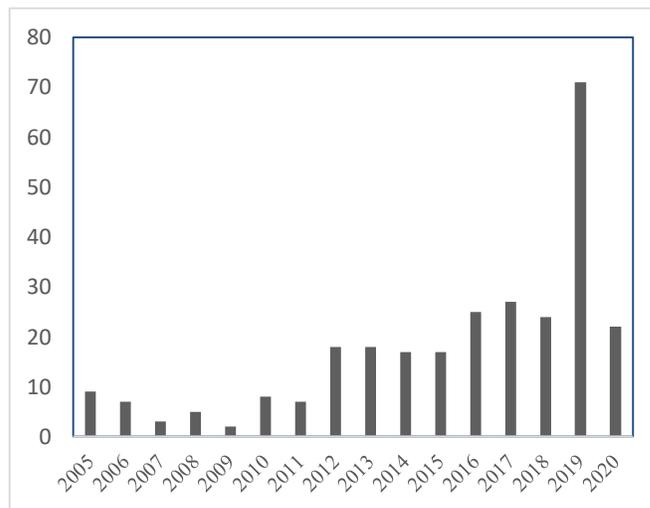


Fig. 3. Annual Scientific Production.

TABLE III. TOP CONTRIBUTING COUNTRIES

S No.	Country	TP	TC	ACPP
1	PEOPLES R CHINA	92	911	9.90
2	USA	42	2759	65.69
3	JAPAN	33	361	10.94
4	INDIA	18	54	3.00
5	CANADA	17	1306	76.82
6	SPAIN	17	280	16.47
7	ENGLAND	16	552	34.50
8	AUSTRALIA	11	709	64.45
9	TAIWAN	10	93	9.30
10	SINGAPORE	9	183	20.33





### V. MANUAL ANALYSIS

The previous section presents a fairly comprehensive computational analysis of TEM literature. To further strengthen and complement this research, this section presents a manual analysis of the top-cited and most used articles of TEM\_DS. To create a more spectacular investigation, the growth of TEM is observed by dividing the observation period into three demi-decades (refer Fig. 9). Period 2005 to 2009 is referred to as the first demi-decade representing the origin of research work followed by second and third demi-decade showing the periods of 2010-14 and 2014-19 respectively. The year 2020 is termed as the latest period of observation. Despite the gargantuan growth observed in this sector, most authors believe that the field remains in a nascent stage. Hence, demi-decade serve as a good observation period since a decade seems too large for such a field.

TEM literature during these years primarily used three different kinds of approaches/methods: lexicon-based, learning-based, or hybrid (lexicon and learning-based). Also, work on emotion analysis has been carried out on a variety of data sources (for example, blogs, microblogs, news headlines and news articles, literary texts, and discussion forums, etc.) creating a list of benchmark datasets and lexicons which can be utilized for further research and experimentation (for a detailed list of these computational resources, refer Naurah and Mohamed[19] recent survey. In line with the above aspects of TEM research, this section reports a thorough

investigation of the text of top-cited and most-used articles(journals and publications) from every demi-decade to identify which of the publications in TEM\_DS use which kind of approach, data source, dataset, and lexicon (refer to Tables VII to IX). Additionally, this section analyses the level at which emotions were mined in the respective publications i.e. word-level, topic-level, sentence-level, paragraph-level, and the document-level and reports the emotion model utilized (Categorical and Dimensional)[30]. We also analyze the latest period (2020) for understanding some of the recent trends of TEM research by analyzing a few latest publications of this on-going year. Table X presents an analysis of the latest papers. After observing the Tables VII to X, the following findings can be reported. First, machine learning seems to be the most popular choice of approach. Second, most works use a categorical model for emotion classification. Also there has been a substantial increase in the number of dataset and lexicons. The latest period (2020) has witnessed a shift from conventional machine learning to deep learning for developing automated emotion recognition systems.

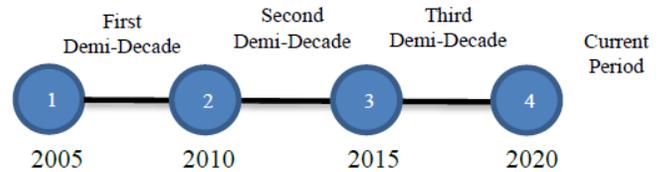


Fig. 9. Timeline view of the TEM Growth Period.

TABLE VII. MANUAL ANALYSIS OF TOP-CITED AND MOST-USED TEM PUBLICATIONS DURING FIRST DEMI-DECADE

S. No.	Reference	Year	Approach	Data Source	Model	Dataset	Lexicon	Level
1	[31]	2005	Keyword-Based	Questionnaire	Categorical	Self-Crawled	-	Sentence
2	[32]	2008	Hybrid	Web Forums, Blogs	Categorical	A	WordNet[33]	Sentence
3	[34]	2006	Learning-based	Heterogeneous	Categorical	MPQA Corpus[35]	-	Sentence
4	[36]	2005	Keyword-based	Text messages	Categorical	Text messages in chat system	WordNetAffect[37]	Word,Sentence
5	[38]	2005	Learning-based	Call Center Dialogs	Categorical	Self-Crawled	-	Sentence
6	[39]	2006	Lexicon-based	Song Lyrics	Categorical	Self-Crawled	-	Sentence
7	[40]	2005	Learning-based	Fairy Tales	Categorical	Alm's Fairy Tales Dataset[4]	-	Sentence
8	[40]	2005	Keyword-based	Text messages	Categorical	Text messages in chat system	WordNetAffect[37]	Word,Sentence
9	[41]	2006	Learning-based	Typed Sentences	Categorical	Typed Sentences	-	Sentence
10	[42]	2006	Keyword-Based	E-drama	Categorical	Self-Crawled	WordNet[33]	Sentence

TABLE VIII. MANUAL ANALYSIS OF TOP-CITED AND MOST-USED TEM PUBLICATIONS DURING SECOND DEMI-DECADE

S. No	Reference	Year	Approach	Data Source	Model	Dataset	Lexicon	Level
1	[43]	2012	Learning-Based	News articles	Categorical	Self-Crawled	-	Word and Topic
2	[44]	2014	Learning-Based	News articles	Categorical	Self-Crawled	-	Word and Topic
3	[45]	2010	Keyword-based	Blogs	Categorical	Self-Crawled	Self-built	Sentence, Document, Paragraph
4	[45]	2014	Learning-Based	News articles	Categorical	Self-Crawled SemEval-2007[46]	Self-built	Word and Topic
5	[47]	2014	Learning-Based	Short Text	Categorical	ISEAR Dataset[48], Stanford Twitter Dataset[49], Personality Dataset[50]	SenticNet[51], EmoSenticNet[52] ConceptNet[53], AffectiveSpace[54], EmoSenticSpace[52]	Sentence
6	[55]	2014	Rule-Based	Microblog	Categorical	Self-Crawled	-	Sentence
7	[56]	2013	Learning-Based	Suicide Notes	Categorical	2011 i2b2 NLP Challenge[57]	SentiWordNet[58], WordNet[33]	Paragraph
8	[59]	2011	Rule-Based	Dairy Like Blogs	Categorical	Self-Crawled	-	Word, Phrase, Sentence
9	[60]	2012	Learning-Based	Customer Experience posts	Categorical	Ren-CECPs[61]	-	Sentence
10	[62]	2014	Learning-Based	News articles	Categorical	Self-Crawled	-	Word and Topic

TABLE IX. MANUAL ANALYSIS OF TOP-CITED AND MOST-USED TEM PUBLICATIONS DURING THIRD DEMI-DECADE

S. No.	Reference	Year	Approach	Data Source	Model	Dataset	Lexicon	Level
1	[63]	2015	Learning-Based	Short Text	Categorical	ISEAR Dataset[48]	WordNetAffect[37]	Sentence
2	[64]	2015	Hybrid	Twitter Microblog	Categorical	Self-Crawled	NRC Emotion Lexicon[65]	Sentence
3	[66]	2015	Learning-Based	Twitter Microblog	Categorical	Self-Crawled	NRC Emotion Lexicon[65]	Sentence
4	[67]	2016	Hybrid	Heterogeneous Short Text, Twitter	Categorical	Self-Crawled	-	Sentence
5	[68]	2016	Learning-Based	Heterogeneous Short Text, Twitter	Categorical	Self-Crawled	-	Topic
6	[69]	2016	Learning-Based	News articles	Categorical	Self-Crawled	-	Sentence, Paragraph
7	[70]	2016	Learning-Based	Reviews	Categorical	Amazon Review Dataset[71]	NRC Emotion Lexicon[65], GALC Lexicon[72], WordNetAffect[37]	Sentence, Paragraph
8	[73]	2019	Learning-Based (Deep Learning)	TV Show "Charmed" Transcripts	Categorical	Self-Crawled	-	Sentence
9	[74]	2019	Learning-Based (Deep Learning)	News articles	Categorical	Self-Crawled	-	Paragraph
10	[75]	2018	Learning-Based (Deep Learning)	Heterogeneous Short Text, Twitter	Categorical	ISEAR Dataset[48], Fairy Tales[4], Election Tweets[66], Affective Text[76], General Tweets[77]	-	Sentence

TABLE X. MANUAL ANALYSIS OF TOP-CITED AND MOST-USED TEM PUBLICATIONS DURING LATEST PERIOD

S. No.	Reference	Approach	Data Source	Model	Dataset	Lexicon	Level
1	[78]	Learning-based (Deep Learning)	News articles	Categorical	Self-Crawled, Emo-Crowd-EN[78], SemEval-2018[79]	-	Sentence
2	[80]	Learning-based (Deep Learning)	Twitter	Categorical	SemEval-2018[79], TASS-2017[81]	-	Sentence
3	[82]	Learning-based	Online discussion Forum	Categorical	Self-Crawled	-	Sentence
4	[83]	Learning-based (Deep Learning)	Info-Graphic	Categorical	Self-Crawled	VADER[84]	Sentence
5	[85]	Learning-based	Twitter	Categorical	Self-Crawled	NRC Emotion Lexicon[65]	Word, Sentence

## VI. CONCLUSIONS

This manuscript presented an overview of TEM research by conducting an exhaustive science mapping analysis based on the dataset of 280 publications obtained from the WoS for the years 2005 to 2020. This analysis was conducted using various bibliometric indicators, taking into account the various dimension of analysis including countries/regions, institutions, authors, and keywords. In succeeding to answer the queries mentioned in section 1, this manuscript brings out the scathing investigation of TEM literature published to date. In this study, two different kinds of analyses are combined to shape the logical structure of the field of TEM. This would provide the community of social scientists and researchers, with the knowledge they need to illuminate the development path and start underpinning strategies to tackle the challenges prevailing till date.

Several findings can be extracted based on the presented work like:

- There is an increase in the annual publications in every demi-decade. The year 2019 recorded the highest peak.
- China is the most influential country recording the highest TP.
- Tokushima University contributed the highest number of papers in this domain.
- Machine Learning emerged out as the favorite approach for the TEM research fraternity with the recent focus on deep learning.
- Most top-cited publications utilized categorical approach for emotion modeling and a variety of datasets and lexicons have been explored to date.

Regardless of its contributions, this study experiences the following limitations. As the analysis depended on the dataset collected from WoS, therefore might be influenced by any inherent impediment of WoS's coverage of publications. Thus, the outcomes may not completely reflect the entire literature on TEM. Another impediment is the search phrases that we utilized, which may lead to a reduction of some relevant data. If an article about emotion detection didn't use the keywords we utilized for search, it doesn't show up in our data collection. Future research may, however, build upon the research work presented and try to address the shortcomings by utilizing data from varied databases, and a larger set of

indicators to assess influence, quality, and inter-connections in the literature.

## REFERENCES

- [1] Heimbach, B. Schiller, T. Strufe, and O. Hinz, "Content Virality on Online Social Networks," Proc. 26th ACM Conf. Hypertext Soc. Media - HT '15, pp. 39–47, 2015, doi: 10.1145/2700171.2791032.
- [2] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current State of Text Sentiment Analysis from Opinion to Emotion Mining," ACM Comput. Surv., vol. 50, no. 2, pp. 1–33, 2017, doi: 10.1145/3057270.
- [3] Y.-S. Seol and H.-W. Kim, "Personalized Emotion Recognition Considering Situational Information and Time Variance of Emotion," IEICE Trans. Inf. Syst., vol. E96D, no. 11, pp. 2409–2416, Nov. 2013, doi: 10.1587/transinf.E96.D.2409.
- [4] C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in Proceedings of the conference on human language technology and empirical methods in natural language processing, 2005, pp. 579–586.
- [5] H. Small and B. C. Griffith, "The structure of scientific literatures I: Identifying and graphing specialties," Sci. Stud. (St. Bonaventure), vol. 4, no. 1, pp. 17–40, 1974.
- [6] H. Small, "Update on science mapping: Creating large document spaces," Scientometrics, vol. 38, no. 2, pp. 275–293, 1997.
- [7] M. J. Cobo, A. G. López - Herrera, E. Herrera - Viedma, and F. Herrera, "Science mapping software tools: Review, analysis, and cooperative study among tools," J. Am. Soc. Inf. Sci. Technol., vol. 62, no. 7, pp. 1382 – 1402, 2011.
- [8] M. M. Kessler, "Bibliographic coupling between scientific papers," Am. Doc., vol. 14, no. 1, pp. 10–25, 1963.
- [9] M. J. Cobo, A. G. López - Herrera, E. Herrera - Viedma, and F. Herrera, "SciMAT: A new science mapping analysis software tool," J. Am. Soc. Inf. Sci. Technol., vol. 63, no. 8, pp. 1609 – 1630, 2012.
- [10] P. R. Kleinginna and A. M. Kleinginna, "A categorized list of emotion definitions, with suggestions for a consensual definition," Motiv. Emot., vol. 5, no. 4, pp. 345–379, 1981, doi: 10.1007/BF00992553.
- [11] S. E. Stumpf, Socrates to Sartre: A history of philosophy. McGraw-Hill New York, 1993.
- [12] K. R. Scherer and H. G. Wallbott, "Evidence for universality and cultural variation of differential emotion response patterning," J. Pers. Soc. Psychol., vol. 66, no. 2, p. 310, 1994.
- [13] E. C. C. Kao, C. C. Liu, T. H. Yang, C. T. Hsieh, and V. W. Soo, "Towards text-based emotion detection: A survey and possible improvements," Proc. - 2009 Int. Conf. Inf. Manag. Eng. ICIME 2009, no. March 2009, pp. 70–74, 2009, doi: 10.1109/ICIME.2009.113.
- [14] H. Binali and V. Potdar, "Emotion detection state of the art," Proc. CUBE Int. Inf. Technol. Conf. - CUBE '12, p. 501, 2012, doi: 10.1145/2381716.2381812.
- [15] M. C. Jain and V. Y. Kulkarni, "TexEmo: Conveying Emotion from Text-The Study," Int. J. Comput. Appl., vol. 86, no. 4, pp. 975–8887, 2014.
- [16] V. Tripathi, A. Joshi, and P. Bhattacharyya, "Emotion Analysis from Text: A Survey," Cfilt.litb.Ac.in, 2015, [Online]. Available:

- <http://www.cfil.itb.ac.in/resources/surveys/emotion-analysis-survey-2016-vaibhav.pdf>.
- [17] K. Sailunaz, M. Dhaliwal, J. Rokne, and R. Alhaji, "Emotion detection from text and speech: a survey," *Soc. Netw. Anal. Min.*, vol. 8, no. 1, pp. 1–26, 2018, doi: 10.1007/s13278-018-0505-2.
- [18] P. Apte and S. S. Khetwat, *Emerging Technologies in Data Mining and Information Security*, vol. 813. Springer Singapore, 2019.
- [19] N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," *Knowl. Inf. Syst.*, doi: 10.1007/s10115-020-01449-0.
- [20] M. A. García-Pérez, "Strange attractors in the Web of Science database," *J. Informetr.*, vol. 5, no. 1, pp. 214–218, 2011.
- [21] M. E. Falagas, E. I. Pitsouni, G. A. Malietzis, and G. Pappas, "Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses," *FASEB J.*, vol. 22, no. 2, pp. 338–342, 2008.
- [22] J. Hou, X. Yang, and C. Chen, "Emerging trends and new developments in information science: a document co-citation analysis (2009–2016)," *Scientometrics*, vol. 115, no. 2, pp. 869–892, 2018, doi: 10.1007/s11192-018-2695-9.
- [23] N. J. Van Eck and L. Waltman, "VOSviewer manual," Leiden: Univeriteit Leiden, vol. 1, no. 1, pp. 1–53, 2013.
- [24] N. Van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010.
- [25] I. V. Marshakova, "System of document connections based on references," *Nauchno-tehnicheskaya informatsiya seriya 2-informatsionnye protsessy I Sist.*, no. 6, pp. 3–8, 1973.
- [26] H. D. White and K. W. McCain, "Visualizing a discipline: An author co-citation analysis of information science, 1972–1995," *J. Am. Soc. Inf. Sci.*, vol. 49, no. 4, pp. 327–355, 1998.
- [27] K. W. McCain, "Mapping economics through the journal literature: An experiment in journal cocitation analysis," *J. Am. Soc. Inf. Sci.*, vol. 42, no. 4, pp. 290–296, 1991.
- [28] M. Callon, J. Law, and A. Rip, "How to study the force of science in Callon M, Law J and Rip A eds Mapping the dynamics of science and technology." Macmillan Press, Basingstoke and London, 1986.
- [29] M. Callon, J.-P. Courtial, W. A. Turner, and S. Bauin, "From translations to problematic networks: An introduction to co-word analysis," *Inf. (International Soc. Sci. Council)*, vol. 22, no. 2, pp. 191–235, 1983.
- [30] R. A. Calvo and S. K. Mac, "Emotions in text: dimensional and categorical models," *An Int. J. Comput. Intell.*, vol. 29, no. 3, pp. 527–543, 2012.
- [31] F. J. M. Laros and J. Steenkamp, "Emotions in consumer behavior: a hierarchical approach," *J. Bus. Res.*, vol. 58, no. 10, pp. 1437–1445, Oct. 2005, doi: 10.1016/j.jbusres.2003.09.013.
- [32] A. Abbasi, H. Chen, S. Thoms, and T. Fu, "Affect analysis of web forums and blogs using correlation ensembles," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 9, pp. 1168–1180, Sep. 2008, doi: 10.1109/TKDE.2008.51.
- [33] T. Pedersen, S. Patwardhan, and J. Michelizzi, "WordNet: Similarity-Measuring the Relatedness of Concepts," in *AAAI*, 2004, vol. 4, pp. 25–29.
- [34] T. Wilson, J. Wiebe, and R. Hwa, "Recognizing strong and weak opinion clauses," *Comput. Intell.*, vol. 22, no. 2, pp. 73–99, May 2006, doi: 10.1111/j.1467-8640.2006.00275.x.
- [35] L. Deng and J. Wiebe, "Mpqqa 3.0: An entity/event-level sentiment corpus," in *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2015, pp. 1323–1328.
- [36] C. L. Ma, H. Prendinger, and M. Ishizuka, "Emotion estimation and reasoning based on affective textual interaction," in *AFFECTIVE COMPUTING AND INTELLIGENT INTERACTION, PROCEEDINGS*, 2005, vol. 3784, pp. 622–628.
- [37] C. Strapparava, A. Valitutti, and O. Stock, "The Affective Weight of Lexicon," in *LREC*, 2006, pp. 423–426.
- [38] L. Vidrascu and L. Devillers, "Real-life emotion representation and detection in call centers data," in *Affective Computing And Intelligent Interaction, Proceedings*, 2005, vol. 3784, pp. 739–746.
- [39] Y. H. Cho and K. J. Lee, "Automatic affect recognition using natural language processing techniques and manually built affect lexicon," *IEICE Trans. Inf. Syst.*, vol. E89D, no. 12, pp. 2964–2971, Dec. 2006, doi: 10.1093/ietisy/e89-d.12.2964.
- [40] C. O. Alm and R. Sproat, "Emotional sequencing and development in fairy tales," in *Affective Computing And Intelligent Interaction, Proceedings*, 2005, vol. 3784, pp. 668–674.
- [41] Z. Teng, F. Ren, and S. Kuroiwa, "Recognition of emotion with SVMs," in *Computational Intelligence, PT 2, Proceedings*, 2006, vol. 4114, pp. 701–710.
- [42] L. Zhang, J. A. Barnden, R. J. Hendley, and A. M. Wallington, "Exploitation in affect detection in improvisational e-drama," in *Intelligent Virtual Agents, Proceedings*, 2006, vol. 4133, pp. 68–79.
- [43] S. Bao et al., "Mining Social Emotions from Affective Text," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 9, pp. 1658–1670, Sep. 2012, doi: 10.1109/TKDE.2011.188.
- [44] Y. Rao, Q. Li, X. Mao, and L. Wenyan, "Sentiment topic models for social emotion mining," *Inf. Sci. (Ny)*, vol. 266, pp. 90–100, May 2014, doi: 10.1016/j.ins.2013.12.059.
- [45] C. Quan and F. Ren, "A blog emotion corpus for emotional expression analysis in Chinese," *Comput. SPEECH Lang.*, vol. 24, no. 4, pp. 726–749, Oct. 2010, doi: 10.1016/j.csl.2010.02.002.
- [46] C. Strapparava and R. Mihalcea, "Semeval-2007 task 14: Affective text," in *Proceedings of the 4th International Workshop on Semantic Evaluations*, 2007, pp. 70–74.
- [47] S. Poria, A. Gelbukh, E. Cambria, A. Hussain, and G.-B. Huang, "EmoSenticSpace: A novel framework for affective common-sense reasoning," *KNOWLEDGE-BASED Syst.*, vol. 69, no. SI, pp. 108–123, Oct. 2014, doi: 10.1016/j.knosys.2014.06.011.
- [48] K. Deaux, "14. Gender and emotion: Notes from a grateful tourist," *Gen. Emot. Soc. Psychol. Perspect.*, p. 301, 2000.
- [49] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *CS224N Proj. report*, Stanford, vol. 1, no. 12, p. 2009, 2009.
- [50] J. A. M. Correa, M. K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," *IEEE Trans. Affect. Comput.*, 2018.
- [51] E. Cambria, D. Das, S. Bandyopadhyay, and A. F. Editors, *A Practical Guide to Sentiment Analysis*, vol. 5, 2017.
- [52] S. Poria, A. Gelbukh, E. Cambria, A. Hussain, and G. Bin Huang, "EmoSenticSpace: A novel framework for affective common-sense reasoning," *Knowledge-Based Syst.*, vol. 69, no. 1, pp. 108–123, 2014, doi: 10.1016/j.knosys.2014.06.011.
- [53] C. Havasi, R. Speer, and J. Alonso, "Conceptnet: A lexical resource for common sense knowledge," *Recent Adv. Nat. Lang. Process. V Sel. Pap. from RANLP*, vol. 309, p. 269, 2007.
- [54] K. Scherer, E. Dan, and A. Flykt, "What determines a feeling's position in affective space? A case for appraisal," *Cogn. Emot.*, vol. 20, no. 1, pp. 92–113, 2006.
- [55] W. Li and H. Xu, "Text-based emotion classification using emotion cause extraction," *Expert Syst. Appl.*, vol. 41, no. 4, 2, pp. 1742–1749, Mar. 2014, doi: 10.1016/j.eswa.2013.08.073.
- [56] B. Desmet and V. Hoste, "Emotion detection in suicide notes," *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6351–6358, Nov. 2013, doi: 10.1016/j.eswa.2013.05.050.
- [57] W. W. Chapman, P. M. Nadkarni, L. Hirschman, L. W. D'Avolio, G. K. Savova, and O. Uzuner, "Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions." *BMJ Group BMA House, Tavistock Square, London, WC1H 9JR*, 2011.
- [58] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining," in *LREC*, 2006, vol. 6, pp. 417–422.
- [59] A. Nevarouskaya, H. Prendinger, and M. Ishizuka, "Affect Analysis Model: novel rule-based approach to affect sensing from text," *Nat. Lang. Eng.*, vol. 17, no. 1, pp. 95–135, Jan. 2011, doi: 10.1017/S1351324910000239.

- [60] F. Ren and C. Quan, "Linguistic-based emotion analysis and recognition for measuring consumer satisfaction: an application of affective computing," *Inf. Technol. Manag.*, vol. 13, no. 4, SI, pp. 321–332, Dec. 2012, doi: 10.1007/s10799-012-0138-5.
- [61] P. A. Conference, Paclitc 30. 2016.
- [62] Y. Rao, Q. Li, L. Wenyin, Q. Wu, and X. Quan, "Affective topic model for social emotion detection," *NEURAL NETWORKS*, vol. 58, no. SI, pp. 29–37, Oct. 2014, doi: 10.1016/j.neunet.2014.05.007.
- [63] S. Poria, E. Cambria, A. Hussain, and G.-B. Huang, "Towards an intelligent framework for multimodal affective data analysis," *NEURAL NETWORKS*, vol. 63, pp. 104–116, Mar. 2015, doi: 10.1016/j.neunet.2014.10.005.
- [64] S. M. Mohammad and S. Kiritchenko, "Using Hashtags To Capture Fine Emotion Categories From Tweets," *Comput. Intell.*, vol. 31, no. 2, pp. 301–326, May 2015, doi: 10.1111/coin.12024.
- [65] S. M. Mohammad and S. Kiritchenko, "Using Hashtags to Capture Fine Emotion Categories from Tweets," *Comput. Intell.*, vol. 31, no. 2, pp. 301–326, 2015, doi: 10.1111/coin.12024.
- [66] S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, "Sentiment, emotion, purpose, and style in electoral tweets," *Inf. Process. Manag.*, vol. 51, no. 4, pp. 480–499, Jul. 2015, doi: 10.1016/j.ipm.2014.09.003.
- [67] I. Perikos and I. Hatzilygeroudis, "Recognizing emotions in text using ensemble of classifiers," *Eng. Appl. Artif. Intell.*, vol. 51, no. SI, pp. 191–201, May 2016, doi: 10.1016/j.engappai.2016.01.012.
- [68] Y. Rao, H. Xie, J. Li, F. Jin, F. L. Wang, and Q. Li, "Social emotion classification of short text via topic-level maximum entropy model," *Inf. Manag.*, vol. 53, no. 8, SI, pp. 978–986, Dec. 2016, doi: 10.1016/j.im.2016.04.005.
- [69] Y. Rao, "Contextual Sentiment Topic Model for Adaptive Social Emotion Classification," *IEEE Intell. Syst.*, vol. 31, no. 1, pp. 41–47, 2016.
- [70] A. Felbermayr and A. Nanopoulos, "The Role of Emotions for the Perceived Usefulness in Online Customer Reviews," *J. Interact. Mark.*, vol. 36, pp. 60–76, Nov. 2016, doi: 10.1016/j.intmar.2016.05.004.
- [71] M. Chen and Y. Sun, "Sentimental Analysis with Amazon Review Data." 2017.
- [72] Velentina Sintsova, "Advancing Fine-Grained Emotion Recognition in Short Text," vol. 7162, 2016, doi: <http://dx.doi.org/10.5075/epfl-thesis-7162>.
- [73] K. Shrivastava, S. Kumar, and D. K. Jain, "An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network," *Multimed. Tools Appl.*, vol. 78, no. 20, pp. 29607–29639, Oct. 2019, doi: 10.1007/s11042-019-07813-9.
- [74] M.-Y. Chen, C.-H. Liao, and R.-P. Hsieh, "Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach," *Comput. Human Behav.*, vol. 101, pp. 402–408, Dec. 2019, doi: 10.1016/j.chb.2019.03.021.
- [75] B. Kratzwald, S. Ilic, M. Kraus, S. Feuerriegel, and H. Prendinger, "Deep learning for affective computing: Text-based emotion recognition in decision support," *Decis. Support Syst.*, vol. 115, pp. 24–35, Nov. 2018, doi: 10.1016/j.dss.2018.09.002.
- [76] C. Strapparava and R. Mihalcea, "Learning to identify emotions in text," *Proc. 2008 ACM Symp. Appl. Comput. - SAC '08*, p. 1556, 2008, doi: 10.1145/1363686.1364052.
- [77] S. Mohammad and S. Kiritchenko, "Wikiart emotions: An annotated dataset of emotions evoked by art," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.
- [78] Z. Ahmad, R. Jindal, A. Ekbal, and P. Bhattacharyya, "Borrow from rich cousin: transfer learning for emotion detection using cross lingual embedding," *Expert Syst. Appl.*, vol. 139, Jan. 2020, doi: 10.1016/j.eswa.2019.112851.
- [79] H. Naderi, B. Haji Soleimani, S. Mohammad, S. Kiritchenko, and S. Matwin, "DeepMiner at SemEval-2018 Task 1: Emotion Intensity Recognition Using Deep Representation Learning," *Proc. 12th Int. Work. Semant. Eval.*, no. 2016, pp. 305–312, 2018, doi: 10.18653/v1/S18-1045.
- [80] M. Graff, S. Miranda-Jimenez, E. S. Tellez, and D. Moctezuma, "EvoMSA: A Multilingual Evolutionary Approach for Sentiment Analysis," *IEEE Comput. Intell. Mag.*, vol. 15, no. 1, pp. 76–88, Feb. 2020, doi: 10.1109/MCI.2019.2954668.
- [81] L.-F. Hurtado, F. Pla, and J.-A. González, "ELiRF-UPV at TASS 2017: Sentiment analysis in Twitter based on deep learning," *Proc. TASS*, pp. 29–34, 2017.
- [82] M. K. Kim and T. Ketenci, "The role of expressed emotions in online discussions," *J. Res. Technol. Educ.*, vol. 52, no. 1, pp. 95–112, Jan. 2020, doi: 10.1080/15391523.2019.1697861.
- [83] A. Kumar, K. Srinivasan, C. Wen-Huang, and A. Y. Zomaya, "Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data," *Inf. Process. Manag.*, vol. 57, no. 1, Jan. 2020, doi: 10.1016/j.ipm.2019.102141.
- [84] A. Dey, M. Jenamani, and J. J. Thakkar, "Senti-N-Gram: An n-gram lexicon for sentiment analysis," *Expert Syst. Appl.*, vol. 103, pp. 92–105, 2018.
- [85] A. K. Rathore and P. V. Ilavarasan, "Pre- and post-launch emotions in new product development: Insights from twitter analytics of three products," *Int. J. Inf. Manage.*, vol. 50, pp. 111–127, Feb. 2020, doi: 10.1016/j.ijinfomgt.2019.05.015.

# Prediction of Heart Diseases (PHDs) based on Multi-Classifiers

Amirah Al Shammari<sup>1</sup>

Department of Computer Science  
College of Computer, Al Jouf University  
Al Jouf, Saudi Arabia

Haneen al Hadeaf<sup>2</sup>, Hedia Zardi<sup>3</sup>

Department of Computer Science  
College of Computer, Qassim University  
Buraydah, Saudi Arabia

**Abstract**—At present, the number of articles on Heart Disease Detection (HDD) based on classification searched by Google Scholar search engine exceeds 17,000. The medical sector is one of the most important fields that benefit from ML. Heart diseases (HDs) are considered to be the leading cause of death worldwide, as it is difficult for doctors to predict them earlier. Therefore, the HDD is highly required. Today, the health sector contains huge data that has hidden information where this information can be considered as essential to make diagnostic decisions. In this paper, a new diagnostic model for the detection of HDs is on a multi-classifier applied to the heart disease dataset, which consists of 270 instances and 13 attributes. Our multi-classifier is composed of Artificial Neural Network (ANN), Naïve Bays (NB), J48, and REPTree classifiers, which select the most accurate of them. In addition, the most effective feature on prediction is determined by applying feature selection using the "GainRatioAttributeEval" technique and "Ranker" method based on the full tainting set. Experimental results show that the NB classifier is the best, and our model yields over 85% accuracy using the WEKA tool.

**Keywords**—Classification; diseases; heart-attack; multi-classifier; heart disease detection

## I. INTRODUCTION

Pumping blood to the whole body is the most critical task in human bodies. Therefore, the heart is the most important organ for humans. All of the Heart Diseases (HDs) concern to categorize this kind of cardiovascular diseases like coronary heart disease, Angina pectoris, coronary heart collapse, Cardiomyopathy, coronary cardiovascular illness, Arrhythmias, along with Myocarditis [1]. HDs are still the main cause of death worldwide. HDs are the leading cause of death in the United Kingdom, the United States, Canada, and Australia. According to the Centers for Disease Control (CDC), about 610,000 people die of HDs in the United States every year. It is estimated that 25% of deaths in the United States occur as a result of HDs. The possibility of detection at an early stage will help prevent the attacks. HD is defined as a variety of diseases, conditions, and disorders that affect the heart and the blood vessels. People die having experienced symptoms that were not taken into consideration. In addition, the quality of services in health care centers implies diagnosing disease correctly and delivers effective handlings for patients. On the other hand, poor diagnosis can lead to disastrous consequences, which are unacceptable.

Moreover, there is a need for medical practitioners to predict HDs before they occur in their patients. In these cases, many studies have been done on predicting heart disease by applying different data mining techniques to predict the accuracy of heart disease from related data sets. So, this study is presented.

Today, most health organizations around the world exploit information systems to manage their healthcare, including data about patients. Usually, these systems store significant amounts of data (numbers, text, charts, and images), especially about patients. This data is a golden (vital) resource to support clinical decision making because it is a rich source of hidden information that is largely unexploited. This is a great reason that motivates researchers to generate datasets from these unused data. Then, the data can be analyzed by employing a variety of data mining techniques in order to detect many diseases, especially HDs.

Knowledge Discovery in Databases (KDD) is the process of determining useful knowledge from a collection of data [2] using data mining and Machine Learning (ML) techniques. KDD includes data integration, data cleansing, data selection, and incorporating prior knowledge on datasets and interpreting accurate solutions from the observed results [3]. Classification is one of the most popular data mining tasks that assigns new, unknown items in a collection to target predefined categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks [4].

In this paper, a model for the prediction of HDs (PHDs) using multi-classifiers is proposed to detect the existence of HD (sick or normal), as it is shown in section IV. The research question of this paper is: "Which is the most efficient technique for the prediction of Heart diseases by considering the factors of accuracy and speed?"

The remainder of this paper is organized as follows. Section II reviews the background of the used ML techniques (J48, Naïve Bays, ANNs, and REPTree). Many HDD based on ML is discussed in Section III, followed by a full description of the proposed PHDS model in Section IV. Next, in Section V, the experiments and results are discussed. Finally, conclusions and suggested future work are given in Section VI, and VII, respectively.

## II. BACKGROUND

PHDs model differentiates between four ML techniques to choose the most accurate of them by applying them using the WEKA tool. In this section, a simple background is introduced WEKA and about each of the used techniques.

WEKA is used as a platform for machine learning as it has a collection of artificial intelligence algorithms in a domain for data mining tasks. It includes specific tools for data preprocessing and preparation, classification, clustering, regression, association rules, and visualization. WEKA gives us the ability to build models in order to detect hidden patterns in data and make a prediction without human interruption. Moreover, it contains a collection of predefined methods to evaluate the results of the techniques.

### A. Artificial Neural Network (ANN)

It consists of an interconnected group of artificial neurons. ANN processes information using a connectionist approach. In most cases, an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Modern neural networks are usually used to model complex relationships between inputs and outputs or to find patterns in data [5]. ANN has the main three key advantages that make it more appropriate for ML problems: It can learn and model non-linear, complicated relationships, generalize, and does not enforce any restrictions on the distribution of input variables.

### B. Naïve Base (NB)

A Naive Bayes (NB) classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. The classifier assumes that the presence (or absence) of a particular feature of a class (attribute) is unrelated to the presence (or absence) of any other feature. Even if these features depend on each other or upon the existence of the other features, the classifier considers all of these properties to contribute to that probability independently. The NB classifier performs reasonably well, even if the underlying assumption is not true [6]. NB is selected in this paper based on many advantages points as it requires less training data, fast to predict the class of the test data set, performs well in the multi-class prediction, and handle either continuous or discrete data.

### C. J48 Decision Tree

It is the implementation of algorithm Iterative Dichotomiser 3 (ID3). J48 is developed by the WEKA project team. J48 classifier is a straightforward C4.5 decision tree for classification, which creates a binary tree. It is the most useful decision tree approach for classification problems [7]. This technique constructs a tree to model the classification process.

In general, decision tree algorithms are [8] robust to errors, handle missing values by observing the data into other attributes, and generate understandable rules. Also, the learning and classification processes are uncomplicated and quick, with accuracy superior to the others.

### D. Reduced Error Pruning Tree (REPTree)

REPTree algorithm uses the regression tree logic. It generates multiple trees in different iterations. Afterward, it chooses the best of them as the representative tree [8]. In pruning the tree, it uses the mean square error on the predictions made by the tree. Fundamentally, REPTree is a fast decision tree learner, which builds a decision/regression tree using information gain as the splitting criterion and prunes it using reduced error. This ML can be effectively exploited in experimental comparisons to find the smallest optimally pruned tree with respect to the test set. Additionally, the property of this method is its linear computational complexity since each node is visited only once to evaluate the opportunity of pruning it.

## III. RELATED WORK

Before research on HDD focused their efforts on applying different data mining techniques, many data mining techniques for the diagnosis of heart disease were implemented following different approaches, such as Decision Tree, NB, ANNs, which give different levels of accuracies [9].

Patel et al. [9] reported the results of the comparison between three different algorithms based on the decision tree looking for the best performance in HDD using WEKA. The tested algorithms were the J48 algorithm, Logistic model tree algorithm, and Random Forest algorithm. They concluded, after experiments, that the winning algorithm for best performance was J48.

Sudhakar and Manimekalai [10] proposed a model that generates a class of data based on association rules from a training data set. Their model classifies the test data set into predefined class labels using the three different data mining classification techniques: ANNs, Decision Tree, and NB. Their overall objective was to study the different data mining techniques available for the prediction of HDs and to compare them in order to identify the best HDD prediction method.

A hybrid algorithm with the ANN (backpropagation) approach for HDs prediction was proposed by Dewan and Sharma [11]. This hybrid algorithm extracts unknown patterns and relations related to heart diseases from a past heart disease database record.

Masethe and Masethe [12] compared the performance of J48, Bayes Net, NB, Simple Cart, and REPTree in the prediction of possible HDs attacks to determine which model gives the highest percentage of correct prediction. They concluded that the most accurate classification techniques were J48 followed by REPTree and Simple Cart algorithms, while Bayes Net and NB algorithms had less accuracy rat.

Kim, Lee, and & Lee [13] proposed a predictive model for coronary heart disease (CHD) based on data collected for Disease Control and Prevention. This model incorporates fuzzy logic and CART-based rule induction to support the prediction of CHD. Rule induction was conducted to generate the rules. The fuzzy logic was used in the prediction model as an inference model. The experimental results showed that the accuracy and receiver operating characteristic curve values of the proposed systems were 69.51% and 0.594.

Krishnaiah, Narsimha, and Chandra [14] built a model to predict heart disease patients based on a fuzzy approach. In this model, the diagnosis was based on historical data. To remove the uncertainty of the data, the Fuzzy K-NN classifier employed, and the results showed the capability to remove the redundancy of the data and the better accuracy of the system.

Choi, Schuetz, Stewart, and Sun [15] explored whether the use of deep learning to model temporal relations between events in electronic health records would improve model performance in predicting the initial diagnosis of heart failure compared to conventional methods that ignore temporality. They used Recurrent neural network models with gated recurrent units to detect relations among time-stamped events. Based on the experiments, deep learning models appear to improve the performance of models for the detection of incident heart failure.

Comparison between the performance of ANN, NB, J48, and REPTree classifiers to the best of our knowledge has not been reported. In this study, we investigate and report such a comparison with our PHDs model, as explained in the following sections.

#### IV. PHDs MODEL

The PHDs model is a classification model based on supervised learning of classifiers and testing. Four classifiers in PHDs learned to select the most accurate of them based on our dataset. After training, the most accurate classifier is considered as the classifier of the PHDs model. After that, the considered classifier is used to detect any new unknown instance. The PHDs model is composed of two stages containing four steps.

##### A. The Learning Stage is Composed of Three Steps

In this stage, the model is built based on the learning of four ML classifiers using one dataset with known instants "classified instants" to choose the most accurate among them.

##### B. The Classification Stage is Composed of a Single Compound Step (Last Step)

In this stage, any new unknown instant can be prepared and then classified to normal or sick using the most accurate classifier.

These two stages are consisting of four steps divided into two stages, as is shown in Fig. 1 and as discussed follows:

##### 1) First stage: Learning and selecting the classifier:

This stage is composed of three steps:

a) *Preprocessing and Preparation*: As the quality of data is a key issue with data mining, so data preprocessing and preparation is a required step for serious, effective data mining. The results of data mining tasks as classification are affected by the quality of the dataset. So, in order to increase the accuracy of the mining, data preprocessing has to be performed. This stage includes handling missing values by using the average of attribute values from the same class. It should be noted that there are no noise data or inconsistencies. The other important tasks of preprocessing data, such as normalizing the attributes before conducting the ANN technique. In addition, the class attribute is transformed into binomial. The selected dataset prepared as follows:

- Convert data text format into comma .csv format for WEKA.
- Name columns (attribute values) by title.
- Choose the class attribute and convert it into binomial some class values were digits which not allowed for ANN classifiers.
- Open \*.csv file from WEKA.
- Select class attribute as a label with importing wizard (important for classifications).
- Handle missing values using the average of instants of each class.
- Normalize the input data (1,0, -1) to process by ANN classifier.

b) *Classifiers' Learning and Testing*: In this step, each of the four classifiers learned and tested using dataset. The accuracy results of each classifier in recorded to be used later in the next step.

2) *Classifier selection*: In this step, the most accurate classifier is selected as the classifier of the model. The selected classifier is used to detect HDs for any new instance.

3) *Second stage: Classification*: This stage is composed of a single compound step. In this stage, any new unknown instant will be prepared and then classified to either normal or sick using the most accurate classifier selected from the *first stage*.

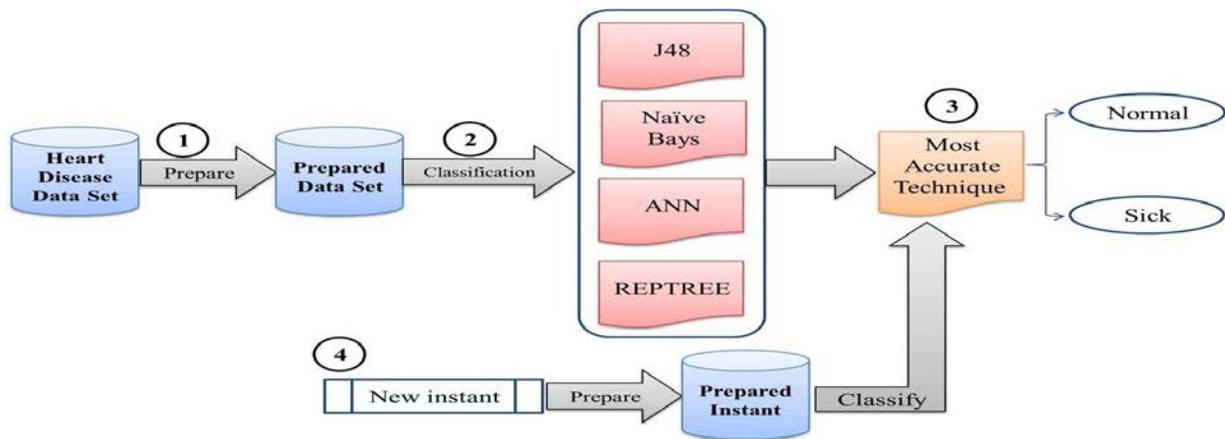


Fig. 1. PHDs Model.

## V. EXPERIMENTS AND DISCUSSION

Experiments were conducted with the WEKA tool using the dataset of heart disease from the UCI Machine Learning Repository [16]. WEKA is chosen to conduct experiments for many reasons, which are: it has many visualization tools and algorithms for data analysis and predictive modeling, which give easy access and use. WEKA supports graphical user interfaces either for process data or to illustrate the results, which makes it easy to understand. Also, it has an extensive collection of data preprocessing and modeling techniques.

On the other hand, the heart disease dataset preferred because this dataset collected and designed for a classification task, has a suitable number of records, and Only 3 of its attributes have missing value. The data set specifications listed in Table I.

This dataset is taken from 270 individuals; the diagnosis of some of them was definite for having heart disease, has 14 attributes. The last attribute is a special one for "class", either the presence or the absence of HDD. These attributes are represented in Table II. The purpose of analyzing the dataset was to detect for the presence of HDs (normal is none and sick is present).

The four algorithms applied to the data set using the percentage of data for learning and the remainder for testing in order to assess the performance of the classification technique for predicting a class. Many experiments with variants of parameters for training and testing data and evaluation options (percentage split, cross-validation) conducted. The best results are shown.

### A. Percentage Split

The experiment's results of NB and ANN algorithms with test mode: 80.0% training and 20.0% testing are illustrated in Tables III and IV, respectively. While results of the REPTree algorithm with test mode: 85.0% training and 15.0% testing are shown in Table V. Finally, results of the J48 algorithm with test mode: 90.0% training and 10.0% testing are represented in Tables VI. All the accuracy and time results are illustrated in Fig. 2 and Fig. 3, respectively.

Using the percentage split, ANNs then J48 algorithm are achieved higher accuracy while NB then REPTree algorithm is less. So overall confusion matrices and Fig. 2 and 3, it is concluded that ANN is the most accurate, and the J48 is the fastest algorithm.

TABLE I. DATASET SPECIFICATIONS

<b>Data Set Characteristics</b>	Multivariate
<b>Attribute Characteristics</b>	Categorical, Integer, Real
<b>Associated Tasks</b>	Classification
<b>Number of Instances</b>	270
<b>Number of Attributes</b>	13
<b>Missing Values?</b>	Yes

TABLE II. DATASET ATTRIBUTES

Symbol	Attribute
A	Age
B	Sex
C	Chest pain type (4 values)
D	Resting blood pressure
E	Serum cholesterols in mg/dl
F	Fasting blood sugar > 120 mg/dl
G	Resting electrocardiographic results (values 0,1,2)
H	Maximum heart rate achieved
I	Exercise induced angina
J	Old peak = ST depression
K	The slope of the peak exercise ST segment
L	Number of major vessels (0-3) colored by fluoroscopy
M	Thallium:3=normal; 6=fixed defect; 7=reversible defect
Class	Class, Absence (Normal) or presence (Sick)

TABLE III. CONFUSION MATRIX OF NB ALGORITHM

	Predicted Sick	Predicted Normal
Actual Sick	TP (22)	FP (4)
Actual Normal	FN (4)	TP (24)

TABLE IV. CONFUSION MATRIX OF ANN ALGORITHM

	Predicted Sick	Predicted Normal
Actual Sick	TP (24)	FP (2)
Actual Normal	FN (3)	TP (19)

TABLE V. CONFUSION MATRIX OF REPTree ALGORITHM

	Predicted Sick	Predicted Normal
Actual Sick	TP (14)	FP (4)
Actual Normal	FN (4)	TP (24)

TABLE VI. CONFUSION MATRIX OF J48 ALGORITHM

	Predicted Sick	Predicted Normal
Actual Sick	TP (8)	FP (4)
Actual Normal	FN (0)	TN (15)

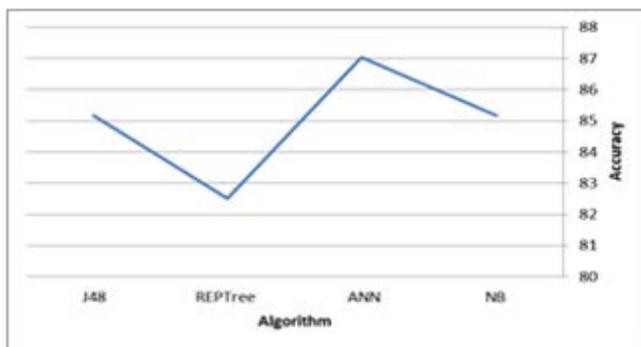


Fig. 2. Accuracy of Models using Percentage Split.

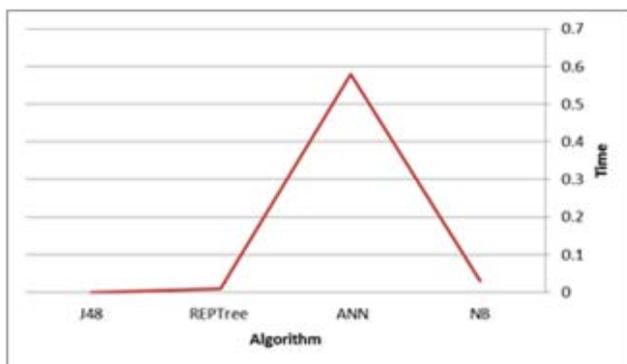


Fig. 3. Conducting Time using Percentage Split (by Seconds) for the Best Result of each Algorithm.

### B. Cross-Validation

The experiment's result of the NB algorithm with 15 folds shown in Tables VII. On the other hand, the ANN and J48 algorithms with test mode: 21 folds are illustrated in Tables VIII and X, respectively. Finally, the results of the REPTree algorithm with 24 folds presented in Tables IX.

All the accuracy and time results based on cross-validation are illustrated in Fig. 4 and Fig. 5, respectively. Using cross-validation, NB, then the REPTree algorithm becomes the most accurate while ANN then J48 algorithm is less.

TABLE VII. CONFUSION MATRIX OF NB ALGORITHM

	Predicted Sick	Predicted Normal
Actual Sick	TP (97)	FP (23)
Actual Normal	FN (20)	TN (130)

TABLE VIII. CONFUSION MATRIX OF ANN ALGORITHM

	Predicted Sick	Predicted Normal
Actual Sick	TP (92)	FP (28)
Actual Normal	FN (24)	TP (126)

TABLE IX. CONFUSION MATRIX OF REPTree ALGORITHM

	Predicted Sick	Predicted Normal
Actual Sick	TP (89)	FP (31)
Actual Normal	FN (18)	TP (132)

TABLE X. CONFUSION MATRIX OF J48 ALGORITHM

	Predicted Sick	Predicted Normal
Actual Sick	TP (93)	FP (27)
Actual Normal	FN (29)	TP (121)

TABLE XI. ACCURACY OF ALGORITHMS

Algorithm	Percentage-Split	Cross-Validation
NB	85.18	84.07
ANN	89.58	81.65
REPTree	80.60	81.55
J48	85.182	79.26

So, we conclude that the NB algorithm is the best one due to its high-performance using a cross-validation method. In addition, it is the fastest.

The comparison between the accuracies of all algorithms based on percentage-split and cross-validation is shown in Table XI and Fig. 6. So, we conclude that the NB algorithm is the best one due to its high-performance using a cross-validation method. Also, it is the fastest.

Predefined instances of the heart disease dataset, 18 new unlabeled instances are supposed with random values to apply the second stage of the PHDs model. After inputting them to PHDs, they are labeled, as shown in Fig. 7.

Like Comparing our results with other research works, our results are inconsistent with them, that is in [9] and [12], J48 is the most accurate, while in [13] and [15] is the ANN. In our opinion, this inconsistency due to many factors as a dataset and an applied tool. Moreover, none of the research work applied the algorithms together on the same dataset and using the same tool. We believe that these results can give a chance to develop a new effective diagnostic tool to help doctors and HDS patients.

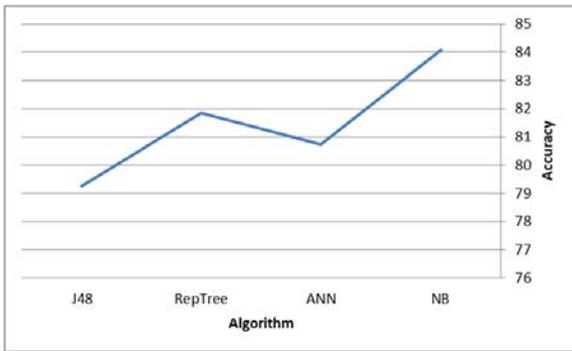


Fig. 4. Accuracy of Models using Cross-Validation.

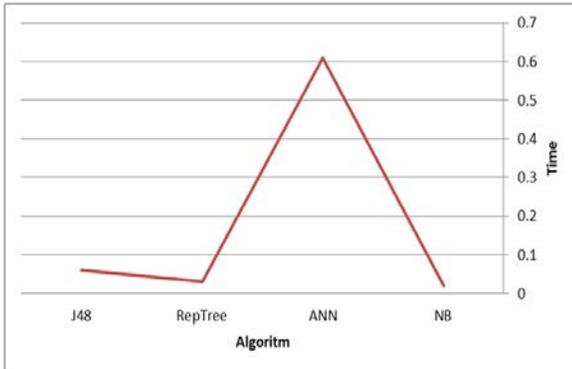


Fig. 5. Conducting Time using Cross-Validation (by Seconds) for the Best Result of each Algorithm.

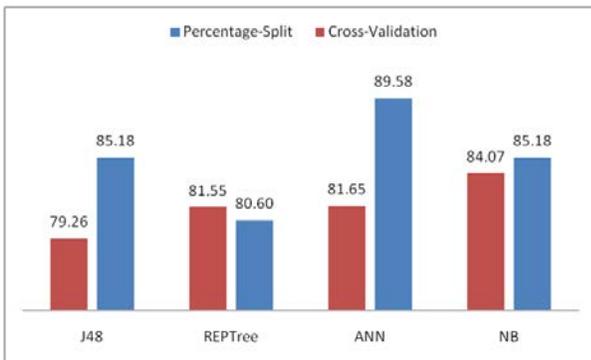


Fig. 6. Comparison of the Accuracy of the Algorithm.

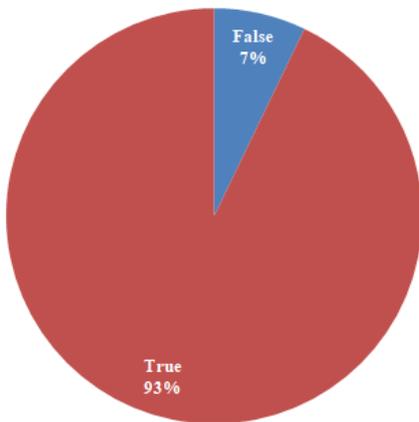


Fig. 7. Predict of New Instances using NB.

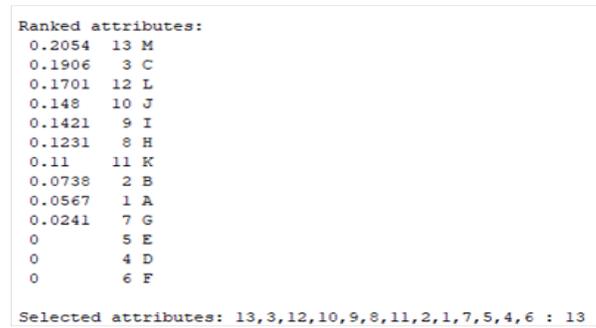


Fig. 8. Results of Feature Selection Technique.

To determine the most effective feature on the prediction, feature selection is applied using the "GainRatioAttributeEval" technique and "Ranker" method based on the full tainting set. As it is shown in Fig. 8, all the (13) attributes have effects on the prediction, but the most three attributes in order are (M,C,L), which are: Thalassemia, Chest pain type, Number of major vessels; respectively.

## VI. CONCLUSION

This paper intends to present a new diagnostic model for the detection of the HDs using the most efficient classifier based on accuracy and time using the WEKA tool. Experiments on PHDs model are conducted using Artificial Neural Network (ANN), Naïve Bays (NB), J48, and REPTree classifiers and are tested based on percentage split and cross-validation. The overall results of experimental results show that the NB classifier is the best, and our model yields over 85% accuracy using it based on the WEKA tool. It can conclude from the analysis of the experimental results that the NB technique turned out to be the most accurate classifier for the HDD. Also, results showed that this technique was the fastest of all. These results are incompatible with many previous works, which conclude that the NB did not give the best accuracy. These results are significant in the field of detecting the HDs also to decrease the reasons for death. Moreover, the results show that the three most effective attributes in order are: Thalassemia, Chest pain type, Number of major vessels.

## VII. FUTURE WORK

In the future, more experiments using many variants of data sets can be conducted to prove the current results or explore new conclusions. In addition, Text mining can be employed to predict and diagnose the HDs.

## REFERENCES

- [1] Purusothaman, G., & Krishnakumari, P. (2015). A survey of data mining techniques on risk prediction: Heart disease. *Indian Journal of Science and Technology*, 8(12), 1.
- [2] Ho, T. B. (2016). *Knowledge Discovery*. In *Knowledge Science*, pp. 70-93. CRC Press.
- [3] Fayyad, U., Piatetsky-Shapiro, G., and Smyth P. "From data mining to knowledge discovery in databases", *AI magazine*, 17 (3), (1996): pp. 37-54.
- [4] Bharathi, A., and E. Deepankumar. "Survey on classification techniques in data mining", *International Journal on Recent and Innovation Trends in Computing and Communication* 2.7 (2014): 1983-1986.
- [5] Gurney, K., *An introduction to neural networks*. CRC press, 2014.

- [6] Lewis, David D. "Naïve (Bayes) at forty: The independence assumption in information retrieval", European conference on machine learning. Springer, Berlin, Heidelberg, 1998.
- [7] Ramesh, D., Pasha, S. N. and Roopa, G. (2017). "A Comparative Analysis of Classification Algorithms on Weather Dataset Using Data Mining Tool", Oriental Journal of Computer Science and Technology, 10 (4).
- [8] Kalmegh, S. (2015). "Analysis of WEKA Data Mining Algorithm REPTree, Simple Cart and Random Tree for Classification of Indian News". International Journal of Innovative Science, Engineering & Technology (IJSET), 2(2), pp. 438-446.
- [9] Patel, J., Tejal Upadhyay, D. and Patel, S. (2015). Heart disease prediction using machine learning and data mining technique. Heart Disease, 7(1), 129-137.
- [10] Sudhakar, K. and Manimekalai, D. M. (2014). Study of heart disease prediction using data mining. International journal of advanced research in computer science and software engineering, 4(1).
- [11] Dewan, A., and Sharma, M. (2015, March). Prediction of heart disease using a hybrid technique in data mining classification. In 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 704-706). IEEE.
- [12] Masethe, H. D. and Masethe, M. A. (2014, October). Prediction of heart disease using classification algorithms. In Proceedings of the world Congress on Engineering and computer Science (Vol. 2, pp. 22-24).
- [13] Kim, J., Lee, J., & Lee, Y. (2015). Data-mining-based coronary heart disease risk prediction model using fuzzy logic and decision tree. Healthcare informatics research, 21(3), 167-174.
- [14] Krishnaiah, V., Narsimha, G., & Chandra, N. S. (2015). Heart disease prediction system using data mining technique by fuzzy K-NN approach. In Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1 (pp. 371-384). Springer, Cham. [16] Blake, C. L. and Merz, C. J. UCI Repository of machine learning databases. Irvine, CA, 1998. <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [15] Choi, E., Schuetz, A., Stewart, W. F., & Sun, J. (2016). Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association, 24(2), 361-370.
- [16] Blake, C. L. and Merz, C. J. UCI Repository of machine learning databases. Irvine, CA, 1998. <http://www.ics.uci.edu/~mllearn/MLRepository.html>

#### AUTHORS' PROFILE

**Amirah Al Shammari** is a Teaching Assistant in the Computer Science department at Aljouf University, Saudi Arabia and is currently completing my Master's degree in the Informatics department at the University of Qassim, Buraydah, Saudi Arabia.

**Haneen Al Hadeaf** is a Teaching Assistant in the Computer Science department at Qassim University, Saudi Arabia. Currently, her studying a master's degree in the Informatics department at Qassim University, Buraydah, Saudi Arabia. She received her B.Sc. in Computer Science from Imam Mohammad Ibn Saud Islamic University in Riyadh, Saudi Arabia. And her area of research in computer vision, machine learning, and software engineering.

**Hedia Zardi** an assistant professor in the computer science department at Qassim University, Saudi Arabia. She received the PhD degree in 2016 from Sorbone University, Paris, France and Manouba University, Tunisia.

# A New Approach to Predicting Learner Performance with Reduced Forgetting

Dagou Dangui Augustin Sylvain Legrand KOFFI<sup>1</sup>, Tchimou N'TAKPE<sup>2</sup>  
Assohon ADJE<sup>3</sup>, Souleymane OUMTANAGA<sup>4</sup>

Unité de Formation et de Recherche des Mathématiques et Informatique (UFR-MI)  
Université Félix Houphouët-Boigny (UFHB), Abidjan, Côte d'Ivoire<sup>1,3</sup>

Laboratoire de Mathématiques et Informatique, Université Nangui Abrogoua, Abidjan, Côte d'Ivoire<sup>2</sup>

Laboratoire de Recherche en Informatique et Télécommunication

Institut National Polytechnique Félix Houphouët-Boigny, Yamoussoukro, Côte d'Ivoire<sup>4</sup>

**Abstract**—The work on predicting learner performance allows researchers through machine learning methods to participate in the improvement of e-learning. This improvement allows, little by little, e-learning to be promoted and adopted by several educational structures around the world. Neural networks, widely used in various performance prediction works, have made several exploits. However, factors that are highly influential in the field of learning have not been explored in machine learning models. For this reason, our study attempts to show the importance of the forgetting factor in the learning system. Thus, to contribute to the improvement of accuracy in performance predictions. The interest being to draw the attention of researchers in this field to very influential factors that are not exploited. Our model takes into account the study of the forgetting factor in neural networks. The objective is to show the importance of attenuation of the forgetting, on the quality of performance predictions in e-learning. Our model is compared to those based on Random Forest and linear regression algorithms. The results of our study show first that neural networks (95.20%) are better than Random Forest (95.15%) and linear regression (93.80%). Then, with the attenuation of forgetting, these algorithms give 96.63%, 95.85% and 93.80% respectively. This work allowed us to show the great relevance of oblivion in neural networks. Thus, the exploration of other unexploited factors will make better performance prediction models.

**Keywords**—Performance prediction; e-learning; artificial neural networks; forgetting factor

## I. INTRODUCTION

Most of the work on e-learning recommendation systems focuses on the construction of recommendation models. The aim of this recommendation model building work is to improve the recommendations of learning objects for learners [1]. The prediction of learner performance in education, particularly in e-learning, is one of the most studied research areas. It is part of the quest to improve recommendations in recommendation systems [2]. Predictions of performance in e-learning systems require a good knowledge of the users in order to provide better accuracy.

However, eLearning systems, despite having large amounts of information about their users, suffer from the lack of some very relevant information [3]. This information would help to improve the results of performance predictions. As a result, the search for highly relevant information is now at the heart of

research in the field of learning [3]. It is a challenge for researchers.

Our study in this paper focuses on improving the accuracy of learner performance predictions. Thus, we start from the fact that the use of relevant data in weak algorithms is less advantageous than the use of relevant data in powerful algorithms. For this, we use neural networks, one of the most efficient learning machine algorithms in the field of learner performance prediction [4] [5]. We use the forgetting factor in our study. It is a very important factor, which is widely used in the field of learning psychology [6]. Taking into account the attenuation of the forgetting factor in learners in neural networks is at the heart of our study.

The paper is organized as follows: Section 2 presents related work. Section 3 presents neural network algorithms and other algorithms for comparison. Section 4 presents our new approach. Section 5 presents an evaluation of our method and a discussion about our approach. Section 6 concludes this manuscript.

## II. RELATED WORK

Neural networks techniques are used in many fields and today are widely used in e-learning because of their high efficiency. Oladokun et al [4], conducted a study to predict the performance of candidates likely to succeed at university using neural networks to improve the quality of university degrees. The results of this study show a performance of more than 70% and thus demonstrate the ability of neural networks to improve university admission systems. However, the authors point out the limits in the search for relevant information to make the model more effective. Arsad et al [7], for their part, also propose a neural network model for the prediction of student performance from the entry level to registration. Following the basic subjects that the student takes in the course, the results of the study showed that there is a direct correlation between the students' results for the first semester subjects with the final academic performance regardless of their gender. Thus, based on these results, the authors believe that a strategic study can be undertaken during study periods to improve students' final performance. Knowing the advantages of predicting educational performance, which helps in decision making to improve educational services, Chen et al [8] conducted a study, allowing them to propose a neural network model to predict

student performance in standardized examinations. Two meta-heuristic algorithms were used separately to form the feed-forward network for prediction and optimized the interlayer weights and biases of the neural network. Given the quality of the results, the model was designed to help students with admission procedures and to strengthen the system of services in educational institutions. Faced with the non-existence of methods for predicting student performance for some higher education institutions, Shahiri et al [5] conducted a study to predict student performance using data mining techniques on courses taught. Using performance prediction algorithms, the aim was to determine the most important factors in the learners' data for better planning of courses during study periods in order to improve the performance of their learning and teaching process. Among the prediction algorithms studied, the neural network (98%) shows the highest performance, followed by the decision tree (91%), the support vector machine (83%), the nearest neighbor cases (83%) and the Naive Bayes (76%). Jishan et al [9] present a study aimed at improving the accuracy of student performance prediction models. To do this, their study focused on data pre-processing using an oversampling technique, Synthetic Minority Over-Sampling (SMOTE) and a discretization method known as Optimal Equal Width Binning, and applied to three classification algorithms, the Naive Bayes, the decision tree and the neural network. The result of this study shows that the accuracy of the prediction models improves when using the two proposed pre-processing methods, with the neural network at the first rank followed by the Naive Bayes classifier. Kouser et al [10] are looking for ways to improve the results provided by the unstructured and low-level information collected, in order to make good and better predictions of student performance. To this end, they use exploratory methods for the analysis of raw data to extract high quality information. Four variables of students' daily activities on the Moodle platform were used in the construction of a neural network model for predicting good student grades for classes. Abu Zohair [11], faced with the difficulties encountered for predicting performance, which is not credible in education because of the small size of the data, conducted a study to find out if it was possible to obtain an accurate prediction rate by training and mobilizing students with a small data set by identifying key indicators in the data set. The results of this study indicate that the media vector machine and the learning of discriminant analysis algorithms give test rates of accuracy and reliability of 98.5%. In response to the problems of drop-out and delayed graduation, Umar [12] is conducting a study to predict poor student performance. These results can then be used for academic follow-up to alleviate these problems. Thus, a neural network capable of predicting a student's overall average using the student's personal data has been designed. The results were 73.68% correctly predicted performance with also an accuracy of 66.67% of the students likely to drop out or experience a delay before graduation. Raga and Raga [13] conducted an experiment to develop a model for predicting student performance in order to measure the non-linear predictive power of neural networks in a blended learning environment. But first, the hyperparameters of the model were determined by a series of experiments. The results of this experiment show, for a single course in the first month, an accuracy of 91.07%

with a ROC\_AUC score of 0.88 which improves as the accumulation progresses. But for the mid-term results, the highest precision was 80.36% and the ROC\_AUC score was 0.70. Azzi et al. en [14], propose an approach based on artificial neural networks in order to address the problems of customization of e-learning systems. In particular, they address the problem of designing courses based on the background of the learners. The role of the proposed model is to mimic the course designer in order to create customized courses for learners. The system is thus able to choose the appropriate content to improve the learner's performance.

The previous literature review presents neural networks that are widely used in performance prediction work in e-learning. Precisely because of the power of its algorithms, which are still proving their worth in this field.

Studies on predicting learner performance and taking into account forgetting have also been carried out. Nguyen et al [2] propose a study for the prediction of student performance. They use tensor factorization methods to implicitly take into account latent factors and temporal effect. The results of their proposed approaches show that they are promising and appropriate in improving prediction results. The authors also argue that a predictive approach could be used to account for the sequential effect. Nedungadi and Remya [15] propose a new PC-BKT model, an improved model of the existing BKT which generally sets the forgetting parameter to 0. The results provided by the PC-BKT show the percentage of classification errors reduced as the algorithm adjusts the learning rate of a skill over the duration between the last uses. Thus, they state that the time it takes a student to start forgetting the skill is 30 days.

The observation made at the end of this second literature review is the scarcity of performance prediction works taking into account forgetting. Even rarer are these works with neural networks.

Studies have also been carried out on its great influence in the field of learning. Ziegler [6] studied the process of forgetting empirically using data collected on the brain, spinal cord and nerves in an experiment in which a group of 58 elderly people took part in training courses followed by examinations. The results are such that the rate of forgetfulness was higher for half, and had no huge change after 118 days. The reason is that during this time, the students had revision times. Krondorfer [16] in his work proposes to differentiate between forgetting and remembering in order to help erode paralyzing traumatic memories. Thus, he argues that there are no impeccable distinctions or clearly marked boundaries between these terms. Remembrance is elevated to the status of unquestionable virtue and forgetting is mocked as undesirable and reprehensible. Casey and Olivera [17] set out in a study to clarify the relationship between organizational memory and forgetting, and to identify areas that need to be developed to improve the understanding of memory constructs. Thus, they argue that the dynamic nature of organizational knowledge, the role of time in the way organizations retain knowledge, and the role of power dynamics in what and how organizations choose to remember and forget could address their concern. Gordon [18] in one essay proposes to reflect on the meaning of the

tension between remembering and forgetting in the context of historical and tragic events. He argued that an ethic of remembering and forgetting could enable victims of trauma to understand not only the sources of their suffering but also to take responsibility for their own liberation. Gan and Zeng [19] are conducting a study with the aim of improving the speed of convergence of iterative learning, controlling it and reducing the fluctuation of system error. For this purpose, they use a class of steady-state linear systems with a variable forgetting factor. The results of the work give the algorithm is efficient and that the convergence speed is improved with a low error rate. Boutis et al [20] conducted a study to determine the rate of knowledge degradation over time using the forgetting curve. Thus, by a test on 106 participants measuring the degradation over time for 12 months, the conclusion drawn was that the degradation of learning was attenuated every two months. These results can thus influence the scheduling of refresher courses.

This third review of the literature presents the study on forgetting in more research work. Which shows its great importance.

### III. MACHINE LEARNING ALGORITHMS

#### A. Artificial Neural Networks (ANN)

Neural networks are very powerful algorithms used in Artificial Intelligence, especially in the field of machine learning. It has been designed to approach problem solving in the same way as the human brain does. It has the ability to solve problems of great complexity [10]. The architecture of neural networks consists of three types of layers, the hidden layers located in between, the input layer, and the output layer. Each layer may consist of at least one node called a neuron. The neurons of the different layers are connected to each other by synaptic weights as shown in Figure 1. Neural networks vary according to the type of configuration. The simplest of these configurations is the perceptron.

Data flows through neural networks from input to output through a process called forward propagation. In this process, neuron outputs are determined by the arithmetic operation of applying a neuron activation function to the sum of all inputs. The activation function is a stimulation threshold which, when reached, causes the neuron to respond. Figure 2 shows a representation of a neuron. There are several types of activation functions, which can be both linear and non-linear depending on the objectives to be achieved. The most commonly used is the sigmoid or logistic function. However, other more powerful activation functions such as the reread [21] exist.

After forward propagation, the outputs and errors of the neural network outputs are determined. Back propagation is performed to adjust the values of the neural network weights randomly initialized at the beginning.

#### B. Random Forest

The Random Forest is also an algorithm in the field of machine learning. It makes it possible to make a classification of whole. It is more effective in the predictions for sets than the particular predictions [22]. Studies have shown that the Random Forest is based on the most powerful nonparametric

classifiers [23] [24]. In this logic, this classifier is adapted to our study for a good measurement of our model. It is based on the principle of decision trees. As its name indicates, it is composed of a large number of decision trees which function as a set. Each individual tree in Random Forest makes a class prediction. Thus, the class with the most votes becomes the class considered for the Random Forest prediction.

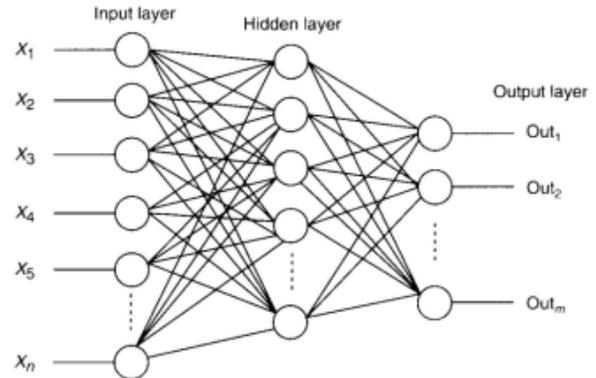


Fig. 1. Neural Network Architecture [10].

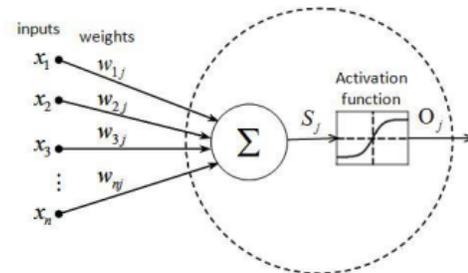


Fig. 2. Representation of a Neuron [10].

#### C. Linear Regression

Linear regression is also a self-learning algorithm [25]. It is an approach for modelling the relationship between a dependent variable and one or more explicit variables. The objective of this algorithm is to determine the hyperparameters of the linear model, formed from a set of data. It allows the design of a rectilinear or curvilinear function that best approximates the elements of the data set. Linear regression is one of the best known and most widely used methods in statistics for the analysis of quantitative data.

### IV. NEW APPROACH

Many performance predictions works have been performed by neural networks because of their high efficiency [5] [9]. The approach proposed in this work is also based on neural networks. It integrates a very important latent factor which is oblivion. This is a very influential factor in the field of learning. Moreover, performance prediction works including this factor are rare in the literature. Even more so with neural networks in predicting learner performance. The objective is to show the impact that this factor has on performance predictions in neural networks. The history of learner performance is taken into account, as well as other general factors at the input of the network. The goal is to bring more precision in performance predictions. (Ebbinghaus, H. (1885/1962)), one of the fathers

of experimental psychology who showed by the results of his experiments on the experimental study of memory and the learning process, that the forgetting curve describes a decreasing exponential form [26] [27], as shown in figure 3.

This Ebbinghaus theory is still relevant today and is widely used by many researchers working on the cognitive. It tends to show that the rate of information loss by the human brain is exponential. The work of L. Averell and A. Heathcote in 2011 [29], focused on determining the mathematical form of forgetting. They proposed three candidate functions that best express forgetting, namely the exponential  $e^{-\alpha t}$ , the pareto  $(1 + \gamma t)^{-\beta}$  and the power  $(1 + t)^{-\beta}$ . The exponential function is the most appreciated for the description of forgetting compared to the other two functions. Their research confirms Ebbinghaus' theory. The proposal of our model for managing forgetting in e-learning is based on the exponential form of forgetting defined in the following equation,

$$R = e^{-\alpha t} \quad (1)$$

where R is the memory retention and  $\alpha$  is the forgetting rate.

The forgetting rate, on the other hand, is a function of the memory strength F, as presented in the following equation.

$$\alpha = \frac{1}{F} \quad (2)$$

[30] Thus, the greater the memory strength, the lower the forgetfulness rate, which increases in (1) memory retention R, thus reducing forgetfulness. Similarly, the smaller the memory strength, the higher the forgetting rate, which normally reduces memory retention, thus increasing forgetting. According to the work of Ebbinghaus, memory depends essentially on the number of repetitions of a learning element and the time spent reviewing what has been learned [27]. From this reflection, memory is strengthened with the high number of repetitions of the learner. Thus, to reduce forgetting in our model, the number of repetitions of learning elements must be increased. However, the set of data available to us for our work does not include a variable signifying the number of repetitions for the learners' learning. However, our data set has a variable that means absences from classes by learners. Repetitions and absences have opposite effects on learning. Thus, a learner's retention capacity increases as the number of repetitions increases. It also increases when the number of absences from classes decreases. However, the two increases are not the same in both cases. To mitigate forgetting in our model, we consider the rate of absences at courses instead of the number of repetitions of learning elements. Thus, by reducing the content of the absence variable, retention capacity increases. However, it is better to take into account the number of repetitions if it is included such as a variable in a dataset for a study. Therefore, the manipulation of the absence rate, free time after school and the learner's study time would be tantamount to acting on forgetting so as to see its impact on the prediction of performance. Thus, likely to lead to good decisions for homework sessions in order to optimize academic performance.

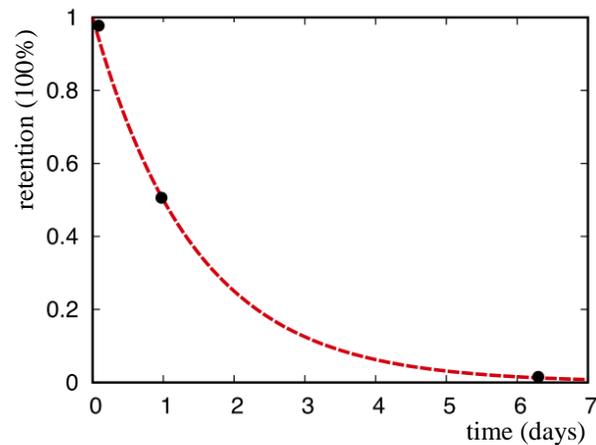


Fig. 3. The Ebbinghaus Forgetting Curve.[28].

Our experiments are done in two steps. In the first step, a study of learner performance predictions is made by designing three prediction models. The first model is done using neural networks, the second using Random Forest and the third using linear regression. A comparison of the performance results obtained from our three models is then performed. The objective of this first step is to confirm or refute the results of the work on the performance of neural networks as the best predictor of e-learning. Assertion made in the literature review.

In the second step, the experience of the first step is experienced again, this time taking into account the reduction of forgetting. To this end, reductions in the rate of absence and free time are made. These are factors on which forgetting depends. The reduction of each of these variables aims to reduce forgetfulness among learners. Thus, for this phase, the objective is to determine approximately the probable performance of the learners with reduced forgetting. The results of the predictions with reduced forgetting will be compared with the results of the experiments in the first stage. Thus, it can be shown the contribution of the attenuation of the forgetting in the predictions of performance of the learners.

The reductions of learners' absences at class and free time, in our experience, are respectively 90% of the content of the variable "Absences" and 20% of the content of the variable "Freetime" of the data set. As for the learner study time variable, it remains unchanged. Ideally, it should be increased in order to intensify the reduction of forgetting. The reduction rates proposed for the absence rate and free time are not based on any theory. They are taken randomly with the sole aim of reducing forgetfulness.

#### A. Data Set

The dataset used in this work is available in the UCI's database. It consists of 395 samples with 30 variables. It also includes three data sets G1, G2 and G3 which are the successive results of evaluations. Each of these variables is a whole score between 0 and 20. The objective of our study will be to predict the G3 assessment scores of learners assumed to be unknown and assuming that G1 and G2 scores are already made. There is no missing value data in this set. The following list gives a description of the 30 variables.

1 <b>school</b>	student's school (binary: "GP" - Gabriel Pereira or "MS" - Mousinho da Silveira)
2 <b>sex</b>	student's sex (binary: "F" - female or "M" - male)
3 <b>age</b>	student's age (numeric: from 15 to 22)
4 <b>address</b>	student's home address type (binary: "U" - urban or "R" - rural)
5 <b>famsize</b>	family size (binary: "LE3" - less or equal to 3 or "GT3" - greater than 3)
6 <b>Pstatus</b>	parent's cohabitation status (binary: "T" - living together or "A" - apart)
7 <b>Medu</b>	mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
8 <b>Fedu</b>	father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
9 <b>Mjob</b>	mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")
10 <b>Fjob</b>	father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")
11 <b>reason</b>	reason to choose this school (nominal: close to "home", school "reputation", "course" preference or "other")
12 <b>guardian</b>	student's guardian (nominal: "mother", "father" or "other")
13 <b>traveltime</b>	home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
14 <b>studytime</b>	weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15 <b>failures</b>	number of past class failures (numeric: n if 1<=n<3, else 4)
16 <b>schoolsup</b>	extra educational support (binary: yes or no)
17 <b>famsup</b>	family educational support (binary: yes or no)
18 <b>paid</b>	extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19 <b>activities</b>	extra-curricular activities (binary: yes or no)
20 <b>nursery</b>	attended nursery school (binary: yes or no)
21 <b>higher</b>	wants to take higher education (binary: yes or no)
22 <b>internet</b>	Internet access at home (binary: yes or no)
23 <b>romantic</b>	with a romantic relationship (binary: yes or no)
24 <b>famrel</b>	quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25 <b>freetime</b>	free time after school (numeric: from 1 - very low to 5 - very high)
26 <b>goout</b>	going out with friends (numeric: from 1 - very

low to 5 - very high)

27 <b>Dalc</b>	workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28 <b>Walc</b>	weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29 <b>health</b>	current health status (numeric: from 1 - very bad to 5 - very good)
30 <b>absences</b>	number of school absences (numeric: from 0 to 93)

Evaluation grades:

31 <b>G1</b>	first period grade (numeric: from 0 to 20)
32 <b>G2</b>	second period grade (numeric: from 0 to 20)
33 <b>G3</b>	final grade (numeric: from 0 to 20, output target)

Dataset source:

(<https://archive.ics.uci.edu/ml/datasets/Student+Performance>)

### B. Splitting the Data Set

Splitting the data set in the machine learning process consists of dividing the data used into two parts. The first part, which will be the largest portion, is reserved for the training phase of the model. It is usually greater than or equal to 60%. The second part, which is obviously the smaller one, is dedicated to the test phase of the trained model. For the experimentation of our model, the data set was divided into two parts. One part for training with a proportion of 70% and the other part for testing with a proportion of 30%.

### C. Configuration of our Neural Network

Our proposed neural network model is shown in Figure 4 and is configured on three layers, one input layer, another as a hidden layer and the third is the output layer. The input layer has 32 neurons for the 30 variables of the dataset and G1 and G2 notes to take into account the learners' prior skills. The hidden layer also has 32 neurons. The output layer has a single neuron. The neuron activation function used is the "relu" because it is more efficient than the sigmoid [21], with "adam" as solver and a constant learning rate, initialized at 0.01. the maximum number of iterations is 500.

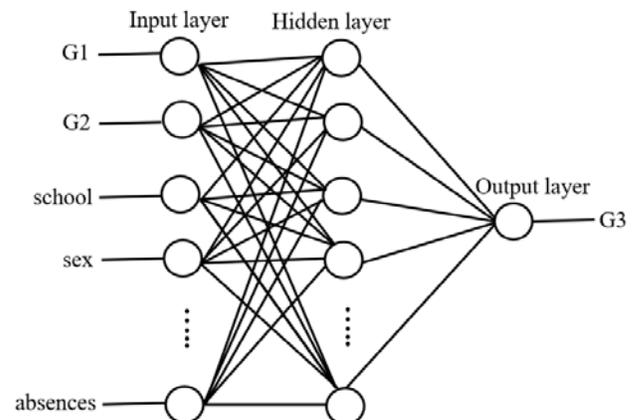


Fig. 4. Our Neural Network Model.

V. EXPERIMENTAL RESULTS AND DISCUSSION

Our experiments were carried out with a four-core I5 computer with a processor speed ranging from 1.70 to 2.40. The computer was equipped with an I5 processor with four cores. This computer works with a 12 GB RAM memory. The programming language used is the python language. After the experiment follows the evaluation phase. In this phase, the root means square error and the confusion matrix are used for the evaluation of the experiments in this study. The error is thus determined by the following formula:

$$RMSE = \sqrt{\frac{\sum_{ui}(r_{ui}-\hat{r}_{ui})^2}{n}} \tag{3}$$

The confusion matrix defines the metrics accuracy, precision, recall and F-measure as shown in equations (4), (5), (6) and (7) respectively. These equations are a function of the information in Table 1, defined as follows:

TABLE I. CONFUSION MATRICES

		Detected	
		Positive	Negative
Current	Positive	True positives (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

TP : Positive class result correctly predicted;

TN: negative class result: negative class result predicted correctly;

FP : result of the incorrectly predicted positive class;

FN : negative class result: result of the negative class incorrectly predicted.

The confusion matrix thus defines the following metrics:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

$$F - Measure = \frac{2*Precision*Recall}{Precision+Recall} \tag{7}$$

**Accuracy:** refers to the proportion of correct predictions;

**Precision:** refers to the proportion of correct predictions among positive predictions;

**Recall:** refers to the proportion of positives that are correctly identified;

**F-Measure:** Used to evaluate a compromise between recall and precision.

A. Results and Discussion of Step 1 of the Experiment

This step consists of comparing the performance of the algorithms used in this study. The goal is to confirm the claims of the literature review indicating neural networks as the best predictor of learner performance in e-learning. Table 2 presents

the results of the confusion matrix metrics for the artificial neural network algorithms, Random forest and Linear regression. Table 3 presents the results of the errors obtained from these algorithms during the experiment.

At the end of these first phase experiments, the artificial neural networks show the best results for each of the accuracy, precision, recall and F-Measure metrics of 94.94%, 95.20%, 97.36% and 96.27% respectively. The error rate is 0.2250. The performance of the neural networks is followed by the performance of the Random Forest with 94.17 %, 95.15 %, 96.23 % and 95.69 % for the accuracy, precision, recall and F-Measure metrics respectively. The error rate for Random Forest is 0.2413. Linear regression comes last with 92.15%, 93.80%, 94.32% and 93.98%, in the same order of accuracy, precision, recall and F-Measure metrics. The error rate is 0.2541. The results of this first phase confirm the performance of neural networks as the best predictors of e-learning performance as indicated in the literature review. On the other hand, the enormous performance of Random Forest should be underlined. Could it be more effective than neural networks under other conditions and circumstances, on other types and scales of data? The question remains for possible studies.

B. Results and Discussion of Step 2 of the Experiment

This step consists of determining the performance of the algorithms used in this study by reducing the effects of forgetting among learners. The performances obtained are compared to the performances obtained during the first step according to the algorithms. Table 4 presents the results of the confusion matrix metrics for the artificial neural network, Random forest and Linear regression algorithms. Table 5 presents the results of the errors obtained from these algorithms during the experiment.

TABLE II. COMPARISON RESULTS OF ALGORITHMS WITH OMISSION

Algorithms	Accuracy	Precision	Recall	F- Measure
Artificial Neural Network	94.94 %	95.20 %	97.36 %	96.27 %
Random forest	94.17 %	95.15 %	96.23 %	95.69 %
Linear regression	92.15 %	93.80 %	94.32 %	93.98 %

TABLE III. ALGORITHM ERRORS WITH OVERLOOKED ALGORITHMS

Algorithms	RMSE
Artificial Neural Network	0,2250
Random forest	0,2413
Linear regression	0,2541

TABLE IV. COMPARISON RESULTS OF ALGORITHMS WITH ATTENUATED FORGETTING

Algorithms	Accuracy	Precision	Recall	F- Measure
Artificial Neural Network	95.95 %	96.63 %	97.36 %	97 %
Random forest	94.43 %	95.85 %	95.85 %	95.85 %
Linear regression	92.15 %	93.80 %	94.32 %	93.98 %

TABLE V. ALGORITHM ERRORS WITH SOFT OVERLOOKED ALGORITHMS

Algorithms	RMSE
Artificial Neural Network	0,2012
Random forest	0,2360
Linear regression	0,2541

At the end of these second phase experiments, the artificial neural networks still show the best results for each of the accuracy, precision, recall and F-Measure metrics of 95.95%, 96.63%, 97.36% and 97% respectively. The error rate is 0.2012. Neural network performance is always followed by Random Forest performance with this time 94.43%, 95.85%, 95.85% and 95.85% for the accuracy, precision, recall and F-Measure metrics respectively. The error rate for the Random forest now increases to 0.2360. Linear regression comes last, but maintains the same performance as in Phase 1 with 92.15%, 93.80%, 94.32% and 93.98%, still in the same order as the accuracy, precision, recall and F-Measure metrics. The error rate is 0.2541. The results of this second phase still confirm the performance of neural networks as the best predictors of e-learning performance as indicated in the literature review, despite the reduction in forgetting.

The finding in this second phase is that there is an improvement in accuracy and precision for neural networks of 1.01% and 1.43% respectively when forgetting is reduced. For Random forest, there is an improvement in accuracy and precision of 0.26% and 0.7% respectively. However, there is no variation in the linear regression. These variations in accuracy and precision for these algorithms have two meanings:

1) Respectively, with respect to accuracy and precision, neural networks have a better ratio of correct predictions and also a better ratio of confirmations of correct positive predictions compared to the other two algorithms.

2) A greater improvement in the ratio of correct predictions and the ratio of confirmations of correct positive predictions compared to the other two algorithms.

As far as recall is concerned, there is no improvement when reducing forgetfulness for neural networks and linear regression. However, there is a regression of -0.38% for Random forest. This means that compared to Random forest, neural networks and linear regression best predict positive cases.

For F-measure, with the reduction of forgetfulness, neural networks, Random forest and linear regression increase by 0.74%, 0.16% and 0% respectively. Thus, neural networks and Random forest reduce the trade-off between recall and accuracy. This result not only confirms the better performance of neural networks compared to other algorithms but also shows an improvement in the quality of neural networks when reducing forgetfulness.

Finally, the error margins were reduced to 0, 0.0053 and 0.0238 respectively for linear regression, Random forest and neural networks with the reduction of forgetting among learners. The neural networks still show their performance with more error reduction.

## VI. CONCLUSION

The objective of our study was to clarify the contribution of taking into account the forgetting factor in models for predicting learner performance in e-learning. During our work, we integrated it into an artificial neural network to create a model in order to see the impact of this factor on performance predictions in these neural networks. In this study, two other classifiers, namely Random Forest and linear regression, were used in order to compare their results with those of our model. At the end of this work, we observe the performance of neural networks. With the process of reducing forgetfulness proposed by our model, neural networks give even better results. This will allow us to better specify the performance of students in order to adopt better strategies in decisions about their training. Our theory also improves the performance of the Random Forest.

In our future work, we will extend this study to a larger scale to see what impact our theory will have on large datasets. We will also rule on the study of stress, another latent factor that has a strong influence on memory in the learning system. We will also conduct another study to identify the most important variables in a list of variables in a data set. This may help to further improve predictions of learner performance.

## ACKNOWLEDGMENT

At the Unité de Formation et de Recherche des Mathématiques et Informatique (UFR-MI) of the Université Félix Houphouët-Boigny (UFHB) and the Laboratoire de Recherche en Informatique et Télécommunication (LARIT). Also, to the UCI for its database.

## REFERENCES

- [1] K. I. Ghauth et N. A. Abdullah, "Learning materials recommendation using good learners' ratings and content-based filtering", *Education Tech Research Dev*, vol. 58, n° 6, p. 711-727, déc. 2010, doi: 10.1007/s11423-010-9155-4.
- [2] N. Thai-nghe, L. Drumond, R. Nanopoulos, et L. Schmidt-thieme, "Matrix and tensor factorization for predicting student performance", in *In Proceedings of the 3rd International Conference on Computer Supported Education (CSEDU)*, 2011.
- [3] H. Al-Samarraie, B. K. Teng, A. I. Alzahrani, et N. Alalwan, "E-learning continuance satisfaction in higher education: a unified perspective from instructors and students", *Studies in Higher Education*, vol. 43, n° 11, p. 2003-2019, nov. 2018, doi: 10.1080/03075079.2017.1298088.
- [4] V. Oladokun, A. Adebajo, B. Sc, et O. Charles-Owaba, "Predicting Students Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course", *The Pacific Journal of Science and Technology*, vol. 9, April 1, 2.
- [5] A. M. Shahiri, W. Husain, et N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques", *Procedia Computer Science*, vol. 72, p. 414-422, janv. 2015, doi: 10.1016/j.procs.2015.12.157.
- [6] L. H. Ziegler, "Learning and Forgetting", *The Journal of Higher Education*, vol. 2, n° 3, p. 144-146, mars 1931, doi: 10.1080/00221546.1931.11772362.
- [7] P. M. Arsad, N. Buniyamin, et J. A. Manan, "Neural network model to predict electrical students' academic performance", in *2012 4th International Congress on Engineering Education*, déc. 2012, p. 1-5, doi: 10.1109/ICEED.2012.6779270.
- [8] J.-F. Chen, H.-N. Hsieh, et Q. H. Do, "Predicting Student Academic Performance: A Comparison of Two Meta-Heuristic Algorithms Inspired by Cuckoo Birds for Training Neural Networks", *Algorithms*, vol. 7, n° 4, p. 538-553, déc. 2014, doi: 10.3390/a7040538.

- [9] S. Jishan, R. Rashu, N. Haque, et R. M. Rahman, "Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique", *Decision Analytics*, vol. 2, n° 1, p. 1- 25, 2015, doi: 10.1186/s40165-014-0010-2.
- [10] R. Kouser, J. D. Joann, et K. Suganya, "Forecasting Student Academic Performance by Decision Tree Learning using Artificial Neural Networks", 2017. <https://www.semanticscholar.org/paper/FORECASTING-STUDENT-ACADEMIC-PERFORMANCE-BY-TREE-Kouser-Joann/5b116400ef849d120602656719888f5c2e3ec0e2>.
- [11] L. M. Abu Zohair, "Prediction of Student's performance by modelling small dataset size", *International Journal of Educational Technology in Higher Education*, vol. 16, n° 1, p. 27, août 2019, doi: 10.1186/s41239-019-0160-3.
- [12] M. A. Umar, "Student Academic Performance Prediction using Artificial Neural Networks: A Case Study", *International Journal of Computer Applications*, vol. 178, p. 24- 29, sept. 2019, doi: 10.5120/ijca2019919387.
- [13] R. C. Raga et J. D. Raga, "Early Prediction of Student Performance in Blended Learning Courses Using Deep Neural Networks", in *2019 International Symposium on Educational Technology (ISET)*, juill. 2019, p. 39- 43, doi: 10.1109/ISET.2019.00018.
- [14] I. Azzi, A. Jeghal, A. Radouane, A. Yahyaouy, et H. Tairi, « Approach Based on Artificial Neural Network to Improve Personalization in Adaptive E-Learning Systems », in *Embedded Systems and Artificial Intelligence*, Singapore, 2020, p. 463- 474, doi: 10.1007/978-981-15-0947-6\_44.
- [15] P. Nedungadi et M. S. Remya, "Incorporating forgetting in the Personalized, Clustered, Bayesian Knowledge Tracing (PC-BKT) model", in *2015 International Conference on Cognitive Computing and Information Processing(CCIP)*, mars 2015, p. 1- 5, doi: 10.1109/CCIP.2015.7100688.
- [16] B. Krondorfer, "IS Forgetting Reprehensible? Holocaust Remembrance and the Task of Oblivion", *Journal of Religious Ethics*, vol. 36, n° 2, p. 233- 267, juin 2008, doi: 10.1111/j.1467-9795.2008.00345.x.
- [17] A. J. Casey et F. Olivera, "Reflections on Organizational Memory and Forgetting", *Journal of Management Inquiry*, vol. 20, n° 3, p. 305- 310, sept. 2011, doi: 10.1177/1056492611408264.
- [18] M. Gordon, "Between Remembering and Forgetting", *Stud Philos Educ*, vol. 34, n° 5, p. 489- 503, sept. 2015, doi: 10.1007/s11217-014-9451-2.
- [19] Y. Gan et Q. Zeng, "Variable Gain Iterative Learning Control with Forgetting Factor", *IOP Conference Series: Materials Science and Engineering*, vol. 394, p. 052082, août 2018, doi: 10.1088/1757-899X/394/5/052082.
- [20] K. Boutis et al., "The effect of testing and feedback on the forgetting curves for radiograph interpretation skills", *Medical Teacher*, vol. 41, n° 7, p. 756- 764, juill. 2019, doi: 10.1080/0142159X.2019.1570098.
- [21] D. Pedamonti, "Comparison of non-linear activation functions for deep neural networks on MNIST classification task", avr. 2018.
- [22] M. Pal, "Random forest classifier for remote sensing classification", *International Journal of Remote Sensing*, vol. 26, n° 1, p. 217- 222, janv. 2005, doi: 10.1080/01431160412331269698.
- [23] T. Han, D. Jiang, Q. Zhao, L. Wang, et K. Yin, "Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery", *Transactions of the Institute of Measurement and Control*, vol. 40, p. 2681 - 2693, mai 2018, doi: 10.1177/0142331217708242.
- [24] M. Liu, M. Wang, J. Wang, et D. Li, "Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar", *Sensors and Actuators B: Chemical*, vol. 177, p. 970-980, févr. 2013, doi: 10.1016/j.snb.2012.11.071.
- [25] I. Naseem, R. Togneri, et M. Bennamoun, "Linear Regression for Face Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, n° 11, p. 2106- 2112, nov. 2010, doi: 10.1109/TPAMI.2010.128.
- [26] "Ebbinghaus, H. (1885/1962). *Memory: A contribution to experimental psychology*. New York: Dover".
- [27] G. A. Kimble et M. Wertheimer, *Portraits of Pioneers in Psychology*. Psychology Press, 2014.
- [28] B. Settles et B. Meeder, "A Trainable Spaced Repetition Model for Language Learning", in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany, août 2016, p. 1848-1858, doi: 10.18653/v1/P16-1174.
- [29] L. Averell et A. Heathcote, "The form of the forgetting curve and the fate of memories", *Journal of Mathematical Psychology*, vol. 55, n° 1, p. 25- 35, févr. 2011, doi: 10.1016/j.jmp.2010.08.009.
- [30] S. G. Hu et al., "Emulating the Ebbinghaus forgetting curve of the human brain with a NiO-based memristor", *Appl. Phys. Lett.*, vol. 103, n° 13, p. 133701, sept. 2013, doi: 10.1063/1.4822124.

# Genetic Algorithm with Comprehensive Sequential Constructive Crossover for the Travelling Salesman Problem

Zakir Hussain Ahmed

Department of Mathematics and Statistics, College of Science  
Al Imam Mohammad Ibn Saud Islamic University (IMSIU)  
Riyadh, Kingdom of Saudi Arabia

**Abstract**—The travelling salesman problem (TSP) is a very famous NP-hard problem in operations research as well as in computer science. To solve the problem several genetic algorithms (GAs) are developed which depend primarily on crossover operator. The crossover operators are classified as distance-based crossover operators and blind crossover operators. The distance-based crossover operators use distances between nodes to generate the offspring(s), whereas blind crossover operators are independent of any kind of information of the problem, except follow the problem's constraints. Selecting better crossover operator can lead to successful GA. Several crossover operators are available in the literature for the TSP, but most of them are not leading good GA. In this study, we propose reverse greedy sequential constructive crossover (RGSCX) and then comprehensive sequential constructive crossover (CSCX) for developing better GAs for solving the TSP. The usefulness of our proposed crossover operators is shown by comparing with some distance-based crossover operators on some TSPLIB instances. It can be concluded from the comparative study that our proposed operator CSCX is the best crossover in this study for the TSP.

**Keywords**—Genetic algorithm; reverse greedy sequential constructive crossover; comprehensive sequential constructive crossover; travelling salesman problem; NP-hard

## I. INTRODUCTION

The travelling Salesman Problem (TSP) is an old and famous combinatorial optimization problem in computer science and operations research which was documented in 1759 by Euler, of course not by that name, whose aim was to solve the Knights' tour problem. A solution of the problem is a knight visit of each of the 64-squares of a chessboard exactly once in its tour. In 1932, the term 'travelling salesman' was first used in a German book 'The travelling salesman, how and what he should do to get commissions and be successful in his business', written by a veteran travelling salesman. In 1948, the problem was formally introduced by RAND Corporation. The problem became popular by the Corporation's reputation and then by introducing linear programming for solving combinatorial optimization problems. The problem aims to obtain a least cost Hamiltonian cycle in a network of nodes [1]. The TSP has applications in automatic drilling of printed circuit boards and circuits, very-large-scale-integrated (VLSI) circuit, X-ray crystallography, movement of people, computer wiring [2]. The problem can be defined as follows:

A network of  $n$  nodes, with 'node 1' as 'depot' and a travel cost (or distance, or travel time etc.,) matrix  $C = [c_{ij}]$  of order  $n$  related to ordered pairs  $(i, j)$  of nodes is given. The aim of the problem is to obtain a least cost Hamiltonian cycle. Symmetric and asymmetric TSPs are two cases of TSP. It is symmetric if  $c_{ij} = c_{ji}$ , for all  $i, j$ ; else, asymmetric. For asymmetric case, there are possibly  $(n-1)!$  solutions out of which at least one provides the minimum cost, and for symmetric case, there are likely  $\frac{(n-1)!}{2}$  solutions along with

same cost opposite cyclic permutations in  $n$ -node network. The possible number of solutions is very huge in both cases and the problem is NP-Hard [3]. The problem is very famous, and it is researched by many researchers because of its difficulty and its usability to model many other difficult real-life problems.

In the literature for the TSP, numerous exact and heuristic methods have been proposed, and most of them are heuristic/metaheuristic methods. Genetic algorithm, tabu search, artificial neural network, ant colony optimization, particle swarm optimization, etc. are some of the effective metaheuristic algorithms [1]. Among them genetic algorithm is seen to be one of the best metaheuristic algorithms for the problem [4].

Genetic algorithm (GA) is a search process that is inspired by natural biological evolution process. It was proposed by John Holland in 1970s [5]. It starts initially with a population of strings, called chromosomes, that encode solutions to a problem, and operates probably three operators - selection, crossover and mutation, to produce new and probably better populations in successive generations. Crossover operator is the leading operator in GAs [1]. Mutation enlarges search space as well as protects GAs from damaging any genetic element resulted from crossover and selection operators. Among three operators, crossover is the primary operator, and hence, numerous crossover operators have been suggested as well as improved for finding better solution to the TSP [5].

In this present study, we propose first reverse greedy sequential constructive crossover (RGSCX) and then comprehensive sequential constructive crossover (CSCX) by combining greedy sequential constructive crossover (GSCX) [6] with RGSCX for the TSP. We then applied the proposed operators manually on two chromosomes to produce offspring(s). Finally, the usefulness of our proposed crossover

operators is shown by comparing with some distance-based crossover operators on some TSPLIB instances. One can conclude from the comparative study that our proposed operator CSCX is the best crossover operator in this study for the TSP.

This paper is organized as follows: A survey on related work for the TSP is reported in Section II. Section III develops proposed crossover operators for the problem, whereas, Section IV reports computational experiments for six crossover operators. Finally, Section V presents conclusion and future works.

## II. RELATED WORK

There are numerous crossover operators suggested for solving the TSP. Two kinds of crossover operators have been suggested for the problem, namely distance-based crossover and blind crossover. In distance-based crossover operators, offspring are created using distance between nodes, whereas, in blind crossover operators, offspring are not created using any information of the problem, rather they bother only about the problem's constraints. The partially mapped crossover (PMX) [7], ordered crossover (OX)[8], order based crossover (OBX) and position based operator (PBX) [9], alternating edges crossover (AEX) [10], cycle crossover (CX) [11], edge recombination crossover (ERX) [12], generalized N-point crossover (GNX) [13], etc. are some of the blind crossover operators, and greedy crossover (GX) [10], heuristic crossover (HX) [14], distance preserving crossover (DPX) [15], sequential constructive crossover (SCX) [4], etc. are some of the popular distance-based crossover operators.

Selecting better crossover operator can lead to successful GA. Though several crossover operators are available in the literature for the TSP, but most of them are not leading good GA. So, our aim is to develop distance-based crossover operators and then compare them with some existing crossover operators for showing the effectiveness of the proposed operators. Osaba et al. [16] pointed that comparisons between heuristic/metaheuristic methods using blind operators and heuristic methods using optimizing functions must be avoided. Otherwise, comparison would not be reliable, because nature of the methods is different. If someone wants to verify the quality of any distance-based crossover operator, suppose SCX, the results found by the operator should be compared against the one found by another distance-based crossover operator, such as HX, GX or DPX. Now, if performance of SCX is compared against other blind operators, such as the PMX, OX, CX, ERX or AEX, the comparison will not be fair. Since, we propose to modify SCX, hence modified SCXs will be compared with only some distance-based crossovers such as HX, GX and SCX. So, we are going to explain some of these operators only through an example of pair of parent chromosomes. To represent solution by chromosome, the path representation that lists the permutation of node is considered here. For example, let {1, 2, 3, 4, 5, 6, 7, 8, 9} is the list of node labels for a 9-node problem example, then the tour {1→7→6→3→8→9→2→4→5→1} is represented by (1, 7, 6, 3, 8, 9, 2, 4, 5). The objective function cost (value) is the total cost of all edges in this tour.

### A. Modified Heuristic Crossover Operator

Liepins et al. [17] proposed a modified HX for the TSP that initially selects a node and copy it to the offspring then its nearest node is copied to the offspring until the offspring is complete. Jog et al. [18] described a modified heuristic crossover (MHX) of the HX [17] that creates an offspring chromosome from two parent chromosomes as follows. Select a node randomly as the starting node for offspring chromosome. Compare two arcs leaving the starting node in both parents and select the cheaper one. Continue to copy the cheaper arc of two arcs in both parents into the offspring. If the cheaper parent arc introduces a cycle (repeating any node) into offspring (illegal chromosome), check whether the other parent arc introduces a cycle. If the second arc does not introduce a cycle, copy this arc into the offspring, otherwise, copy the cheaper arc into the offspring from a group of maximum 20 randomly selected arcs that do not introduce a cycle. Continue until a complete offspring is created and the first parent chromosome is replaced by this offspring. This operator generates an offspring using two parent chromosomes.

We illustrate the MHX using the 9-node problem example shown in Table I as the cost matrix. Suppose parent chromosomes P1: (1, 2, 3, 4, 6, 9, 5, 7, 8) and P2: (1, 3, 5, 7, 8, 9, 4, 2, 6) with costs 83 and 75 respectively are selected as parent chromosomes. We use these same chromosomes to illustrate all crossover operators. Also, as we fixed 'node 1' as depot node (first gene), so, we always start the procedure for all crossover operators from the 'node 1'.

Since the procedure starts from the node 1, so, initially the offspring is (1). The arcs in both parents going out from the node 1 are considered, i.e. 1→2 and 1→3 with costs 7 and 15 respectively. So, node 2 is added and the incomplete offspring becomes: (1, 2). Next, the parent arcs leaving the node 2 are considered, i.e. 2→3 and 2→6 with costs 8 and 3 respectively. Among them 2→6 is cheaper, So, node 6 is added and the incomplete offspring becomes: (1, 2, 6). Then the arcs 6→9 and 6→1 must be considered, but the second one is illegal it leads to already visited node, so, 6→9 is chosen which leads to the incomplete offspring: (1, 2, 6, 9). Continuing in this way, one can obtain a complete offspring as: (1, 2, 6, 9, 4, 8, 5, 7, 3) having cost 56.

TABLE I. THE COST MATRIX

Node	1	2	3	4	5	6	7	8	9
1	999	7	15	9	10	6	8	9	10
2	11	999	8	7	11	3	6	4	3
3	15	5	999	16	12	5	8	13	4
4	2	5	11	999	9	13	14	4	2
5	8	6	3	5	999	6	7	10	9
6	6	13	8	11	5	999	5	4	5
7	5	15	3	7	12	6	999	8	9
8	9	3	9	14	3	11	8	999	10
9	11	16	3	9	10	7	9	10	999

### B. Very Greedy Crossover Operator

Julstrom [19] proposed a greedy extension of HX [18], named very greedy crossover (VGX), for solving the TSP, which is as follows. It selects a starting node randomly, then constructs a tour by investigating the four arcs leaving that node in both parents, according to their costs. It copies the first arc that does not make a cycle. If all four parent arcs from any node make cycles, then it uses the cheapest arc to an unvisited node. If the parents share an edge from the current node to one unvisited node, the operator copies that arc, even if it is not cheapest. The VGX operator creates an offspring using two parents. We illustrate the VGX using same example shown above.

Initially, the offspring is (1). The arcs in both parents leaving the node 1 are considered, i.e. 1→2 and 1→8 with costs 7 and 9 respectively in  $P_1$  and 1→3 and 1→6 with costs 15 and 6 respectively in  $P_2$ . So, node 6 is copied that gives the incomplete offspring as: (1, 6). Next, the parent arcs leaving the node 6 are considered, i.e. 6→9, 6→4, 6→1 and 6→2 having their respective costs 5, 11, 6 and 13. Among them 6→9 is cheaper, So, node 9 is copied that gives an incomplete offspring as: (1, 6, 9). Continuing in this way, one can obtain a complete offspring as: (1, 6, 9, 4, 2, 3, 5, 7, 8) having cost 69.

### C. Adaptive Sequential Constructive Crossover Operator

The sequential constructive crossover (SCX) is proposed in [4] for solving the TSP which found very good solution for symmetric and asymmetric TSPLIB instances. In [20], SCX is modified as follows: after current node, if no any legitimate node is available in any parent, then it searches from the starting of the parent chromosome and the first legitimate node is selected as next node. In [21], a comparative study is reported that shows that SCX is the best among eight crossover operators. A modified SCX, named bidirectional circular SCX (BCSCX) is developed in [22]. In [23], an adaptive SCX (ASCX) is proposed that creates an offspring adaptively, either in forward or in backward or in mixed direction that depends on next node's cost. Hence, eight neighbour nodes of any current node is considered, four for each of the two nodes (genes).

Since in a chromosome number of genes is  $n$ , the 'node 1' is selected as the first as well the  $(n+1)^{th}$  genes. The algorithm for the ASCX is stated as follows [23].

Step 1: Start from the first gene, 'node 1' (i.e., current node  $p=1$  in position  $i=1$ ) in forward direction and from the  $(n+1)^{th}$  gene, 'node 1' (it is not shown in the chromosome), (i.e., current node  $q=1$  in position  $j=n+1$ ) in backward direction.

Step 2: Sequentially search both parent chromosomes in right direction and consider the first 'legitimate node' (the node that is not yet visited) appeared after 'node  $p$ ' in each parent. If no 'legitimate node' after 'node  $p$ ' is present in any of the parents, search sequentially from the starting of the parent (wrap around) and consider the first 'legitimate node'. Suppose the 'node  $\alpha$ ' and the 'node  $\beta$ ' are found in 1<sup>st</sup> and 2<sup>nd</sup> parent respectively. Go to Step 3.

Step 3: Sequentially search both parent chromosomes in left direction and consider the first 'legitimate node' appeared after 'node  $p$ ' in each parent. If no 'legitimate node' after 'node

$p$ ' is present in any of the parents, search sequentially from the end of the parent (wrap around) and consider the first 'legitimate node'. Suppose the 'node  $\gamma$ ' and the 'node  $\delta$ ' are found in 1<sup>st</sup> and 2<sup>nd</sup> parent respectively. Now, suppose among four nodes, 'node  $u$ ' is the cheapest with cost  $s=\min. \{c_{p\alpha}, c_{p\beta}, c_{p\gamma}, c_{p\delta}\}$ . Go to Step 4.

Step 4: Sequentially search both parent chromosomes in left direction and consider the first 'legitimate node' appeared after 'node  $q$ ' in each parent. If no 'legitimate node' after 'node  $q$ ' is present in any of the parents, search sequentially from the end of the parent (wrap around) and consider the first 'legitimate node'. Suppose the 'node  $w$ ' and the 'node  $x$ ' are found in 1<sup>st</sup> and 2<sup>nd</sup> parent respectively. Go to Step 5.

Step 5: Sequentially search both parent chromosomes in right direction and consider the first 'legitimate node' appeared after 'node  $q$ ' in each parent. If no 'legitimate node' after 'node  $q$ ' is present in any of the parents, search sequentially from the beginning of the parent (wrap around) and consider the first 'legitimate node'. Suppose the 'node  $y$ ' and the 'node  $z$ ' are found in 1<sup>st</sup> and 2<sup>nd</sup> parent respectively. Now, suppose among four nodes, 'node  $v$ ' is the cheapest with cost  $t=\min. \{c_{wq}, c_{xq}, c_{yq}, c_{zq}\}$ . Now, for selecting the next node as well as adding it in a position in the offspring chromosome go to Step 6.

Step 6: If  $s \leq t$ , then add 'node  $u$ ' in position ' $i$ ' in the partially constructed offspring chromosome and set  $p=u, i=i+1$ . Otherwise, add 'node  $v$ ' in position ' $j$ ' in the partially constructed offspring chromosome and set  $q=v, j=j-1$ . Now, If the offspring is a complete chromosome, then stop, otherwise, go to Step 2.

We illustrate the ASCX using same example shown above. As number of genes in the chromosomes is 9, the 'node 1' is the first as well as the 10<sup>th</sup> gene (not displayed in the chromosomes). After 'node 1' (first gene), the legitimate nodes in  $P_1$  in forward direction is 2 and in backward direction (after wrapping around) is 8, and in  $P_2$  they are 3 and (after wrapping around) 6, having their respective costs 7, 9, 15 and 6. So, the cheapest is node 6 having cost 6. From the end, before 'node 1' (10<sup>th</sup> gene), the legitimate nodes in  $P_1$ , in backward direction is 8 and in forward direction (after wrapping around) is 2, and in  $P_2$  they are 6 and (after wrapping around) 3, having their respective costs 9, 7, 6 and 15. So, the cheapest is node 6 having cost 6. As both cheapest nodes are 6, it is added as the second gene in the current offspring that leads the incomplete offspring to (1, 6, \*, \*, \*, \*, \*, \*, \*).

After 'node 6' (second gene), the legitimate nodes in  $P_1$  in forward direction is 9 and in backward direction is 4, and in  $P_2$  they are (after wrapping around) 3 and 2, having their respective costs 5, 11, 8 and 13. So, the cheapest is node 9 having cost 5. From the end, before 'node 1' (10<sup>th</sup> gene), the legitimate nodes in  $P_1$ , in backward direction is 8 and in forward direction (after wrapping around) is 2, and in  $P_2$  they are 2 and (after wrapping around) 3, having their respective costs 9, 11, 11 and 15. So, the cheapest is node 8 having cost 9. As node 9 is cheaper between the cheapest nodes, it is added as the third gene in the current offspring that leads the incomplete offspring to (1, 6, 9, \*, \*, \*, \*, \*, \*). Continuing in this way, one can obtain a complete offspring as: (1, 6, 9, 4, 8, 2, 3, 5, 7) having cost 59.

#### D. Greedy Sequential Constructive Crossover Operator

Recently, Ahmed [6] proposed the greedy SCX (GSCX) by introducing a greedy method, which is as follows.

Step 1: Start from 'node 1' (i.e., current node  $p = 1$ ).

Step 2: Sequentially search both parent chromosomes and consider the first 'legitimate node' (the node that is not yet visited) appeared after 'node  $p$ ' in each parent. If 'legitimate node' after 'node  $p$ ' is found in both parents, then go to Step 3, otherwise, consider the cheapest 'legitimate node' from the group of remaining legitimate nodes and concatenate it to the partially constructed offspring chromosome. If the offspring is a complete chromosome, then stop, otherwise, rename this present node as 'node  $p$ ' and repeat this Step 2

Step 3: Suppose the 'node  $\alpha$ ' and the 'node  $\beta$ ' are found in 1<sup>st</sup> and 2<sup>nd</sup> parent respectively, then for selecting the next node go to Step 4.

Step 4: If  $c_{pa} < c_{pb}$ , then select 'node  $\alpha$ ', otherwise, 'node  $\beta$ ' as the next node and concatenate it to the partially constructed offspring chromosome. If the offspring is a complete chromosome, then stop, otherwise, rename the present node as 'node  $p$ ' and go to Step 2.

We illustrate the GSCX using same example shown above. As 'node 1' is the first gene, after this node, the legitimate nodes in  $P_1$  is 2 and in  $P_2$  is 3 having  $c_{12}=7$  and  $c_{13}=15$ . As  $c_{12} < c_{13}$ , the node 2 is added as the second gene in the current offspring that leads the incomplete offspring to (1, 2).

After 'node 2', the legitimate nodes in  $P_1$  is 3 and in  $P_2$  is 6 having  $c_{23}=8$  and  $c_{26}=3$ . As  $c_{26} < c_{23}$ , the node 6 is added as the third gene in the current offspring that leads the incomplete offspring to (1, 2, 6).

After 'node 6', the legitimate nodes in  $P_1$  is 9 and in  $P_2$  is nothing. So, we search and find the cheapest legitimate node as 8, which is added as the fourth gene in the current offspring that leads the incomplete offspring to (1, 2, 6, 8). Continuing in this way, one can obtain a complete offspring as: (1, 2, 6, 8, 5, 7, 3, 9, 4) having cost 42.

### III. PROPOSED CROSSOVER OPERATORS

We propose two crossover operators - reverse greedy sequential constructive crossover operator and comprehensive sequential constructive crossover operator.

#### A. Reverse Greedy Sequential Constructive Crossover Operator

In this proposed operator, we apply the GSCX in reverse direction and we name it as reverse GSCX (RGSCX). We construct the offspring in reverse direction, that is, from the last node (gene) of the offspring back to the first node (gene) of the same. So, we define RGSCX as follows.

Step 1: Suppose the 'node  $\alpha$ ' and the 'node  $\beta$ ' are the last nodes in 1<sup>st</sup> and 2<sup>nd</sup> parent respectively. Since 'node 1' is the first node (gene), then for selecting the last node, we check whether  $c_{a1} < c_{\beta 1}$ . If yes, then select 'node  $\alpha$ ', otherwise, 'node  $\beta$ ' as the last node and concatenate it to the partially constructed offspring chromosome. Then rename this present node as 'node  $p$ ' and go to Step 2.

Step 2: Sequentially search both parent chromosomes in reverse direction and consider the first 'legitimate node' (the node that is not yet visited) appeared before 'node  $p$ ' in each parent. If 'legitimate node' before 'node  $p$ ' is found in both parents, then go to Step 3, otherwise, consider the cheapest 'legitimate node' from the group of remaining legitimate nodes and concatenate it to the partially constructed offspring chromosome. If the offspring is a complete chromosome, then stop, otherwise, rename this present node as 'node  $p$ ' and repeat this Step 2.

Step 3: Suppose the 'node  $\alpha$ ' and the 'node  $\beta$ ' are found in 1<sup>st</sup> and 2<sup>nd</sup> parent respectively, then for selecting the previous node go to Step 4.

Step 4: If  $c_{op} < c_{pp}$ , then select 'node  $\alpha$ ', otherwise, 'node  $\beta$ ' as the previous node and concatenate it to the partially constructed offspring chromosome. If the offspring is a complete chromosome, then stop, otherwise, rename this present node as 'node  $p$ ' and go to Step 2.

We illustrate the RGSCX using same example shown above. By default, the 10<sup>th</sup> node is 1. The last nodes (9<sup>th</sup> genes) are 8 and 6 in  $P_1$  and  $P_2$  respectively having  $c_{81}=9$  and  $c_{61}=6$ . As  $c_{61} < c_{81}$ , the node 6 is considered as the 9<sup>th</sup> gene that initiated the incomplete offspring as (6).

Before 'node 6', the legitimate nodes in  $P_1$  is 4 and in  $P_2$  is 2 having  $c_{46}=13$  and  $c_{26}=3$ . As  $c_{26} < c_{46}$ , the node 2 is added as the 8<sup>th</sup> gene in the current offspring that leads the incomplete offspring to (2, 6).

Before 'node 2', the legitimate node in  $P_1$  is nothing. So, we search and find the cheapest legitimate node as 8, which is added as the 7<sup>th</sup> gene in the current offspring that leads the incomplete offspring to (8, 2, 6).

Before 'node 8', the legitimate nodes in both  $P_1$  and in  $P_2$  are node 7, so it is added as the 6<sup>th</sup> gene in the current offspring that leads the incomplete offspring to (7, 8, 2, 6).

Also, before 'node 7', the legitimate nodes in both  $P_1$  and in  $P_2$  are node 5, so it is added as the fifth gene in the current offspring that leads the incomplete offspring to (5, 7, 8, 2, 6).

Before 'node 5', the legitimate nodes in  $P_1$  is 9 and in  $P_2$  is 3 having  $c_{95}=10$  and  $c_{35}=12$ . As  $c_{95} < c_{35}$ , the node 9 is added as the fourth gene in the current offspring that leads the incomplete offspring to (9, 5, 7, 8, 2, 6).

Before 'node 9', the legitimate nodes in  $P_1$  is 4 and in  $P_2$  is 3 having  $c_{49}=2$  and  $c_{39}=4$ . As  $c_{49} < c_{39}$ , the node 4 is added as the third gene in the current offspring that leads the incomplete offspring to (4, 9, 5, 7, 8, 2, 6). Continuing in this way, one can obtain a complete offspring as: (1, 3, 4, 9, 5, 7, 8, 2, 6) having cost 70.

#### B. Comprehensive Sequential Constructive Crossover Operator

We propose a comprehensive SCX (CSCX) by combining two crossover operators GSCX and RGSCX that produces two offspring. So, by using above example parents, it produces both offspring (1, 2, 6, 8, 5, 7, 3, 9, 4) and (1, 3, 4, 9, 5, 7, 8, 2, 6) with cost 42 and 70 respectively which are less than costs of both the parent chromosomes.

Our GA is non-hybrid, simple, which uses basic GA processes and operators, but does not incorporate any other heuristic algorithm. In our simple GA, starting with random chromosome population, good chromosomes are selected by stochastic remainder selection technique, then population passes through one selected crossover operator and swap mutation operator. Our simple GA may be designed as follows.

```
SimpleGA ()  
{ Initialize random population of size Ps;  
  Evaluate the population;  
  Generation = 0;  
  While stopping condition is not satisfied  
  { Generation = Generation + 1;  
    Select good chromosomes by selection operator;  
    Select a crossover operator and do crossover with crossover  
      probability Pc;  
    Do swap mutation with mutation probability Pm;  
    Evaluate the population;  
  }  
}
```

#### IV. COMPUTATIONAL EXPERIMENTS

The simple GAs using six crossover operators (MHX, VGX, ASCX, GSCX, RGSCX and CSCX) have been encoded in Visual C++. To compare the competence of these operators, simple GAs are applied on twenty eight TSPLIB instances [24] and then executed on a Laptop with specification i3-3217U CPU@1.80 GHz and 4 GB RAM under MS Windows 7. Among the twenty eight problem instances, instances ftv33, ftv35, ftv38, p43, ftv44, ftv47, ry48p, ft53, ftv55, ftv64, ft70, ftv70, kro124p, ftv170, rbg323, rbg358, rbg403 and rbg443 are asymmetric, and instances gr21, fri26, bayg29, dantzig42, eil51, berlin52, pr76, lin105, d198 and a280 are symmetric. For all simple GAs, the parameters are set as follows: 50 is population size, 1.0 is crossover probability, 0.20 is mutation probability, and 1,000 is maximum generations that is set as the stopping condition. For each instance, the experiments were repeated 50 times. Figures 1 shows results for rbg443 (considering only 100 generations) by all simple GAs. Each graph is for one crossover operator that shows the improvement of the solution as the number of generations increases. In the figure, the label on the left margin denotes the percentage of excess (Excess (%)) to the best known solution reported in TSPLIB website, which is calculated by the formula.

$$Excess (%) = \frac{Solution\ Obtained - Best\ Known\ Solution}{Best\ Known\ Sol} \times 100.$$

It is seen for the Figure 1 that MHX has some deviations, but not the best. GSCX has limited deviation but gets stuck very quickly in local minimum. Though ASCX and CSCX have less deviations and they are competing, however CSCX finds best results.

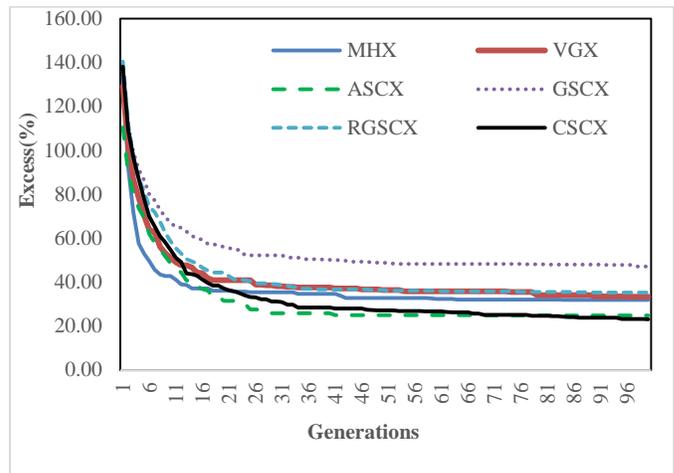


Fig. 1. Excess(%) by GAs using different Crossover Operators.

We summarize the results of our experiments using six crossover operators in Tables II and IV. We have organized the tables as follows: a row corresponds to the summarized results for a problem instance using variant GAs, first column reports a problem instance and its best-known solution (within brackets), second column reports the size of the instance, third column reports title of the summarized results and remaining each column is for GA using the mentioned crossover operator. The result using each crossover operator is designated by its best solution cost (Best Sol), average solution cost (Avg. Sol), percentage of excess of average solution to the best-known solution (Avg. Exc(%)), standard deviation of solution costs (S.D.), and average convergence time (Avg. Time) (in seconds). The best result over these six crossover operators for an instance is marked by bold face.

The Table II reports results using the GAs for the asymmetric instances. The crossover operators MHX and GSCX could not obtain either lowest best solution cost or average solution cost for any asymmetric instance. The crossover operators VGX and RGSCX obtain lowest best solution cost for the instance p43, whereas CSCX obtains lowest best solution costs for remaining seventeen instances. The crossover operator ASCX obtains lowest average solution cost for the instance p43 with lowest S.D., whereas CSCX obtains lowest average solution cost with lowest S.D. for remaining seventeen instances. By looking at average of Avg. Exc (%), one can make rank of the crossover operators. Accordingly, CSCX produce the best results, while ASCX is the second best, MHX and RGSCX are competing for the third best, and VGX is the worst. The results are also depicted in Figure 2, which also demonstrates the usefulness of our proposed crossover CSCX.

TABLE II. RESULTS BY THE CROSSOVER OPERATORS FOR ASYMMETRIC TSPLIB INSTANCES

Instance	n	Results	MHX	VGX	ASCX	GSCX	RGSCX	CSCX
ftv33 (1286)	34	Best Sol	1376	1404	1371	1380	1396	<b>1341</b>
		Avg. Sol	1479.56	1501.18	1394.72	1458.48	1464.16	<b>1382.86</b>
		Avg. Exc (%)	15.05	16.73	8.45	13.41	13.85	7.53
		S.D.	50.71	39.65	2.85	47.24	43.98	14.42
		Avg. Time	0.18	0.09	0.18	0.05	0.08	0.10
ftv35 (1473)	36	Best Sol	1520	1543	1586	1531	1583	<b>1499</b>
		Avg. Sol	1623.68	1649.28	1657.08	1631.32	1705.56	<b>1551.44</b>
		Avg. Exc (%)	10.23	11.97	12.50	10.75	15.79	5.33
		S.D.	55.93	46.01	31.98	47.09	56.93	32.17
		Avg. Time	0.18	0.15	0.07	0.05	0.08	0.17
ftv38 (1530)	39	Best Sol	1604	1618	1679	1613	1672	<b>1550</b>
		Avg. Sol	1678.08	1714.04	1748.64	1690.50	1722.22	<b>1605.72</b>
		Avg. Exc (%)	9.68	12.03	14.29	10.49	12.56	4.95
		S.D.	40.17	44.37	25.88	39.65	59.15	32.38
		Avg. Time	0.23	0.13	0.12	0.05	0.10	0.22
p43 (5620)	43	Best Sol	5631	<b>5625</b>	5631	5631	<b>5625</b>	5627
		Avg. Sol	5640.96	5636.52	<b>5635.70</b>	5641.20	5640.18	5639.30
		Avg. Exc (%)	0.37	0.29	0.28	0.38	0.36	0.34
		S.D.	5.91	5.35	1.96	6.97	8.73	7.09
		Avg. Time	0.29	0.34	0.37	0.17	0.14	0.27
ftv44 (1613)	45	Best Sol	1725	1686	1733	1706	1627	<b>1613</b>
		Avg. Sol	1843.98	1863.24	1796.12	1853.28	1793.12	<b>1669.48</b>
		Avg. Exc (%)	14.32	15.51	11.35	14.90	11.17	3.50
		S.D.	59.21	68.81	28.08	60.57	71.07	37.33
		Avg. Time	0.41	0.37	0.73	0.12	0.21	0.19
ftv47 (1776)	48	Best Sol	1860	1902	2054	1864	1919	<b>1833</b>
		Avg. Sol	2046.38	2065.80	2111.08	2021.72	2102.62	<b>1936.26</b>
		Avg. Exc (%)	15.22	16.32	18.87	13.84	18.39	9.02
		S.D.	84.14	91.77	21.98	70.92	69.51	43.12
		Avg. Time	0.57	0.46	0.74	0.26	0.23	0.27
ry48p (14422)	48	Best Sol	15629	15204	15290	15469	15293	<b>14983</b>
		Avg. Sol	16120.54	16062.36	15744.88	16150.78	15664.34	<b>15479.70</b>
		Avg. Exc (%)	11.78	11.37	9.17	11.99	8.61	7.33
		S.D.	287.99	306.73	200.95	278.65	186.04	278.44
		Avg. Time	0.45	0.36	0.50	0.17	0.18	0.13
ft53 (6905)	53	Best Sol	8061	7899	7631	7882	7973	<b>7486</b>
		Avg. Sol	8617.90	8529.00	8127.34	8614.86	8427.90	<b>7816.04</b>
		Avg. Exc (%)	24.81	23.52	17.70	24.76	22.06	13.19
		S.D.	278.54	291.14	156.51	277.13	224.61	194.86
		Avg. Time	0.49	0.59	0.38	0.35	0.31	0.34
ftv55 (1608)	56	Best Sol	1773	1753	1749	1723	1705	<b>1639</b>
		Avg. Sol	1872.60	1846.58	1798.86	1841.82	1773.14	<b>1712.58</b>
		Avg. Exc (%)	16.46	14.84	11.87	14.54	10.27	6.50
		S.D.	56.81	60.22	20.58	50.89	48.93	39.74
		Avg. Time	0.56	0.54	1.05	0.36	0.34	0.28
ftv64 (1839)	65	Best Sol	2010	2079	2145	1990	1999	<b>1879</b>
		Avg. Sol	2196.10	2228.14	2236.32	2140.28	2178.94	<b>1921.62</b>
		Avg. Exc (%)	19.42	21.16	21.61	16.38	18.49	4.49
		S.D.	85.06	71.79	44.42	76.32	89.40	39.74
		Avg. Time	0.66	0.90	1.01	0.29	0.53	0.49
ft70 (38673)	70	Best Sol	40976	40926	41592	41129	41445	<b>40050</b>
		Avg. Sol	42208.68	42210.74	42447.92	42185.60	42283.14	<b>41080.98</b>
		Avg. Exc (%)	9.14	9.15	9.76	9.08	9.34	6.23
		S.D.	517.87	480.29	292.36	571.46	442.49	376.74
		Avg. Time	0.99	0.91	1.80	0.64	0.72	0.53
ftv70 (1950)	71	Best Sol	2145	2154	2276	2118	2068	<b>1975</b>
		Avg. Sol	2326.06	2350.86	2332.22	2296.32	2294.38	<b>2065.54</b>
		Avg. Exc (%)	19.29	20.56	19.60	17.76	17.66	5.93
		S.D.	80.56	90.05	41.54	69.83	91.22	59.63
		Avg. Time	0.76	1.08	1.16	0.48	0.63	0.68
kro124p (36230)	100	Best Sol	41199	41764	41246	41251	40956	<b>38432</b>
		Avg. Sol	43371.14	43167.84	42471.12	42829.16	42967.98	<b>40303.68</b>
		Avg. Exc (%)	19.71	19.15	17.23	18.21	18.60	11.24
		S.D.	990.53	781.36	462.23	780.6	1057.48	877.77
		Avg. Time	0.94	1.12	0.58	0.42	0.92	0.79

(CONTD.) RESULTS BY THE CROSSOVER OPERATORS FOR ASYMMETRIC TSPLIB INSTANCES

Instance	n	Results	MHX	VGX	ASCX	GSCX	RGSCX	CSCX
ftv170 (2755)	171	Best Sol	3303	3551	3232	3656	3517	<b>2968</b>
		Avg. Sol	3607.68	3835.46	3393	3799.50	3767.42	<b>3178.74</b>
		Avg. Exc (%)	30.95	39.22	23.16	37.91	36.75	15.38
		S.D.	130.91	159.74	95.42	130.15	214.40	79.22
		Avg. Time	1.79	4.05	0.93	1.74	2.74	2.99
rbg323 (1326)	323	Best Sol	1553	1558	1611	1597	1617	<b>1400</b>
		Avg. Sol	1594.20	1644.42	1618.8	1677.12	1677.76	<b>1443.04</b>
		Avg. Exc (%)	20.23	24.01	22.08	26.48	26.53	8.83
		S.D.	19.15	28.06	17.7	34.19	28.94	15.82
		Avg. Time	13.45	20.24	23.53	15.76	17.71	24.28
rbg358 (1163)	358	Best Sol	1481	1495	1327	1522	1514	<b>1325</b>
		Avg. Sol	1541.4	1555.36	1387.92	1591.04	1650.14	<b>1373.36</b>
		Avg. Exc (%)	32.54	33.74	19.34	36.80	41.89	18.09
		S.D.	26.05	27.79	24.05	36.93	48.64	22.04
		Avg. Time	8.86	25.67	30.77	18.16	24.36	19.5
rbg403 (2465)	403	Best Sol	3033	3104	2922	3149	2833	<b>2636</b>
		Avg. Sol	3110	3172.66	2983.38	3214.22	2980.9	<b>2704.58</b>
		Avg. Exc (%)	26.17	28.71	21.03	30.39	20.93	9.72
		S.D.	38.51	35.02	21.43	42.23	46.77	30.76
		Avg. Time	42.8	33.66	38.93	28.96	32.33	27.67
rbg443 (2720)	443	Best Sol	3399	3517	3252	3573	3188	<b>2932</b>
		Avg. Sol	3498.92	3604.94	3321.58	3678.86	3329.06	<b>2993.72</b>
		Avg. Exc (%)	28.64	32.53	22.12	35.25	22.39	10.06
		S.D.	41.53	33.71	20.95	48.04	60.93	28.09
		Avg. Time	53.28	45.97	50.82	30.78	38.33	33.07
<b>Average of Avg. Exc (%)</b>			<b>18.00</b>	<b>19.49</b>	<b>15.58</b>	<b>19.07</b>	<b>18.09</b>	<b>8.20</b>

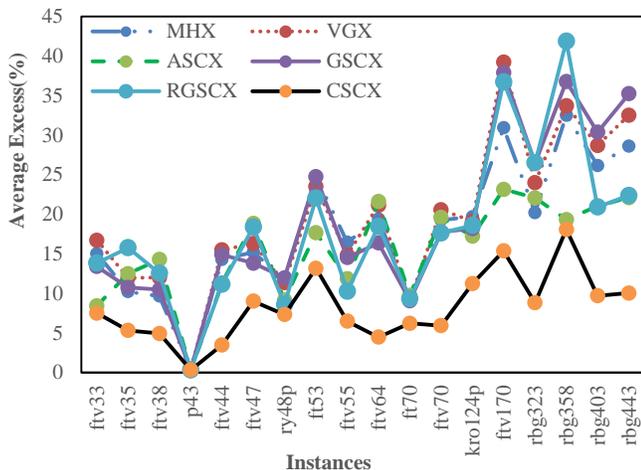


Fig. 2. Average Excess(%) by different GAs for Asymmetric Instances.

To validate the above observations, we also carried out an adequate statistical analysis. By considering that reported results in Table II are random and independent samples, a set of Student's t-tests were conducted. Indeed, for every pair of crossover operator, the hypothesis is tested whether one of the operators is better than the other. The efficiency of an operator is categorized by its average of average excess (%) computed over the all problem instances with best-known solutions.

The results of our hypotheses testing are summarized in Table III. In the table, each row contains two columns, where the first lists a crossover operator and the second column lists

its inferior crossover operators. The results are statistically significant at the significance level 0.05 [25]. In Table III, each crossover is ranked according to its number of inferior crossover operators. It is found that there is statistically significant difference between CSCX and other crossover operators at level 0.05, and so, as expected the best ranked crossover is CSCX. Also, the second best is ASCX. No significant difference is found between RGSCX and MHX, as expected, they share the third rank. Also, no significant difference is found between GSCX and VGX, and hence, they share the worst rank.

The Table IV reports results by the GAs for the symmetric TSPLIB instances. The crossover MHX and ASCX obtain lowest best solution cost only for two instances - gr21 and fri26; VGX obtains lowest best solution cost for three instances - dantzig42, eil51 and lin105; GSCX and RGSCX obtain lowest best solution cost for three instances - gr21, fri26 and bayg29; and CSCX obtains lowest best solution cost for six instances - gr21, fri26, berlin52, pr76, pr226 and a280. The crossover operators MHX, VGX, ASCX and GSCX could not obtain lowest average solution cost for any symmetric instance. The proposed RGSCX obtains lowest average solution cost with lowest S.D. for the only one instance bayg29, and CSCX finds lowest average costs along with lower S.D. for remaining 9 instances. So, our proposed crossover CSCX is the best. Also, by looking at average of Avg. Exc (%), one can make it clear that CSCX produces best results, while VGX, RGSCX and GSCX are competing for the second best, and MHX is the worst. The results are also depicted in Figure 3, which also demonstrates the usefulness of our proposed crossover CSCX.

TABLE III. RESULTS OF STATISTICAL HYPOTHESES TESTING ON ASYMMETRIC INSTANCES

Crossover	Inferior crossovers
CSCX	MHX, VGX, ASCX, GSCX, RGSCX
ASCX	MHX, VGX, GSCX, RGSCX
RGSCX	VGX, GSCX
MHX	VGX, GSCX
GSCX	-----
VGX	-----

TABLE IV. RESULTS BY THE CROSSOVER OPERATORS FOR SYMMETRIC TSPLIB INSTANCES

Instance	n	Results	MHX	VGX	ASCX	GSCX	RGSCX	CSCX
gr21 (2707)	21	Best Sol	<b>2707</b>	2754	<b>2707</b>	<b>2707</b>	<b>2707</b>	<b>2707</b>
		Avg. Sol	2874.78	2927.70	2825.2	2845.28	2829.58	<b>2806.88</b>
		Avg. Exc(%)	6.20	8.15	4.37	5.11	4.53	3.69
		S.D.	109.00	89.03	61.29	86.33	87.07	55.14
		Avg. Time	0.03	0.02	0.04	0.01	0.02	0.04
fri26 (937)	26	Best Sol	<b>937</b>	953	<b>937</b>	<b>937</b>	<b>937</b>	<b>937</b>
		Avg. Sol	989.30	987.30	944.04	972.62	969.62	<b>937.00</b>
		Avg. Exc(%)	5.58	5.37	0.75	3.80	3.48	0.00
		S.D.	29.86	24.51	11.27	17.72	20.05	0.00
		Avg. Time	0.08	0.03	0.13	0.02	0.03	0.04
bayg29 (1610)	29	Best Sol	1642	1646	1686	<b>1634</b>	<b>1634</b>	1639
		Avg. Sol	1741.78	1767.18	1756.52	1719.42	<b>1718.42</b>	1719.54
		Avg. Exc(%)	8.19	9.76	9.10	6.80	6.73	6.80
		S.D.	69.49	55.81	39.51	56.77	49.15	47.16
		Avg. Time	0.08	0.04	0.02	0.03	0.05	0.1
dantzig42 (699)	42	Best Sol	753	<b>714</b>	754	723	724	723
		Avg. Sol	807.44	788.60	813.36	781.80	792.60	<b>774.26</b>
		Avg. Exc(%)	15.51	12.82	16.36	11.85	13.39	10.77
		S.D.	29.64	33.92	22.3	26.27	26.83	23.89
		Avg. Time	0.35	0.24	0.29	0.08	0.14	0.02
eil51 (426)	51	Best Sol	444	<b>432</b>	444	436	442	437
		Avg. Sol	470.76	463.76	466.66	463.94	462.40	<b>458.78</b>
		Avg. Exc(%)	10.51	8.86	9.54	8.91	8.54	7.69
		S.D.	13.31	10.64	8.24	12.01	10.81	9.96
		Avg. Time	0.46	0.49	0.90	0.38	0.37	0.30
berlin52 (7542)	52	Best Sol	7885	7919	7910	7926	7891	<b>7646</b>
		Avg. Sol	8346.32	8217.40	8429.48	8156.70	8162.72	<b>7995.52</b>
		Avg. Exc(%)	10.66	8.96	11.77	8.15	8.23	6.01
		S.D.	247.93	224.64	154.44	189.50	237.35	137.83
		Avg. Time	0.62	0.39	0.70	0.22	0.29	0.22
pr76 (108159)	76	Best Sol	118331	117411	120729	116844	117724	<b>113676</b>
		Avg. Sol	129036.50	123427.82	128392.50	127293.76	127868.12	<b>123337.58</b>
		Avg. Exc(%)	19.30	14.12	18.71	17.69	18.22	14.03
		S.D.	4517.92	3464.34	2791.62	5273.80	5243.46	4268.20
		Avg. Time	0.65	0.78	1.05	0.48	0.53	0.18
lin105 (14379)	105	Best Sol	16369	<b>15245</b>	15978	15921	15627	15622
		Avg. Sol	17635.34	16616.80	16920.84	17118.08	17068.94	<b>16575.84</b>
		Avg. Exc(%)	22.65	15.56	17.68	19.05	18.71	15.28
		S.D.	772.10	635.92	320.38	557.25	637.83	413.79
		Avg. Time	1.34	1.79	1.97	1.10	1.03	0.23
pr226 (80369)	226	Best Sol	93260	88768	91723	92428	90724	<b>87477</b>
		Avg. Sol	103462.00	93261.56	95511.32	95315.32	94946.68	<b>90411.98</b>
		Avg. Exc(%)	28.73	16.04	18.84	18.60	18.14	12.50
		S.D.	6228.71	2996.66	1326.34	4740.22	5643.82	1395.34
		Avg. Time	4.70	5.05	5.09	3.18	3.30	2.02
a280 (2579)	280	Best Sol	3022	2905	3059	2980	2891	<b>2833</b>
		Avg. Sol	3197.00	3059.62	3179.67	3111.00	3094.10	<b>2958.34</b>
		Avg. Exc(%)	23.96	18.64	23.29	20.63	19.97	14.71
		S.D.	69.14	72.46	59.45	99.19	106.69	53.85
		Avg. Time	2.80	4.07	12.56	3.34	4.80	8.08
<b>Average of Avg. Exc (%)</b>			<b>15.13</b>	<b>11.83</b>	<b>13.04</b>	<b>12.06</b>	<b>11.99</b>	<b>9.15</b>

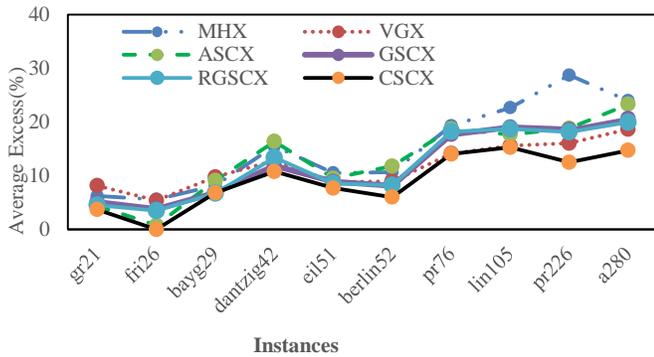


Fig. 3. Average Excess(%) by different GAs for Symmetric Instances.

To validate the above observations, we carried out statistical analysis for these instances also, and the results are summarized in Table V. It is seen that there is statistically significant difference between CSCX and other crossover operators at the significance level 0.05, and so, as expected the best ranked crossover is CSCX. However, there is no significant difference found among VGX, RGSCX, GSCX and ASCX, so, they share the second rank, and MHX is the worst in the rank. From this whole study one can conclude that the proposed crossover CSCX is the best.

TABLE V. RESULTS OF STATISTICAL HYPOTHESES TESTING ON SYMMETRIC INSTANCES

Crossover	Inferior crossovers
CSCX	MHX, VGX, ASCX, GSCX, RGSCX
VGX	MHX
RGSCX	MHX
GSCX	MHX
ASCX	MHX

### V. CONCLUSION AND FUTURE WORKS

The crossover operators are classified as distance-based crossover operators and blind crossover operators. There are several crossover operators available in the literature. In this study, we proposed reverse greedy sequential constructive crossover (RGSCX) and then comprehensive sequential constructive crossover (CSCX) for the TSP. To show the usefulness of our proposed crossover operators, we compared with four distance-based crossover operators, such as MHX, VGX, ASCX and GSCX. We applied these crossover operators manually on two chromosomes to produce offspring(s) and found that our proposed crossover CSCX is the best. After that, GAs using all six crossover operators are developed and performed comparative study among them on eighteen asymmetric and ten symmetric TSPLIB instances. In terms of solution quality, it is found that our proposed crossover CSCX is the best. The observation is confirmed by Student's t-test at the significance level 0.05. So, CSCX might be worked good for other associated combinatorial optimization problems. However, the proposed RGSCX could not obtain good solutions and it is competing for third position on asymmetric instances.

In this study, our aim was to propose crossover operators and compare them against some existing crossover operators. It

was not aimed to improve solution quality using them, and so, no local search procedure was used for developing state-of-art algorithm for the problem. Also, highest crossover probability was used to display the exact characteristics of operators. Though our proposed CSCX obtains best solutions, still it gets stuck in local minima in the first half of the generations. Hence, good local search along with immigration methods [26-30] may be incorporated to it to develop hybrid genetic algorithm to find better quality solutions to the problem instances, which is under our investigation.

### ACKNOWLEDGMENT

The author is very much thankful to the honourable anonymous reviewers for their constructive comments and constructive suggestions which helped the author to improve this paper.

### REFERENCES

- [1] Z. Michalewicz, "Genetic Algorithms + Data Structures = Evolution Problems," Second Edition, Springer-Verlag, New York, 1994.
- [2] C.P. Ravikumar, "Solving large-scale travelling salesperson problems on parallel machines," *Microprocessors and Microsystems*, vol. 16, no. 3, pp. 149-158, 1992.
- [3] S. Arora, "Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems," *Journal of ACM*, vol. 45, no. 5, pp. 753-782, 1998.
- [4] Z.H. Ahmed, "Genetic algorithm for the traveling salesman problem using sequential constructive crossover operator," *International Journal of Biometrics & Bioinformatics*, vol. 3, pp. 96-105, 2010.
- [5] D.E. Goldberg, "Genetic algorithms in search, optimization, and machine learning," Addison-Wesley, New York, 1989.
- [6] Z.H. Ahmed, "Solving the traveling salesman problem using greedy sequential constructive crossover in a genetic algorithm," *IJCSNS International Journal of Computer Science and Network Security*, vol. 20, no. 2, pp. 99-112, 2020.
- [7] D.E. Goldberg, and R. Lingle, "Alleles, loci and the travelling salesman problem," In J.J. Grefenstette (ed.) *Proceedings of the 1st International Conference on Genetic Algorithms and Their Applications*. Lawrence Erlbaum Associates, Hilldale, NJ, 1985.
- [8] L. Davis, "Job-shop scheduling with genetic algorithms," *Proceedings of an International Conference on Genetic Algorithms and Their Applications*, pp. 136-140, 1985.
- [9] G. Syswerda, "Schedule optimization using genetic algorithms," In Davis, L. (ed.) *Handbook of Genetic Algorithms*, New York: Van Nostrand Reinhold, pp. 332-349, 1991.
- [10] J. Grefenstette, R. Gopal, B. Rosmaita, and D. Gucht, "Genetic algorithms for the traveling salesman problem," In *Proceedings of the First International Conference on Genetic Algorithms and Their Applications*, (J. J. Grefenstette, Ed.), Lawrence Erlbaum Associates, Mahwah NJ, pp. 160-168, 1985.
- [11] I.M. Oliver, D. J. Smith and J.R.C. Holland, "A study of permutation crossover operators on the travelling salesman problem," In J.J. Grefenstette (ed.). *Genetic Algorithms and Their Applications: Proceedings of the 2nd International Conference on Genetic Algorithms*. Lawrence Erlbaum Associates, Hilldale, NJ, 1987.
- [12] D. Whitley, T. Starkweather and D. Shaner, "The traveling salesman and sequence scheduling: quality solutions using genetic edge recombination," In L. Davis (Ed.) *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, New York, pp. 350-372, 1991.
- [13] N.J. Radcliffe and P.D. Surry, "Formae and variance of fitness," In D. Whitley and M. Vose (Eds.) *Foundations of Genetic Algorithms 3*, Morgan Kaufmann, San Mateo, CA, pp. 51-72, 1995.
- [14] J. J. Grefenstette, "Incorporating problem specific knowledge into genetic algorithms," In L. Davis (Ed.), *Genetic algorithms and simulated annealing*, London, UK: Pitman / Pearson, pp. 42-60, 1987.

- [15] B. Freisleben and P. Merz, "A genetic local search algorithm for solving symmetric and asymmetric traveling salesman problems," in Proc. the 1996 IEEE International Conference on Evolutionary Computation, (Nagoya, Japan), pp.616-621, 1996.
- [16] E. Osaba, R. Carballedo, F. Diaz, E. Onieva, A.D. Masegosa, and A.Perallos, "Good practice proposal for the implementation, presentation, and comparison of metaheuristics for solving routing problems," *Neurocomputing*, vol. 271, no. 3, pp. 2-8, 2018.
- [17] G. E. Liepins, M. R. Hilliard, M. Palmer and M. Morrow, "Greedy genetics," in Proc. the Second International Conference on Genetic algorithms and their application, pp.90-99, October 1987.
- [18] P. Jog, J.Y. Suh and D. Van Gucht, "The effects of population size, heuristic crossover and local improvement on a genetic algorithm for the traveling salesman problem," in Proc. 3rd Int. Conf. Genetic Algorithms, Morgan Kaufmann Publishers, pp. 110–115, 1989.
- [19] B.A. Julstrom, "Very greedy crossover in a genetic algorithm for the traveling salesman problem," In: ACM symposium on Applied computing, pp. 324-328, 1995.
- [20] Z.H. Ahmed, "Improved genetic algorithms for the traveling salesman problem," *International Journal of Process Management and Benchmarking*, vol. 4, no. 1, pp. 109-124, 2014.
- [21] I. H. Khan, "Assessing different crossover operators for travelling salesman problem," *IJISA International Journal of Intelligent Systems and Applications*, vol. 7, no. 11, pp. 19-25, 2015.
- [22] S. Kang, S.-S. Kim, J.-H. Won, and Y.-M. Kang, "Bidirectional constructive crossover for evolutionary approach to travelling salesman problem," 2015 5th IEEE International Conference on IT Convergence and Security (ICITCS), pp. 1-4, 2015.
- [23] Z.H. Ahmed, "Adaptive sequential constructive crossover operator in a genetic algorithm for solving the traveling salesman problem," *IJACSA International Journal of Advanced Computer Science and Applications*, vol. 11, no. 2, pp. 593-605, 2020.
- [24] G. Reinelt, TSPLIB, <http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/>
- [25] M. Spiegel and L. Stephens, "Schaum's Outline of Statistics," Fourth Edition, McGraw-Hill, New York, 2011.
- [26] Z.H. Ahmed, "A hybrid genetic algorithm for the bottleneck traveling salesman problem," *ACM Transactions on Embedded Computing Systems*, vol. 12, Art. No. 9, 2013.
- [27] Z.H. Ahmed, "An experimental study of a hybrid genetic algorithm for the maximum travelling salesman problem," *Mathematical Sciences*, vol. 7, pp. 1-7, 2013.
- [28] Z.H. Ahmed, "The ordered clustered travelling salesman problem: A hybrid genetic algorithm," *The Scientific World Journal*, vol. 2014, Art ID 258207, 13 pages, 2014.
- [29] Z.H. Ahmed, "The minimum latency problem: a hybrid genetic algorithm," *IJCSNS International Journal of Computer Science and Network Security*, vol. 18, no. 11, pp. 153-158, 2018.
- [30] Z.H. Ahmed, "Performance analysis of hybrid genetic algorithms for the generalized assignment problem," *IJCSNS International Journal of Computer Science and Network Security*, vol. 19, no. 9, pp. 216-222, 2019.

# A Workflow Scheduling Algorithm for Reducing Data Transfers in Cloud IaaS

Jean Edgard GNIMASSOUN<sup>1</sup>,  
Souleymane OUMTANAGA<sup>4</sup>  
Laboratoire de Recherche en  
Informatique et Télécommunication  
Institut National Polytechnique-HB  
Yamoussoukro, Côte d'Ivoire

Tchimou N'TAKPE<sup>2</sup>  
Laboratoire de Mathématiques et  
Informatique  
Université Nangui Abrogoua  
Abidjan, Côte d'Ivoire

Gokou Hervé Fabrice DIEDIE<sup>3</sup>  
Laboratoire de Recherche en  
Mathématiques et Informatique  
Université Peleforo Gon Coulibaly  
Korhogo, Côte d'Ivoire

**Abstract**—The cloud IaaS easily offers to have homogeneous multi-core machines (whether they are "bare metal" machines or virtual machines). On each of these machines, there can be high-performance input-output SSD disks. That allows to distribute the files produced during the execution of the workflow to different machines in order to minimize the additional costs associated with transferring these files. In this paper, we propose a scheduling algorithm called WSRDT (Workflow Scheduling Reducing Data Transfers) whose purpose is to minimize the makespan (execution time) of data-intensive workflows by reducing transfers data between dependent tasks on the network. Intermediate files produced by tasks are stored locally on the disk of the machine where the tasks were executed. We experimentally verify that the increase in the number of cores per machine reduces the additional cost due to data transfers on the network. Experiences with a veritable workflow show those advantages of the algorithms presented. Data-driven scheduling significantly reduces the execution time and the volume of data transferred on the network, our approach outperforms one of the best state-of-the-art algorithms that we have adapted with our hypotheses.

**Keywords**—Workflow scheduling; makespan reduction; multi-cores virtual machine; data-intensive workflows; IaaS cloud

## I. INTRODUCTION

Scientists, to run their different parallel applications, generally used clusters and grids computing. These different execution platforms quickly have showed their limits giving the ever-increasing demands for computing, storage resources, and so on. To solve this issue, cloud computing offers an illusion of infinite resources where scientists can request the resources needed to run a parallel application. Cloud computing typically offers three (03) types of services, SaaS (Software as a Service), PaaS (Platform as a Service), IaaS (Infrastructure as a Service). The use of these different services is flexible and scalable from the request of the user, via a pay-as-you-go model. With three (03) basic services, the most suitable for running parallel applications is the IaaS cloud. The providers of this service offer computing and storage resources essential for running all parallel applications that require a significant resource due to its complex structure.

Parallel applications come from several research fields such as biology, astronomy, physics, agriculture, etc., and have in common, on the one hand, their complex structure with

dependencies between the different tasks, and on the other hand a need for high computing and storage service, given the large volume of data to be processed and transferred. These scientific applications are very often modeled as scientific workflow. These scientific workflows require a High-Performance Computing (HPC) environment for their execution.

The evolution of the computing environment from grid to cloud computing has always considered scientific workflows. However, with this new paradigm, scientific workflows are now executed on virtual, dynamic and scalable resources as an instance in cloud computing. The challenge of mapping workflow tasks, which is a task scheduling problem in a cloud computing environment, is the subject of several scientific studies to find algorithms to execute workflows in a reasonable time and budget. This problem of scheduling on IaaS infrastructures of cloud computing is known as NP-hard [1], for this purpose, several heuristics [2][3][4] and metaheuristics [5][6] have been proposed in the literature in order to minimize either the total execution time of the workflow (makespan), the cost of using IaaS resources in the cloud, or both.

Solutions for workflow scheduling on cloud IaaS infrastructures exist, but these algorithms generally consider the execution of a task on a VM with a single computational core. And depending on the complex structure of the workflows, this could lead to several data transfers (communications) in networks thus constituted. According to the Amazon EC2's VM deployment model, users will be able to order VMs with a maximum of ninety-six (96) parallel computing cores in the same VM<sup>1</sup>. However, minimizing the makespan of a workflow, one must consider in addition to the execution time of the task on the computing resource, the time of transfer from a task to its successor(s), because large volumes of data are must be transferred. Using multi-core VMs for scheduling could give better results for the makespan, as it could reduce the amount of data exchanged in the network. Indeed, if two dependent tasks running on the same VM, the communication time (between these two tasks) is assumed to be zero.

Most of the algorithms in the literature are not clairvoyant, i.e., they do not consider the location of data coming from

<sup>1</sup><https://aws.amazon.com/fr/ec2/instance-types/m5/>

predecessors and data going to successors of a task. In order to improve the execution time of a workflow, the execution time of each task in the workflow and the data transfer time between dependent tasks must be considered. The works in the literature consider only the data coming from the predecessor tasks, i.e. from top to bottom, since the scientific application is modeled as a DAG. However, an improvement can be done on the location of the data. Moreover, the algorithms in the literature do not exploit multi-core machines for the simultaneous execution of several tasks in the same machine and the distributed storage of the data produced during the execution of the workflow.

Reducing the execution time of a scientific application means considering the execution time of each task of the application, but also the file transfer time between the different dependent tasks through the network. The main problem addressed in this paper is how to do a good mapping of the different tasks from a data-intensive application by reducing the files to be transferred in the network in order to obtain a better execution time.

The remainder of this paper is organized as follows. Section II introduces the related work in this field and section III present the platform and application models. Sections IV and V describes the proposed approach: WSRDT and section V validates the effectiveness WSRDT. Concluding remarks are given in section VI.

## II. RELATED WORK

Two main approaches exist for scheduling tasks in the cloud, which are list scheduling algorithms [7][8][9] and clustering algorithms [10][11][12]. Most of the list scheduling algorithms are inspired by HEFT [13], which aims to minimize the makespan and was originally proposed for computational grid environment and has long been studied and adapted for the cloud environment.

In this section present a review of the literature on algorithms whose objective is to minimize makespan. Running a scientific workflow application in the cloud requires efficient mapping so that tasks do not have to wait too long, which could result in a very long execution time. Typically, the resources provided in the cloud to run workflows are VMs with computational units, storage, etc. Reducing the execution time of a workflow consisting of hundreds or even thousands of tasks in the cloud is a challenge, given the flexibility of available resources. The key part of resource management in a cloud environment is the mapping of these tasks to these on-demand computing resources. Rimal et al. [14] propose a model based on the public cloud (Amazon EC2), in order to minimize the makespan, the proposed algorithm is based on the principle of critical path (Critical Path: CP). Critical path tasks are assigned to different resources in the cloud, and to maximize the use of these leased resources, other tasks that are not part of the critical path are assigned to those resources already leased, taking into account the billing that is done per unit of time; knowing that a VM used during 01H01mn would be charged for 02H. This approach based on the "multi-tenant cloud" consists in deploying the tasks of the same CP at the cloud provider whose resources allow to finish these different tasks at the earliest possible time in order to reduce the

completion time of the workflow. The study of Rimal et al. showed that their approach gives better results compared to the FCFS algorithm, which is not a clear-sighted algorithm because it does not take into account all the dependencies that would exist between the different tasks of the workflow. In addition to the critical path approach, Gamal et al. [10] propose a new approach based on task classification. Their task clustering approach, where groupings are done according to a certain neighborhood, minimizes the makespan based on the Min-Min [9] algorithm for mapping tasks to cloud resources. Min-Min algorithm can be used in cloud computing. Min-Min algorithm depends on execution time for scheduling tasks. Tasks with minimum execution time will be scheduled first. Tasks with long execution time have high delay. The Min-Min algorithm is not suitable for running a parallel application where the tasks are dependent just like the FCFS algorithm. Almi'ani and Lee [15] proposed a three-step approach to minimizing the makespan: (i) the partitioning step; in this step, the number of tasks assigned to each partition is first determined taking into consideration the execution time for the CP in the workflow. Since the sum of task execution times along CP (i.e., critical path length) represents the lower bound of makespan (i.e., the optimal solution), this step tries to ensure the total execution time for tasks belonging to the same partition to be less than CP length. While critical path length only includes execution times as tasks along CP are meant to be assigned to the same resource, the length of tasks in any other partition should include execution times and communication times. To ensure that partitions created in the partitioning step are at the optimal granularity for the final resource allocation, tasks of different partitions are (ii) rearranged/adjusted. The optimality here primarily concerns the number of tasks in each partition with respect to the capacity of potentially assigned resource and data locality. As partitions are expected to have dependency relationships due to task precedence constraints primarily dictated by data dependencies, rearranging tasks between different partitions involves the recalculation of timing values. To execute each task the (iii) resource assignment step consists of the resource set identification to identify types of resource set allocated to partitions such that the amount of time partitions are required to wait due to the presence of the data dependencies between partitions is minimized. To assess their approach Almi'ani et al. compared their approach to HEFT, but HEFT provided better makespan compared to their approach which gives better cost of using cloud resources. The most suitable algorithms for scheduling a parallel application where tasks are dependent are list algorithms because this type of algorithm takes into account the dependencies between all the tasks in the workflow.

The Heterogeneous Earliest Finish Time Algorithm (HEFT) is a popular list-based heuristic scheduling algorithm for optimizing the makespan [13] in workflow applications, whose pseudo-code is very close to algorithm 1. The method consists of two phases: ranking and mapping. In the ranking phase (line 1) based on the (1), the order in which the tasks are being mapped is computed using the bottom-level metric (distance of the beginning task to the end task of the workflow). The idea of this ranking is to execute before those tasks having more dependent tasks than others. Further details about how to sort the tasks can be found in [13]. Once the

execution order is determined, the second phase consists in assigning each task to the resources following the order computed in the first phase. For each task and for each resource, the completion time of that task on that resource is computed. Finally, the task is mapped onto the resource where it is finished earlier. After all tasks have been mapped, the workflow can be executed.

$$bl_i = \omega_i + \max_{j \in succ(i)} (c_{i,j} + bl_j) \quad (1)$$

Where  $succ(i)$  is the set of immediate successors of task  $v_i$ ,  $c_{i,j}$  is the data transfers time from task  $v_i$  to task  $v_j$ , and  $\omega_i$  is the execution time of task  $v_i$ . Since the *bottom-level* is computed recursively by traversing the DAG upward, starting from the end task. For the end task  $v_{end}$ , the *bottom-level* value is equal to.

$$bl_{end} = \omega_{end} \quad (2)$$

HEFT is a very popular list scheduling algorithm that aims at minimizing the makespan when resources are fixed, but HEFT does not perform well in minimizing the makespan of a workflow when the volumes of data exchanged between tasks are large. Indeed HEFT, after having sorted the tasks according to their priorities, tries to minimize the end date of execution of each task in the order of this list. It is therefore a ‘‘blind’’ algorithm through which the decision taken for a task is final and can have a negative impact on lower priority tasks. In addition, these algorithms use a naive adaptation of HEFT for Cloud IaaS platforms using a single centralized storage service for data exchanges between tasks. This significantly increases the additional cost of inter-task data exchange.

In all these studies, the VMs considered are heterogeneous and are in fact distinguished by their differences in terms of the number of cores. The authors therefore assume that the workflow tasks are parallel tasks that can run on any number of cores. However, in reality, the workflows on which they make their assessments are inspired by real workflows studied by Juve et al. [16]. In their study, Juve et al. show that almost all tasks in real workflows are single-core. There is only one task in one of the studied workflows that can use up to two cores. In this study, the workflows considered are therefore comprised of single-core tasks only.

All the studies in the literature do not take into account that one can take advantage of the rental of multi-core machines containing local storage disks in order to reduce the makespan by reducing data transfers. However, storing all the files used and produced by a workflow on a single central storage service can cause contention on the network. This study is based on the use of local VM disks to propose an algorithm to minimize makespan. The algorithm will also take advantage of the fact that the same multicore VM can be used to execute several tasks in parallel.

### III. PLATFORM AND APPLICATION MODELS

In this paper, the platform model is based on a typical IaaS cloud configuration. Multiple virtual machine (VM) instances are deployed on physical servers within a single datacenter. More precisely, a set of VMs like Amazon EC2 M5 instances is considered. Specifically, these are the M5d instances that are provided with local storage on the NVMe SSD, while regular M5 instances must rely on Amazon Elastic Block Storage (EBS) to store the data. Table I details the characteristics of the available M5d instances. The indicated costs in dollars per hour correspond to on-demand Linux instances in the US-East region (Ohio) at the time of writing of this article.

The number of virtual cores (vCPUs) in this instance series ranges from 2 to 96, with a constant amount of memory per core of 4GiB. These instances are typically deployed by Amazon on nodes featuring an Intel Xeon Platinum 8000 series processor. The specific feature of the M5d instances is to attach a fast block-level storage on SSD drives that is coupled to the lifetime of the instance. This work, aim at leveraging this fast storage that is shared by the vCPUs of an instance to store the intermediate files produced during the execution of a workflow, hence reducing the number of data transfer over the network for tasks scheduled on the same virtual machine. Only the entry and exit files of the workflow will be stored on an external storage node.

In terms of network connectivity with other instances or the Elastic Block Storage (EBS) service, the available bandwidth depends on the size of the instance. Only the largest instances that can exploit a full node, i.e., with 64 or 96 vCPUs, have a guaranteed network bandwidth of 20 and 25 Gbps respectively. For smaller instances, i.e., from 2 to 16 cores, the bandwidth is proportional to the vCPUs.

TABLE I. CHARACTERISTICS OF THE AWS M5D INSTANCE TYPES

Model	vCPU	Memory (GiB)	Instances Storage (GiB)	Network Bandwidth (Gbps)	EBS Bandwidth (Mbps)	Cost (\$/H)
M5d.large	2	8	1 x 75 NVMe SSD	Up to 10	Up to 3,500	0,113
M5d.xlarge	4	16	1 x 150 NVMe SSD	Up to 10	Up to 3,500	0,226
M5d.2xlarge	8	32	1 x 300 NVMe SSD	Up to 10	Up to 3,500	0,452
M5d.4xlarge	16	64	2 x 300 NVMe SSD	Up to 10	3,500	0,904
M5d.8xlarge	32	128	2 x 600 NVMe SSD	10	5,000	1,808
M5d.12xlarge	48	192	2 x 900 NVMe SSD	10	7,000	2,712
M5d.16xlarge	64	256	4 x 600 NVMe SSD	20	10,000	3,616
M5d.24xlarge	96	384	4 x 900 NVMe SSD	25	14,000	5,424

In this study, large VMs are preferred in each platform, because they allow multiple tasks to be executed in parallel. In their study Juve et al. [16] have shown that each task in a real scientific workflow is a single-core activity, i.e. can only be executed on a single computing core, rather than using all the cores of a VM. It is on this same principle that this study is based. Thus, a user who wants to run his parallel application in the cloud must rent a number of cores in total, the proposed approach provides him with a platform that would minimize the application completion time. For example, if the user wants to use 100 cores in total, the platform will consist of a VM with 96 cores and a VM with 4 cores. In the case of 200 cores total, the platform will consist of three VMs, two VMs of 96 cores and one of 8 cores and so on.

The scientific workflows (cf. Fig. 1) to schedule are represented by Directed Acyclic Graphs (DAGs)  $G = \{V, \mathcal{E}\}$  where  $V = \{v_i \mid i = 1, \dots, V\}$  is a set of vertices representing the computational tasks of the workflow and  $\mathcal{E} = \{e_{i,j} \mid (i,j) \in \{1, \dots, V\} \times \{1, \dots, V\}\}$  is a set of edges between vertices, representing either a data dependency, i.e., a file transfer, or a flow dependency between two tasks. Each of the task composing the workflow has a predefined (estimated) duration, requires a set of *input files* to start its execution, and will produce a set of *output files* upon completion.

Notations such as  $Input_i^k$  (resp.  $Output_i^k$ ), represent the  $k^{th}$  input (resp. output) file of a given task  $v_i$ . When an output file produced by a task  $v_i$  is consumed as input by another task  $v_j$ , this creates a data dependency between  $v_i$  and  $v_j$ , represented by the edge  $e_{i,j}$ .

The input files that are not produced by any of the tasks in the workflow are called the *entry files* of the workflow. Conversely, the output files that are not consumed by any task are called the *exit files* of the workflow. Finally, two quantities associated with each task of the workflow that will be used during the planning process have been defined. The *Local Input Volume* of task  $v_i$  on machine  $M_j$ , or  $LIV_{i,j}$ , as the sum of the size of the files that  $v_i$  takes as input that are locally stored on  $M_j$ . Respectively, the *Local Output Volume*, or  $LOV_{i,j}$  as the sum of the sizes of the files produced by  $v_i$  that are used by successors of  $v_i$  also scheduled on  $M_j$ .

Note that if a file is used by more than one successor, its size is accounted for as many times as successors. The  $LIV$  (resp.  $LOV$ ) of an entry (resp. exit) task is by definition set to zero. Bandwidth be proportional to the number of cores and equal to 208.33 Mbps per core. All the virtual machine instances started for the execution of a given workflow are connected through a single switch.

According to the description of the M5d instances, the connection from a VM to EBS goes through a dedicated network connection, which is taken into account in the simulated infrastructure. As for the network connections between VMs, One of the assumptions made in this study is that the bandwidth of the dedicated connection between VM and EBS is proportional to the number of cores for small VMs with up to 16 cores (i.e., 218.75 Mbps per core).

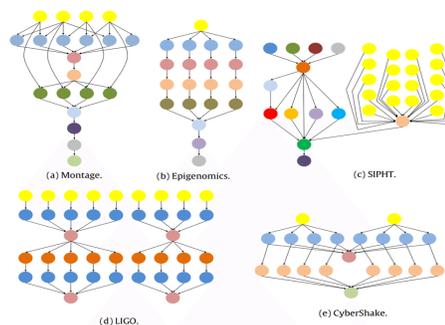


Fig. 1. Some Examples of Scientific Workflows.

During the execution of the workflow, all the intermediary files, i.e. those that are produced by a task and consumed by another, will be stored locally on the SSD storage of one or several machines. Only the entry and exit files of the workflow will be stored on an external storage service accessible by all the machines. The time to transfer a file from one machine to another includes the time to read the file on the disk of the source machine, the duration of the data transfer over the network and the time to write the file on disk at destination.

#### IV. A PLANNING ALGORITHM TO MINIMIZE DATA TRANSFER OVER THE NETWORK

The proposed planning algorithm aims at leveraging two main characteristics of the target IaaS cloud platform, i.e. multi-core instances and a fast-local storage space shared among cores, to minimize the impact of data transfers on the execution of data-intensive scientific workflows.

In this section, the assumption is that the provisioning of virtual machine instances has been done. Then, the objective of this algorithm is to schedule the set  $V$  of  $V$  tasks composing the workflow on a set  $M$  of  $n$  VMs instances. These instances can have different sizes. Each of them has a unique id, the largest instances having the smallest ids. How the set of instances is defined will be explained in Section V.

Algorithm 1 starts by building a sorted scheduling list that contains all the tasks of the workflow (lines 1-2). The tasks are sorted by decreasing *bottom level* value [13]. The bottom level of a task  $v_i$ , or  $bl_i$ , is the length of the longest path from  $v_i$  to the end of the workflow. This ordering gives the highest priorities to the most critical tasks and ensures the respect of the dependencies between tasks.

Then, the algorithm determines a first mapping for each task  $v_i$  in  $V$  (line3-7). The selected machine  $M_j$  in  $M$  is the one that first minimizes the start time of  $v_i$  (denotes as  $st_j(v_i)$ ) and then maximizes the volume of the input files needed by  $v_i$  for its execution that are already locally stored on  $M_j$ . The rationale is that between two virtual machines able to start  $v_i$  start its execution at the same time, the algorithm favor the one that minimizes the amount of data transfer over the network.

As all the considered virtual machine instances have multiple cores, scheduling a task  $v_i$  on a machine  $M$  implies to maintain a local schedule inside the virtual machine. In order to maximize the utilization of the cores within a virtual machine, each machine is managed as a job and resource manager will

do. In particular, this study leverage the available information on the (estimated) duration of each task to implement a conservative backfilling mechanism [17] when building the local schedule. Keeping such *usage profile* of a virtual machine up to date is mandatory to determine the time when a new task can start on this particular machine (i.e.  $st_j(v_i)$ ). Then, after selecting  $M$  of the execution of  $v_i$ , it is essential to update the *usage profile* of  $M$  (line 6). These usage profiles of the virtual machines are also used in second step of Algorithm 1 in which the tasks in this initial schedule are rearranged to further reduce the amount of data transfers over the network.

This rearrangement step (lines 8 to 11) browses the workflow DAG *level by level* from the bottom to the top. The motivation of this second step is that during the initial placement that proceeds from top to bottom, only the volume of data coming from the direct predecessors of a task is considered. It is indeed impossible to account for the locality of the data needed by the direct descendants of a task when scheduling it at their placement is not determined yet. This may lead to avoidable data movements.

Level 0 is the topmost level of the DAG that comprises all the entry tasks of the workflow. For each of the other tasks, the level is recursively computed as the maximum level of its predecessors plus one. Finally,  $L$  denote the number of levels in the workflow.

**Algorithm 1** Mapping workflow tasks without rearrangement

```
1   Compute  $bl_i$  of each task  $v_i$ 
2   Sort  $V$  by decreasing  $bl_i$  values
3   for all  $v_i \in V$  do
4      $M \leftarrow \{M_j \in M \mid st_j(v_i) \text{ is minimal and } LIV_{i,j} \text{ is maximal}\}$ 
5     Map  $v_i$  on  $M$ 
6     Update the usage profile of  $M$ 
7   end for
8   for  $l = L$  to 0 do
9      $V_l \leftarrow$  tasks in level  $l$  sorted by decreasing  $bl$  values
10    Rearrange ( $V_l$ ) ▶ see Algorithm 2
11  end for
```

The principle of the rearrangement step is described in Algorithm 2. It start by saving the current start time and mapping (denoted as  $st^c(v_i)$  and  $M^i$ ) for each task  $v_i$  in  $V_l$  (lines 2 and 3). Then, the local volume  $LIV_{i,j}$  for task  $v_i$  on machine  $M_j$  (lines 4 to 7) is determined. Also, the local volume for the current mapping of  $v_i$  (line 7) is saved before cancelling this mapping (line 8). This last action creates some idle slots in the usage profiles of different machines that can be used to improve data locality by “migrating” some tasks from one machine to another. The conditions to migrate a task  $v_i$  from its former mapping to a new mapping on  $M_k$  are that it would improve the data locality, i.e.  $LIV_{i,k} \geq LIV_{i,j}^c$ , and reduce the starting time of the task, i.e.  $st_k(v_i) \leq st^c(v_i)$ . Where  $st_k(v_i)$  is the new start time of  $v_i$  on  $M_k$ .

The main loop in Algorithm 2 (lines 11 to 32) aims at iteratively improving the mappings for tasks in  $V_l$ . At each step, the algorithm first try to find a better mapping (lines 15 to 21) for each task by considering the machine that leads to the greatest increase the local volume first. If the task can also start earlier on this machine, it is selected for a new tentative mapping.

There are three exit cases to this while loop: (i) there exists a better mapping for  $v_i$  on another machine  $M_j$ ; (ii)  $v_i$  has been remapped on the same machine  $M_i$  with a better or equal start time; or (iii) no better mapping was found.

**Algorithm 2** Rearrangement of tasks at level  $l$

```
1   for all  $v_i \in V_l$  do
2      $st^c(v_i) \leftarrow$  current start time of  $v_i$ 
3      $M^i \leftarrow$  current mapping of  $v_i$ 
4     for all  $M_j \in M$  do
5        $LIV_{i,j} \leftarrow LIV_{i,j} + LOV_{i,j}$ 
6     end for
7      $LIV_i^c \leftarrow$  current local volume of  $v_i$ 
8     cancel the current mapping of  $v_i$ 
9     end for
10    level_is_rearranged  $\leftarrow$  FALSE
11    while  $\neg$  level_is_rearranged do
12      level_is_rearranged  $\leftarrow$  TRUE
13      for all  $v_i \in V_l$  do
14        Sort  $M$  by decreasing  $LIV_{i,j}$  value
15        while  $LIV_{i,j} \geq LIV_i^c$  do
16          if  $st_j(v_i) \leq st^c(v_i)$  then
17            map  $v_i$  on  $M_j$ 
18            update the usage profile of  $M_j$ 
19          break
20        end if
21      end while
22      if  $v_i$  is mapped on  $M^i$  or
23         $st_{M^i}(v_i) > st^c(v_i)$  then ▶ no better mapping
24         $V_l \leftarrow V_l \setminus \{v_i\}$  ▶ mapping is definitive
25      level_is_rearranged  $\leftarrow$  FALSE
26    end if
27  end for
28  if  $\neg$  level_is_rearranged then
29    for all  $v_i \in V_l$  do
30      cancel the current mapping of  $v_i$ 
31    end for
32  end while
```

This last case means that a task with a higher priority has been mapped on  $M_i$  and  $st^c(v_i)$  can no longer be guaranteed. In both cases,  $v_i$  is set back to its original mapping, which becomes definitive (lines 22 to 26). However, this decision may invalidate some of the migrations (e.g., the task with higher priority mapped on  $M_i$ ). Then, all the tentative mappings determined in this step (lines 28 to 30) are cancelled and another rearrangement of the remaining tasks is searched. Algorithm 2 ends when only migration decisions are taken during the current step. The level is then considered as fully rearranged and the decided mappings become definitive.

## V. RESULTS AND DISCUSSION

To evaluate this approach with HEFT, a simulator based on the WRENCH project<sup>2</sup>[18] was wrote, a Cyber-Infrastructure simulation framework that provides high-level simulation abstractions for building accurate and scalable full-fledged simulators with minimal software development efforts. WRENCH is an open-source C++ library composed of two layers: the core simulation models and base abstractions (computing, communicating, storing) are provided by SimGrid [19][20] on top of which services to simulate the execution of

<sup>2</sup> <https://wrench-project.org>

computational workloads (compute services, storage services, network proximity services, data location services, etc.) are defined. By leveraging SimGrid's accurate models and their scalable implementations, WRENCH simulators can yield nearly identical behaviours when compared to actual systems.

#### A. Determining the Data Transferred through Simulation

The planning produced by Algorithms 1 and 2 minimizes the amount of data transferred over the network during the execution of the workflow. However, the quality of that planning strongly depends on the set of multi-core virtual machines that share a fast storage space given as input.

In Table II, VS is the size of VM used; mksp is the makespan; TVF is Total volume of files and VFT is the volume of file transferred. This table provides a detailed description of the set of files for the five literature's workflows and the volumes of file to be transferred if either Algorithm 1 (without rearrangement) or Algorithm 2 (with rearrangement) was applied. These transferred files are obtained after simulation on platforms where each VM has 16 or 96 cores. With platforms with 96 cores per VM, there's a slight difference in the volume of files transferred compared to platforms with 16 cores. On the other hand, compared to the total volume of each workflow, the rearrangement approach allows to transfer fewer files across the network. This is explained by the fact that before executing a task, algorithms 1 and 2 have to search on one hand for each task the VM on which its parent tasks have stored the maximum amount of files (because after its execution, each task stores locally i.e. on the VM where it has executed all the output files) and on the other hand, with rearrangement minimizing file transfers to the child tasks.

The use of large VMs allows several tasks to run in parallel and favours the execution of dependent tasks on the same VM, thus allowing negligible communication time between these tasks (i.e. for dependent tasks running on the same VM).

Table II shows that when the rearrangement approach is applied, 21.33%, 0.9% and 49.36% of files are respectively avoided being transferred for Epigenomics, CyberShake and Montage workflows. The platforms used in this study have a total number of cores of 384, 288 and 972 for the CyberShake, Epigenomics and Montage workflows, respectively. These platforms are multiples of 96 and are greater than the total number of cores used in parallel for each workflow.

To generate the different platforms and avoid wasting resources, i.e. avoid leasing resources that will not be used, the number of tasks that can be executed in parallel for each of the workflows is determined. To do this, the platform is oversized, i.e. this platform has as many cores as there are tasks in the workflow. In the case of the workflows used in this study, there are 1000 tasks, so the platform with 1000 cores is considered. This allowed to determine the total number of cores that could be used in parallel for each of the workflows. Thus, for CyberShake, 374 cores are used in parallel while for Epigenomics and Montage it is respectively 246 and 662 cores used in parallel. This preliminary study will be used in section V in order to determine the limit of platforms to be used for each workflow in experiments.

TABLE II. IMPACT OF REARRANGEMENT

wf	VS	Rearrangement		No Rearrangement		TVF
		mksp	VFT	mksp	VFT	
Cyb	16	593.83	382	1247.98	386	400.39
	96	311.03	315	351.1	317	
Epi	16	34227.2	1221.65	34231.6	1222.13	1230.93
	96	34330.4	1219.81	34314.2	1219.84	
Mont	16	380.312	10.56	380.89	10.81	17.32
	96	375.44	8.77	376.1	9.33	

#### B. Impact of Rearrangement's Step

To evaluate the contributions, three real-world scientific workflows from the five scientific workflows in the Pegasus Gallery<sup>3</sup> are used, as these three workflows (mentioned above) are data-intensive compared to the other two. These applications are:

- CyberShake: is an application of the Southern California Earthquake Center to characterize earthquake hazards;
- Epigenomics: is a data processing pipeline to automate the execution of various genome sequencing operations;
- Montage: is an astronomy application that creates custom mosaics of the sky from multiple images.

This assessment start by evaluating the impact of the size of the virtual machine on the execution time of the scientific workflow. For each workflow, we consider infrastructures where we vary (i.e. increase) the maximum number of cores per VM (from 2 cores to 96 cores maximum). For a total number of cores to be used, we generate platforms with 2 cores per VM, 4 cores per VM, ..., 96 cores per VM.

On Figs. 2, 3, and 4, platforms composed of 2 cores per VM provide poor execution times compared to platforms of 32 cores per VM and themselves provide poor execution times compared to platforms of 96 cores per VM. The increasing of the total number of cores per VM provides good execution times. It is for this reason that this study favor large VMs (i.e. VMs with several cores) because these VMs can execute several tasks in parallel and considerably reducing the makespan.

In Figs. 2, 3 and 4, max\_ft represents the minimum bound if there was no communication during the execution of the workflow. The execution time obtained on platforms with large VMs (i.e. having 96 cores) is approaching this theoretical minimum limit. Here, the rearrangement step (Algorithm 2) is applied to measure the performance of the initial offline planning.

Different behaviors for each of the three considered workflows are observed. First, the number of total cores used has almost no influence on the execution time for the CyberShake application while for Epigenomics, a plateau is

<sup>3</sup> [https://pegasus.isi.edu/workflow\\_gallery](https://pegasus.isi.edu/workflow_gallery)

observed from two total cores used. For Montage, the execution time decreases up to seven total cores used. This evolution of the execution time is directly related to the level of parallelism a workflow can exploit, i.e., how many tasks can be executed concurrently. Second, the execution time decreases when the size of the virtual machine instances grows, but that the improvement becomes very limited for sizes above 32. More interestingly, the CyberShake workflow which produces much more intermediate data than the two other workflows, relying on the local storage of small instances (i.e., with up to eight cores) leads to execution times worse than the solution with one core per VM where all the intermediate data are stored on the EBS service. This is because using too many small VMs on a single host (i.e., up to 48 instances with two cores) increases the number of data transfers between instances and cause contention on the network. Conversely, in the baseline configuration, each VM benefits of a dedicated network connection to the shared storage service.

Fig.5 shows the impact of the rearrangement step on the execution time. For the Montage workflow, the structure of the workflow is such that rearrangement has no influence on the offline planning hence neither on the execution time. For the two other workflows, rearranging the offline planning to further reduce the amount of transfers over the network can only improve the execution time.

In this study, two offline scheduling algorithms are proposed whose objective is to minimize makespan by reducing file transfers over the network. Algorithm 1 performs the mapping by only considering the files comes from the parent tasks. As for algorithm 2, it rearranges the mapping of algorithm 1 by considering the tasks level by level in order to improve the mapping of algorithm 1 by reducing the transfer of files to the child tasks.

After showing the gain of rearrangement on CyberShake and Epigenomics workflows, Fig. 6 shows more details of rearrangement on CyberShake. Approach with rearrangement gives better results compared to the approach without rearrangement. In the next part of this work we will compare this approach with the HEFT algorithm.

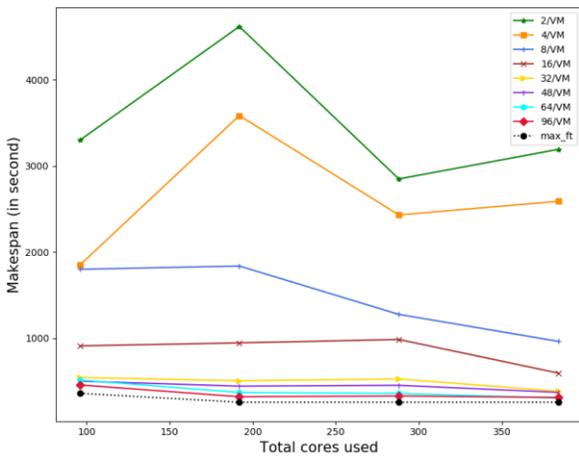


Fig. 2. Evolution of the Makespan of the CyberShake Workflow with Variation of the Total Number of Cores per VM.

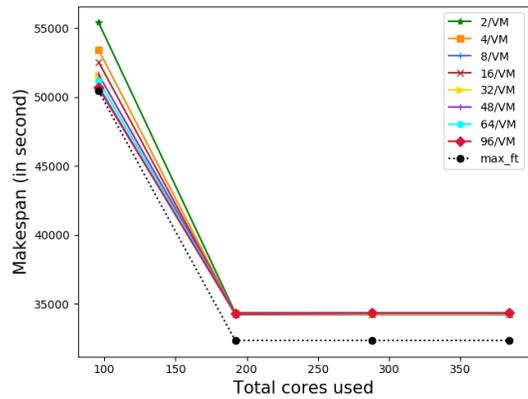


Fig. 3. Evolution of the Makespan of the Epigenomics Workflow with Variation of the Total Number of Cores per VM.

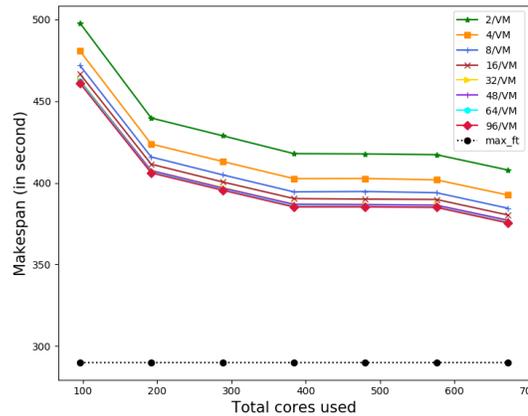


Fig. 4. Evolution of the Makespan of the Montage Workflow with Variation of the Total Number of Cores per VM.

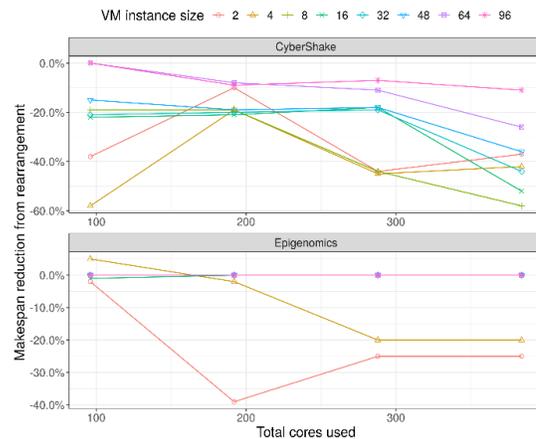


Fig. 5. Impact of Rearrangement Step on the Makespan (Execution Time) for different VM Instance Size, using Algorithm 1 vs Algorithm 2.

For data-intensive applications, such as the scientific workflow CyberShake it is essential to perform a good mapping of the different tasks, in order to have a good execution time. For this type of application, approach with rearrangement gives better results (cf. Fig.6). Since the use of large VMs has an impact on the makespan, the evaluation of both approaches is based on platforms whose total number of cores is a multiple of 96.

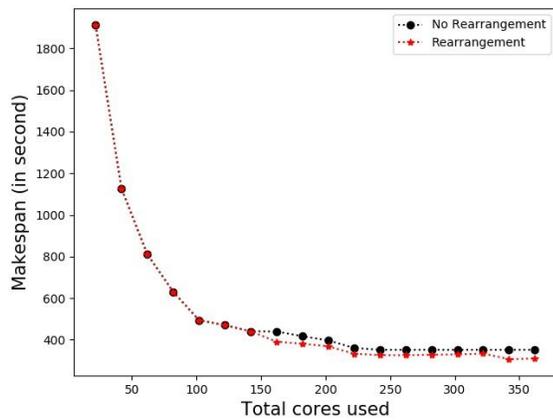


Fig. 6. Impact of Rearrangement on CyberShake Scientific Workflow.

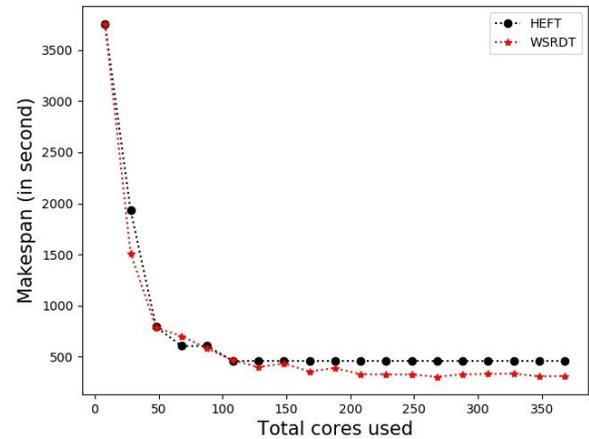


Fig. 7. Evaluation Results for the CyberShake Workflow.

### C. Comparison of HEFT and WSRDT Algorithms

The main objective of this study is to minimize the execution time of an application. To achieve this goal, the major contribution of this paper is the reduction of files to be transferred during the execution of the application, which have a considerable impact on the execution time. In order to evaluate the performance of this algorithm, the proposed approach is compared to HEFT which is a very popular heuristic in the scheduling of parallel applications. HEFT is adapted to the IaaS cloud resources and to the simulation environment. The results of simulations show that approach proposed provides good results compared to HEFT.

In the results of Figs.7, 8 and 9, we use platforms whose total number of cores used varies from 2 cores to 374 cores for CyberShake, from 2 to 246 cores for Epigenomics and up to 662 cores for Montage, by increments of 2. The principle of platform generation was explained in section III.

If there are more VMs in a platform, then there will be several files to transfer on the network. It is true that large VMs are prioritized for the execution of applications, but the important element for which this choice is do, is the use full bandwidth for this type of machine that we have summarized in Table I. With the HEFT algorithm, the more VMs there are in the platform, the more files will be transferred in the network. On the other hand, with approach proposed, files are reduced on two levels for each task, i.e. the reduction coming from the parent tasks (Algorithm 1) and the reduction going to the child tasks (rearrangement of the mapping of the Algorithm 2). These two approaches allow us to have good results compared to HEFT as is the case in Figures 7, 8 and 9.

These experiences allow to show that our approach provides better results summarized through the following gains. The CyberShake workflow allows to obtain a gain of 32.22% on the makespan compared to HEFT, as for the Epigenomics workflow, a gain of 44.54% on the makespan and a gain of 18.62% on the makespan with the workflow Montage are obtained.

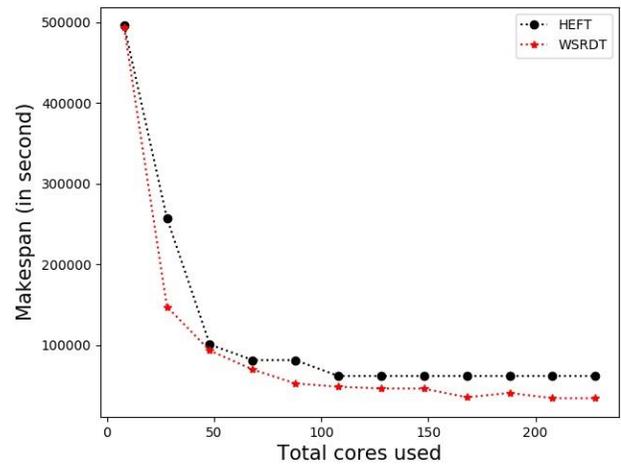


Fig. 8. Evaluation Results for the Epigenomics Workflow.

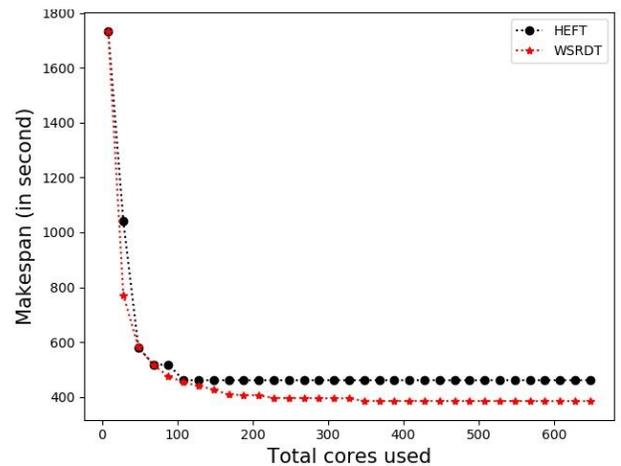


Fig. 9. Evaluation Results for the Montage Workflow.

## VI. CONCLUSION

Infrastructure as a Service Clouds now allows scientists to execute their data intensive workflows infrastructures that match the computing and storage requirements of these applications. Determining the set of virtual machine instances that have to compose these infrastructures is a complex task, usually delegated to Workflow Management Systems. A key to performance is to be able to leverage the characteristics of virtual machines instances.

In this paper, first showed the interest of using multi-core machines, because by increasing the number of cores per machine, several tasks are executed on this machine and the bandwidth linking a machine to the switch is proportional to the number of cores of the machine. Then we proposed scheduling algorithms that minimize the makespan by reducing data transfers between dependent tasks on the network. Finally, the results of the experiments showed that the proposed approach gives better results than HEFT, which is one of the best list-scheduling algorithms.

As part of our future work, we plan to compare the simulated executions with actual runs on the AWS computing cloud with M5d instances in order to confirm the impact of the proposed algorithms. We also plan to study the multi-objective aspect of the scheduling problem so that users can favor either a shorter execution time or a lowest cost by proposing a complementary approach where one of the objective is fixed, i.e., either a given budget or a fixed deadline.

## REFERENCES

- [1] J. K. Lenstra and A. H. G. Rinnooy Kan, "Complexity of Scheduling under Precedence Constraints," *Oper. Res.*, vol. 26, no. 1, pp. 22–35, 1978, doi: 10.1287/opre.26.1.22.
- [2] V. Arabnejad, K. Bubendorfer, B. Ng, and K. Chard, "A Deadline Constrained Critical Path Heuristic for Cost-Effectively Scheduling Workflows," in *Proc. IEEE/ACM 8th Int. Conf. Utility and Cloud Computing (UCC)*, 2015, pp. 242–250, doi: 10.1109/UCC.2015.41.
- [3] V. Arabnejad, K. Bubendorfer, and B. Ng, "A budget-aware algorithm for scheduling scientific workflows in cloud," *IEEE Int. Conf. High Perform. Comput. Commun.*, pp. 1188–1195, 2007.
- [4] P. Lu, G. Zhang, Z. Zhu, X. Zhou, J. Sun, and J. Zhou, "A Review of Cost and Makespan-Aware Workflow Scheduling in Clouds," *J. Circuits, Syst. Comput.*, p. 1930006, 2018.
- [5] X. Wang, B. Cao, C. Hou, L. Xiong, and J. Fan, "Scheduling budget constrained cloud workflows with particle swarm optimization," in *2015 IEEE Conference on Collaboration and Internet Computing (CIC)*, 2015, pp. 219–226.
- [6] L. Singh and S. Singh, "A genetic algorithm for scheduling workflow applications in unreliable cloud environment," in *International Conference on Security in Computer Networks and Distributed Systems*, 2014, pp. 139–150.
- [7] K. Almi'ani, Y. C. Lee, and B. Mans, "On efficient resource use for scientific workflows in clouds," *Comput. Networks*, vol. 146, pp. 232–242, 2018, doi: 10.1016/j.comnet.2018.10.003.
- [8] X. Zhou, G. Zhang, J. Sun, J. Zhou, T. Wei, and S. Hu, "Minimizing cost and makespan for workflow scheduling in cloud using fuzzy dominance sort based HEFT," *Futur. Gener. Comput. Syst.*, vol. 93, pp. 278–289, 2019, doi: 10.1016/j.future.2018.10.046.
- [9] Z. G. Chen, K. J. Du, Z. H. Zhan, and J. Zhang, "Deadline constrained cloud computing resources scheduling for cost optimization based on dynamic objective genetic algorithm," in *Proc. IEEE Congress Evolutionary Computation (CEC)*, 2015, pp. 708–714, doi: 10.1109/CEC.2015.7256960.
- [10] H. Gamal El Din Hassan Ali, I. A. Saroit, and A. M. Kotb, "Grouped tasks scheduling algorithm based on QoS in cloud computing network," *Egypt. Informatics J.*, vol. 18, no. 1, pp. 11–19, 2017, doi: 10.1016/j.eij.2016.07.002.
- [11] S. Abrishami, M. Naghibzadeh, and D. H. J. Epema, "Deadline-constrained workflow scheduling algorithms for Infrastructure as a Service Clouds," *Futur. Gener. Comput. Syst.*, vol. 29, no. 1, pp. 158–169, 2013, doi: 10.1016/j.future.2012.05.004.
- [12] A. Deldari, M. Naghibzadeh, and S. Abrishami, "CCA: a deadline-constrained workflow scheduling algorithm for multicore resources on the cloud," *J. Supercomput.*, vol. 73, no. 2, pp. 756–781, 2017.
- [13] H. Topcuoglu, S. Hariri, and Min-You Wu, "Performance-effective and low-complexity task scheduling for heterogeneous computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 13, no. 3, pp. 260–274, 2002, doi: 10.1109/71.993206.
- [14] B. P. Rimal and M. Maier, "Workflow Scheduling in Multi-Tenant Cloud Computing Environments," *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 1, pp. 290–304, 2017, doi: 10.1109/TPDS.2016.2556668.
- [15] K. Almi'ani and Y. C. Lee, "Partitioning-Based Workflow Scheduling in Clouds," in *Proc. IEEE 30th Int. Conf. Advanced Information Networking and Applications (AINA)*, 2016, pp. 645–652, doi: 10.1109/AINA.2016.83.
- [16] G. Juve, A. Chervenak, E. Deelman, S. Bharathi, G. Mehta, and K. Vahi, "Characterizing and profiling scientific workflows," *Futur. Gener. Comput. Syst.*, vol. 29, no. 3, pp. 682–692, 2013, doi: 10.1016/j.future.2012.08.015.
- [17] D. G. Feitelson and A. M. Weil, "Utilization and predictability in scheduling the IBM SP2 with backfilling," in *Parallel Processing Symposium, 1998. IPPS/SPDP 1998. Proceedings of the First Merged International... and Symposium on Parallel and Distributed Processing 1998, 1998*, pp. 542–546.
- [18] H. Casanova, S. Pandey, J. Oeth, R. Tanaka, F. Suter, and R. F. da Silva, "WRENCH: A Framework for Simulating Workflow Management Systems," in *2018 IEEE/ACM Workflows in Support of Large-Scale Science (WORKS)*, 2018, pp. 74–85.
- [19] H. Casanova, "Simgrid: A toolkit for the simulation of application scheduling," in *Proceedings First IEEE/ACM International Symposium on Cluster Computing and the Grid*, 2001, pp. 430–437.
- [20] H. Casanova, A. Giersch, A. Legrand, M. Quinson, and F. Suter, "Versatile, scalable, and accurate simulation of distributed applications and platforms," *J. Parallel Distrib. Comput.*, vol. 74, no. 10, pp. 2899–2917, 2014.

# Minimization of Spectrum Fragmentation for Improvement of the Quality of Service in Multifiber Elastic Optical Networks

Boris Stephane ZOUNEME<sup>1</sup>, Georges Nogbou ANOH<sup>2</sup>, Souleymane OUMTANAGA<sup>3</sup>  
Institut National Polytechnique Félix Houphouët-Boigny, Yamoussoukro, Côte d'Ivoire<sup>1,3</sup>  
Université Virtuelle de Côte d'Ivoire, Abidjan, Côte d'Ivoire<sup>2</sup>  
Laboratoire de Recherche en Informatique et Télécommunication, Abidjan, Côte d'Ivoire<sup>1,2,3</sup>

**Abstract**—Internet data traffic is still growing considerably in recent decades. In view of this exponential and dynamic growth, elastic optical networks are emerging as a promising solution for today's flexibly allocated bandwidth transmission technologies. The setup and release of dynamic connections with different spectrum bandwidths and data rates leads over time to spectrum fragmentation in the network. However, single-fiber elastic optical networks are faced with the problem of optical spectrum fragmentation. Spectrum fragmentation refers to small blocks, isolated, non-aligned spectrum segments which is a critical issue for elastic optical network researchers. With the advent of multifiber, this fragmentation ratio has become more pronounced, resulting in a high blocking ratio in multifiber elastic optical networks. In this paper, we propose a new routing and spectrum allocation algorithm to minimize fragmentation in multifiber elastic optical networks. In the first step, we define different virtual topologies as many as there is fiber, for each virtual topology, the  $k$  shortest paths are determined to find the candidate paths between the source and the destination according to the minimization of a proposed parameter called allocation cost. In the second step, we apply the resource allocation algorithm followed by the choice of the optimal path with a minimum energy cost. Blocking probability and spectrum utilization are used to evaluate the performance of our algorithm. Simulation results show the effectiveness of our proposed approach and algorithm.

**Keywords**—Routing and spectrum allocation; fragmentation; multifiber elastic network; quality of service

## I. INTRODUCTION

The remarkable development of wireless communication, especially 5G and services such as IPTV, Big Data, Cloud Computing, distance learning, IOTs, etc., introduces significant challenges in transmission technologies and systems for routing and switching large volumes of data, as well as network costs and energy consumption. According to Cisco VNI 2017, consumer Internet traffic is expected to increase by 27% between 2016 and 2021 [1][2], requiring a high-capacity network infrastructure with optical resources. Wavelength Division Multiplexing (WDM) technology has been proposed to support this exponential traffic growth. However, in WDM networks, due to the fixed 50 Ghz grid and the rigidity of resource allocation, it leads to abusive and inefficient use of bandwidth. Elastic Optical Networks were then proposed to support variable bandwidth allocation using O-FDM (Optical

Frequency Division Multiplexing) technology according to demand with a more flexible spectral grid. O-OFDM allows continuous optical carriers and variable numbers of subcarriers on a channel. In elastic optical networks based on OFDM, the optical spectrum is divided into a number of frequency slot intervals with a width of 6.25 or 12.5 GHz [3][4].

Several themes related to this new high-speed network technology are of interest for in-depth research. Continuity and contiguity constraints in the routing process and resource allocation (optical paths and spectrum) in single-fiber elastic optical networks; as well as the continuous establishment and release of optical connections with different spectrum bandwidths and data rates lead over time to spectrum fragmentation which is a crucial problem for elastic optical network research specialists. This spectrum fragmentation refers to small blocks of small, isolated, non-aligned spectrum segments with serious consequences, as it can lead to misuse of bandwidth and a high probability of blocking. Several routing and spectrum allocation algorithms have been proposed in monofiber. However, with the advent of multifiber, which provides greater flexibility in the allocation of spectrum slots in elastic optical networks, this fragmentation has been accentuated, leading to a high blocking probability and a lower QoS in the network. In this paper, we propose a new routing and spectrum allocation solution that significantly improves the state of multifiber elastic optical networks by taking into account power saving and a proposed parameter called allocation cost in the choice of the optimal path.

The rest of this paper is organized as follows. Section II introduces the architecture of elastic optical gratings and defines the problem. Section III presents the mathematical modelling of the multifiber elastic optical grating. Section IV describes the spectrum management strategy and the routing and spectrum allocation algorithm taking into account power consumption. The results of the performance evaluation of our routing and spectrum allocation strategy are presented in section V. Section VI concludes this paper.

## II. RELATED WORK

Optical elastic networks are varied, high-speed data transport networks that require fairly large resources. This new type of optical transport network technique is based on optical modulation OFDM (Orthogonal Frequency Division

Multiplexing) [5][6] in which the frequency grids are flexible from one connection request to another, unlike the fixed frequency grids of conventional WDM optical networks.

Through its flexibility, elastic optical network (EON) offers new features in terms of segmentation and aggregation of spectral resources (subwavelengths and super-wavelengths), efficient adaptation of multiple data rates, as well as elastic variation of the allocated resources. OFDM technology allows the channels of neighbouring subcarriers to overlap, as shown in Fig. 1.

This increases the spectral efficiency of the transmission. For this purpose, an optical signal is produced by a variable bandwidth OFDM transponder using the technology with just the right spectral resources to meet the customer's demand. This technique makes it possible to provide spectrum frequencies according to the needs of traffic from source to destination. Advances in optical transmission techniques and devices have led to the emergence of elastic optical networks. The introduction of advanced modulation formats and Wavelength Cross-Connects (WXC) optical cross-connects (WXC) allows the increasing volume of traffic to be carried over long distances without opto-electronic-to-optical (OEO) conversion. Paths with bandwidths determined by the customer's traffic volume are allocated by flexible ratio transponders from the transmitter and are sent via Wavelength Cross-Connects (WXC) with variable bandwidth (Bandwidth-Variable BVT) to the receiver as shown in Fig. 2.

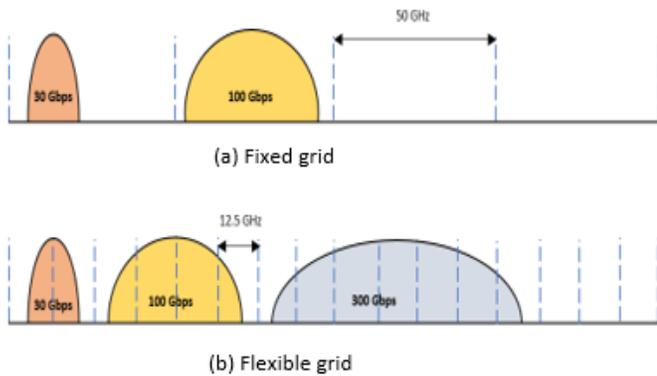


Fig. 1. Fixed Frequency Grid/Flexible Frequency Grid.

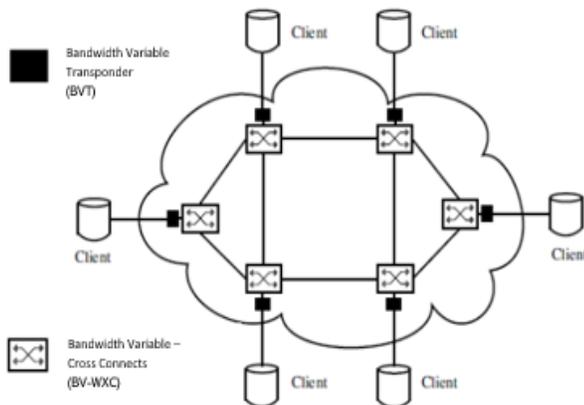


Fig. 2. Architecture of Elastic Optical Networks.

The system is based on Orthogonal Optical Frequency Division Multiplexing (OFDM). A frequency-locked multicarrier generator is used to generate equally spaced subcarriers.

The generated subcarriers are first separated by a wavelength division demultiplexer (DMUX), then individually modulated with parallel modulators, and finally coupled to generate a superimposed spectral superchannel. There are three transponder models, namely the Mixed Line Ratio (MLR) model, the Multi Flow (MF) model and the Bandwidth Variable (BV) model. The MLR model uses a few types of transponders, each with a different bit ratio, e.g. 40, 100 and 400 Gb/s transponders to meet a wide range of traffic demands [7].

The MF model uses an MF transponder with several subtransmitter-receivers, which can be allocated to different traffic requests, each of which has a fixed throughput capacity. The BV model supports all types of traffic requests with a single BV transponder, which allocates as few spectrum resources as possible to traffic requests with a maximum bit ratio of 400 Gbps. As shown in the study presented in [7], the BV model provides better spectrum and the lowest consume ratio. It uses different modulation formats such as 16-QAM, 64-QAM and QPSK (Quadrature Phase Shift Keying). In addition, it offers a better balance between spectral efficiency and transmission range. Due to the reduction of active resources, a new generation of optical transponders (S-BVT) has been investigated in [8]. Flexible grating optical networks allow efficient use of spectrum resources by using the 12.5 GHz frequency instead of the traditional fixed 50 GHz spacing, however introducing spectrum fragmentation (FS).

In the literature, SF is considered a crucial and major problem in EON, especially in a context of dynamic traffic. In this work, we present a comprehensive analysis and an accurate assessment of spectrum fragmentation in elastic optical networks.

The problem of fragmentation already exists in areas other than optical networks. It has been addressed for the management of synchronous optical networks (SONET) and code division multiple access (CDMA) systems in wireless networks [9] [10]. Data fragmentation has been studied for some context of EON (routing, number of locations, spectrum continuity constraint, network size, etc.) makes the MSDS issue different. MSDS in optical networks can be approached in two dimensions: horizontal fragmentation and vertical fragmentation as shown in Fig. 3.

In EON, connection requests are dynamically allocated according to continuity and contiguity constraints [11][12]. The continuity constraint must guarantee the use of the same wavelength or spectrum slot from end to end of an optical connection and leads to so-called horizontal fragmentation.

However, in flexible optical networks, lightpaths may be established without taking into account the spectrum continuity constraint. In this case, wavelength converters are required. However, service providers want to avoid the use of wavelength converters because of their additional deployment costs.

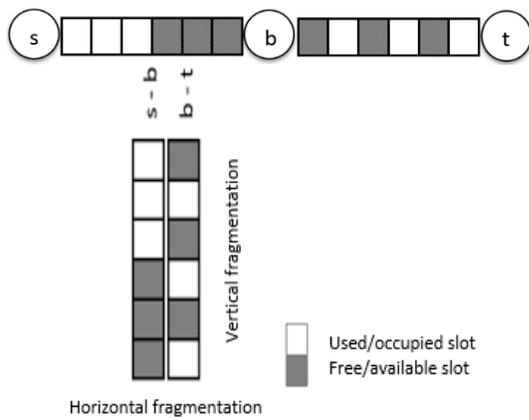


Fig. 3. Example of Vertical and Horizontal Fragmentation, s-b and b-t links are Horizontally Fragmented and Available/Free Slots are Vertically Fragmented.

It is clear that all these facts inevitably lead to the fragmentation of the optical spectrum which is a serious problem in ROE which is more accentuated in the dynamic context of the network in which requests arrive and disappear randomly. In such networks, new requests are therefore more likely to be blocked due to fragmentation [13]. Poor management of resource allocation significantly increases the ratio of fragmentation both in the path, in the link and in the network as a whole, leading to a deterioration of spectrum usage and huge losses on the operator side [14]. Several methods have been proposed to solve the problem of spectrum fragmentation, the simplest method is a fragmentation-aware Routing and Spectrum Allocation (RSA) algorithm that specifies how traffic requests are routed to minimize fragmentation [15]. Approaches to divide blocked demand into sub-demands in order to use finer granularities have also been proposed [16]. Another intuitive method is to defragment the spectrum so that free blocks are contiguous [17]. These methods are either complex, inefficient or difficult to implement. A new spectrum allocation strategy and a path selection scheme based on both the new network state after demand establishment and the probability of path selection have been proposed to eliminate fragmentation in multifiber elastic optical networks [18]. A quantitative parameter for evaluating link fragmentation is proposed to eliminate fragmentation by supporting a larger number of connections [19].

### III. MATHEMATICAL MODELING OF THE MULTIFIBER ELASTIC OPTICAL NETWORK

The topology of our network is defined by the graph

$G(E, L, SF, D)$  where :

- $E$  represents all the nodes of the network by a set of optical distribution frames (OXCs).
- $L = \{(i,j) \text{ with } i \neq j\}$  represents the set of links in the network where each link contains several fibers assuming that the number of fibers on each link is identical and  $i$  and  $j$  represent the vertices of the graph.

- $SF = \{f_{s_1}^l, f_{s_2}^l, \dots, f_{s_n}^l\}$  with  $n \in \mathbb{N}$  such as  $1 \leq l \leq F$  and  $n \geq 1$  where  $F$  is the number of fiber per link and  $n$  is the number of slots.
- $D = \{d_1, d_2, \dots, d_n\}$  represents the set of network connection requests.

A connection request  $cr$  between a source  $s$  and a destination  $d$  with a ratio  $\omega$  (Gb/s) is represented by:

$$cr = (s_{cr}, d_{cr}, \omega_{cr}) \quad (1)$$

The number of frequency slots  $NFS$  for a connection request  $cr$  on the path  $p$  is calculated according to the following expression in [2]:

$$NFS(p(cr)) = \frac{\omega_{cr}}{M(p(cr)) \cdot \omega_s} + gb \quad (2)$$

where  $gb$  represents the guard band

- $M(p(cr))$  is a function that returns the coefficient of the appropriate modulation format:

$$M(p(cr)) = Modform(\sum_{(i,j) \in p} l(i,j)) \quad (3)$$

- $\omega_s$  is the optical signal speed for a frequency slot

where  $l$  is the length of the link  $(i,j)$  on the path  $p$ .

In the case of adaptive routing, the capacity of a slot and the range of the optical signal depends on the modulation format chosen for a connection. In addition, the use of transponders at different linear rates based on the optical technique of orthogonal frequency division multiplexing without dispersion (CO-OFDM) in EON can exploit several modulation formats for different subcarriers and provide flexibility and heterogeneous handling of connection requests.

Therefore, the transmission ratio for a subcarrier may vary depending on the modulation format, namely BPSK, QPSK, 4-QAM, 16-QAM, 32-QAM and 64-QAM.

The table I is a summary of the energy consumed by a slot according to the modulation format.

TABLE I. ENERGY CONSUMPTION A SLOT IN FUNCTION OF MODULATION FORMAT AND OPTICAL RANGE

Optical modulation format	Slot capacity (Gb/s)	Energy consumed by a transponder (w)	Reach (km)
BPSK M=1	12.5	112.374	4000
QPSK M=2	25	133.416	2000
8-QAM M=3	37.5	154.457	1000
16-QAM M=4	50	175.498	500
32-QAM M=5	62.5	196.539	250
64-QAM M=6	75	217.581	125

#### IV. DYNAMIC ALGORITHM FOR MINIMIZING FRAGMENTATION IN MULTIFIBER ELASTIC OPTICAL NETWORKS

In our approach of routing and allocation spectrum named AMF (Algorithm for Minimizing Fragmentation), we assume that a given connection request must be established on the same fiber between the source and the destination. Therefore, being in a multifiber context, we define as much virtual topology as there are fibers on a given link. When a *cr* connection request arrives in the network; several virtual topologies are created. For each virtual topology, we determine a physical path between the source node *s* and the destination node *d*. Assuming the physical topology with two fibers per link *f*<sub>1</sub> and *f*<sub>2</sub> described by Fig. 4. we will obtain two virtual topologies describe by Fig. 5 and Fig.6 respectively.

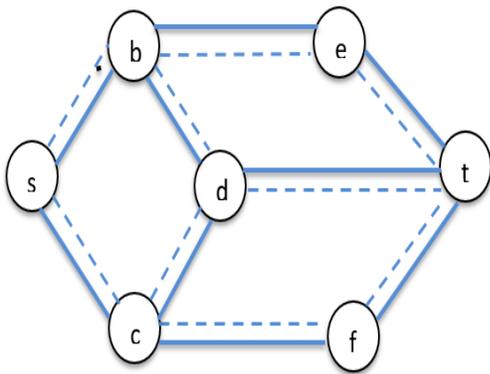


Fig. 4. Multifiber Network.

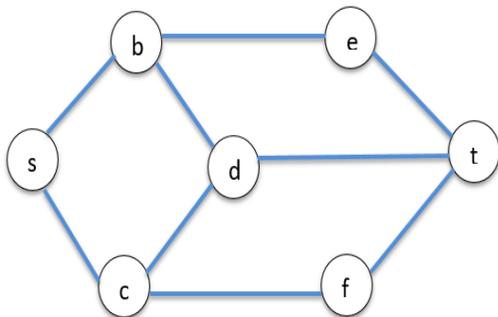


Fig. 5. Virtual topology 1 from *f*<sub>1</sub>.

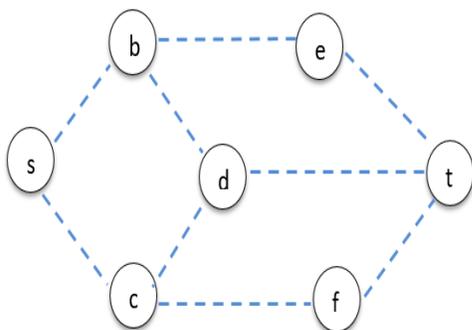


Fig. 6. Virtual Topology 2 from *f*<sub>2</sub>.

An optical path is characterized by source, destination and optical fiber as follows:

$p(s,d,f_i)$  where *s* is the source, *d* is the destination and *i* represents the fiber.

After the representation of the different virtual topologies, for each virtual topology, we calculate the shorter paths from *s* to *d*. For each path found, we calculate the number of frequency slots needed to establish the connection from (2). From the resources available on the path, we determine whether this path has the resources for the establishment of the connection request. If so, we try to determine the lightpaths (physical paths + spectral resources) which are eligible. It is then assumed that for each of the *N* topologies and for each of the shorter paths from source to destination, resources are determined for each physical path.

We assume that for *k* physical paths, we obtain *k'* optical paths with  $k' \leq k$  where *k* is a natural number for each virtual topology. We will have at most  $N' * k'$  eligible optical paths for the network ; where  $N' \leq N$ .

However, the question related to this approach that can be raised is which optical path will be chosen from the  $N' * k'$  paths ?

To answer this question, after determining the  $N' * k'$  eligible optical paths ; we are seeking to determine the fragmentation ratio.

In this case, the light path that will be chosen as the optimal path will be the one that consumes less energy and has the lowest fragmentation ratio, or if all light paths consume the same energy then the one with the lowest fragmentation ratio will be chosen. If this is not possible, the connection request is blocked.

##### A. Fragmentation Parameter

We can consider a fragmentation parameter that we will define as a criterion for choosing the best optimal path with the objective of minimizing fragmentation.

This fragmentation parameter, which we will name Allocation Cost AC, depends on the presence of the total number of frequency slots available on the path and the number of slot blocks.

This parameter is defined as follows:

$$AC = \sum_{i=1}^n \exp\left(\frac{1}{S_i}\right) \quad (4)$$

where, *S<sub>i</sub>*, the number of frequency slots in block *i* available on the path.

We illustrate this parameter as follows with three links *L*<sub>1</sub>, *L*<sub>2</sub> and *L*<sub>3</sub> for a connection request requiring three contiguous frequency slots :

*L*<sub>1</sub>

1	2	3	4	5	6	7	8	9	10	11	12
---	---	---	---	---	---	---	---	---	----	----	----

$$AC(L_1) = 4e^1 = 10.873$$

$L_2$



$$AC(L_2) = e^{\frac{1}{3}} = 1.395$$

$L_3$



$$AC(L_3) = e^{\frac{1}{3}} = 1.395$$

We note that  $AC(L_3) \leq AC(L_2) \leq AC(L_1)$ , when the cost of allocation is high this leads to the creation of small slot blocks with a low probability of being used for future connection requests and thus increases network fragmentation. So with regard to this fragmentation parameter when the value of this parameter is low the fragmentation of the link is reduced.

### B. Energy Cost of a Lightpath

The energy cost of a lightpath for a  $cr$  connection request is calculated according to the following parameters: the energy consumption of a slot, the energy consumption of an optical splitter along the link and the energy consumption of an Erbium Doped Fiber Amplifier (EDFA). Let the following expression be used:

$$ECL(C(cr)) = (NFS(C(cr)) - gb) \times C_{slot} \times PC_{slot} + \left( \sum_{(i,j) \in C} C_{Amp} \times PC_{Amp} \times (\rho(i,j) + 2) + \sum_{(i,j) \in C} C_{OXC} \times PC_{OXC} \right) \times \frac{NFS(C(cr))}{SUM_{slot}} \quad (5)$$

$$with \quad \rho(i,j) = \begin{cases} 0, & if \frac{l(i,j)}{d_{Amp}} \leq 1 \\ \left\lceil \frac{l(i,j)}{d_{Amp}} \right\rceil, & otherwise. \end{cases} \quad (6)$$

The principle of determining physical paths by our ADCP algorithm, for each connection request that arrives on the network; we create as many virtual topologies as there are fibers on the network links defined by the  $G$  graph. After creating the different topologies, we determine the shortest paths and for each path found, we determine the path that has resources available for the connection request.

If there is a path in the network that has resources available for the connection request, then the physical path and the spectral resources, i.e. the lightpath, are saved.

Then we calculate the energy cost for each path in order to determine the optimal path that will have the minimum energy cost.

Our method (AMF) selects the path and slots that use the minimum spectrum based on our AC cost of allocation parameter and a minimum energy cost, thus reducing bandwidth blocking for the overall network.

Network topology ( $G(E, L, SF, D)$ )

Location of the OXC network nodes, flexible links, fibers and connection requests, the guard band are given as input.

The output will be the best lightpath with the lowest spectrum utilization and the lowest blocking probability.

For a given connection request and topology AMF algorithm will create the virtual topologies from the graph ( $G(E, L, SF, D)$ ), then for each virtual topology will determine the  $k$ -shortest paths using ADCP algorithm and then find the available slots using the ADRS algorithm. If the available slots are found then the physical paths with available slots are saved. Then the best lightpath that has the minimum power cost and a low allocation cost is chosen according to ADMC algorithm to serve the connection request.

### Algorithm to determine the physical path (ADCP)

Input: Graph  $G$

Output: Physical paths and resources

- 1: Create the virtual topologies from graph  $G$
- 2: For each connection request that arrives in the network
- 3: For each virtual topology
- 4: Determine the  $k$ -shortest paths
- 5: For each shortest path
- 6: Determine the resources for the connection request
- 7: If resource found then
- 8: Save the physical path and its resources
- 9: End if
- 10: End for
- 11: If lightpath found then
- 12: Choose the optimal path
- 13: Establish the connection request
- 14: Otherwise
- 15: Go to step 3
- 16: End if
- 17: End for
- 18: End for

### Algorithm to Minimize Fragmentation (AMF)

Input: Graph  $G$

Output: Physical paths and resources

- 1: Create the virtual topologies from graph  $G$
- 2: For each connection request that arrives in the network
- 3: For each virtual topology
- 4: Determine the  $k$ -shortest paths using *algorithm ADCP*
- 5: For each shortest path
- 6: Determine the resources for the connection request using *ADRS algorithm*
- 7: If resource found then
- 8: Save the physical path and its resources
- 9: End if
- 10: End for
- 11: If lightpath found then
- 12: Choose the best optimal lightpath that has the minimum energy cost using *algorithm ADMC*
- 13: Save the optimal lightpath
- 14: End if
- 15: End for
- 16: If optimal lightpaths
- 17: Choose the best optimal lightpath that has the lowest fragmentation ratio
- 18: End if
- 19: End for

**Algorithm to determine spectral resource (ADRS)**

Input: Physical paths

Output: Resources to be used (free frequency slots)

- 1: Determine the number of slots for the connection request
- 2: Determine the blocks of slots for the path
- 3: Determine the free blocks of slots which are eligible for the connection
- 4: Determine the smallest eligible block among all the eligible blocks
- 5: Extract the resource blocks from this block for the connection request

In the principle of determining the best lightpath by our ADMC algorithm, for a given connection request, the lightpath with minimal energy cost is chosen. In the case where several lightpaths have the same minimum cost, the lightpath that reduces the fragmentation ratio is chosen.

**Algorithm to determine the best lightpath (ADMC)**

Input: All paths and the connection request

Output: Optimal lightpath for the connection request

- 1: Calculate the energy cost of each path
- 2: Determine the path with the minimum energy cost
- 3: If two paths have the same minimum energy cost then
- 4: Choose the path that reduces the fragmentation ratio
- 5: End if

V. RESULTS

A. Assumptions

In this section, we discuss the simulation results and evaluate the performance efficiency of our proposed AMF algorithm with the algorithm proposed in [20] in NSFNET and US-backbone topologies.

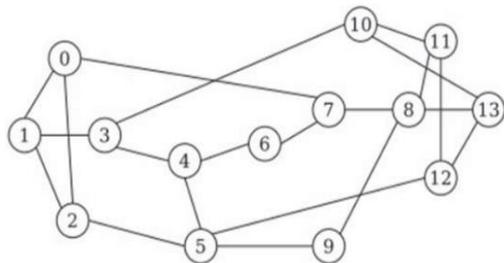


Fig. 7. NSFNET Topology.

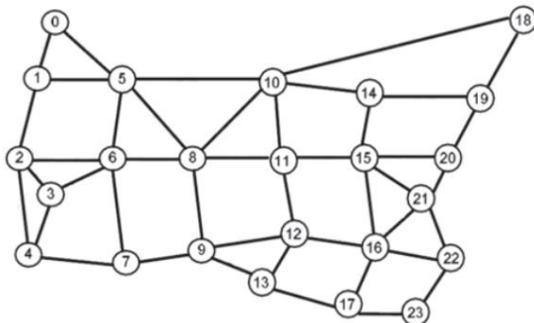


Fig. 8. US-Backbone Topology.

The NSFNET topology consists of 14 nodes and 22 links while the US-backbone topology consists of 24 nodes and 43 links illustrated in Figs. 7 and 8. We assume that each link has an identical number of fibers  $F=10$ . The total number of slots per fiber and the bandwidth of each slot are 352 and 12.5 Ghz respectively.

Note also that in this paper the frequency conversion is not taken into account. Table II. shows the simulation parameters.

Between  $10^4$  et  $10^5$  connection requests arrive as dynamic traffic in the network following a sequential and random process. Each source-destination node pair is randomly selected. We estimate the capacity of a fiber link on the network at 4400 Gbps.

Thus, the performance of our algorithm is evaluated in terms of blocking probability and spectrum utilization ratio.

The blocking probability is defined as the ratio between the number of rejected connection requests and the total number of requests arriving on the network. Whereas the spectrum or spectrum resource usage ratio is the ratio of the number of occupied slots to the total number of available slots. For better network performance, the probability of blocking and spectrum usage rates should be as low as possible.

B. Performance Evaluation

Fig. 9 and 10 show the simulation results in terms of blocking probability in the NSFNET and US-backbone topologies respectively. Thus, we can observe that the curves in both figures follow roughly the same trend. The results of our AMF algorithm compared to those of the NSA algorithm show that the AMF strategy results in a significantly lower blocking probability for all traffic loads in both topologies.

TABLE II. SIMULATION PARAMETERS

Parameter	Value
Degree	3
Number of slots/fiber	352
Bandwidth of a slot	12.5 Ghz
Capacity of a slot of $M = 1$	12.5 Gbps
Range debit	12.5 - 400 Gbps
Capacity of a fiber link	4400 Gbps

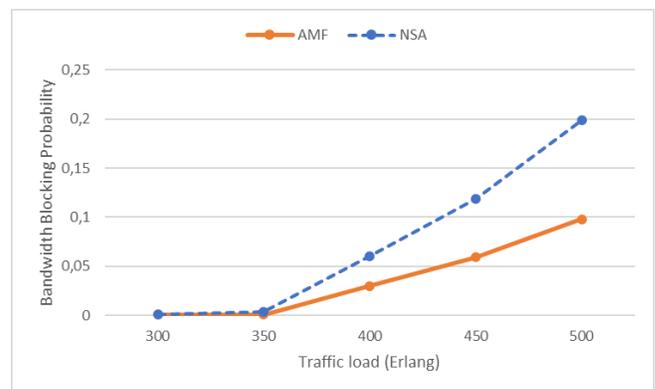


Fig. 9. Comparison of Bandwidth Blocking Probability in NSFNET Topology.

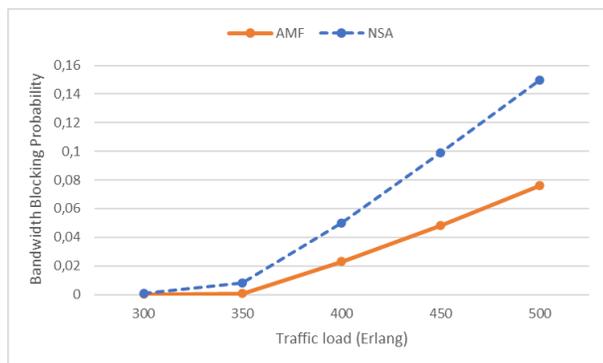


Fig. 10. Comparison of Bandwidth Blocking Probability in US-Backbone Topology.

Several observations can be made from the results obtained, namely: between 300 and 350 Erlang the probability of blocking is approximately the same for both approaches in both topologies. However, we find that a reduction of 16% and 19.4% is achieved in terms of blocking probability by our proposed AMF approach compared to the NSA approach for the US-backbone and NSFNET topologies respectively. This is due to the fact that as the unused spectrum becomes larger and larger, nodes reject incoming connection requests due to the limit of spectrum capacity. This shows the efficiency of both routing and spectrum allocation of our proposed algorithm.

The spectrum utilisation ratio as a function of network traffic load is shown in Figs. 11 and 12 for the NSFNET and US-backbone topologies. Note that the number of resources increases as the network load increases.

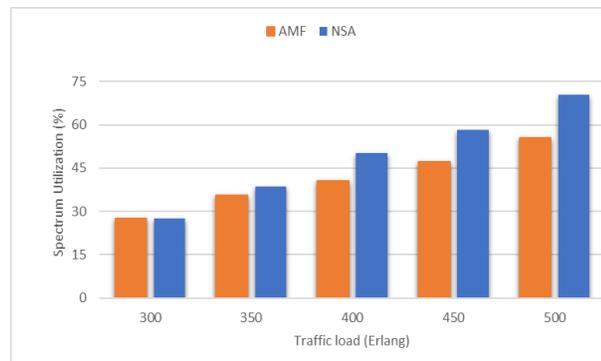


Fig. 11. Comparison of Spectrum utilization Ratio in NSFNET Topology.

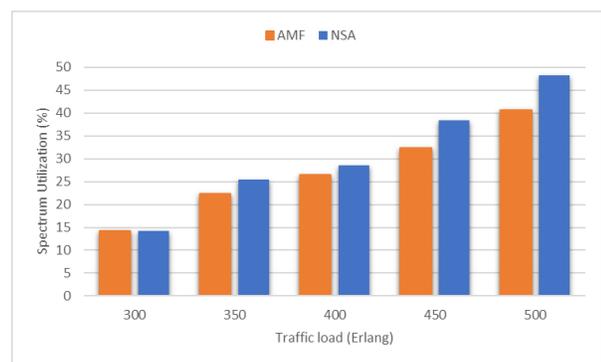


Fig. 12. Comparison of Spectrum utilization Ratio in US-Backbone Topology.

As we observe from these figures, our AMF approach requires a minimum ratio of use of spectrum resources. Simulation results show a gain of 18.7% and 38.9% in terms of spectrum usage efficiency for US-backbone and NSFNET topologies respectively. It shows that the NSA algorithm has a slight impact on the quality of service which is of paramount importance for the profitability of the service provider and especially for the customer satisfaction. Hence the significant improvement of our AMF algorithm in terms of service quality. Note that taking into account the traffic load, [20] calculates the set of paths and the probability of path selection for each node in offline mode, which gives a better algorithmic complexity to our approach. However, in our AMF approach, the routing and spectrum allocation is done dynamically and allows to obtain a good performance in spectrum utilization ratio.

## VI. CONCLUSION

Routing and spectrum allocation is a major challenge in elastic optical networks. Dynamic and continuous establishment and withdrawal leads to fragmentation of the spectrum resulting in small, isolated and non-contiguous fragments, resulting in inefficient use of spectrum and increased probability of blocking. In this paper, we proposed a new algorithm to minimize fragmentation (AMF) with a new cost of allocation parameter (AC) for dynamic traffic demands in elastic optical networks with multiple fibers per link. Assuming that the connection is established on the same fiber between the source and the destination, we define as many virtual topologies as there are fibers on a given link taking into account minimum power consumption of network components in the choice of the optimal optical path. The performance of our proposed approach is evaluated by simulating NSFNET and US-backbone network topologies in terms of bandwidth blocking probability and spectrum utilization efficiency to achieve better performance than existing algorithms in the literature.

In the future work, we plan to optimize bandwidth by increasing network capacity and reduce the complexity of network components by developing routing and spectrum allocation algorithms by integrating into multifiber elastic optical networks the Software Defined Optical Networks (SDONs) considered as next-generation networks with centralized controllers based on the Software Defined Network (SDN).

## REFERENCES

- [1] Cisco, Cisco Visual Networking Index: Forecast and Methodology, 2016 - 2021. Cisco white paper, June 2017.
- [2] T. Ahmed, S. Rahman, M. Tornatore, X. Yu, K. Kim, and B. Mukherjee, "Dynamic Routing and Spectrum Assignment in Co-Existing Fixed/Flex-Grid Optical Networks", IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), 2018.
- [3] J. Wu, M. Xu, S. Subramaniam, and H. Hasegawa, "Routing, fiber, band, and spectrum assignment (RFBSA) for multigranular elastic optical networks" in IEEE Int. Conf. On Communications (ICC), 2017, pp. 1-6.
- [4] J. Wu, S. Subramaniam and H. Hasegawa, (2015). " Comparison of OXC Node Architectures for WDM and Flex-Grid Optical Networks", 24th International Conference on Computer Communication and Networks (ICCCN), 2015.

- [5] N. G. Anoh., M. Babri, D. K. Ahmed, M. F. Roger, B. Aka and C. Lishou, "An efficient hybrid protection scheme with shared/dedicated backup paths on elastic optical networks", *Digital Communications and Networks*, March 2017.
- [6] C. C. Bijoy, N. Sarma, E. Oki, "Routing and Spectrum Allocation in Elastic Optical Networks: A Tutorial" *IEEE Communication Surveys and Tutorial*, vol.17, no.3, 2015.
- [7] G. Zhang, M. D. Leenheer, A. Morea, B. Mukherjee", A survey on OFDM-based elastic core optical networking", *IEEE Communications Surveys & Tutorials*, 2013.
- [8] T. Tanaka, A. Hirano, and M. Jinno, "Impact of transponder architecture on the scalability of optical nodes in elastic optical networks," *IEEE Communications Letters*, vol. 17, 2013.
- [9] M. S. Johnstone, P. R. Wilson "The memory fragmentation problem: solved?", In: *Proceedings of the 1st international symposium on Memory management*, 1998.
- [10] G. Bouyer. *Les réseaux synchrones étendus PDH et SDH*. Hermes, 1997.
- [11] M. S. Johnstone, P. R. Wilson, "The memory fragmentation problem: solved?", In: *Proceedings of the 1st international symposium on Memory management (ISMM)*, 1998.
- [12] J. Wu, M. Xu, S. Subramaniam, and H. Hasegawa, "Joint Banding-Node Placement and Resource Allocation for Multigranular Elastic Optical Networks", *Journal of Optical Communications and Networking*, 10(8), C27, 2018.
- [13] Y. Wang et al, "Spectrum consecutiveness-based routing and spectrum allocation in flexible bandwidth networks", *Chinese Optics Letters*, 2012.
- [14] M. S. Johnstone, Wilson, P.R. "The memory fragmentation problem: solved?", In: *Proceedings of the 1st international symposium on Memory management*, 1998.
- [15] Y. Sone, A. Hirano, A. Kadohata, M. Jinno, O. Ishida," Routing and spectrum assignment algorithm maximizes spectrum utilization in optical networks", *Optical Communication*, 2011.
- [16] A. Pages, J. Perello, S. Spadaro, J. Comellas, "Optimal route, spectrum, and modulation level assignment in split-spectrum enabled dynamic elastic optical networks", *Opt. Commun. Netw. IEEE/OSA*, 2014.
- [17] R. Wang, , B. Mukherjee, "Provisioning in elastic optical networks with non-disruptive defragmentation", *Lightwave Technol.*, 2013.
- [18] W. Jingxin, S. Suresh and H. Hiroshi, "Dynamic Routing and Spectrum Assignment for Multi-fiber Elastic Optical Networks", *Conference Paper*, 2018.
- [19] S. Krishan, S. Shree and S. Sujata, "Fragmentation Suppressed RSA Algorithm for Elastic Optical Network: A Quantitative approach", *IEEE British and Irish Conference on Optics and Photonics London*, 2018.
- [20] W. Jingxin, S. Suresh and H. Hiroshi, "Efficient Dynamic Routing and Spectrum Assignment for Multifiber Elastic Optical Networks", *J. Opt Commun. Netw*, vol. 11, 2019.

# The Effect of Firm's Size on Corporate Performance

Meiryani<sup>1</sup>

Accounting Department  
Faculty of Economics and Communication  
Bina Nusantara University  
Jakarta, Indonesia 11480

Olivia<sup>2</sup>

Accounting Department  
Faculty of Economics and Communication  
Bina Nusantara University, Jakarta, Indonesia 11480

Jajat Sudrajat<sup>3</sup>

BINUS Entrepreneurship Center, Management Departement  
BINUS Business School, Undergraduate Program  
Bina Nusantara University  
Jakarta, Indonesia 11480

Zaidi Mat Daud<sup>4</sup>

Department of Accounting and Finance  
Faculty of Economics and Management  
Universiti Putra Malaysia

**Abstract**—The purpose of this study is to determine the effect of capital structure on firm's financial performance that is conducted on 55 manufacturing sector listed companies in Indonesia Stock Exchange. The data analysis is conducted using R Studio software. Study is used data panel analysis with random effect model. The result of this study are (1) firm's size has no effect on firm's financial performance which is proxied by return-on-assets; (2) firm's size has no effect on firm's financial performance which is proxied by market-to-book-value.

**Keywords**—Firm size; financial performance; return on assets; market to book value

## I. INTRODUCTION

In evaluating the company's financial performance, it can be assisted with certain measurement tools, one of which is by using profitability ratios and market ratios. Profitability ratios are ratios used to determine a company's ability to manage its assets [1]. An assessment of a company's profitability can be measured through Return-on-Asset (ROA). Return-on-assets are used to measure the effectiveness of a company in generating profits by utilizing its assets. The company's ability to utilize assets effectively and productively can generate net profit which is the result of the capital that has been invested in an asset. Return-on-assets as a form of company effectiveness in managing its assets and capital [2]. The return-on-asset measurement avoids direct comparisons with market value but relates to several factors used by market makers in their valuation of the company [3]. Return-on-assets are able to measure the company's ability to generate profits in the past and then projected in the future [4]. Company performance through profitability is one of the most important areas that is focused by shareholders and also debt holders if the company uses debt to operate [5].

Market-to-book ratios indicate investors' views on the value of the company, shares sold higher than the book value reflect good companies [6]. Therefore, the company's market value reflects how the market sees the company so that it will affect investors' valuation of the company in making its investment. One of the motivations of investors to invest their capital is return [7]. Firm's size can also affect a company's financial performance. Firm's Size is a size or scale that

shows the size of a company. Firm's size can usually be seen from the company's total assets and total sales. In general, large companies will get more attention or better known to the public. The greater the size of the company can give a good signal to the public. This is in accordance with the signaling theory that the greater the firm's size will give a positive signal for the company's financial performance is getting better [8].

Research conducted by [9] in manufacturing sector companies in the country of Sri Lanka shows that there is no effect of company size which is proxied by total assets and total sales on profitability proxied by ROA and NPM.

Based on the background described above, the following are the problems identified by the researcher (then the problem can be formulated to be investigated):

- 1) Does Firm's Size affect corporate performance which is proxied by return-on-assets?
- 2) Does firm's size have a positive effect on corporate performance that is proxied by market-to-book value?

Based on this background, this study takes the title of the thesis "The effect of firm's size on firm's financial performance".

## II. LITERATURE REVIEW

### A. Signaling Theory

Signaling theory states that there is information asymmetry or information mismatch between internal (company managers) with external parties or users of financial statements (stakeholders). With this, signaling theory explains why companies have the incentive to provide financial statement information to external parties, namely investors and creditors [10]. With the disclosure of information made by the company, users of financial statements such as investors and potential investors can make an appropriate consideration or analysis to invest their capital so as to increase market views regarding the company's financial performance. Disclosure of information carried out by company managers can give signals to users of financial statements so that accurate information is very important in the process of

making investment decisions or providing loans by external parties. With the existence of signaling theory, it is hoped that financial statement users can obtain accurate information in order to reduce the problem of information asymmetry.

The choice of debt-to-equity-ratio indicating the quality of a company [11]. Companies by adding more debt to the company's capital structure can function as a credible signal of higher expected future cash flows. That is because the company is committed to meeting interest payments from debt holders and managing the remaining cash flow more effectively [12]. Investors consider debt issuance as a better signal than the issuance of ordinary shares because investors understand the company's management goals in issuing shares that are considered as overvalued [13]. In addition, the greater the size of the company will give a positive signal which means the company's financial performance is also getting better [14]. Signaling theory explains that to increase market views on company performance, companies have a large size or scale of companies and will increase debt.

### B. Firm's Size

Firm size is a major factor in determining company profitability because of a concept known as economies of scale that can be found in the traditional view of the company [15]. It can be interpreted that companies can produce goods at a much lower cost by large companies. The size of the company is an increase in the company's employees who have a large market capitalization, and describes the size of a company. The higher the total assets that indicate the assets or assets owned by the company. Company size can be measured using total assets, sales or company capital [16]. Companies that have large total assets are considered to have good prospects in a relatively stable period and are able to generate profits compared to companies that have small total assets [17]. Large-scale companies have a higher competitiveness than small companies, because large companies have a large market so they have a great opportunity to obtain large profits [18].

## III. THEORETICAL FRAMEWORK

### A. Effect of Firm's Size on Corporate Performance

One of the factors influencing company performance (return on assets) using panel data is firm's size [19]. Companies large enough to be able to take advantage of their size to negotiate the value of their inputs and then reduce their average costs. This will result in increased profitability for the company. Intellectual capital, debt-to-equity-ratio (DER), company size, and asset turnover have positive effects while debt-to-Assets-ratio has a negative effect on ROA [19]. Based on signaling theory, the greater the firm's size will give a positive signal to the public or market which means the company has better financial performance. There is a positive and significant effect on financial performance which is proxied by ROA. He stated that the greater the assets or assets, the more capital invested, more money circulation and greater market capitalization, would improve the company's financial performance. a significantly larger firm's size and positively increased the performance of companies proxied by ROA

[20]. Based on the description above, the researchers propose a hypothesis, as follows:

Ha1: Firm size has a positive effect on firm's financial performance which is proxied by return on assets (ROA).

### B. Positive Effect of Firm's Size on Firm's Financial Performance (MtBV)

Firm size can indicate that the company is experiencing growth and growth so that the market will respond positively. The greater the total assets and sales, the greater the size or scale of a company [21]. If a company has a large firm's size, then there is a possibility that it will have a better market value and financial performance compared to smaller companies [22]. Conventional banks listed on the Indonesia Stock Exchange shows that there is a positive influence on bank size on shareholder value which is proxied by market-to-book value [23]. The higher the total assets, the higher the market-to-book value. There is a positive influence on bank size with shareholder value which is proxied by market to book value. This is due to the large banks getting more attention from the public so that they are viewed more positively by investors. Based on the description above, the researchers propose a hypothesis, as follows:

Ha2: Firm size has a positive effect on firm's financial performance which is proxied by market to book value (MtBV).

## IV. RESEARCH METHODOLOGY

The data analysis method used in this study is multiple linear regression analysis techniques. Before testing on multiple linear regression methods, this method must be tested by classic assumptions to get the best results so that the data is consistent and unbiased. Types of classic assumption tests in this study are: normality test, Multicollinearity test, Heteroscedasticity test, Autocorrelation test. The regression model test using multiple linear regression analysis aims to choose the most appropriate model to be used in predicting this research. Here are some tests that can be done consisting of a chow test, the hausman test, and the lagrange multiplier test to determine which model is more appropriate in predicting the multiple linear regression equation model in this study.

### A. Operationalization of Variables

Research variables are research objects that are the center of attention of researchers to be observed, studied and drawn conclusions. Variables are objects that have variations between one object and another. The variables studied were divided into two namely independent variables and dependent variables. Following is the definition and measurement of each variable will be explained as follows:

The dependent variable (or often referred to as the dependent variable) is a variable that is of primary concern to the researcher and is influenced or is due, due to the independent variables. The dependent variable in this study is corporate performance measured by profitability ratios, namely Return-on-Assets (ROA) and market ratios, Market-to-Book-Value (MtBV).

1) *Return-on-Assets (Y1)*: In analyzing financial statements, the ratio of the most widely used by users of financial statements is Return-on-Asset (ROA) because the ratio is able to show the company's success in generating profits. ROA is used to measure overall management effectiveness in generating profits with available assets [24]. ROA is able to measure the company's ability to generate profits in the past and then can be proxied in the future (Rosikah et al, 2018).

2) *Market-to-Book-Value (Y2)*: Market-to-Book-Value (MtBV) is the ratio of the total market capitalization that is shared with the book value of net assets [25].

### B. Variable Independent

The independent variable (or often referred to as the independent variable) is a variable that affects the dependent variable both positively and negatively. If there are independent variables, the dependent variable will also be present, and with each unit increase in the independent variable, there will also be an increase or decrease in the dependent variable. So the independent variable influences or causes the change or the emergence of the dependent variable. The independent variables used in this study are:

Firm's size is the size of a company that can be measured by assets, sales, market capitalization. Firm's size can be calculated using the following formula:

$$Firm's\ Size = \ln (Total\ Asset)$$

## V. RESULT AND DISCUSSION

In this study, the financial statements issued by manufacturing companies listed on the Indonesia Stock Exchange (IDX) will be the object of research by researchers. After determining the research object, the sample is selected based on purposive sampling. Purposive sampling is a technique used to determine samples with a non-random sampling technique where the researcher determines sampling by setting criteria that are appropriate to the design and purpose of the study. The following is the procedure for determining the sample criteria for this research sample, as follows (Table I):

TABLE I. POPULATION SCREENING TABLE

No.	Criteria	Total
1	The population of manufacturing companies listed on the Stock Exchange during the 2014-2018 period	152
2	Companies experiencing delisting from the stock exchange or experiencing liquidation during the 2014-2018 period	(6)
3	Not served in rupiah	(28)
4	Does not always publish audited financial statements (audited financial statements) in full for periods that end for 5 years in a row during the period 2014-2018	(14)
5	Manufacturing companies that suffered losses during the 2014-2018 period	(43)
	Amount of cross-sectional data (per year)	61
	Outlier	(50)
	Amount of selected cross-sectional data (per year)	11
	Research period 2014-2018 (time-series data)	5 tahun
	Number of samples selected (total panel data)	55

Source: www.idx.co.id and processed

Through this screening process, 305 companies (61 cross-sectional data with 5 years' time-series data) were obtained on the Indonesia Stock Exchange which were allowed to be selected as samples. By using a confidence level of 95% to obtain an error rate of 5%. Furthermore, the authors screened outlier data based on box plot observations and standardized scores (z-scores) with a value of  $\geq 2.5$  which were declared as outliers so that 55 panel data were obtained that would be used for the five-year study period from 2014-2018, consisting of combined 11 cross-sectional data and 5 time series data.

### A. Research Data Collection

Most of this research uses numbers contained in accounts that appear on financial statements. The following will explain how researchers obtain these numbers:

a) Total Assets is a symbol of the position of total assets in the current period. In the Statement of Financial Position presented comparatively, the author will use the numbers on "Total Assets" as at 31 December in the current year period;

b) Market Price is a symbol of the market price of shares in each company obtained using the Close figure every December 31 Historical Data section contained on the website www.finance.yahoo.com.

c) Number of Outstanding Shares is a symbol of the number of shares outstanding during the period. If a company has Treasury Shares, then those shares must be excluded from the calculation of the number of shares outstanding because the number of shares outstanding is shares circulating to the market or the public that are not shares that were recovered by the Company. In the Statement of Financial Position of the Equity section presented comparatively, the author will use the number of shares written on "Issued and fully paid share capital" as of December 31 and reaffirm the Notes to the Report Finance (Notes to the Financial Statements) by looking at the number of shares outstanding by subtracting the numbers on the "Treasury Shares" as of December 31 in the current period;

d) Firm's Size is the size of the company with the natural logarithm symbol of total assets in the current period. In the Statement of Financial Position presented comparatively, the author will use the numbers on "Total Assets (Total Assets)" as of December 31, which will then be performed natural logarithms.

### B. Research Data Calculation

After getting all the data needed, then some calculations will be performed to get the desired information:

a) To meet the regression equation used, the researcher will do a regression model to find the estimated coefficients of each variable;

b) Return-on-Assets are calculated by dividing Net Income After Tax by Total Assets as ROA variable;

c) Market-to-book value is calculated by dividing Market Value with Book Value as an MtBV variable. Market Value is obtained by multiplying Market Price by Outstanding Number of Shares and Book Value is obtained by subtracting Shareholder's Equity with Paid in capital;

d) Firm's Size is obtained from total assets which then the logarithm of its natural search is called Ln (Total Assets).

### C. Descriptive Statistical Analysis

Descriptive statistical analysis aims to provide an overview of the data used in this study. Descriptive statistics consist of minimum values, maximum values, mean values, and standard deviations of each variable (Table II).

Descriptive statistical analysis aims to provide an overview of the data used in this study. Descriptive statistics consist of minimum values, maximum values, mean values, and standard deviations of each variable. In these results, *n* indicates the number of samples for each variable used in this study. Mean is the average value of research data. Standard deviation is a measure of the spread of data used to measure the amount of variation or distribution of a number of data values. Standard deviation is a large difference from the sample value to the Mean. The lower the standard deviation, the closer it will be to the average value, whereas if the standard deviation value is higher, the wider the variation of the data range. Maximum is the highest value of the amount of data analyzed. Minimum is the lowest value of the amount of data analyzed.

Descriptive statistical tests consist of ROA (Return-on-Asset) and MtBV (Market-to-Book-Value) as the dependent variable. The number of sample data used in the calculation of descriptive statistics is the entire sample data, which is 55 data. Return on Assets (ROA) is the first dependent variable that measures a company's profitability based on information contained in the company's financial statements. ROA has an average value of 0.08; which means that the ability of companies to get profits from their assets is 0.08 or 8%. ROA has a standard deviation value of 0.04 which means that the level of size distribution of ROA variable data is 0.04. The lowest and highest value of ROA is 0.02 and 0.16, respectively in PICO companies in 2018 and BATA in 2015. That is, BATA is a company that is able to generate profits by managing its assets effectively while PICO is a valued company less able to effectively generate profits by managing the assets they have.

Market-to-Book-Value (MtBV) is the second dependent variable that measures a company's market value based on information provided on the market and the company's financial history. MtBV has an average value of 1.71 with the lowest and highest values of 0.29 and 4.63 respectively in PICO companies in 2015 and STTP in 2014. Based on market or public views, stakeholders will be interested in investing in an company when the company reflects good performance so that the stock price will show a higher price and have a relatively lower additional paid in capital because the stock is sold not to overprice from face / nominal value. That is, STTP is a company deemed to have an appeal by the market or the public which is reflected in the stock price of Rp. 2,880 with a total of circulating shares of 1,310,000,000 shares, whereas the opposite is for PICO with having a stock price of Rp 128 with a total number of outstanding shares of 568,375. 000 sheets.

TABLE II. DESCRIPTIVE STATISTICAL ANALYSIS

	N	Minimum	Maximum	Mean	Std. Deviation
ROA	55	0,02	0,16	0,08	0,04
MtBV	55	0,29	4,63	1,71	1,05
Firm's Size	55	26,57	30,40	27,93	1,15

Source: R Studio Application Processing Results

Firm's size is one of the independent variables from a sample of banking companies. Firm's size is calculated from the logarithm of the company's total assets. The average value of the firm's size shows an average value of 27.93. The standard deviation of the company size is 1.15 with the lowest value of 26.57 which is the ALDO stock code in the 2014 period, while the highest value is 30.40, which is AUTO in 2018, which means AUTO has the largest company size compared to with manufacturing sector companies listed on other Indonesian Stock Exchanges measured by the logarithm value of total assets of banking companies while company size on ALDO has the smallest company size compared to manufacturing sector companies listed on other Indonesian Stock Exchanges in 2014-2018 measured from natural logarithm value of total assets.

### D. Estimation of the Regression Model Coefficient

The multiple linear regression model used to estimate ROA (Return-on-Asset) and MtBV (Market-to-Book-Value) is carried out in a panel by combining 11 cross-sectional data multiplied by 5 years of the 2014-2018 research period so that 55 observation (Table III).

TABLE III. REGRESSION MODEL COEFFICIENT ESTIMATION RESULTS

Independent Variable		Expected Sign	ROA
Intercept	Coeff.	-	-2,706223
FIRM'S SIZE	Coeff.	+	1,077067

Source: R Studio Application Processing Results

The estimated coefficient of -0.242502 can be interpreted that ROA will have a value of -0.242502 if the DER variable has a value equal to 1, and the other independent variables are constant or can be interpreted as each DER increase of 1 unit will decrease ROA by -0.242502. While the negative sign (-) indicates the opposite relationship between the DER variable with ROA; and the estimated coefficient of 1.077067 can be interpreted that ROA will have a value of 1.077067 if the Firm's Size variable has a value equal to 1, and the other independent variables are constant or can be interpreted as each Firm's Size increase of 1 unit will increase ROA by 1.077067. While the positive sign (+) indicates a direct relationship between the Firm's Size variable and ROA [28].

### E. Partial Test of Regression Model 2

In this section, we will show the partial test results on the MtBV (Market-to-Book-Value) multiple linear regression equation, as follows (Table IV):

TABLE IV. PARTIAL TEST RESULT OF REGRESSION MODEL 2

Uji Parsial	Sig.
DER	0,69437
FIRM'S SIZE	0,32109

Source: R Studio Application Processing Results

Based on the partial test results, namely, z-test statistics, the following conclusions can be obtained:

Firm's Size variable has the result of a significance value (Sig.) Of 0.32109; where the value is greater than 0.05; thus indicating that the independent variable Firm's Size does not significantly influence the MtBV dependent variable [30].

#### F. Discussion

1) Firm's Size influences Return-on-Asset: Based on the hypothesis carried out in the previous chapter, the first hypothesis proposed in this study is as follows:

Ha1 = Firm's Size has a positive effect on Return-on-Assets.

Firm's Size variable produces a significance value of 0.17367; which means that the significance value is greater than 0.05 then Firm Size does not have a significant effect on ROA (Return-on-Asset). While the coefficient value of the Firm's Size variable regression model for ROA (Return-on-Asset) is 1.077067. Therefore, it can be concluded that the Firm's Size variable has no significant and positive effect on ROA (Return-on-Asset) with a 95% confidence level. So from testing this hypothesis, Ha is rejected and H0 is accepted.

The results of this study are also supported by the research of [29] which shows that firm's size does not affect company performance consisting of return-on-assets, market-to-book-value or tobin's q. This is because the firm's size measured by the natural logarithm of the total assets does not reflect the actual value of the assets so the firm's size cannot be a guarantee of good company performance.

2) Firm's Size affects Market-to-Book-Value: Based on the hypothesis carried out in the previous chapter, the 8th hypothesis proposed in this study is as follows:

Ha2 = Firm's Size has an influence on Market-to-Book-Value.

Firm's Size variable produces a significance value of 0.32109; which means that the significance value is greater than 0.05, Firm Size does not have a significant effect on MtBV (Market-to-Book-Value) and has a positive effect based on the sign of the Firm Size Variable regression model coefficient which is equal to 3.43163. Therefore, it can be concluded that the Firm's Size variable does not have a significant positive effect on MtBV (Market-to-Book-Value) with a 95% confidence level. So from testing this hypothesis, Ha is rejected and H0 is accepted.

Based on the results of the study, it can be interpreted that the firm's size variable does not support signaling theory which states that the greater the size of a company can improve the company's financial performance. The reason why firm's size does not have a significant positive effect on

MtBV (Market-to-Book-Value), because the public does not see the size of the manufacturing sector in Indonesia as a determinant of the success of a company's performance [26].

There are several reasons why firm's size does not have a significant positive effect on MtBV (Market-to-Book-Value) because even though large-sized companies have relatively large assets, they may not necessarily explain high returns to shareholders [27].

#### VI. CONCLUSION

Based on the results of the rainfall and the discussion carried out, a number of conclusions is obtained, as follows:

1) Based on the test results, firm's size has no effect on firm's financial performance which is proxied by return-on-assets.

2) Based on the test results, firm size does not affect firm's financial performance which is proxied by market-to-book-value.

#### ACKNOWLEDGMENT

Thanks to BINUS University Jakarta Indonesia.

#### REFERENCES

- [1] Aghekyan-Simonian, M., Forsythe, S., Kwon, W. S., & Chattaraman, V. (2012). The role of product brand image and online store image on perceived risks and online purchase intentions for apparel. *Retailing and Customer Services*, 325-331.
- [2] Alawad, A., Aljoufie, M., Tiwari, A., & Daghestani, L. (2015). Beyond Geographical and Cultural Barriers: The Concept of a Virtual Gallery for Arts, Design & Architecture Schools in Saudi Arabia. *Art and Design Review*, 87-93.
- [3] Att Petcharit, Puris Sornsaruht, Paitoon Pimdee. (2020). An Analysis of Total Quality Management (TQM) within the Thai Auto Parts Sector. Vol 16, No 02 (2020). *International Journal of Online and Biomedical Engineering*.
- [4] Arshi Naim, Fahad Alahmari. (2020). Reference Model of E-Learning and Quality to Establish Interoperability in Higher Education Systems. Vol 15, No 02 (2020). *International Journal of Emerging Technologies in Learning*.
- [5] Abeywardhana, D. K. Y. (2016). Impact of Capital Structure on Firm Performance: Evidence from the Manufacturing Sector of SMEs in the UK. *SSRN Electronic Journal*.
- [6] Ahangar, R. G. (2011). The relationship between intellectual capital and financial performance: An empirical investigation in an Iranian company. *African journal of business management*, 5 (1), 88-95.
- [7] Alusia, Y. H. & Pangestuti, I. R. D. (2015). Analysis of the Effect of Non-Performing Loans, Investment Proportion, Loan To Deposit Ratio, Liability to Equity Ratio and Bank Size Against Shareholder Value in Conventional Banks Listing on the Indonesia Stock Exchange in 2011 to 2014. *Diponegoro Journal of Management*, 4 (4), 1-9.
- [8] Ashraf, M., Ameen, A., & Shahzadi, K. (2017). The Impact of Capital Structure on Firm's Profitability: A Case of Cement Industry of Pakistan. *International Journal of Business and Social Science*, 8 (4), 140-147.
- [9] Atidhira, A.T. & Yustina, A.I. (2017). The Influence of Return on Assets, Debt to Equity Ratio, Earnings per Share, and Company Size on Share Returns in Property and Real Estate Companies. *JAAF (Journal of Applied Accounting and Finance)*, 1 (2), 128-146.
- [10] Azzahra, A. S. & Fate. (2019). Effect of Firm Size and Leverage Ratio on Financial Performance in Mining Companies. *JWEM STIE MICROCIL*, 9 (1), 13-20.
- [11] Azhar Susanto & Meiryani. 2019. The Quality of Accounting Information system and its impact on the quality of accounting information: user ability and top management support. *Journal of Engineering and applied sciences* 13(2), pp. 384-387.

- [12] Azhar Susanto & Meiryani. 2019. Antecedents of environmental management accounting and environmental performance : Evidence from Indonesian small and medium enterprises. *International Journal of Energy Economics and Policy*. 9(6), pp. 401-407.
- [13] Central Statistics Agency (2018). Indonesia Gross Domestic Product Quarterly 2014-2018. ISSN: 1907-4557, No. Publication: 07130.1803.
- [14] Bchini, B. (2015). Intellectual capital and value creation in the Tunisian manufacturing companies. *Procedia Economics and Finance*, 23, 783-791.
- [15] Indonesia stock exchange. (2019). Financial and Annual Report. Retrieved October-November-December 2019, from the Indonesia Stock Exchange: [www.idx.co.id](http://www.idx.co.id)
- [16] CIMA (2003). *Understanding Corporate Value: Managing and Reporting Intellectual Capital*. Carnfield University School of Management. UK
- [17] Chen, M. C., Cheng, S. J., & Hwang, Y. (2005). An empirical investigation of the relationship between intellectual capital and firms' market value and financial performance. *Journal of intellectual capital*, 6 (2), 159-176.
- [18] Darmawan, M. & Toro, J. S. (2012). The Impact of Intellectual Capital on Banks Go Public's Market Value and Financial Performance Listed in Indonesia Stock Exchange (IDX). *Managerial Focus*, 11 (2), 164 - 182.
- [19] Devi, B. E., Khairunnisa, & Budiono, E. (2017). Effect of Intellectual Capital on Company Financial Performance. *e-Proceeding of Management*, 4 (1), 491-500.
- [20] Dey, R. K., Hossain, S. Z., & Rahman, R. A. (2018). Effect of Corporate Financial Leverage on Financial Performance: A Study on Publicly Traded Manufacturing Companies in Bangladesh, *Asian Social Science*, 14 (12), 124-133.
- [21] Dewi, Y. T. & Hatane, S. E. (2015). The Role of Employee Stock Ownership Plans in the Effect of Intellectual Capital Value Added on Financial Performance. *Business Accounting Review*, 3 (1), 478-488.
- [22] Dwipayani, C. C. & Prastiwi, A. (2014). Effect of Intellectual Capital on Profitability and Market Performance. *Diponegoro Journal of Accounting*, 3 (3), 1-9.
- [23] Egbunike, C. F. & Okerekeoti, C. U. (2018). Macroeconomic factors, firm characteristics and financial performance, *Asian Journal of Accounting Research*, 3 (2), 142-168.
- [24] Fu, X., Lin, Y., & Molyneux, P. (2014). Bank Efficiency and Shareholder Value in Asia Pacific. *Journal of International Financial Markets, Institutions & Money*, 33, 200-222.
- [25] Ghozali, I. (2013). *Application of Multivariate Analysis with SPSS 21*. Diponegoro University. Semarang.
- [26] Ghozali, I. (2016). *Multivariate Analysis Application with SPSS Program*. Semarang: Diponegoro University Publisher Agency, Semarang.
- [27] Gitman, L. J. & Zutter, C. J. (2015). *Principles of Managerial Finance*. Fourteenth Edition. United States of America: Pearson.
- [28] Godfrey, J., Hodgson, A., Tarca, A., Hamilton, J., & Holmes, S. (2010). *Accounting Theory*. Milton, Qld .: John Wiley & Sons, Inc.
- [29] Gupta, N. K. & Gupta, H. (2014). Impact of Capital Structure on Financial Performance in Indian Construction Companies. *International Journal of Economics, Commerce and Management United Kingdom*, 2 (5), 1-14.
- [30] Harjito, D. A. (2011). Pecking Order Theory and Trade-Off in Capital Structure Asnalysis on the Indonesia Stock Exchange. *SIASAT BUSINESS*, 15 (2), 187-196. 109.

# Image Captioning using Deep Learning: A Systematic Literature Review

Murk Chohan<sup>1</sup>, Adil Khan<sup>2</sup>, Muhammad Saleem Mahar<sup>3</sup>  
Saif Hassan<sup>4</sup>, Abdul Ghafoor<sup>5</sup>, Mehmood Khan<sup>6</sup>

Department of Computer Science  
Sukkur IBA University  
Pakistan

**Abstract**—Auto Image captioning is defined as the process of generating captions or textual descriptions for images based on the contents of the image. It is a machine learning task that involves both natural language processing (for text generation) and computer vision (for understanding image contents). Auto image captioning is a very recent and growing research problem nowadays. Day by day various new methods are being introduced to achieve satisfactory results in this field. However, there are still lots of attention required to achieve results as good as a human. This study aims to find out in a systematic way that what different and recent methods and models are used for image captioning using deep learning? What methods are implemented to use those models? And what methods are more likely to give good results. For doing so we have performed a systematic literature review on recent studies from 2017 to 2019 from well-known databases (Scopus, Web of Sciences, IEEEExplore). We found a total of 61 prime studies relevant to the objective of this research. We found that CNN is used to understand image contents and find out objects in an image while RNN or LSTM is used for language generation. The most commonly used datasets are MS COCO used in all studies and flicker 8k and flicker 30k. The most commonly used evaluation matrix is BLEU (1 to 4) used in all studies. It is also found that LSTM with CNN has outperformed RNN with CNN. We found that the two most promising methods for implementing this model are Encoder Decoder, and attention mechanism and a combination of them can help in improving results to a good scale. This research provides a guideline and recommendation to researchers who want to contribute to auto image captioning.

**Keywords**—Image Captioning; Deep Learning; Neural Network; Recurrent Neural Network (RNN); Convolution Neural Network (CNN); Long Short Term Memory (LSTM)

## I. INTRODUCTION

Auto image captioning is the process to automatically generate human like descriptions of the images. It is very dominant task with good practical and industrial significance [62]. Auto Image captioning has a good practical use in industry, security, surveillance, medical, agriculture and many more prime domains. It is not just very crucial but also very challenging task in computer vision [1]. Traditional object detection and image classification task just needed to identify objects within the image where the task of Auto image captioning is not just identifying the objects but also identifying the relationship between them and total scene understanding of the image. After understanding the scene it is also required to generate a human like description of that

image. Since the boost of automation and Artificial Intelligence lots of research is going on to give machine human like capabilities and reduce manual work. For machines acquiring results and accuracy as good as human in image captioning problem has always been a very challenging task.

Auto image captioning is performed by following key tasks in order. At first features are extracted after proper extraction of features different objects from an image are detected, after that the relationship between objects are to be identified (i.e. if objects are cat and grass it is to be identified that if cat in on grass). Once objects are detected and relationships are identified now it is required to generate the text description, i.e. Sequence of words in orderly form that they make a good sentence according to the relationship between the image objects.

To perform above key tasks using deep learning different deep learning networks are used. For Example to get visual features and objects CNN with different region proposing models like RCNN, Faster RCNN can be used and to generate text description in sequence RNN or LSTM can be used. Using these networks various different methods are developed to perform auto image captioning in various different domains. However, still, there is room for the machine to make capable enough to generate descriptions like a human [61]. . After training the Deep Learning network for image captioning to evaluate its performance various evaluation matrices like BLEU, CIDEr, and ROUGE-L exists.

The purpose of this Systematic Literature Review is to study all newest Articles from 2017 to 2019 to find different methods to achieve auto image captioning in different domains, what different datasets are used to achieve the task, In which different practical domains this task is used, which technique Outperforms others and finally attains to describe the technicalities behind different networks, methods and evaluation matrices. Our study will help new researchers who want to work in this domain to attain better accuracy. We specially focused and the collection of quality articles which have been published till now. We attempt to find our different techniques presented in [1- 60] articles, find their methods strengths and weakness. Finally we attempt to summarize them to explain which technique has better performance in its particular domain. Our work mostly focuses on identifying the most popular techniques. The areas in which yet there is

attention require and in result section we also attempt to explain the technical concepts behind the used approaches.

## II. METHODOLOGY

The planning conducting and reporting of this Systematic literature review is done step by step. First in planning section we identified the need of conducting this research its importance. Identifying the research questions and design search strategy, designing quality assessment criteria and finally designing data extraction strategy is also planned during this stage. After proper planning we have conducted

the research. In alignment with our research problem we have come up with research questions for which we try to find answers during this research.

### A. Research Questions

Before conducting this study we kept the following research questions to measure the quality of our work. This study basically provide a detailed knowledge related to these research questions. Table I provides the list of research questions.

TABLE I. LIST OF RESEARCH QUESTIONS

RQ#	Research Question	Motivation
RQ 1	How image captioning recognizes the important objects, attributes of objects and their relationships in an image?	Identifying DL techniques for object detection and relation finding mechanism
RQ 2	How Deep learning-based techniques are capable of handling the complexities and challenges of image captioning?	Identifying DL methods to handle challenges of image captioning
RQ 3	What deep learning techniques are used for image captioning?	Identifying DL techniques for Language generation as well as object detection
RQ 4	Which techniques outperform other techniques?	Comparison between several techniques
RQ 5	What datasets are used for Image Captioning?	Identifying different datasets used for image captioning
RQ 6	What evaluation mechanisms are used in literature for image captioning?	Identifying different methods to evaluate image captioning models

### B. Search Results

According to our research questions we came up with our search keywords and we categorized them in two different groups, shown in the Table II.

Using scientific approach for searching the results from different academic databases. We composed the query string from the keywords cited in Table II.

Query String: ("Image Captioning") AND ("Deep Learning" OR "Neural Network" OR "RNN" OR "LSTM" OR "CNN")

We applied the cited search query string on three well known academic databases namely IEEE Xplore, Web of Sciences and Scopus to search the articles. We adopted the most recent articles published during 2017-2019 from the journals, and our initial search results are illustrated in the Table III.

TABLE II. KEYWORDS IN TWO DIFFERENT GROUPS (GROUP 1 AND GROUP 2)

Keywords: Group 1	Keywords: Group 2
Image Captioning	Deep Learning, Neural Network, RNN, CNN, LSTM

TABLE III. INITIAL STAGE RESULTS FROM IEEE XPLORE, WEB OF SCIENCES AND SCOPUS

Database	Original search results
IEEE Xplore	247
Web of Science	167
Scopus	313
<b>Total</b>	<b>727</b>

Since an article can be indexed in many databases we removed the duplicate articles from either one of the database. After duplicate removal total number of studies from all three databases are shown in Table IV.

Abstract screening is also important to filter the searched studies to keep valuable studies that are more related to someone's work. We performed abstract screening on the 577 articles which were remained after duplicate removal to check out the relevance of studies with our work. We found many studies not relevant to our topic like some were about audio captioning or video captioning. After the abstract screening, we had a total of 308 studies out of 577 studies, Table V illustrates the total number of studies from each database after the abstract screening.

TABLE IV. NUMBER OF STUDIES OF DUPLICATE REMOVAL

Database	Duplicate removal results
IEEE Xplore	162
Web of Science	167
Scopus	248
<b>Total</b>	<b>577</b>

TABLE V. NUMBER OF STUDIES AFTER ABSTRACT SCREENING RESULTS

Database	Abstract screening results
IEEE Xplore	63
Web of Science	92
Scopus	143
<b>Total</b>	<b>308</b>

**C. Quality Assessment Criteria**

The quality of 308 articles was assessed for quality assessment criteria. We assessed the quality of selected 308 studies to ensure the quality assessment of our study. We went through the full text screening of those studies which were ambiguous and was not clear from abstract screening. The process of quality assessment criteria (QAC) was done with full text screening. All four authors agreed to make some quality assessment questions (QAQ) to ensure the quality of our work.

- QA Q1** The article must be published in journal
- QA Q2** Article has proposed a proper method to implement image captioning using deep learning.
- QA Q3** The article must have clear and unambiguous results.
- QA Q4** Article must discuss the applications and challenges of image captioning.
- QA Q5** Article must discuss the evaluation strategy of the built model.

We assessed the quality of 308 studies on the basis of quality assessment criteria (QAC) questions and through full text screening, we found total 61 studies from all three databases. Number of each studies from all three databases shown in Table VI.

The result which we found above illustrated in PRISMA diagram (see Fig. 1). All this process we this dissipated in following diagram.

**D. Data Extraction and Synthesis**

After selection of final 61 primary studies we extracted data from those studies for performing final synthesis. We defined our data extraction strategy based on our research questions. We have extracted following parameters from our primary studies for further synthesis, year or article published, title, models use for language generation and object detection, methods use to implement models, datasets used, evaluation matrices used for evaluation purpose and finally accuracy of proposed model.

The purpose of synthesis is to summarize the facts extracted in data extraction and give a clear picture of work done in past and directions to new researchers.

TABLE VI. NUMBER OF STUDIES AFTER QUALITY ASSESSMENT CRITERIA

Database	Quality assessment results
IEEE Xplore	12
Web of Science	16
Scopus	33
<b>Total</b>	<b>61</b>

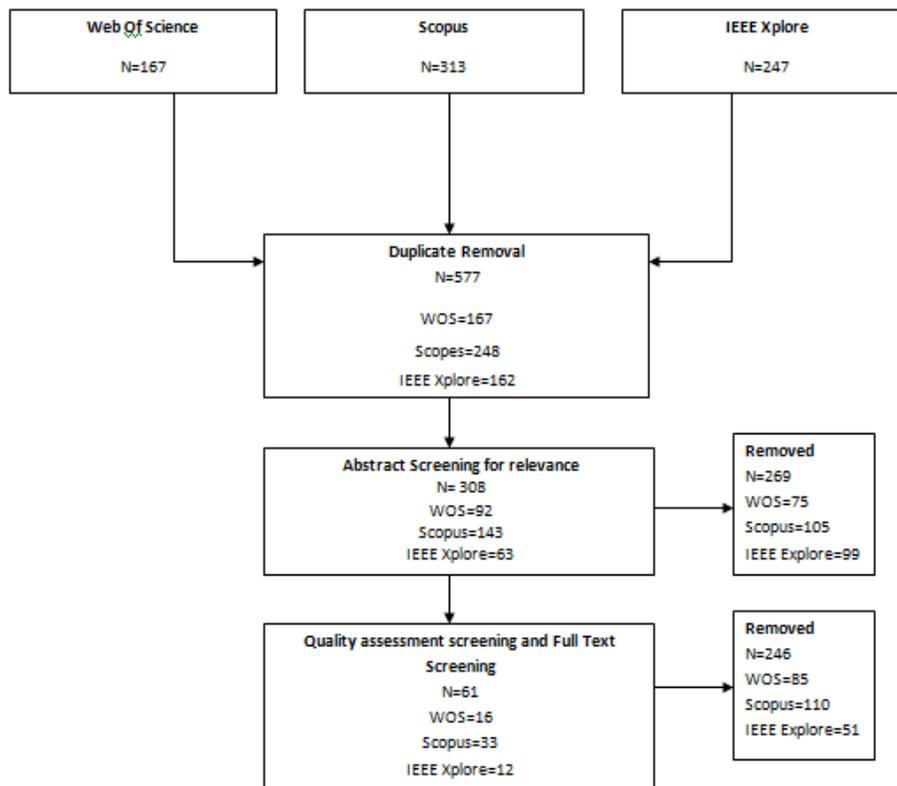


Fig. 1. PRISMA Diagram.

### III. RESULTS AND DISCUSSION

#### A. Datasets

There are many datasets available for performing image captioning. In literature most common used data sets are MS COCO and flicker 8k and 30k. Moreover for a text description of specific task like in medical or traffic movement description their own dedicated datasets are created. Fig. 2 below show the datasets along with their frequency in our selected studies.

1) *MSCOCO*: MS COCO stands for common object in context. It is very large dataset which contains 330k images, 1.5 object instances and 5 captions per image. MS COCO is found to be very widely used dataset in literature. It is very best suited for image captioning because unlike other datasets it contains non iconic images. Iconic images are those images which contains only one object with a background where as non-iconic images contains various objects overlapping. Object layout plays an important role in understanding context of scene and that is very carefully taken care of while labeling images. Fig. 3 shows some images taken from MS COCO dataset.

2) *Deep learning networks*: Deep learning network used for images is Convolution neural network. CNN has been proved best to map image data into output variable. There are various prebuilt model that take advantage from this feature of CNN i.e. RCNN faster RCNN etc. these models are used for object detection and localization in images which is very necessary task in image captioning since it's not just classification task and understanding image contents is necessary. Once image data is understood there is need of predicting the sequence of words to generate the text for that particular image. For sequence prediction two most famous networks are Recurrent Neural Network (RNN) and long short term memory (LSTM). For image captioning generation task CNN is either used with RNN or LSTM where CNN is used for understanding image contents and RNN or LSTM for text description generation. Fig. 4 and Table VII represents the number of studies that have used RNN or LSTM with CNN. Fig. 5 shows use of CNN and RNN networks for image captioning in year 2017 to 2019.

In terms of performance we Compared BLEU-1 performance of both text prediction networks and found out that LSTM outperforms RNN in terms of accuracy. Fig. 6 shows the result of top 5 highest accuracy achieving papers for both networks.

3) *Convolution Neural Network (CNN)*: Convolution Neural Network is an algorithm of Deep Learning which is normally used to process images. CNN is an evolution of simple ANN that gives better result on images. Simple dense network is best for classification tasks where some features are used to classify the image. CNN performs best with more features in an image. It is used to process the local features as well. Because images contain repeating patterns of particular

thing (any image). It takes images as an input and understands it to perform assigned tasks. Two main functions of CNN are convolution and pooling. Convolution is used in CNN to detect the edges of an image and pooling is used to reduce the size of an image. It is a method in which we take a small number matrix called kernel or filter then move it over our picture and convert it depending on the filter values. Following formula is used to calculate the feature map, where  $f$  is used to denote input image and  $h$  is used to denote filter. The outcome matrix rows and column indexes are labeled with  $m$  and  $n$ , respectively.

$$G[M, N] = (F * H)[M, N] \\ = \sum_J \sum_K (H[J, K] F[M - J, N - K])$$

Calculation of convolution layer done in two steps. First step is used to calculate the intermediate value  $Z$ , and its addition with bias. Second step is to apply non-linear activation  $g$  with intermediate value.

$$Z^{[l]} = W^{[l]} * A^{[l-1]} + b^{[l]}$$

$$A^{[l]} = g^{[l]}(Z^{[l]})$$

4) *Recurrent Neural Network (RNN)*: CNNs commonly do not do well in a sequential fashion when the input data is interrelated. CNNs have no connection of any kind between previous input and next data. So all of the outputs depend on themselves. Depending on the trained model, CNN takes input and gives output. For doing above task RNN is used. RNN have its memory, so that it is able to remind what happened earlier in data. Earlier means previous inputs. RNN performs best on textual data because text is interrelated (sequential data). Basic formula for RNN is written below.

$$h(t) = f(h^{[t-1]}, x(t); \theta)$$

$f$  is a function of current hidden state  $h$ .  $h^{(t-1)}$  is a previous hidden state,  $x(t)$  is current input, and  $\theta$  is a parameter of function.

5) *Long Short Term Memory (LSTM)*: LSTM is a variant of RNN. It is better than simple RNN because it solves the issues faced by simple RNN. Two major issues faced by simple RNN is (i) exploding gradient and vanishing gradient and (ii) long term dependency. LSTM uses gates to remember the past and gates are the heart of LSTM. Gates which are available in LSTM are (i) input gate (ii) forget gate and (iii) output gate. They all are sigmoid activation function. Sigmoid means output between 0 and 1, mostly 0 or 1. When output is 0, it means gate is blocking. If output is 1 then pass everything. Below is the equation for above defined gates.

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i).$$

#### B. Evaluation Mechanism

Evaluating the trained model is quiet difficult task in image captioning for this purpose various evaluation matrices are created. Most common evaluation mechanisms found in literature are BLEU, ROUGE-L, CIDEr, METEOR, and

SPICE. It is found that BLEU score is most popular method of evaluation used by almost all of the studies. You can verify this from given Fig. 7 and Table VIII.

6) *BLEU*: BLEU stands for bilingual evaluation understudy. It is an evaluation mechanism widely use in text generation. It is a mechanism for comparing the machine generated text with one or more manually written text. So basically it summarizes that how close a generated text is to an expected text. BLEU score is majorly prevalent in automated machine translation but it can be also used in image captioning, text summarization, speech recognition etc. Particularly in image captioning the BLEU score is accuracy that how close a generated caption is to a manual human generated caption of that particular image. The score scale lies between 0.0 to 1.0. Where 1.0 is perfect score and 0.0 is worst score.

We found that almost all studies used bleu as their evaluation matrix and they calculated BLEU-1 to 4 where BLEU-1 is calculating accuracy only on 1 gram, BLEU-2 for 2 grams, BLEU-3 for 3 grams and BLEU-4 for 4 grams.

The BLEU score can be calculated from following formula.

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N w_n \log p_n\right)$$

7) *METEOR*: METEOR stands for metric for evaluation and translation with explicit ordering. While BLEU takes account of entire text generated overshadowing the score of each and individual sentence generated the METEOR takes care of that. For doing so METEOR enhances the precision and recall functions. Instead of precision and recall the meteor utilizes weighted F-score for mapping unigram and for incorrect word order it uses penalty function.

Formula for weighted function is:

$$F = \frac{PR}{\alpha P + (1 - \alpha)R}$$

Where P and R stands for precision and recall calculated as m/c and m/r, where c and r are candidate and reference length and m is number of mapped unigrams among two texts.

Formula for Penalty function is:

$$Penalty = \gamma \left(\frac{c}{m}\right)^\beta, \text{ where } 0 \leq \gamma \leq 1$$

Where c is number of matched chunks and m is total number of matches.

Over all meteor score is found by:

$$M = (1 - Penalty) * F$$

8) *ROUGE-L*: ROUGE stands for recall oriented understudy for gisting evaluation. As clear from its name ROUGE is only based on recall but ROUGE-L is based on its F score which is harmonic mean of its precision and recall values. Following are the formulas for calculating precision, recall and F values

$$P = \frac{LCS(A, B)}{m} \text{ and } R = \frac{LCS(A, B)}{n}$$

Here A and B are candidate and reference generated text and m and n are their lengths and LCS stands for longest common sequence since ROUGE-L depends on longest common sequence.

Now for calculating F their harmonic means are calculated.

$$F = \frac{(1 + b^2)RP}{R + b^2P}$$

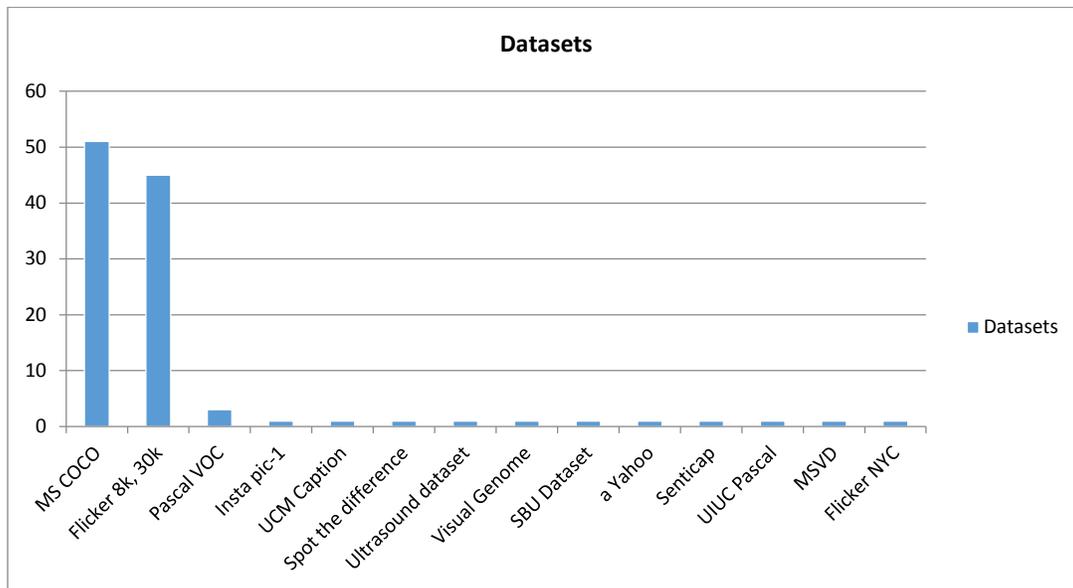


Fig. 2. Datasets used for Image Captioning in Selected Studies.



Fig. 3. MS COCO Dataset Images.

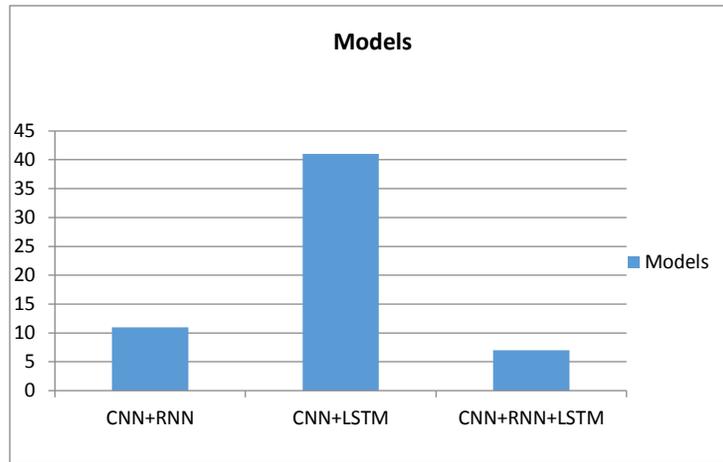


Fig. 4. Deep Learning Model used for Image Captioning in Literature.

TABLE VII. DEEP LEARNING MODEL USED FOR IMAGE CAPTIONING IN LITERATURE

SR#	STUDIES	NETWORKS	
		LSTM	RNN
1	[1],[2],[4],[6],[8],[10],[11],[13],[14],[15],[16],[18],[19],[20],[22],[23],[26],[27],[28],[29],[31],[32],[33],[36],[38],[37],[40],[41],[42],[43],[44],[45],[46],[47],[49],[50],[52],[53],[54],[57],[58]	✓	
2	[5],[7],[9],[12],[17],[30],[34],[35],[51],[56],[21]		✓
3	[18],[24],[25],[39],[48],[59],[60]	✓	✓

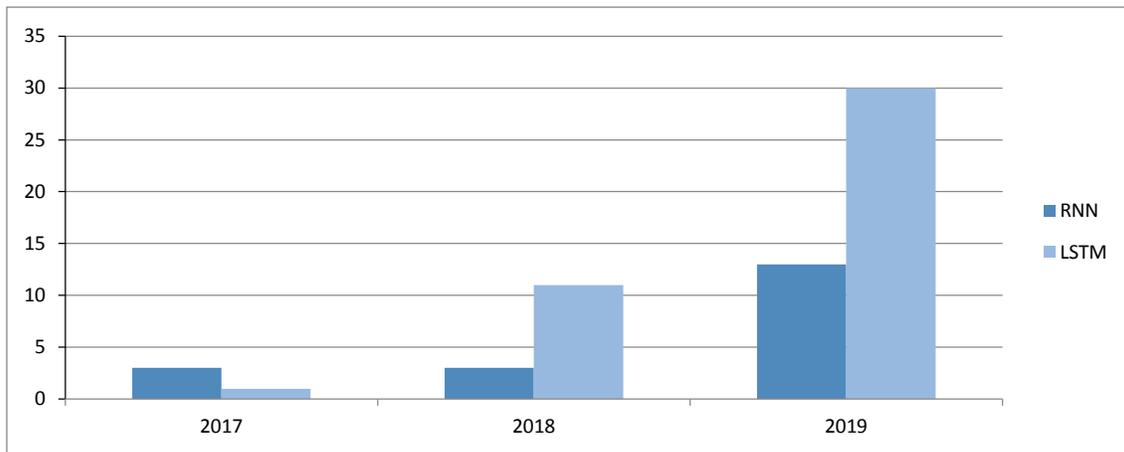


Fig. 5. Use of CNN and RNN Networks through the Years.

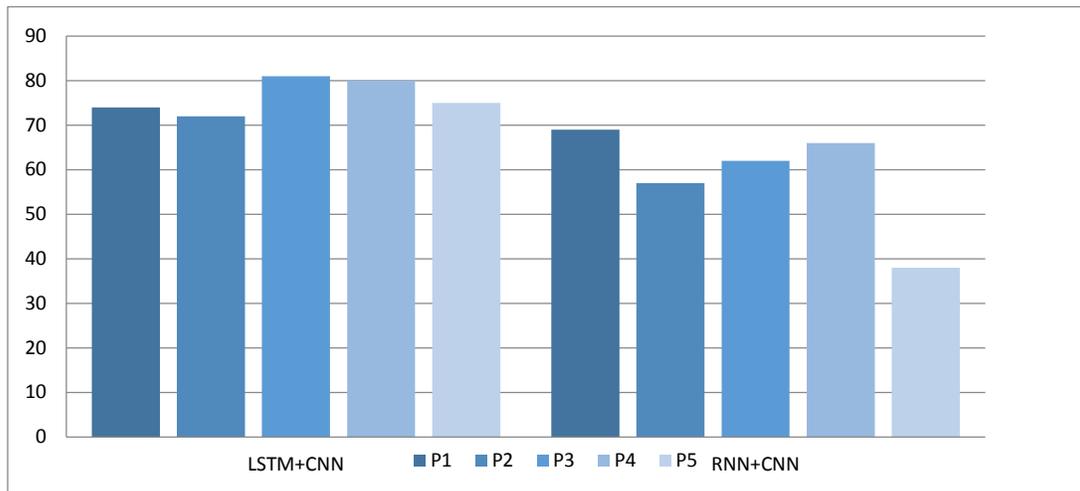


Fig. 6. Comparison of Best Score Achieved by RNN and LSTM (B1 Result Comparison).

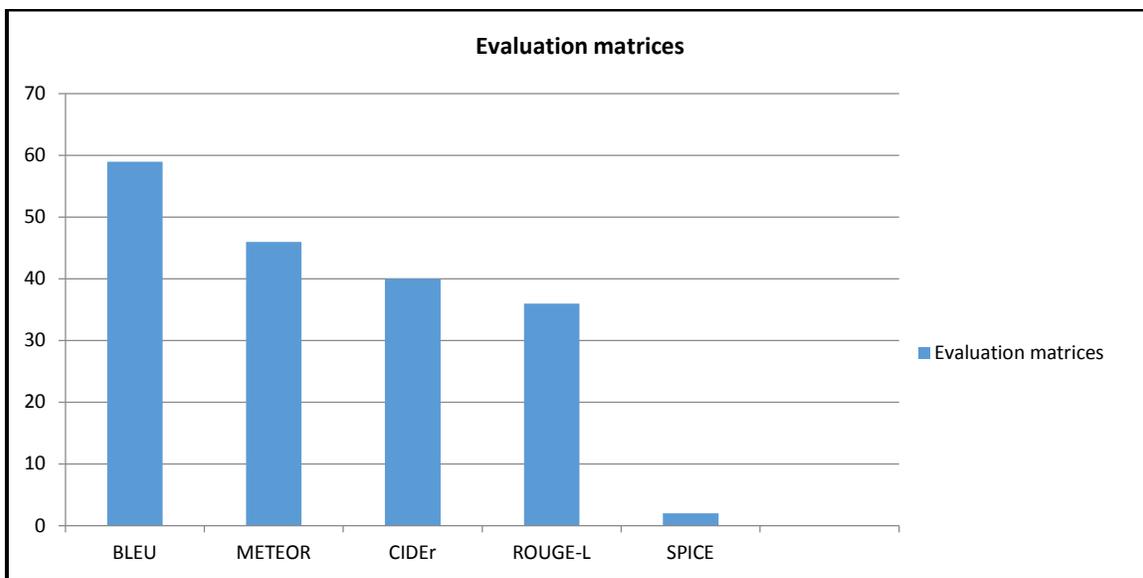


Fig. 7. Evaluation Metrics used in Literature.

TABLE VIII. EVALUATION METRICS USED IN LITERATURE

Sr.No.	STUDIES	EVALUATION METRICS				
		B	M	C	R	S
1	[1], [2], [3], [4], [6], [8], [10], [11], [13], [15], [16], [18], [20], [22], [26], [29], [30], [31], [32], [33], [38], [39], [40], [42], [43], [44], [49], [50], [52], [53], [54], [55], [56], [57],	✓	✓	✓	✓	
2	[5], [12], [19], [23], [34], [27], [28], [36], [37], [45], [51], [58], [59]	✓	✓	✓		
3	[7], [35]	✓	✓			
4	[9], [41], [47], [48]	✓				
5	[17]	✓	✓		✓	✓
6	[25]	✓	✓		✓	
7	[34]			✓		
8	[46]	✓		✓	✓	✓
9	[60]	✓	✓	✓	✓	✓

#### IV. CONCLUSION

This systematic literature review (SLR) presents a detailed analysis of different deep learning models used for image captioning. To perform the study we searched articles from three academic databases, after applying inclusion and exclusion criteria on all article and we selected 61 primary studies to perform a literature review. Using data extraction mechanism we extracted the data and analyzed it deeply. We found various different models and techniques used for image captioning. For image content extraction CNN is the best-suited model and for language generation two frequently used models are RNN and LSTM. It is found that LSTM has outperformed RNN. We also found different studies have used several different mechanisms for scene understanding like encoder-decoder mechanism and attention mechanism. The most suitable dataset for image captioning is MSCOCO because it contains non-iconic images, unlike other datasets.

Throughout our review, we have observed that image captioning is mostly used generally. There are various domains that can take advantage of image captioning to automate their tasks.

1) A model can be trained in medical ultrasound or MRI images or angiographic videos to generate a complete report of a person without any consent from a doctor. Image captioning can be used to generate an automatic report by looking at those medical images of a person.

2) Image captioning can also be used in industries to automate various tasks. A model can be trained on images of a company product manufacturing environment to find out an anomaly in the environment or product automatically. It can also be used also used to detect any mishap in a company like fire or security issues.

3) Image captioning can also be used in agriculture to generate the report of crops for owners by looking at images of crops.

4) Image captioning can also be used in traffic analysis report generation by using CCTV cameras installed on streets and thus guide drivers which is the best suitable path to take and where parking is available.

#### REFERENCES

- [1] Yang, L., & Hu, H. (2019). Adaptive syncretic attention for constrained image captioning. *Neural Processing Letters*, 50(1), 549-564.
- [2] Fu, K., Jin, J., Cui, R., Sha, F., & Zhang, C. (2016). Aligning where to see and what to tell: Image captioning with region-based attention and scene-specific contexts. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2321-2334.
- [3] Li, J., Yao, P., Guo, L., & Zhang, W. (2019). Boosted Transformer for Image Captioning. *Applied Sciences*, 9(16), 3260.
- [4] Oluwasanmi, A., Aftab, M. U., Alabdulkreem, E., Kumeda, B., Baagyere, E. Y., & Qin, Z. (2019). CaptionNet: Automatic end-to-end siamese difference captioning model with attention. *IEEE Access*, 7, 106773-106783.
- [5] Wu, J., & Hu, H. (2017). Cascade recurrent neural network for image caption generation. *Electronics Letters*, 53(25), 1642-1643.
- [6] Tan, J. H., Chan, C. S., & Chuah, J. H. (2019). COMIC: Toward A Compact Image Captioning Model With Attention. *IEEE Transactions on Multimedia*, 21(10), 2686-2696.
- [7] Huang, G., & Hu, H. (2019). c-RNN: A Fine-Grained Language Model for Image Captioning. *Neural Processing Letters*, 49(2), 683-691.
- [8] Xiao, F., Gong, X., Zhang, Y., Shen, Y., Li, J., & Gao, X. (2019). DAA: Dual LSTMs with adaptive attention for image captioning. *Neurocomputing*, 364, 322-329.
- [9] Dhir, R., Mishra, S. K., Saha, S., & Bhattacharyya, P. (2019). A Deep Attention based Framework for Image Caption Generation in Hindi Language. *Computación y Sistemas*, 23(3).
- [10] Xiao, X., Wang, L., Ding, K., Xiang, S., & Pan, C. (2019). Deep Hierarchical Encoder-Decoder Network for Image Captioning. *IEEE Transactions on Multimedia*, 21(11), 2942-2956.
- [11] Zeng, X., Wen, L., Liu, B., & Qi, X. (2019). Deep learning for ultrasound image caption generation based on object detection. *Neurocomputing*.
- [12] Karpathy, A., & Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3128-3137).
- [13] Xiao, X., Wang, L., Ding, K., Xiang, S., & Pan, C. (2019). Dense semantic embedding network for image captioning. *Pattern Recognition*, 90, 285-296.
- [14] Han, M., Chen, W., & Moges, A. D. (2019). Fast image captioning using LSTM. *Cluster Computing*, 22(3), 6143-6155.
- [15] Dash, S. K., Saha, S., Pakray, P., & Gelbukh, A. (2019). Generating image captions through multimodal embedding. *Journal of Intelligent & Fuzzy Systems*, 36(5), 4787-4796.
- [16] Li, L., Tang, S., Zhang, Y., Deng, L., & Tian, Q. (2017). Gla: Global-local attention for image description. *IEEE Transactions on Multimedia*, 20(3), 726-737.
- [17] Kinghorn, P., Zhang, L., & Shao, L. (2019). A hierarchical and regional deep learning architecture for image description generation. *Pattern Recognition Letters*, 119, 77-85.
- [18] Su, Y., Li, Y., Xu, N., & Liu, A. A. (2019). Hierarchical deep neural network for image captioning. *Neural Processing Letters*, 1-11.
- [19] Zhang, Z., Wu, Q., Wang, Y., & Chen, F. (2018). High-quality image captioning with fine-grained and semantic-guided visual attention. *IEEE Transactions on Multimedia*, 21(7), 1681-1693.
- [20] Shetty, R., Tavakoli, H. R., & Laaksonen, J. (2018). Image and video captioning with augmented neural architectures. *IEEE MultiMedia*, 25(2), 34-46.
- [21] Dong, J., Li, X., & Snoek, C. G. (2018). Predicting visual features from text for image and video caption retrieval. *IEEE Transactions on Multimedia*, 20(12), 3377-3388.
- [22] Ding, S., Qu, S., Xi, Y., Sangaiah, A. K., & Wan, S. (2019). Image caption generation with high-level image features. *Pattern Recognition Letters*, 123, 89-95.
- [23] Wu, Q., Shen, C., Wang, P., Dick, A., & van den Hengel, A. (2017). Image captioning and visual question answering based on attributes and external knowledge. *IEEE transactions on pattern analysis and machine intelligence*, 40(6), 1367-1381.
- [24] Yang, J., Sun, Y., Liang, J., Ren, B., & Lai, S. H. (2019). Image captioning by incorporating affective concepts learned from both visual and textual components. *Neurocomputing*, 328, 56-68.
- [25] Guo, R., Ma, S., & Han, Y. (2019). Image captioning: from structural tetrad to translated sentences. *Multimedia Tools and Applications*, 78(17), 24321-24346.
- [26] Zhang, X., He, S., Song, X., Lau, R. W., Jiao, J., & Ye, Q. (2019). Image captioning via semantic element embedding. *Neurocomputing*.
- [27] Cao, P., Yang, Z., Sun, L., Liang, Y., Yang, M. Q., & Guan, R. (2019). Image Captioning with Bidirectional Semantic Attention-Based Guiding of Long Short-Term Memory. *Neural Processing Letters*, 50(1), 103-119.
- [28] Wang, C., Yang, H., & Meinel, C. (2018). Image captioning with deep bidirectional LSTMs and multi-task learning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(2s), 1-20.
- [29] Chen, H., Ding, G., Lin, Z., Guo, Y., Shan, C., & Han, J. (2019). Image captioning with memorized knowledge. *Cognitive Computation*, 1-14.

- [30] He, C., & Hu, H. (2019). Image captioning with text-based visual attention. *Neural Processing Letters*, 49(1), 177-185.
- [31] Zhu, X., Li, L., Liu, J., Li, Z., Peng, H., & Niu, X. (2018). Image captioning with triple-attention and stack parallel LSTM. *Neurocomputing*, 319, 55-65.
- [32] Guo, R., Ma, S., & Han, Y. (2019). Image captioning: from structural tetrad to translated sentences. *Multimedia Tools and Applications*, 78(17), 24321-24346.
- [33] Fu, K., Li, J., Jin, J., & Zhang, C. (2018). Image-text surgery: Efficient concept learning in image captioning by generating pseudopairs. *IEEE transactions on neural networks and learning systems*, 29(12), 5910-5921.
- [34] W.-Y. Lan, X.-X. Wang, G. Yang, X.-R. Li (2019) Improving Chinese Image Captioning by Tag Prediction
- [35] Li, X., & Jiang, S. (2019). Know more say less: Image captioning based on scene graphs. *IEEE Transactions on Multimedia*, 21(8), 2117-2130.
- [36] Chen, X., Zhang, M., Wang, Z., Zuo, L., Li, B., & Yang, Y. (2018). Leveraging unpaired out-of-domain data for image captioning. *Pattern Recognition Letters*.
- [37] Fang, F., Wang, H., Chen, Y., & Tang, P. (2018). Looking deeper and transferring attention for image captioning. *Multimedia Tools and Applications*, 77(23), 31159-31175.
- [38] Wang, E. K., Zhang, X., Wang, F., Wu, T. Y., & Chen, C. M. (2019). Multilayer dense attention model for image caption. *IEEE Access*, 7, 66358-66368.
- [39] Xu, N., Zhang, H., Liu, A. A., Nie, W., Su, Y., Nie, J., & Zhang, Y. (2019). Multi-Level Policy and Reward-Based Deep Reinforcement Learning Framework for Image Captioning. *IEEE Transactions on Multimedia*.
- [40] Yu, J., Li, J., Yu, Z., & Huang, Q. (2019). Multimodal transformer with multi-view visual representation for image captioning. *IEEE Transactions on Circuits and Systems for Video Technology*.
- [41] Cui, W., Zhang, D., He, X., Yao, M., Wang, Z., Hao, Y., ... & Huang, J. (2019). Multi-Scale Remote Sensing Semantic Analysis Based on a Global Perspective. *ISPRS International Journal of Geo-Information*, 8(9), 417.
- [42] Yang, M., Zhao, W., Xu, W., Feng, Y., Zhao, Z., Chen, X., & Lei, K. (2018). Multitask learning for cross-domain image captioning. *IEEE Transactions on Multimedia*, 21(4), 1047-1061.
- [43] Ding, G., Chen, M., Zhao, S., Chen, H., Han, J., & Liu, Q. (2019). Neural image caption generation with weighted training and reference. *Cognitive Computation*, 11(6), 763-777.
- [44] Su, J., Tang, J., Lu, Z., Han, X., & Zhang, H. (2019). A neural image captioning model with caption-to-images semantic constructor. *Neurocomputing*, 367, 144-151.
- [45] Poleak, C., & Kwon, J. (2019). Parallel Image Captioning Using 2D Masked Convolution. *Applied Sciences*, 9(9), 1871.
- [46] Tan, Y. H., & Chan, C. S. (2019). Phrase-based image caption generator with hierarchical LSTM network. *Neurocomputing*, 333, 86-100.
- [47] Wu, L., Xu, M., Wang, J., & Perry, S. (2019). Recall What You See Continually Using GridLSTM in Image Captioning. *IEEE Transactions on Multimedia*.
- [48] Guan, J., & Wang, E. (2018). Repeated review based image captioning for image evidence review. *Signal Processing: Image Communication*, 63, 141-148.
- [49] Xu, N., Liu, A. A., Liu, J., Nie, W., & Su, Y. (2019). Scene graph captioner: Image captioning based on structural visual representation. *Journal of Visual Communication and Image Representation*, 58, 477-485.
- [50] Xian, Y., & Tian, Y. (2019). Self-Guiding Multimodal LSTM—When We Do Not Have a Perfect Training Dataset for Image Captioning. *IEEE Transactions on Image Processing*, 28(11), 5241-5252.
- [51] Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2016). Show and tell: Lessons learned from the 2015 mscoco image captioning challenge. *IEEE transactions on pattern analysis and machine intelligence*, 39(4), 652-663.
- [52] Ding, S., Qu, S., Xi, Y., & Wan, S. (2019). Stimulus-driven and concept-driven analysis for image caption generation. *Neurocomputing*.
- [53] Lu, S., Hu, R., Liu, J., Guo, L., & Zheng, F. (2019). Structure Preserving Convolutional Attention for Image Captioning. *Applied Sciences*, 9(14), 2888.
- [54] Yu, N., Hu, X., Song, B., Yang, J., & Zhang, J. (2018). Topic-oriented image captioning based on order-embedding. *IEEE Transactions on Image Processing*, 28(6), 2743-2754.
- [55] Park, C. C., Kim, B., & Kim, G. (2018). Towards personalized image captioning via multimodal memory networks. *IEEE transactions on pattern analysis and machine intelligence*, 41(4), 999-1012.
- [56] Yang, L., & Hu, H. (2017). TVPRNN for image caption generation. *Electronics Letters*, 53(22), 1471-1473.
- [57] Zhang, Z., Zhang, W., Diao, W., Yan, M., Gao, X., & Sun, X. (2019). VAA: Visual Aligning Attention Model for Remote Sensing Image Captioning. *IEEE Access*, 7, 137355-137364.
- [58] He, X., Yang, Y., Shi, B., & Bai, X. (2019). VD-SAN: Visual-densely semantic attention network for image caption generation. *Neurocomputing*, 328, 48-55.
- [59] Li, X., Yuan, A., & Lu, X. (2019). Vision-to-language tasks based on attributes and attention mechanism. *IEEE transactions on cybernetics*.
- [60] Yang, L., & Hu, H. (2019). Visual Skeleton and Reparative Attention for Part-of-Speech image captioning system. *Computer Vision and Image Understanding*, 189, 102819.
- [61] Staniūtė, R., & Šešok, D. (2019). A Systematic Literature Review on Image Captioning. *Applied Sciences*, 9(10), 2024.

# A Trait-based Deep Learning Automated Essay Scoring System with Adaptive Feedback

Mohamed A. Hussein<sup>1</sup>

National Center for Examination and Educational  
Evaluation- NCEEE  
Cairo, Egypt

Hesham A. Hassan<sup>2</sup>, Mohammad Nassef<sup>3</sup>

Faculty of Computers and Information  
Cairo University  
Cairo, Egypt

**Abstract**—Numerous Automated Essay Scoring (AES) systems have been developed over the past years. Recent advances in deep learning have shown that applying neural network approaches to AES systems has accomplished state-of-the-art solutions. Most neural-based AES systems assign an overall score to given essays, even if they depend on analytical rubrics/traits. The trait evaluation/scoring helps to identify learners' levels of performance. Besides, providing feedback to learners about their writing performance is as important as assessing their level. Producing adaptive feedback to the learners requires identifying the strengths/weaknesses and the magnitude of influence of each trait. In this paper, we develop a framework that strengthens the validity and enhances the accuracy of a baseline neural-based AES model with respect to traits evaluation/scoring. We extend the model to present a method based on essay traits prediction to give trait-specific adaptive feedback. We explored multiple deep learning models for the automatic essay scoring task, and we performed several analyses to get some indicators from these models. The results show that Long Short-Term Memory (LSTM) based system outperformed the baseline study by 4.6% in terms of quadratic weighted Kappa (QWK). Moreover, the prediction of the traits scores enhance the efficiency of the prediction of the overall score. Our extended model is used in the *iAssistant*, an educational module that provides trait-specific adaptive feedback to learners.

**Keywords**—AES system; trait evaluation; adaptive feedback; deep learning; neural networks; ASAP

## I. INTRODUCTION

“Nothing we do to, or for our students is more important than our assessment of their work and the feedback we give them on it [1].” It is widely acknowledged that feedback is a critical element of learning [2]. Both scores and feedback are fundamental aspects of the learning process. Accurate scoring of learners' answers creates a fair way to assess learners' work, which is a very important aspect. However, giving feedback to learners about their answers helps them identify their weaknesses and improve their performance as well.

Rubrics are widely used in evaluating learners' answers to essay questions. Brookhart (2013) defines a rubric as “a coherent set of criteria for learners' work that includes descriptions of levels of performance quality on the criteria [3].” The definition identifies two significant aspects of a good rubric: coherent sets of criteria and descriptions of levels of performance for these criteria. There are two types of rubrics: analytic and holistic rubrics. An analytic rubric evaluates each

criterion separately, and a holistic rubric evaluates all criteria simultaneously. Each type has its advantages and disadvantages. Analytic rubrics give formative feedback to learners and are easier to link to instruction. Nevertheless, they take more time to score and achieve acceptable inter-rater reliability than holistic rubrics. Holistic rubrics are faster and suitable for summative assessment (assessment of learning). On the other hand, a single overall score does not communicate information about what to do to improve learning and is not useful for formative assessment (assessment for learning) [4]. It is also interesting to know that research showed that learners prefer AES feedback over peer feedback [5].

Over the past years, various AES systems have been developed to evaluate learners' responses to a given prompt (essay). AES systems automatically assess the quality of the written text and assign a score to each text. The efficiency of these systems depends on the agreement between the human-rater scores and the AES scores[6]. Research in deep learning has led to the development of neural network models for automatic essay scoring task moving away from feature engineering and found that utilizing neural networks to automatic essay scoring task has achieved state-of-the-art outcomes [7]. Utilizing the automatically learned features has added significant benefits to the efficiency of such systems as well [8] [9].

The vast majority of existing Neural based AES systems were developed for holistic scoring to given essays even if they depend on analytical rubrics/traits [10]. The trait evaluation/scoring helps to identify learners' levels of performance. Besides, providing feedback to the learners about their writing requires identifying the strengths/weaknesses and the magnitude of influence of each trait. Based on that, our goal is to develop a framework that strengthens the validity and enhances the accuracy of neural-based AES approaches with respect to traits evaluation/scoring. Using this framework should help in providing effective adaptive feedback to learners as well.

The following part of the paper is organized as follows: Section 2 describes a brief overview of related work. Section 3 describes the methods and materials, including the AES models (baseline and the augmented), dataset description, training, and testing, in addition to the evaluation metric. Reporting and discussion of results are in Section 4. Then, our conclusion and future improvements are in sections 5.

## II. RELATED WORK

PEG is the earliest AES system that was developed by Ellis Page in 1966. PEG was the starting spark for decades of research into AES. Then, many AES systems have been developed that analyze the quality of text and assign a score to it. AES systems use various manually tuned shallow and deep linguistic features [5].

AES systems can be classified into two main types: i) handcrafted discrete features-based type that is bounded to specific domains, which usually uses natural language processing, latent semantic analysis, or Bayesian network, etc. and ii) automatic feature extraction-based type which usually uses neural networks [5].

Several AES systems include automated scoring alongside providing feedback, e.g., for the first type, Criterion, MY Access, and Writing Pal. Criterion provides an overall score and a learner's feedback using E-rater and Critique as an AES component. Where the E-rater module performs the given essay automatic scoring task and Critique consists of a set of modules that detect mistakes/errors in mechanics, grammar, and usage. Then, it identifies the issues of discourse and style in writing. MY Access offers instant score and diagnostic feedback based on the IntelliMetric AES system to stimulate the learners to improve their writing ability [8]. Moreover, Writing Pal is classified as an intelligent tutoring system that is mainly concerned with learning tasks and provides the service of evaluating writing tasks with feedback [11]. It targets learners' writing strategies within providing automated feedback. However, it classified as a handcrafted discrete features-based system; the automatic essay scoring model is separate from the feedback part. It uses specific algorithms for each feedback category.

In particular, a few of the other type systems consider scoring the traits and providing the appropriate feedback for each essay. Woods et al. [12] established a new ordinal essay scoring model with extension to use essay traits prediction to give a formative trait-specific feedback to learners. Nevertheless, one of the concerns of their system that their Ordinal Logistic Regression (OLR) model does not perform accurately with large scoring ranges essays (like prompts 1 and 7 in ASAP dataset).

## III. MATERIALS AND METHODS

### A. Baseline Model

Taghipour and Ng [6], developed an AES system (AES<sub>T&N</sub>) based on neural networks, which automatically predicts the overall score of a given essay [10]. AES<sub>T&N</sub> takes the sequence of words in an essay as input; their model first uses a convolution layer to extract n-gram level features. These features, which capture the local textual dependencies among the words in an n-gram, are then passed to a recurrent layer composed of an LSTM network. It was trained and given state-of-the-art results on the Kaggle's ASAP dataset. The evaluation metric, which is used to evaluate the efficiency of the system, is Quadratic Weighted Kappa (QWK) [6], [8]. They used a 5-fold cross-validation, and for each fold, they distributed the dataset into 60%, 20%, and 20%; training,

development, and testing sets, respectively. AES<sub>T&N</sub> model architecture is illustrated in Fig. 1.

AES<sub>T&N</sub> results show that all model variations (Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Gated Recurrent Units (GRU), and LSTM) succeed to learn the task properly and its performance comparable to or better than the baseline (AES system called 'Enhanced AI Scoring Engine' (EASE)<sup>1</sup>). The authors reported that the LSTM based AES<sub>T&N</sub> system outperformed other neural networks (RNN, GRU, and CNN) systems significantly and outperformed the baseline by (4.1%).

AES<sub>T&N</sub> system has significantly outperformed the other AES systems, yet there is always an area for improvement to increase the accuracy of scoring. AES<sub>T&N</sub> system has predicted only the overall scores, although some of the essays have analytical rubrics/traits. Moreover, it has not provided any feedback to learners.

### B. Proposed Model

Our model (AES<sub>AUG</sub>) is inspired by the baseline model AES<sub>T&N</sub> of Taghipour and Ng [6]. We extend and utilize the AES<sub>T&N</sub> model to predict not only the overall score for essays but also the traits scores. Besides, we aim to utilize the traits scores to provide adaptive feedback to learners. Fig. 2 presents the AES<sub>AUG</sub> model architecture, which is described.

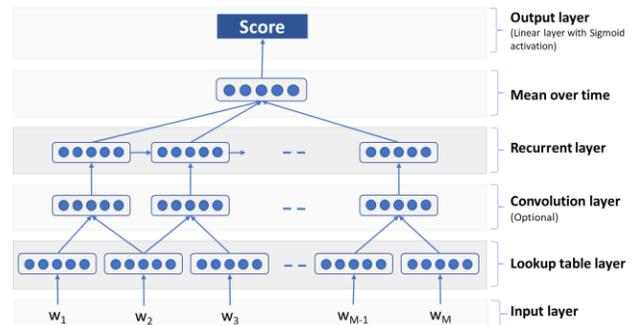


Fig. 1. AEST&N Model Architecture of Taghipour and Ng [6], where the Output Layer Predicts Only the Overall Score.

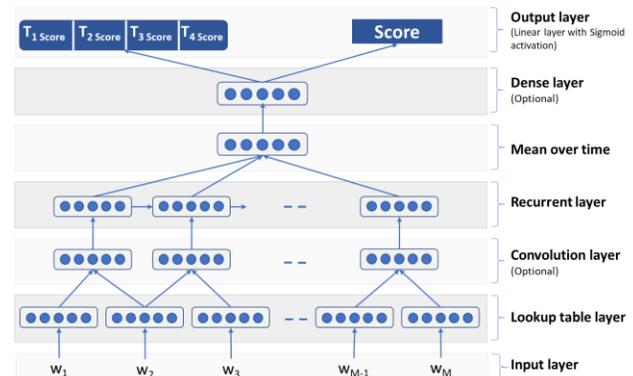


Fig. 2. AES<sub>AUG</sub> Model Architecture, where the Output Layer Predicts both the Overall Score and Traits' Score.

<sup>1</sup> EASE is an open source handcrafted features-based AES system. It depends on Bayesian linear ridge regression and vector regression techniques. It was the third in the ASAP competition (among 154 systems).

1) *The Lookup Table Layer*; first layer/step of the model transforms each word into dimensional space  $d_{LUT}$ . Given a sentence  $S = (c_1, c_2, \dots, c_L)$ , the output of the lookup table operation  $LUT(S)$  represented in Equation 1.

$$LUT(S) = (Ec_1, Ec_2, \dots, Ec_L) \quad (1)$$

where  $c_i$ : one-hot representation of the  $i$ -th word in the sentence, and  $E$ : is the embedding matrix (learned in the training stage).

2) *The Convolution Layer (optional)*; extracts feature vectors from  $n$ -grams. It can capture local contextual dependencies in writing and, therefore, enhance the efficiency of the system. In order to extract local features from the sequence, the convolution layer applies a linear transformation to all  $M$  windows in the given sequence of vectors.

3) *The Recurrent Layer*; processes the input (whether from the convolution layer or directly from the lookup table layer) to generate a representation for the given essay. This representation should encode all the information required for scoring the given essay. Since certain essays are usually long, the proposed model preserved all the intermediate states of the recurrent layer to keep track of the important bits of information. We also experimented with basic RNN vs. GRU vs. LSTM.

In order to control the flow of information during the processing of the input sequence, LSTM units use three gates to discard (forget) or pass the information through time. The following equations formally describe the LSTM function:

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f) \quad (3)$$

$$\tilde{c}_t = \tanh(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c) \quad (4)$$

$$c_t = i_t \circ \tilde{c}_t + f_t \circ c_{t-1} \quad (5)$$

$$o_t = \sigma(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o) \quad (6)$$

$$h_t = o_t \circ \tanh(c_t) \quad (7)$$

where  $\sigma$ : represents the sigmoid function,  $\circ$ : denotes multiplication (element-wise),  $x_t$  and  $h_t$ : the input and output vectors at time  $t$ , respectively,  $W_i, W_f, W_c, W_o, U_i, U_f, U_c$  and  $U_o$ : weight matrices, and  $b_i, b_f, b_c$  and  $b_o$ : bias vectors.

4) *The Mean over Time (MoT)*; this layer input is  $V$  vectors (the output of the recurrent layer) with variable length,  $\mathcal{H} = (h_1, h_2, \dots, h_V)$ . This layer aggregates these inputs into a fixed-length vector and fed it to the dense layer. Equation 8 describes the function of this layer:

$$MoT(\mathcal{H}) = \frac{1}{M} \sum_{t=1}^V h_t \quad (8)$$

5) *The Dense layer (optional)*; gives more depth and enhances the efficiency of the model to predict the traits scores in addition to the overall score in the output layer. The mathematical form of the layer is shown in Equation 9:

$$Y = f(W^T X + b) \quad (9)$$

where  $W$  is weight matrix (with mini-batch size 32),  $b$  is bias vector,  $f$  is activations of the previous layer,  $X$  is the input of the layer (from MoT layer), and  $Y$  is the dense layer output.

6) *The Output layer (Linear Layer with Sigmoid Activation)*; maps the dense layer generated output vector to a scalar value. Equation 10 describes applying the sigmoid activation function on the linear layer mapping:

$$s(x) = \text{sigmoid}(v \cdot x + b) \quad (10)$$

where: the input vector ( $Y$ ),  $v$ : the weight vector, and  $b$ : the bias value. In order to predict the traits scores, we extend the baseline model architecture layers by adding further linear units to the output layer that performs a linear regression to predict traits scores.

We minimized the Mean Squared Error (MSE) between the predicted score and the reference score (human-raters' scores). The  $AES_{T\&N}$  MSE loss function is designed only for the overall score prediction. To fit with predicting the overall and traits scores in our  $AES_{AUG}$  model, we adjusted the  $AES_{T\&N}$  MSE loss function (shown in Equation 11) to compute the overall loss function as a linear combination of multi loss functions (shown in Equation 12), back-propagating the error gradients to the embedding matrix.

$$MSE(s, s^*) = \frac{1}{N} \sum_{i=1}^N (\mathcal{S}_i - \mathcal{S}_i^*)^2 \quad (11)$$

$$MSE(s, s^*) = \frac{1}{N} (\sum_{i=1}^N (\mathcal{S}_i - \mathcal{S}_i^*)^2 + \sum_{j=1}^T \sum_{i=1}^N (t_{ij} - t_{ij}^*)^2) \quad (12)$$

where  $T$ : a number of a specific prompt traits, given  $N$ : number of training essays and their corresponding normalized reference overall scores  $\mathcal{S}_i^*$ , and  $t_{ij}^*$ : traits normalized reference scores. The model computes the predicted overall scores  $\mathcal{S}_i$  and traits scores  $t_{ij}$  for all training essays.

### C. Dataset

AES research has been dominated for the last eight years by the dataset from the 2012 Automated Student Assessment Prize (ASAP) competition [13]. It was established by Kaggle and funded by the Hewlett Foundation. ASAP competition has provided the data and all the required information (hand-crafted features), which can help to evaluate AES systems that use machine learning algorithms. ASAP consists of 12,976 essays, with average length 150-to-550 words per essay, each double scored (Cohen's  $\kappa = 0.86$ ) [8]. The dataset consists of eight tasks/prompts; each task is an essay that has learners' responses. ASAP provided the scoring guides, raters' exemplars, and practice sets for each task. Five tasks employed a holistic scoring rubric, one was scored with a two-trait analytic rubric, and two were scored with a multi-trait analytic rubric but reported as an overall score [14]. Shermis [15] provides a summary of the competition, and most of the recent papers report their results using the same public dataset [16][6][12][17][18][6][19].

In this research, we have used the ASAP data and specifically task 7 data. Task 7 was selected because it has a multi-trait analytic rubric that can be used for formative feedback to learners, and it has the largest dataset (1,569

essays) on the multi-trait analytic rubric-based tasks. The type of writing in task 7 is persuasive/narrative/expository. The prompt asks learners to write a story about patience. The scoring rubric has four traits: ideas, organization, style, and conventions. Each trait score ranges from 0-3. Each score in each trait has a description that guides the rater to identify the appropriate score (level) to each text. In ideas, for example, if the ideas are clearly focused on the topic and are thoroughly developed with specific relevant details, a score of 3 should be assigned. If the ideas are somewhat focused on the topic and are developed with a mix of specific and/or general details, a score of 2 should be assigned. If the ideas are minimally focused on the topic and developed with limited and/or general details, a score of 1 should be assigned. If the ideas are not focused on the task and/or are undeveloped, a score of 0 should be assigned. For objectivity and accuracy, two raters should score the response of each learner for each trait. Then, the scores were summed independently for Rater1 and Rater2 to form the resolved score (0-30) by adding the sum of the two raters.

#### D. Training and Testing

We have followed the dataset split by Taghipour and Ng [6], so we used a 5-fold cross-validation model to assess our proposed system. Data, in each fold, is distributed into 60%, 20%, and 20%; training, development, and test sets, respectively. For prompt no. 7 and each of its four traits, the fold predictions have been aggregated and evaluated together. In order to evaluate the system efficiency, the results are averaged across the four traits. See Fig. 3. The essays have been tokenized by the NLTK<sup>2</sup> tokenizer that lowercases the letters and normalizes the reference scores to the range of [0, 1]. For the system performance evaluation, we rescaled the system-predicted normalized scores to the original range of scores.

In some experimental scenarios, we used a different split ratio in each fold to maximize the training data size for the best training: 80% of the data as a training set, and 20% as the test set.

We followed the AES<sub>T&N</sub> by using the RMSProp optimization algorithm [20] to minimize the MSE loss function over the training data. We also used dropout regularization to avoid overfitting. If the norm of the gradient is larger than a threshold, it will be clipped. We did not use any early stopping method. We trained the model for a fixed 50 epochs, and after each epoch, we monitored the model efficiency on the development set.

The system hyper-parameters are several: To train the network, we have used RMSProp optimizer with the decay rate ( $\rho$ ) set to 0.9. We used pre-trained word embeddings<sup>3</sup>, released by Zou et al. [21] to initialize the lookup table layer. The hyper-parameter settings are listed in Table I. We used Nvidia GEFORCE GTX 1050 GPU to perform our experiments in parallel.

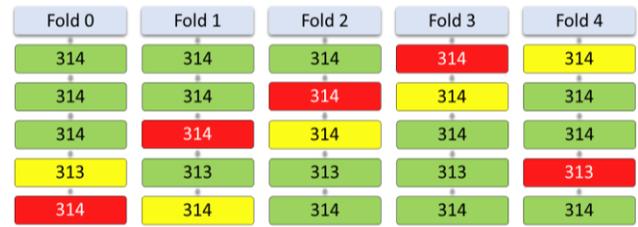


Fig. 3. Prompt no. 7 Dataset Folds Distribution, Green is Training, Yellow is Validation, and Red is Test Set.

TABLE I. AES<sub>AUG</sub> MODEL HYPER-PARAMETERS

Parameter	Parameter meaning/description	Value
$d_{LT}$	Word embedding dimension	50
$d_r$	Output dimension of the recurrent layer	300
$l$	Word context window size	3
$d_c$	Word convolution units	50
$drop-rate$	Dropout probability	0.5
$batch-size$	Mini-batch size <sup>a</sup>	32
$Learn-rate$	Base learning rate	0.001

<sup>a</sup>a fixed 50 epochs.

#### E. Evaluation

The evaluation of AES systems is always done by comparing the AES scores to the scores assigned by human raters. Various statistics tests of correlation or agreement are used for this purpose, including Pearson’s correlation, Spearman’s correlation, and QWK [22]. QWK was identified as the official evaluation metric for ASAP. In this paper, we used the QWK to evaluate our system to the well-established baseline (AES<sub>T&N</sub>) that used the same dataset. The QWK is a commonly used measure of the degree of agreement among raters (a.k.a. inter-rater reliability). The following part illustrates how QWK is computed.

A weight matrix  $W$  is created based on Equation 13:

$$W_{i,j} = \frac{(i-j)^2}{(N-1)^2} \quad (13)$$

where  $i$  and  $j$  are the reference scores, and the hypothesis scores (AES scores), respectively.  $N$  refers to the number of all possible scores.  $O$  is a matrix calculating like  $O_{i,j}$  refers to the number of texts which are given a score  $i$  by the rater and an AES score  $j$ . A count matrix  $E$  is computed to represent the outer product of histogram vectors of the two scores. The sum of elements in  $O$  is equal to the sum of elements in  $E$  as the matrix  $E$  is normalized. Lastly, based on matrices  $O$  and  $E$ , the QWK is computed as of Equation 14:

$$k = 1 - \frac{\sum_{i,j} W_{i,j} O_{i,j}}{\sum_{i,j} W_{i,j} E_{i,j}} \quad (14)$$

Our comparison between the AES<sub>AUG</sub> and AES<sub>T&N</sub> is always by using the QWK values. A one-tailed paired t-test is always used to check the significance of the differences between the two systems.

<sup>2</sup> <http://www.nltk.org>

<sup>3</sup> <http://ai.stanford.edu/~wzou/mt>

#### IV. RESULTS AND DISCUSSION

We describe in this part our experiments and results. In the case of overall scores, we mention the results and then evaluate our system to the baseline system (AES<sub>T&N</sub>). In the case of traits scores, we present only the results of our AES<sub>AUG</sub> system, and its QWK evaluation as the AES<sub>T&N</sub> system did not predict traits scores.

We started our experiments by replicating the AES<sub>T&N</sub><sup>4</sup> model results over the ASAP dataset. Taghipour and Ng [6] (using AES<sub>T&N</sub>) experimented and explored a variety of neural network model architectures like CNN, basic RNN, GRU, and LSTM without using an MoT layer. After replicating the AES<sub>T&N</sub> systems (CNN, RNN, GRU, and LSTM) and producing the same QWKs results, we extended the model to the AES<sub>AUG</sub> model architecture. We trained the model with the training data (described in section 2.3), including the overall score and the four traits reference-scores (by 2 human raters as described in section 2.3). We started by simulating the human approach in scoring traits that every rater gives a score, and the trait score is the summation of the two raters' scores, so AES<sub>AUG</sub> systems predicted two scores for every trait, and we summed them. We got the same QWK (0.805) for the overall score (on Fold 4) and QWK [0.715, 0.623, 0.581, 0.443] for the first predicted traits scores and [0.723, 0.656, 0.568, 0.476] for the second predicted traits scores, with an average [0.598]<sup>5</sup>.

We found that the predicted traits scores have low QWK values, so we analyzed the case by calculating the QWK among the first human rater (H-R1), the second human rater (H-R2), and each of AES<sub>AUG</sub> predicted scores (A-R3 & A-R4). Table II shows QWK for traits scores of the human raters and the AES<sub>AUG</sub> system (using the best model, which is LSTM). We noticed that the agreement (QWK) between the human raters (0.64) is lower than the agreement (QWK) between any AES<sub>AUG</sub> prediction and any of the human raters (0.66, 0.67, 0.68 and 0.68); All the QWKs are shown in Table II. In our attempt to understand the logic behind this low agreement, we examined the prompt content and rubrics with the help of two English language specialists. They confirmed that the definitions of the level descriptors in the rubrics are not clear and definite, which may lead to different interpretations between raters, which accordingly may lead to a low agreement between raters. They also added that using the summation of the two raters on each trait (as described on the ASAP scoring guide) will provide a more accurate and objective indicator for a learner's performance.

In order to enhance the traits QWK scores for AES<sub>AUG</sub> systems, we changed our score calculation approach, i.e., before training the system, we calculated one score for each trait by summing the two human scores. Then, we calculated the QWK score for each trait between one reference-score and one AES<sub>AUG</sub> system predicted score. As a result of that change in score calculation methods, we got higher QWKs for the traits scores [0.820, 0.767, 0.767, 0.733], respectively, with an average QWK of [0.771]. We also noticed that the traits scores prediction within AES<sub>AUG</sub> model architecture enhanced the

accuracy of predicting the overall score [0.851] (on Fold 2) to outperform the baseline AES<sub>T&N</sub> best model (LSTM) which was [0.805] with 4.6% improvement. It even outperformed the best result for prompt no. 7, which is LSTM ensembles (10 runs), which QWK was [0.811] with a 4% improvement. As shown in Table III, predicting traits scores always leads to improvement in the AES<sub>AUG</sub> overall score.

Table III shows the QWKs of our AES<sub>AUG</sub> models on prompt no. 7 overall score and four traits scores. It also shows the AES<sub>T&N</sub> systems replicated results for the overall score. The statistical significance of improvements is marked with '\*'.

We produced the AES<sub>AUG</sub> systems for all models (CNN, RNN, GRU, and LSTM)<sup>6</sup>; all results are shown in Table III. Based on Table III, all models can predict the overall and traits scores competitively compared to the baseline. However, we agree with Taghipour and Ng [6] findings that LSTM has performed better than the other models significantly, and it has outperformed the baseline model by (4.6%). Nevertheless, the least accurate model is basic RNN, which does not work precisely as GRU or LSTM. Such a finding can be due to the moderately long sequences of words in texts. Both LSTM and GRU demonstrate efficient learning of long-term dependencies and sequences. Therefore, we believe this is of the RNN's poor performance points. The CNN model is the fastest in the training and the evaluation compared to other models.

We further investigated the overall and traits scores predicted by our best model (AES<sub>AUG</sub> LSMT), for the predicted and original in ASAP dataset. We presented the results in Fig. 4((a) for overall score, (b), (c), (d), and (e) for the traits). The graphs show the system predictions are less varied and positively contribute to the performance of our proposed approach.

TABLE II. QWK AMONG HUMAN RATERS (H-R1 & H-R2) AND EACH OF AES<sub>AUG</sub> PREDICTED SCORES (A-R3 & A-R4)

Raters	Average QWK score
H-R1 vs. H-R2	0.641
A-R3 vs. A-R4	0.906
H-R1 vs. A-R3	0.684
H-R1 vs. A-R4	0.680
H-R2 vs. A-R3	0.669
H-R2 vs. A-R4	0.670

TABLE III. THE QWK OF THE AES<sub>AUG</sub> SEVERAL NEURAL NETWORK MODELS AND THE AES<sub>T&N</sub>

Systems	AES <sub>T&amp;N</sub> QWK	AES <sub>AUG</sub> QWK	AES <sub>AUG</sub> traits QWK				
			1	2	3	4	average
<b>CNN</b>	0.746	<b>0.822*</b>	0.793	0.717	0.714	0.700	0.731
<b>RNN</b>	0.743	<b>0.760*</b>	0.733	0.656	0.652	0.641	0.671
<b>GRU</b>	0.752	<b>0.827*</b>	0.837	0.749	0.728	0.700	0.754
<b>LSTM</b>	0.805	<b>0.851*</b>	0.820	0.767	0.767	0.733	<b>0.771</b>

\*p < .0001.

<sup>4</sup> <https://github.com/nusnlp/nea>

<sup>5</sup> We tried to add L2 regularization and 256 dense layers, but the model extracted was not better than the one that was concluded.

<sup>6</sup> All the mentioned neural network models are unidirectional and include the MoT layer. The convolution layer is included in the CNN model only.

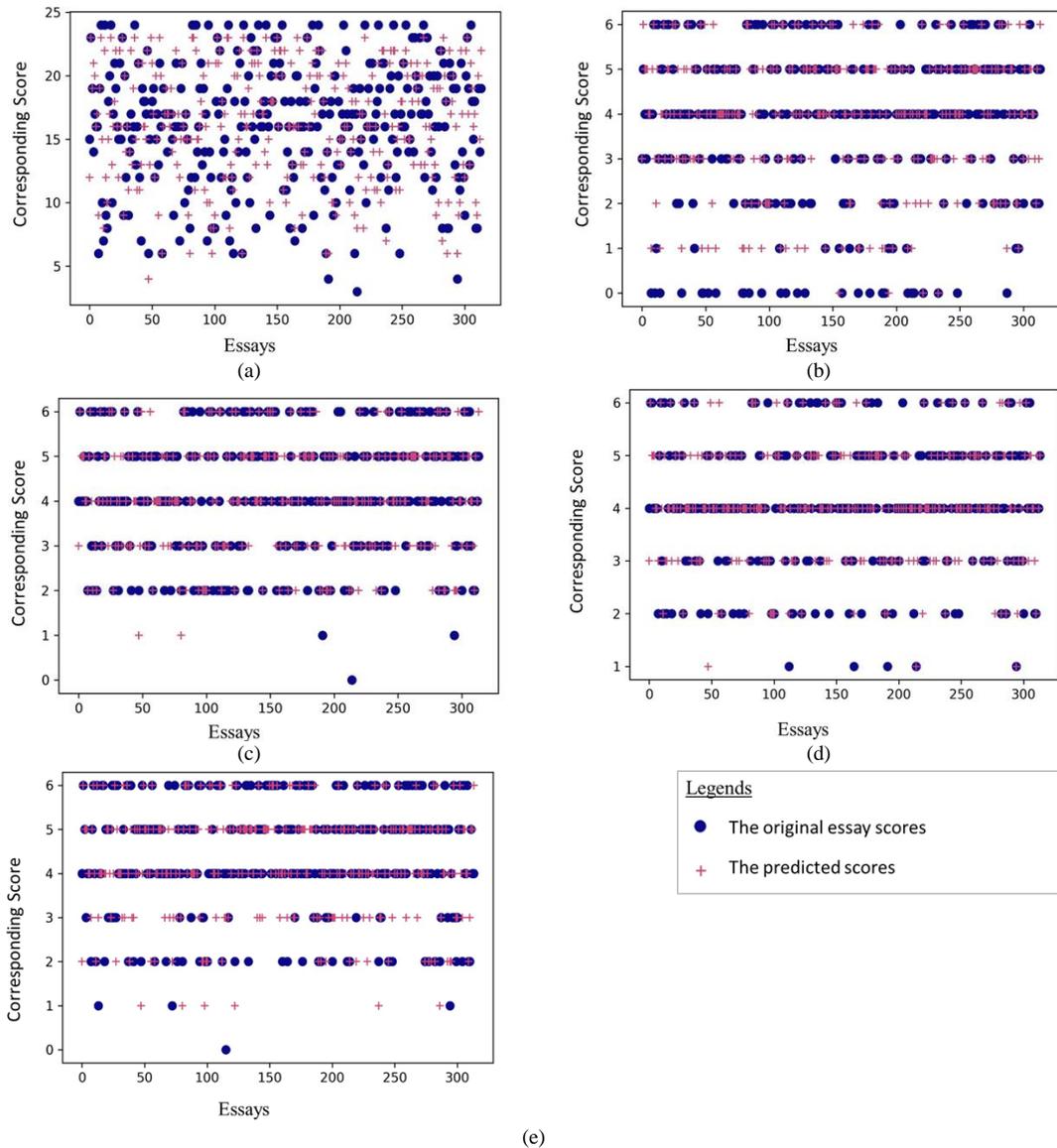


Fig. 4. The Graphs show for Prompt no. 7 and its Traits, the System Predictions are Less Varied and Positively Contribute to the Efficiency of our Proposed Approach. (a) Representing the Overall Score, While (b), (c), (d), and (e) Represent the Four Traits' score, respectively. The Blue Circles Represent the Original Essay Scores, and the Red Pluses the Predicted Scores. All Predicted Scores are Mapped to their Original Scoring Scale.

In the end, we experimented with using a different split to the dataset from the one described in Section III-D (which is 60% training, 20% validation, and 20% testing). Thus, we merged the training set with the validation set to be 80% training and 20% testing. It has achieved better QWK scores for the overall score to be [0.858] instead of [0.851], which means that the availability of a bigger training set will improve the results.

Finally, we used the above method, and its results in the iAssistant, an educational module that provides trait-specific adaptive feedback to learners. As shown in Fig. 5, iAssistant provides learners with predicted scores on multiple rubric traits and levels of performance per each trait. In addition to that, it helps learners to evaluate the length of their essay on a scale of 3 levels (short, good, and long).

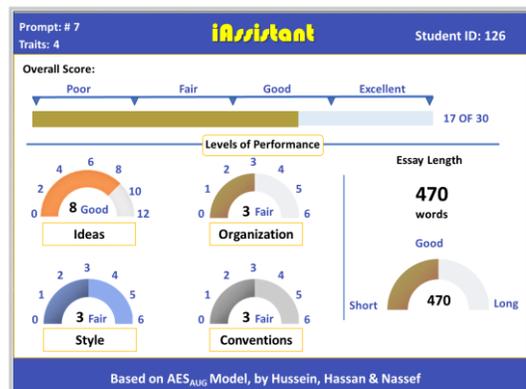


Fig. 5. An Example of iAssistant in use: Predicted Scores on Multiple Rubric Traits and Levels of Performance. In Addition to Representing the Overall Score and Length of the Essay.

## V. CONCLUSIONS

In this paper, we have proposed a framework, based on deep learning models that strengthens the validity and enhances the accuracy of a baseline system with respect to traits' evaluation/scoring. Our method does not rely only on overall score prediction but also on essay traits prediction to give trait-specific adaptive feedback. We explored multiple deep learning models for the automatic essay scoring task.

Based on our experiments, we can conclude that our proposed AES<sub>AUG</sub> model outperformed all the previously used AES models (CNN, RNN, GRU, and LSTM). Including traits in training has significantly improved the learning process. Thus, our AES<sub>AUG</sub> system has significantly increased the accuracy of the overall and traits scores for essays using analytic-rubrics. This point highlights the contributions of our model over all the previous models.

It is also found that the LSTM<sub>AUG</sub> model, like the AES<sub>T&N</sub> system, proves to be the best model to predict scores for essays that include relatively long sequences of words which is consistent with the nature of the LSTM models. However, adding a dense layer between the MoT layer and the output layer did not improve the results of our AES<sub>AUG</sub> model. We can also assume, based on our experiments, that increasing the training data has a positive effect on the accuracy of AES<sub>AUG</sub> scores.

Additionally, it is very important to note that the clarity of the definition of the scoring rubrics strongly influences the accuracy of both human and AES<sub>AUG</sub> scores, which accordingly affects the quality of the adaptive feedback that can be given to the learners. In other words, the more the rubric is clear and definite, the more the AES<sub>AUG</sub> scores are accurate, and the feedback is more specific.

Finally, our proposed AES<sub>AUG</sub> model offers a new methodology that may be interesting to the users, and it provides more accurate results without requiring a high configuration of hardware.

## VI. FUTURE WORK

The future directions of this work may be to highlight the words and sentences that made the AES system give a specific score for further analysis and adaptive feedbacking, in addition to training and testing the model on a larger dataset with well-defined rubrics.

## REFERENCES

- [1] S. Brown, 500 Tips on Assessment. Routledge, 2004.
- [2] T. C. Stephen, "Using Automated Essay Scoring to Assess Higher-Level Thinking Skills in Nursing Education," 2019.
- [3] S. M. Brookhart, How To Create and Use Rubrics. Ascd, 2013.
- [4] A. J. N. and S. M. Brookhart, Educational assessment of students. Pearson Merrill Prentice Hall, 2007.
- [5] M. Lu, Q. Deng, and M. Yang, "EFL writing assessment: Peer assessment vs. automated essay scoring," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2020.
- [6] K. Taghipour and H. T. Ng, "A Neural Approach to Automated Essay Scoring," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 1882–1891.
- [7] F. Nadeem, H. Nguyen, Y. Liu, and M. Ostendorf, "Automated Essay Scoring with Discourse-Aware Neural Models," 2019.
- [8] M. A. Hussein, H. Hassan, and M. Nassef, "Automated language essay scoring systems: A literature review," PeerJ Comput. Sci., vol. 2019, no. 8, 2019.
- [9] N. W. Solomons et al., "[Use of tests based on the analysis of expired air in nutritional studies]," Arch. Latinoam. Nutr., vol. 28, no. 3, pp. 301–17, 1978.
- [10] Z. Ke and V. Ng, "Automated Essay Scoring: A Survey of the State of the Art," 2019.
- [11] R. D. Varner, L. K. Crossley, and S. A. McNamara, "Developing pedagogically-guided algorithms for intelligent writing feedback," 2013.
- [12] B. Woods, D. Adamson, S. Miel, and E. Mayfield, "Formative Essay Feedback Using Predictive Scoring Models," dl.acm.org, vol. Part F1296, pp. 2071–2080, Aug. 2017.
- [13] S. Dikli, "Automated essay scoring," Turkish Online Journal of Distance Education, vol. 7, no. 1, pp. 45–56, 2006.
- [14] M. Shermis and B. H. Education, "Contrasting state-of-the-art automated scoring of essays: Analysis," Annual national council on measurement in education meeting. 2012.
- [15] M. D. Shermis, "State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration," Assess. Writ., vol. 20, pp. 53–76, 2014.
- [16] T. Dasgupta, A. Naskar, R. Saha, and L. Dey, "Augmenting Textual Qualitative Features in Deep Convolution Recurrent Neural Network for Automatic Essay Scoring," aclweb.org, pp. 93–102, 2018.
- [17] F. Dong and Y. Zhang, "Automatic Features for Essay Scoring-An Empirical Study," 2016.
- [18] P. Phandi, K. A. Ming Chai, and H. Tou Ng, "Flexible Domain Adaptation for Automated Essay Scoring Using Correlated Linear Regression," Association for Computational Linguistics, 2015.
- [19] D. Alikaniotis, H. Yannakoudakis, and M. Rei, "Automatic Text Scoring Using Neural Networks," 2016.
- [20] Y. N. Dauphin, H. De Vries, and Y. Bengio, "Equilibrated adaptive learning rates for non-convex optimization," in Advances in Neural Information Processing Systems, 2015, vol. 2015-Janua, pp. 1504–1512.
- [21] W. Y. Zou, R. Socher, D. Cer, and C. D. Manning, "Bilingual Word Embeddings for Phrase-Based Machine Translation," Association for Computational Linguistics, 2013.
- [22] H. Yannakoudakis and R. Cummins, "Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications," 2015.

# Mutual Coexistence in WBANs: Impact of Modulation Schemes of the IEEE 802.15.6 Standard

Marwa BOUMAIZ<sup>1\*</sup>, Mohammed EL GHAZI<sup>2</sup>, Mohammed FATTAH<sup>3</sup>  
Anas BOUAYAD<sup>4</sup>, Moulhime EL BEKKALI<sup>5</sup>

Laboratory of Artificial Intelligence, Data Sciences and Emerging Systems  
Sidi Mohammed Ben Abdellah University, FEZ, Morocco<sup>1, 2, 4, 5</sup>  
EST, Moulay Ismail University, Meknes, Morocco<sup>3</sup>

**Abstract**—Due to the mobility of subjects carrying wireless Body Area Networks (WBANs), a BAN may be found in an environment that contains other adjacent BANs, which may influence its proper functioning. The purpose of this paper is to study the effect of interference between adjacent BANs on the performance of a reference BAN in terms of packet loss rate (PLR), while considering the following four parameters: the distance separating adjacent BANs, the number of nodes and traffic payload of an interferer BAN, and the transmission data rate. The study is conducted for the two modulation schemes proposed by the IEEE 802.15.6 standard in the 2.4 GHz narrow band, which are: Differential Binary Phase Shift Keying (DBPSK) modulation and Differential Quadrature Phase Shift Keying (DQPSK) modulation. Simulation results have shown that the adoption of a lower-order modulation such as DBPSK can reduce the effect of interference among adjacent BANs.

**Keywords**—Body Area Network (BAN); mutual coexistence; interference; Differential Binary Phase Shift Keying (DBPSK); Differential Quadrature Phase Shift Keying (DQPSK)

## I. INTRODUCTION

Wireless body area networks, commonly known as BANs, are a novel technology that has emerged with the recent growth of low-power and low-cost microelectronic systems [1]. In the medical field, these systems can be integrated in a wide range of applications [2], as they are able to remotely monitor patients' physiological conditions, and send these data to a nearby coordinator. Equipped with more resources than a simple sensor, the coordinator is able to transfer the information collected by sensors directly to healthcare centers to be analyzed by a medical staff. This remote monitoring not only saves time, but also allows early detection of health problems without invading the patient's privacy or employing full-time medical staff [3].

As a matter of fact, BANs extend over several fields of applications, other than medical ones. They are also present in the fields of cognitive biometry, military, learning and serious gaming, and sports [4]. This variety of applications envisioned for BANs, gave rise to the IEEE 802.15.6 standard [5], that aims to provide an international norm for highly reliable, body-wide, short-range wireless communications, by supporting a broad range of data rates, ranging from 75.9 Kbps (narrowband) to 15.6 Mbps (ultra-wideband). The standard provides a sophisticated MAC layer with three access modes [6] [8] [7], serving three physical layers which are selected

according to the intended application. These are the Narrowband (NB), Ultra-Wideband (UWB) and Human Body Communication (HBC) layers. The NarrowBand physical layer alone provides seven different frequency bands, including the 2.4 GHz band (2400-2483.5 MHz), which uses DBPSK and DQPSK to code the useful information. This band is often preferred over the others due to its worldwide availability[7]. It is also the most mature band [8] offering a greater bandwidth. In addition, it is based on well-known PHY components, which are already widely used in WiFi and Bluetooth [9], and is also characterized by the use of small antennas [10], which makes it perfectly adapted to most on body BAN applications.

However, a BAN can be found in environments where several other BANs or wireless technologies (wifi, bluetooth, zigbee...) coexist. This is the case of interference. Therefore, a BAN can face two kinds of interference: Intra-BAN and or inter-BAN interference. In the first type, interference occurs between nodes of the same BAN. However, this can easily be avoided by using time division channel access [11]. As for the second type, interference is either generated by adjacent BANs [12], or other wireless technologies operating at the same frequency band (2.4 GHz in particular). When multiple adjacent BANs use the same communication channel (the same frequency), interference can take place, as the active superframes may overlap [13]. However, according to the IEEE 802.15.6 standard, packets will be retransmitted in case of interference for a certain period of time, and therefore there is a trade-off to be made between throughput and energy efficiency at the relevant BAN [13].

There are numerous research studies that focus on interference between adjacent BANs, some of which are presented in Section II. In addition, many of them are interested in reducing or cancelling interference by proposing several solutions at both physical and MAC layers. For example, at the physical layer level, power control solutions may be a possible option to maintain link quality [14]. But when the interference level becomes important, these solutions cannot be suitable for WBANs, because if a WBAN coordinator cannot receive signals coming from a sensor node for example, and by applying power control policies, this sensor will increase its transmit power to be able to send its packets correctly. However, by doing this, it is possible to impact eventual adjacent BANs, which in turn will increase their transmission power, making it difficult for any link to operate with an acceptable quality. Therefore, other adaptive

\*Corresponding Author

methods [15], [11] have been proposed which aim to take advantage of other physical layer parameters (modulation, duty cycle, rate, etc.) to reduce interference in BAN networks. However, these parameters must first be studied in different interference scenarios (change in inter-BAN distance, change in bit rates, change in number of nodes, etc.) for a better evaluation.

This research paper evaluates and analyzes the impact of modulation scheme ( $\pi/2$  DBPSK and  $\pi/4$  DQPSK as described by the IEEE 802.15.6 standard in the 2.4 GHz narrowband) on BAN transmission performance in a mutual interference environment between adjacent BANs. In this manuscript, a detailed study is evaluated by using three scenarios, and three relevant parameters which are the inter-BAN distance (scenario 1), the number of nodes (scenario 2) and the traffic payload (scenario 3) of the interferer BAN. All three scenarios investigate the BAN transmission performance, in terms of packet loss rate (PLR), for three data rate levels (low, medium and high) and two modulation schemes  $\pi/2$  DBPSK and  $\pi/4$  DQPSK.

The remainder of this article is organized as follows: Section II introduces a state of the art on co-channel interference between adjacent BANs. Then, in Section III the three scenarios studying mutual coexistence are defined, along with the adopted configuration at both the physical and mac layer levels. Next, Numerical study results are reported and analyzed in Section IV. Finally, the paper is concluded in Section V.

## II. CO-CHANNEL INTERFERENCES BETWEEN ADJACENT BANS: STATE OF THE ART

Mutual interference occurs when a BAN's coordinator receives signals from other nodes of other BANs in its vicinity. Interfering BANs are then said to use the same transmission channel, leading to a packet transmission conflict as the active transmission periods overlap [13].

Research in literature dealing with interference between adjacent BANs can be divided into two categories: interference analysis studies and interference mitigation/cancellation studies. In the following, most relevant works in both of the two categories are presented.

### A. Interference Analysis Research

Interference analysis studies can be divided into three sub-groups: simulation-based studies, empirical based studies, and analytical modelling based studies [3].

1) *Simulation based studies*: Using simulations, Wang et al. [16] analyzed mutual interference for the ultra-wideband IEEE 802.15.6 standard. In this work, two types of receivers were considered, "duty-cycled sampling receiver" and "chirp receiver". Authors observed that the "chirp receiver" guarantees better results than the cycled sampling receiver in terms of PLR when the traffic density increases (Ten adjacent users).

2) *Empirical based studies*: Davenport et al. [17] used a test-bed study and simulations to investigate the effect of interference from adjacent BANs on the PLR. For this

purpose, ten persons carrying BANs were considered in a state of movement in an office area. The PLR measurement in this scenario included different communication techniques: listen before talk (LBT), frequency hopping (FH), and automatic repeat request (ARQ) [3]. It was demonstrated in this work that the combination of LBT and ARQ gives better results.

The assessment of interference between adjacent BANs in case of mobility in indoor/outdoor environments was further discussed in [18]. A scenario of five persons moving in an office was considered, highlighting the effect of distance between nodes and their orientations on interference. The findings of this study showed that distance between nodes does not always have a significant impact on interference in mobility conditions, which is not the case for stationary BANs. However, the orientation of the transceivers has a remarkable effect in reducing interference, especially in outdoor environments which experience fewer reflections and multipath effects.

3) *Analytical modelling-based studies*: Wang et al. [19] proposed a mathematical model for the characterization of adjacent channel interference in BANs based on Gamma distribution. Authors found the minimum distance between sensors that guarantees an acceptable Signal to Interference Ratio (SIR).

Zhang et al. [20] proposed a mathematical approach to model the average SIR and the probability of collision in the case of co-channel interference. In this study several access methods were considered, including TDMA (Time Division Multiple Access), FDMA (frequency division multiple access), FH (frequency hopping), and CDMA (code division multiple access). Authors showed that when there is no coordination between adjacent BANs, TDMA, FDMA and FH guarantee a similar performance in terms of PLR and BER, which remains better than CDMA. However, in the presence of coordination between adjacent BANs (synchronization), the FDMA and FH protocols have shown better results in interference cancellation.

### B. Interference Mitigation Studies

Research works aimed at reducing or avoiding interference between adjacent BANs tend to propose the following solutions [3]:

1) *Time spacing*: Interference mitigation techniques (mutual and cross interference) proposed in this category are essentially based on TDMA. This is to prevent nodes of the same or several BANs, or other networks sharing the same transmission channel, from sending their data simultaneously, which leads to collisions. But they are rather suitable solutions for reducing mutual interference [21], [22], [23], [24], [25], [26]. The concept of these solutions consists in rescheduling data packets to avoid interference, by managing transmissions in empty time slots. However, this process can be quite challenging, as frequent coordination exchanges between adjacent BANs are required to know their schedule, which is a real challenge, given the energy constraint in BAN nodes.

2) *Frequency spacing*: Frequency spacing solutions make judicious use of the available frequency channels for BANs.

The solutions reported in this category reduce interference levels by implementing channel allocation algorithms that specifically address multi-channel networks.

As discussed, for example, in [27], [28], [29], [30], where authors proposed some solutions to reduce interference in BANs based on frequency spacing. The common point between these studies is that they all rely on the detected interference level, to assign different channels to nodes, to mitigate the impact of interference. Nevertheless, the problem of insufficient channels can be faced in these solutions, particularly when interference occurs with networks having high-bandwidth channels [3].

3) *Code diversity*: This category mainly targets CDMA-based BANs. the idea here is to choose orthogonal codes to what is used by other adjacent networks to reduce interference [31], [32], [33], [34]. However, these solutions are sometimes slightly complex in terms of estimating the interference level.

4) *Standard modification*: This category aims to improve the MAC mechanisms implemented by BAN standards. The following research [35] [36] are examples of interference mitigation studies, that have revised and restructured what is proposed in zigbee low power and IEEE 802.15.6 standards respectively, in order to improve coexistence. Therefore, these solutions cannot be applied until they are formally added to the standards [3].

5) *Standards adaptation*: Unlike standard modification solutions, the solutions in this category reduce interference by adapting to BAN standards without modifying them [37], [38], [39], [40].

6) *Hybrid solutions*: Same examples of hybrid solutions to reduce interference in BANs are developed in [41] and [42]. Authors used a combination of the above-mentioned methods to exploit the advantages and reduce the shortcomings of one technique over another.

### III. DESCRIPTION OF THE STUDY SCENARIOS AND CONFIGURATION OF THE PHYSICAL AND MAC LAYERS

#### A. The Study Scenarios

In this study, to evaluate the behavior of modulation schemes in a mutual coexistence environment, two BAN models based on the IEEE 802.15.6 standard are considered. The first one is the *reference* model, called “R-BAN” in this paper. This is the model on which the transmission performance analysis will be carried out and the second one will act as an interferer BAN named “I-BAN”. Both of these models operate in the NB layer at 2.45GHz, using a single hop star topology, in which sensor nodes transmit their data directly to the coordinators with no need for relays [43]. This topology has also been chosen in several studies for its better performance [44].

R-BAN consists of eleven sensor nodes, including a coordinator (or hub) placed toward the center of a patient’s belly as shown in Fig. 1. The ten other sensors are distributed over the different parts of the body (positions from A to J). The choice of these node positions is driven by the need to monitor the majority of vital signs described in on-body BAN medical

applications for the IEEE 802.15.6 standard [45][46]. As for I-BAN, it assumes the same number and node positions in scenarios 1 and 3, while in scenario 2 the number of nodes varies.

1) *Scenario 1: varying the distance between R-BAN and I-BAN*: The purpose of this first scenario is to present the effect of varying the inter-BAN distance [12] on mutual interference at low, medium and high data rates, i.e. 3kbps, 44 kbps and 72 kbps respectively, as well as evaluating the behavior of  $\pi/2$  DBPSK and  $\pi/4$  DQPSK modulation schemes in this scenario. For this, the following three cases are considered:

- Case 1: absence of mutual interference: R-BAN operates alone.
- Case 2: presence of I-BAN at 6 m distance from the reference BAN.
- Case 3: presence of the I-BAN at 3 m distance from the reference BAN.

In the first case packet loss rate (PLR) at R-BAN is simulated in the absence of mutual interference using both DBPSK and DQPSK modulations. This is for the three mentioned data rates (3kbps, 44kbps and 72kbps). In the second case, the interferer BAN is added at 6m from the R-BAN and then this distance is reduced to 3m in the third case, so that the effect of this distance reduction on the performance of R-BAN can be evaluated in terms of PLR at low, medium and high data rates. Fig. 2 shows the three cases of scenario 1.

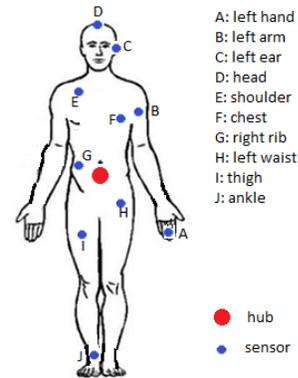


Fig. 1. Reference BAN Model "R-BAN".

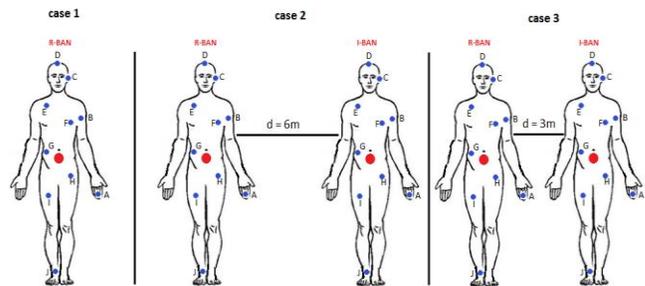


Fig. 2. Scenario 1 Study Cases.

### 2) Scenario 2 : Changing the number of nodes of I-BAN:

The purpose of this scenario is to evaluate the performance of R-BAN by studying the impact of the number of I-BAN nodes, as well as that of DBPSK and DQPSK modulation schemes acting under these conditions, while fixing the distance between the two BANs at 6m. The number of nodes in R-BAN is maintained unchanged, to 11 sensors. In this study scenario the following settings for the I-BAN will be examined:

- Case 1: I-BAN with 2 body sensors.
- Case 2: I-BAN with 5 body sensors (one coordinator and four sensor nodes).
- Case 3: I-BAN with 10 body sensors (one coordinator and nine sensor nodes).
- Case 4: I-BAN with 15 body sensors (one coordinator and 14 sensor nodes).
- Case 5: I-BAN with 21 body sensors (one coordinator and 20 sensor nodes).

### 3) Scenario 3 : Varying the traffic payload of I-BAN: To

assess the impact of traffic payload on the performance of the reference BAN, for the case of co-channel interference, the same positions and number of nodes in BANs are kept as in Scenario 1 but varied the traffic payload ( in Bytes) of I-BAN, while maintaining the R-BAN payload set at 100 Bytes. The inter-BAN distance is fixed at 6m and the behavior of DBPSK and DQPSK modulations is also evaluated in this scenario. For this purpose, the following three cases are taken into account:

- Case 1: I-BAN traffic payload is 10 Bytes.
- Case 2: I-BAN traffic payload is 100 Bytes.
- Case 3: I-BAN traffic payload is 200 Bytes.

### B. DBPSK and DQPSK Modulations

According to the IEEE 802.15.6 standard, in the 2.4 GHz NarrowBand, binary information (superframe payload) is modulated using one of the two following differential phase modulations:  $\pi/2$ DBPSK or  $\pi/4$ DQPSK[47], allowing data to be encoded in the phase of the reference signal. These differential modulations have the advantage of being compatible with non-coherent receivers.

$\pi/2$  DBPSK modulation uses two relative phases to encode the data ( $+\pi/2$  and  $-\pi/2$ ) [48]. For this reason, One-bit symbols are transmitted in the two phases of the same carrier ( $A\sin(2\pi ft)$ ) but spaced by  $180^\circ$ , as follows:

- For a bit 0  $\rightarrow S(t) = A\sin(2\pi ft + 90^\circ)$
- For a bit 1  $\rightarrow S(t) = A\sin(2\pi ft - 90^\circ)$

Regarding  $\pi/4$ DQPSK modulation, two bits per symbol can be encoded using 4 phases shifted by  $\pi/4$ . Moreover, there is always a phase shift even when the adjacent symbols are exactly the same. This results in the following 8 signals:

- For 00 :

$$S(t) = A \sin(2\pi ft) \text{ or } A \sin(2\pi ft + 45^\circ)$$

- For 01 :

$$S(t) = A \sin(2\pi ft + 90^\circ) \text{ or } A \sin(2\pi ft + 135^\circ)$$

- For 10 :

$$S(t) = A \sin(2\pi ft + 180^\circ) \text{ or } A \sin(2\pi ft + 225^\circ)$$

- For 11 :

$$S(t) = A \sin(2\pi ft + 270^\circ) \text{ or } A \sin(2\pi ft + 315^\circ)$$

Among the main differences between the two types of modulations are data rate and robustness against bit errors:

- Data rate: DBPSK allows the transmission of one bit per symbol, instead of two in DQPSK modulation, which means that DQPSK modulation can double the data rate, whilst using the same bandwidth, this is shown in the equation [44] :

$$R_d = \left( \frac{R_s \cdot N}{S} * \frac{k}{n} \right) (\text{kbps}) \quad (1)$$

Where  $R_d$  is the information rate,  $R_s$  is the symbol rate,  $S$  is the spreading factor,  $k/n$  is the BCH code rate, and  $N$  is a variable related to the modulation order  $M$  by  $M=2N$ . Therefore, as the modulation order increases, the bit efficiency increases as well.

- Modulation robustness: the constellation diagram of the two modulations shows that the separation between the phases in DBPSK is greater than in DQPSK: the two phases in DBPSK are separated by  $180^\circ$ , whereas in DQPSK the separation is only  $45^\circ$ , which makes DBPSK modulation much more robust against bit errors.

### C. Physical Layer Configuration

1) On-body channel model : In on-body medical applications, signal propagation takes place on the surface of the human body. In addition to this propagation, which may include a combination of surface waves, creeping waves, diffracted waves, refracted waves and free space propagation waves depending on the position of the antenna [49] [44], Consideration should also be given to the effects of antenna-human body interaction resulting from the placement of body sensors on or near the patient's body, including near-field coupling effects, radiation pattern distortion and antenna impedance changes. These effects impact the functional performance of sensors by degrading their efficiency and reducing the reliability of the collected physiological signals [50]. Due to these issues, a good characterization of the propagation channel is necessary before any design of WBAN solutions.

Assuming that signal propagation in the reference BAN takes place at the body surface, between body sensors, two types of on-body propagation channels can be distinguished:

- The line of sight on-body channel (LOS): free of any kind of obstruction between each sensor and the coordinator,
- The non-line of sight on-body channel (NLOS) where obstacles may exist in the link of each two communicating nodes.

Therefore, IEEE 802.15.6 standard has defined, for the 2.4 GHz narrowband, two path loss models, describing propagation with and without line-of-sight, called CM3A and CM3B, respectively [51]. In this work, the CM3B model is adopted as a path loss model. According to this model, path loss decreases exponentially around the perimeter of the body when the BAN nodes are not in line-of-sight. It flattens over large distances due to the addition of multipath components from indoor environments [51]. Path loss in CM3B model is defined by equation 2:

$$PL(d)[dB] = -10\log_{10}(P_0e^{-m_0d} + P_1) + \sigma_P n_P \quad (2)$$

Where :

- PL (d) refers to the path loss in dB at a distance d,
- P0 is a component related to the average losses occurring near the transmitter and depends on the type of antenna.
- P1 is the average attenuation of the components in an indoor environment irradiated from the body and reflected towards the receiving antenna.
- m0 represents the average exponential decay rate in dB / cm of the creeping wave component diffracting around the body.
- n<sub>p</sub> is a Gaussian random variable of mean and unit zero.
- σ<sub>p</sub> is the log-normal variance in dB around the mean, representing variations measured at different locations in the body and room.

The distances separating each node from the hub were measured and assuming that the proposed network is subject to the same environmental conditions as those of a hospital room, path loss values were also calculated for all the links of the BANs considered. The parameters values of the CM3B model, considered in a hospital room conditions, are presented in Table I [51].

2) *Radio module configuration*: At the physical layer, a radio chip proposal developed by Alan Wong et al. [52] is considered, which is compatible with the recommendations of the IEEE 802.15.6 standard for the 2.4 GHz NB physical layer. In Table II, the adopted configuration for the radio modules of the adjacent BANs in the three scenarios of the study is presented.

#### D. MAC Layer Configuration

The IEEE 802.15.6 MAC layer provides three access modes to the transmission channel [5]. In this paper, the beacon mode with superframe boundaries is chosen in this mode, the Hub (coordinator node) divides its time axis into several access phases. As a result, the R-BAN model is based

on a hybrid MAC layer configuration, combining a scheduled access based on the TDMA, and a contention access based on the CSMA/CA. Thus, a beacon period of 32 allocation intervals of 10 ms each is defined, i.e. 320 ms beacon period. If each node of the network is allocated three slots (30 ms) for scheduled access (TDMA), we end up with 30 allocation slots (300 ms) corresponding to the 10 nodes of the BAN reserved for scheduled access, the two remaining allocation slots (20 ms) will be dedicated to random access (CSMA/CA). The polling mechanism is also activated.

TABLE I. PARAMETERS VALUES OF CM3B MODEL

Parameter	Value
Frequency (GHz)	2.45
P <sub>0</sub> [dB]	-25.8
m <sub>0</sub> [dB/cm]	2.0
P <sub>1</sub> [dB]	-71.3
σ <sub>p</sub> [dB]	3.6

TABLE II. R-BAN AND I-BAN RADIO MODULE CONFIGURATION [12][43]

Parameters	R-BAN	I-BAN
Packet transmission rate (kbps)	3, 44 and 72	3,44 and 72
Modulation	<ul style="list-style-type: none"> <li>• DBPSK</li> <li>• DQPSK</li> </ul>	<ul style="list-style-type: none"> <li>• DBPSK à 3kbps et 44kbps.</li> <li>• DQPSK à 72kbps.</li> </ul>
hub sensitivity	<ul style="list-style-type: none"> <li>• -104 dBm for DBPSK.</li> <li>• -96.5 dBm for DQPSK.</li> </ul>	<ul style="list-style-type: none"> <li>• -104 dBm for DBPSK.</li> <li>• -96.5 dBm for DQPSK.</li> </ul>
Noise bandwidth	1MHz	1MHz
Noise floor	-104dBm	-104dBm
Transmit power	-10 dBm	-10dBm
Clear Channel Assessment Duration	1ms	1ms
CCA threshold	-95	-95
Power consumption for Tx mode	5.9 mW	5.9 mW
Idle mode supply	0.05 mW	0.05 mW
Frequency	2450 Mhz	2450 Mhz
Startup delay	1.5s	1s

#### IV. NUMERICAL STUDY AND RESULTS ANALYSIS

The numerical study of the three study scenarios is carried out using the new version of Castalia simulator (3.3) [53]. This framework is based on OMNeT++(4.6) platform and supports the IEEE 802.15.6 standard, which justifies its use in the numerical study.

In each scenario, packet loss rate (PLR) is evaluated according to the modulation scheme used in R-BAN (π/2 DBPSK and π/4 DQPSK modulation) at low, medium and high data rate.

For all performed simulations, R-BAN transmits its packets, for each one of the data rates levels, using successively the two modulation schemes: DBPSK and DQPSK. As for I-BAN, it uses DBPSK for low and medium data rate rates and DQPSK for high data rate.

*A. Scenario 1: Effect of DBPSK and DQPSK Modulation Schemes on the Transmission Performance (PLR) of R-BAN, as a Function of the Inter-BAN Distance.*

In Fig. 3, 4 and 5, PLR at R-BAN level is presented for low, medium and high data rate respectively, for the three inter-BAN distance cases (case 1: absence of I-BAN, case 2: distance = 6m, case 3: distance = 3m). A first visual analysis of the three figures (cases 2 and 3 in the three figures) clearly shows, as expected a priori, that the global transmission performance of the R-BAN network deteriorates (increase of PLR) as the level of mutual interference increases (smaller distance between adjacent BANs). In the absence of I-BAN (case 1, absence of inter-BAN interference) and when R-BAN nodes transmit their packets at a low rate (3 kbps, Fig. 3), a slightly lower packet loss rate (PLR  $\approx$  7.63%) is noticed with DBPSK modulation than with DQPSK modulation (PLR  $\approx$  10.81%). However, in the presence of I-BAN (cases 2 and 3, inter-BAN interference) and for both of the modulation schemes studied, the PLR increases significantly especially when the inter-BAN distance decreases from 6m to 3m. However, the influence of modulation schemes is inverted this time to designate DQPSK as a very suitable modulation choice for low data rates. Indeed, the results in Fig. 3 show, for an inter-BAN distance of 6 m, a PLR (R-BAN) approaching 73.8%, (DBPSK modulation), which is significantly higher than that obtained with DQPSK modulation (11.63%). This difference in transmission performance is even more pronounced as the level of inter-BAN interference is high (case 3: inter-BAN distance = 3 m). This difference in the performance of the two studied modulations can be related to the difference in sensor sensitivity linked to the adopted modulation scheme. The receiver sensitivity is much lower with DBPSK modulation (-104 dBm) than with DQPSK modulation (-96.5 dBm). This allows the R-BAN coordinator, when configured with DBPSK modulation instead, to pick up even very low-level signals coming from the various nodes of I-BAN. the coexistence between adjacent BANs is more severely affected with high data rate transmissions [3] [4]. Indeed, it is noticed (Fig. 5) that the behavior of the modulation schemes changes compared to the case of low and medium data rate transmissions (Fig. 3 and 4). This time, it is rather the DBPSK modulation that gives a lower PLR than in DQPSK. This can be explained by the robustness of DBPSK modulation against bit errors due to the significant separation of the modulation phases compared to DQPSK modulation. In fact, the constellation diagram of the two modulations [48] shows that in DBPSK, there is a 180° separation between each two transmitted symbols, for only 45° between the four symbols in DQPSK. Thus, much more packet loss can occur with DQPSK modulation if the transmission channel is subject to a high level of interference (in the case of high data rates).

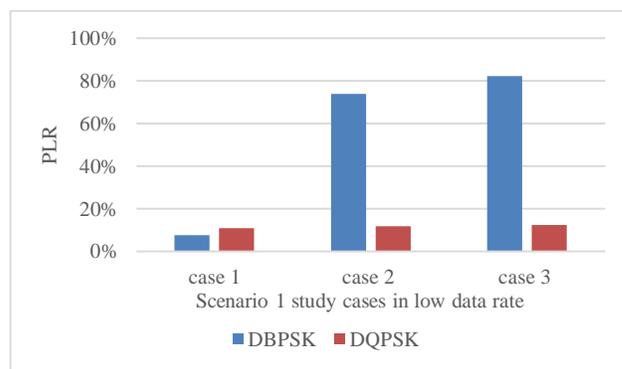


Fig. 3. Scenario 1: PLR (R-BAN) for DBPSK and DQPSK Modulation Schemes at Low Data Rate (3kbps).

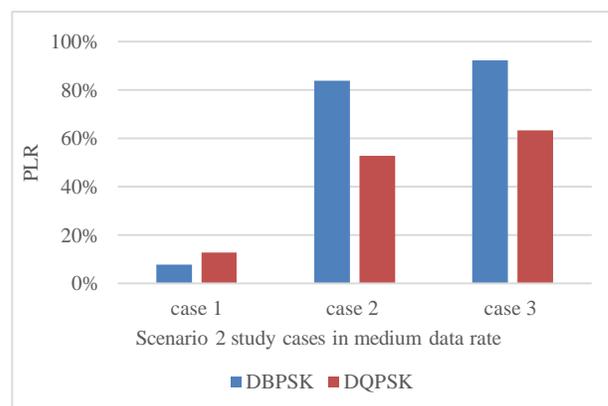


Fig. 4. Scenario 1: PLR (R-BAN) for DBPSK and DQPSK Modulation Schemes at Medium Data Rate Transmission (44 kbps).

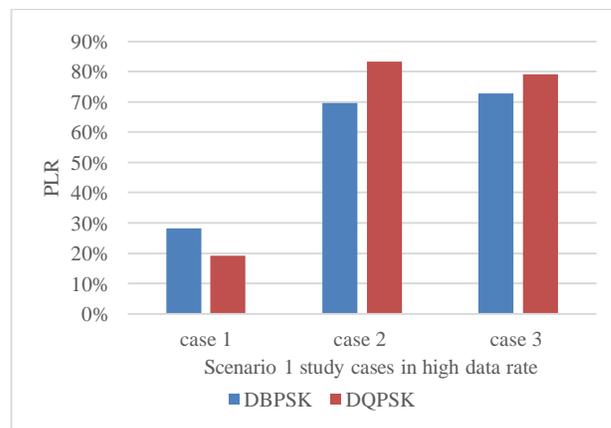


Fig. 5. Scenario 1: PLR (R-BAN) for DBPSK and DQPSK Modulation Schemes for High Data Rate Transmission (72 kbps).

*B. Scenario 2: Influence of DBPSK and DQPSK Modulation Schemes on the Transmission Performance (PLR) of R-BAN, Depending on the Number of Nodes in I-BAN*

In the present study, the scenario 2 described earlier in Section III is simulated, by studying for each of the two modulation schemes (DBPSK and DQPSK), the impact of the number of I-BAN nodes on transmission performance of R-BAN in the three transmission data rates: low, medium and high data rate.

Fig. 6, 7 and 8 successively present the packet loss rate (PLR) of R-BAN, at low medium and high data rate respectively and for different values of the number of nodes N in I-BAN (Case 1: N=2, case 2: N=5, case 3: N=10, case 4: N=15 and case 5: N=21).

An initial analysis of Fig. 6, 7 and 8 shows that, at low data rates, R-BAN transmission performance (PLR) in DQPSK modulation has a relatively correct PLR ( $\approx 10$  to 12%) and is practically unaffected by variations in the number of nodes of I-BAN. This is certainly not the case with DBPSK modulation which, under the same study conditions, clearly shows poor performance and seems to worsen further with a larger number of I-BAN nodes. This performance degradation obtained with DBPSK modulation is due once again to the sensitivity of the receiver which, in DBPSK, allows even weak signals from I-BAN to be received.

For medium and high data rates, and when DQPSK modulation is used, PLR deteriorates significantly, especially at high data rates. This performance degradation with DQPSK is mainly related to the weakness of this modulation scheme in front of bit error occurrences, which are much more frequent in the case of strong interference in the R-BAN transmission channel.

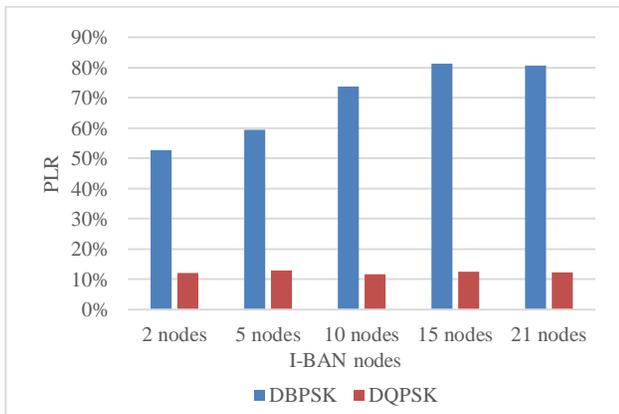


Fig. 6. Scenario 2: PLR (R-BAN) Evaluation, at Low Data Rates (3kbps) for DBPSK and DQPSK as a Function of the Number of nodes in I-BAN.

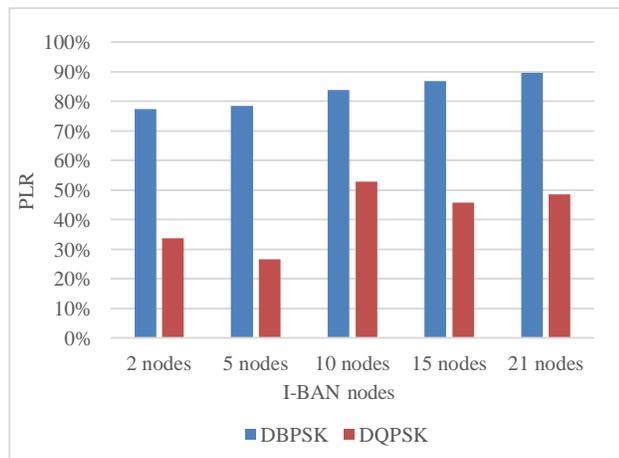


Fig. 7. Scenario 2: PLR (R-BAN) Evaluation, at Medium Data Rates (44kbps) for DBPSK and DQPSK as a Function of the Number of Nodes in I-BAN.

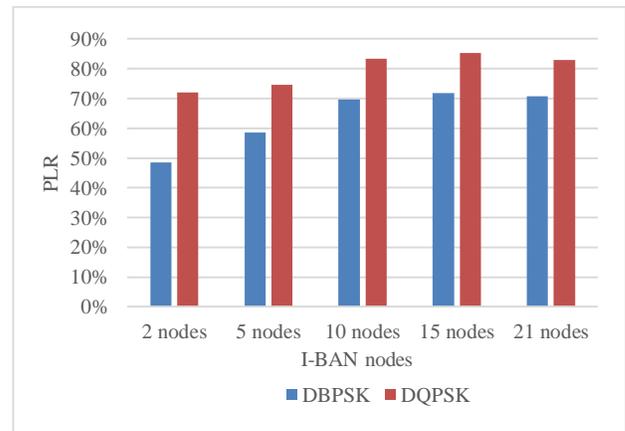


Fig. 8. Scenario 2: PLR (R-BAN) Evaluation, at High Data Rates (72kbps) for DBPSK and DQPSK as a Function of the Number of Nodes in I-BAN.

C. Scenario 3: Influence of the DBPSK and DQPSK Modulation Schemes on the Transmission Performance (PLR) of R-BAN, as a Function of Traffic Payload at I-BAN Level

To present the effect of traffic payload of I-BAN on mutual interference, the packet loss rate reached, at low, medium and high data rates is shown in Fig. 9, 10 and 11, respectively, in the cases where I-BAN transmits its packets with payloads of 10, 100 and 200 Bytes. The reference BAN payload is always maintained at 100 Bytes.

The results of this scenario show, once again, as for the two previous scenarios, that the traffic payload of I-BAN traffic has an impact on the PLR received at R-BAN. Moreover, for the same reasons of receiver sensitivity discussed above, the PLR reached at low and medium data rates (Fig. 9 and 10), with DQPSK modulation, is better than in DBPSK. However, at high data rates (Fig. 11), this behavior is inverted in favor of DBPSK modulation when the interference level increases (high data rate).

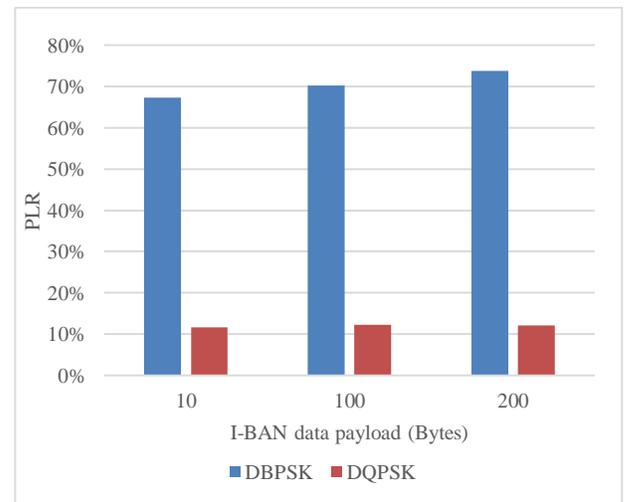


Fig. 9. Scenario 3: PLR (R-BAN) Evaluation at Low Data Rate (3kbps) for DBPSK and DQPSK as a Function of I-BAN Traffic Payload.

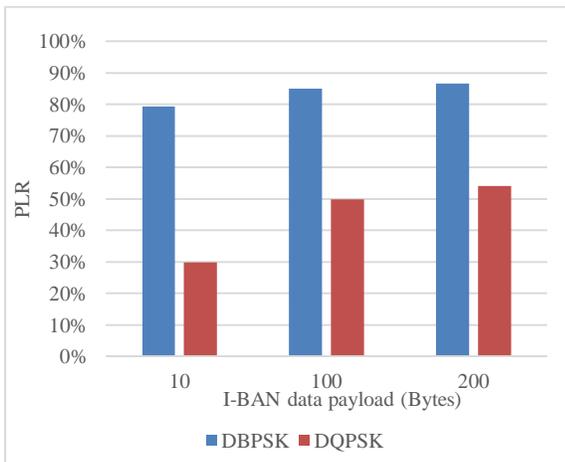


Fig. 10. Scenario 3: PLR (R-BAN) Evaluation at Medium Data Rate (44kbps) for DBPSK and DQPSK as a Function of I-BAN Traffic Payload.

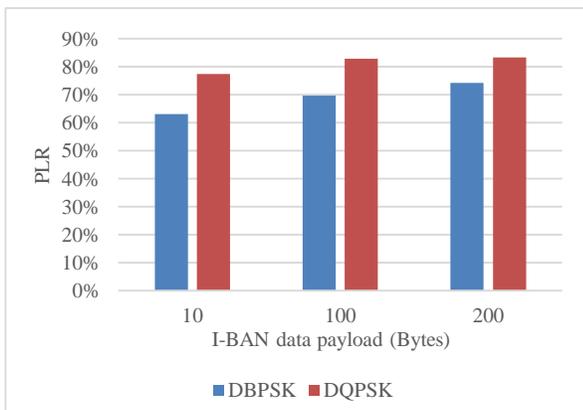


Fig. 11. Scenario 3: PLR (R-BAN) Evaluation at High Data Rate (72kbps) for DBPSK and DQPSK as a Function of I-BAN Traffic Payload.

The obtained results show that in the absence of adjacent BANs, the choice of DBPSK modulation in low data rate on-body medical applications such as: body temperature sampling (120 bps) or blood pressure measurement (<10 bps), and in medium data rate applications such as: EEG signal monitoring (43.2 kbps) or pulse oximetry (32 kbps), guarantees better PLR performance compared to DQPSK modulation. However, in these applications, switching to DQPSK modulation in the case of mutual interference is necessary in order to reduce the packet loss rate, without increasing the transmission power, which will reduce the energy efficiency of the nodes.

For high data rate on-body medical applications, such as ECG monitoring (71 kbps), the use of DQPSK modulation may be a good choice only in an ideal condition, where no other BAN is nearby. Nevertheless, switching to DBPSK modulation is required in the case of interference from other adjacent BANs to lower the packet loss rate, due to its robustness to maintain an acceptable link quality even when the interference level becomes more pronounced.

## V. CONCLUSION

This paper investigated and analyzed the impact of the choice of DBPSK and DQPSK modulation schemes (described in the 2.4 GHz narrow band of the IEEE 802.15.6 standard), on

the transmission performance of a BAN in a mutual interference environment. Performance is evaluated by using Packet Loss Rate (PLR) at various data rate levels. In this context, three study scenarios were examined, which investigate the effects of inter-BAN distance, number of nodes and traffic payload of the interferer BAN respectively, on the performance of a reference BAN.

The numerical study carried out on the three scenarios was able to show the impact of the three cited parameters on the studied performance for each of the two considered modulation schemes. At low and medium data rates, DQPSK modulation has shown better results in terms of PLR, however at high data rates the BAN transmission performance degrades for both modulation schemes, but DBPSK remains relatively more efficient.

The battery power supplying the nodes in a WBAN can deteriorate due to co-channel interference, resulting in a decrease in SINR which will cause throughput degradation. Therefore, Green communication based energy efficient techniques for co-channel interference BANs, is planned as a perspective to finalize this work.

## REFERENCES

- [1] ES-SAQY et al., "Very Low Phase Noise Voltage Controlled Oscillator for 5G mm-wave Communication Systems," 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, 2020, pp. 1-4, doi: 10.1109/IRASET48871.2020.9092005.
- [2] R. Negra, I. Jemili, and A. Belghith, "Wireless Body Area Networks: Applications and Technologies," *Procedia Comput. Sci.*, vol. 83, no. May, pp. 1274–1281, 2016, doi: 10.1016/j.procs.2016.04.266.
- [3] T. Hayajneh, G. Almashaqbeh, S. Ullah, and A. V. Vasilakos, A survey of wireless technologies coexistence in WBAN: analysis and open research issues, vol. 20, no. 8. 2014, doi: 10.1007/s11276-014-0736-8
- [4] M. BOUMAIZ et al., "Energy harvesting based WBANs: EH optimization methods," *Procedia Comput. Sci.*, vol. 151, pp. 1040–1045, 2019, doi: 10.1016/j.procs.2019.04.147.
- [5] L. A. N. Man, S. Committee, and I. Computer, IEEE Standard for Local and metropolitan area networks Y Part 15 . 6 : Wireless Body Area Networks IEEE Computer Society, no. February. 2012, doi: 10.1109/IEEESTD.2012.6161600.
- [6] S. Movassaghi, M. Abolhasan, J. Lipman, D. Smith, and A. Jamalipour, "Wireless Body Area Networks: A Survey," *Ieee Commun. Surv. Tutorials*, vol. 16, no. 3, pp. 1658–1686, 2014, doi: 10.1109/surv.2013.121313.00064.
- [7] R. Cavallari, F. Martelli, R. Rosini, C. Buratti, and R. Verdone, "A survey on wireless body area networks: Technologies and design challenges," *IEEE Commun. Surv. Tutorials*, vol. 16, no. 3, pp. 1635–1657, 2014, doi: 10.1109/SURV.2014.012214.00007.
- [8] P. Mathew, L. Augustine, and D. Selvakumar, "Hardware Implementation of NB PHY Baseband Transceiver for IEEE 802 . 15 . 6 WBAN," pp. 64–71, 2014 , doi: 10.1109/MedCom.2014.7005977
- [9] Y. Wang, Q. Wang, "Evaluating the IEEE 802.15.6 2.4GHz WBAN Proposal on Medical Multi-Parameter Monitoring under WiFi/Bluetooth Interference", *International Journal of E-Health and Medical Communications*, vol. 2, no. 3, July 2011, doi: 10.4018/jehmc.2011070103.
- [10] Wang Z., Jiang H. and Chen H., "WBAN Transceiver Design" In: *CMOS IC Design for Wireless Medical and Health Care*. Springer, New York, NY, pp 73-120,2004, doi: 10.1007/978-1-4614-9503-1\_4.
- [11] W. Yang and K. Sayrafian-pour, "Interference Mitigation for Body Area Networks," pp. 2193–2197, 2011, doi: 10.1109/PIMRC.2011.6139905
- [12] M. Boumaiz and M. Fattah, "The impact of distance between neighboring WBANs," 2019 7th Mediterr. Congr. Telecommun., pp. 1–4, 2019, doi: 10.1109/CMT.2019.8931341

- [13] P. R. Grassi, V. Rana, I. Beretta, and D. Sciuto, "2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks B 2 IRS : a Technique to Reduce BAN-BAN Interferences in Wireless Sensor Networks," 2012, doi: 10.1109/BSN.2012.30.
- [14] S. Movassaghi, M. Abolhasan, and D. Smith, "Interference Mitigation in WBANS : Challenges and Existing solutions," *Work. Adv. Real-time Inf. Networks*, 2013, doi: 10.5130/aaa.b
- [15] M. Barbi, K. Sayrafian and M. Alasti, "Low complexity adaptive schemes for energy detection threshold in the IEEE 802.15.6 CSMA/CA," 2016 IEEE Conference on Standards for Communications and Networking (CSCN), Berlin, 2016, pp. 1-6, doi: 10.1109/CSCN.2016.7785182.
- [16] I. Dotlic, "Interference performance of IEEE 802 . 15 . 6 Impulse – Radio Ultra – Wideband Physical Layer," no. September 2011, 2014, doi: 10.1109/PIMRC.2011.6139895.
- [17] D. M. Davenport, B. Deb, and F. J. Ross, "Wireless Propagation and Coexistence of Medical Body Sensor Networks for Ambulatory Patient Monitoring," pp. 8–12, 2009, doi: 10.1109/BSN.2009.8.
- [18] L. W. Hanlen, D. Miniutti, D. Smith, D. Rodda, and B. Gilbert, "Co-Channel Interference in Body Area Networks with Indoor Measurements at 2 . 4 GHz : Distance-to-Interferer is a Poor Estimate of Received Interference Power," pp. 113–125, 2010,doi:10.1007/s10776-010-0123-z.
- [19] X. Wang and L. Cai, "Interference Analysis of Co-existing Wireless Body Area Networks," pp. 0–4, 2011, doi: 10.1109/GLOCOM.2011.6133624
- [20] A. Zhang, D. B. Smith, D. Miniutti, L. W. Hanlen, D. Rodda, and B. Gilbert, "Performance of Piconet Co-existence Schemes in Wireless Body Area Networks," 2010, doi: 10.1109/WCNC.2010.5506746
- [21] E. K. Sungkwan, Y. Taeshik, and C. Kang, "Asynchronous inter-network interference avoidance for wireless body area networks," pp. 562–579, 2013, doi: 10.1007/s11227-012-0840-4.
- [22] Kim, S., Kim, S., Kim, J.W., Eom, D.S.: A beacon interval shifting scheme for interference mitigation in body area networks. *Sensors* 12(8), 10,930–10,946 (2012).
- [23] L. Wang, C. Goursaud, N. Nikaein, L. Cottatellucci, and J. Gorce, "Coexisting Body Area Networks," vol. 12, no. 1, pp. 123–133, 2013, , doi: 10.1109/TWC.2012.120412.112073
- [24] J. Mahapatro, S. Misra, S. Member, M. Manjunatha, and N. Islam, "Interference Mitigation Between WBAN Equipped Patients," 2012 , doi: 10.1109/WOCN.2012.6331909
- [25] T. T. T. Le and S. Moh, "Link Scheduling Algorithm with Interference Prediction for Multiple Mobile WBANs," 2017, doi: 10.3390/s17102231.
- [26] Z. Xie et al., "An Optimal Backoff Time-Based Internetwork Interference Mitigation Method in Wireless Body Area Network," vol. 2020, 2020, , doi: 10.1155/2020/4365191
- [27] J. Han, H. Kim, J. Bang, and Y. Lee, "Interference Mitigation in IEEE 802.15. 4 Networks," no. 1, 2011, doi: 10.1109/GLOCOM.2011.6134377
- [28] W. Lee and S. H. Rhee, "Channel Management Protocol for Wireless Body Area Networks," pp. 494–497, 2009, doi: 10.1007/978-3-642-04492-2\_61
- [29] Han, S., Lee, S., Lee, S., Kim, Y.: Coexistence perform mance evaluation of ieee 802.15. 4 under ieee 802.11 b in terference in fading channels. In: *Proc of IEEE PIMRC*, pp. 1–5 (2007)
- [30] G. Wu, J. Ren, F. Xia, L. Yao, and Z. Xu, "DISG : Decentralized Inter-User Interference Suppression in Body Sensor Networks with Non-cooperative Game," 2010, doi: 10.1109/UIC-ATC.2010.6.
- [31] K. Ghanem and P. S. Hall, "Interference cancellation using CDMA multi-user detectors for on-body channels," pp. 2152–2156, 2009, doi: 10.1109/PIMRC.2009.5450218
- [32] J. N. Bae, Y. H. Choi, J. Y. Kim, J. W. Kwon, and D. I. Kim, "Efficient Interference Cancellation Scheme for Wireless Body Area Network," vol. 13, no. 2, pp. 167–174, 2011, doi: 0.1109/JCN.2011.6157416
- [33] W. Yang and K. Sayrafian-Pour, "A DS-CDMA Interference Cancellation technique for body area networks," 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Istanbul, 2010, pp. 752-756, doi: 10.1109/PIMRC.2010.5671929
- [34] A. Tawfiq, J. Abouei and K. N. Plataniotis, "Cyclic orthogonal codes in CDMA-based asynchronous Wireless Body Area Networks," 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, 2012, pp. 1593-1596, doi: 10.1109/ICASSP.2012.6288198
- [35] W.-B. Yang and K. Sayrafian-Pour, "Interference Mitigation Using Adaptive Schemes in Body Area Networks," *Int. J. Wirel. Inf. Networks*, vol. 19, no. 3, pp. 193–200, 2012, doi: 10.1007/s10776-012-0192-2.
- [36] C. M. Liang, N. B. Priyantha, J. Liu, and A. Terzis, "Surviving Wi-Fi Interference in Low Power ZigBee Networks," pp. 309–322, 2010, doi: 10.1145/1869983.1870014.
- [37] J. Hou, B. Chang, D. Cho, and M. Gerla, "Minimizing 802.11 Interference on Zigbee Medical Sensors," 2009, doi: 10.4108/ICST.BODYNETS2009.6029
- [38] J. Huang, G. Xing, G. Zhou, and R. Zhou, "Beyond Co-existence : Exploiting WiFi White Space for ZigBee Performance Assurance," pp. 305–314, 2010, doi: 10.1109/ICNP.2010.5762779
- [39] K. Lee, C. Chae, T. Sung, and J. Kang, "Cognitive Beamforming Based Smart Metering for Coexistence with Wireless Local Area Networks," vol. 14, no. 6, pp. 619–628, 2012, doi: 10.1109/JCN.2012.00028
- [40] Y. Wang, Q. Wang, Z. Zeng, G. Zheng, and R. Zheng, "WiCop : Engineering WiFi Temporal White-Spaces for Safe Operations of Wireless Body Area Networks in Medical Applications," pp. 170–179, 2011, doi: 10.1109/RTSS.2011.23.
- [41] B. De Silva, A. Natarajan, and M. Motani, "Inter-User Interference in Body Sensor Networks : Preliminary Investigation and an Infrastructure-Based Solution," pp. 3–8, 2009, doi: 10.1109/BSN.2009.36.
- [42] W. Sun, Y. Ge, W. Wong, and S. Member, "A Lightweight Inter-User Interference Mitigation Method in Body Sensor Networks," pp. 34–40, 2012, doi:1 0.1109/WiMOB.2012.6379098
- [43] M. Boumaiz et al., "The impact of transmission power on the performance of a WBAN prone to mutual interference," 2019 Int. Conf. Syst. Collab. Big Data, Internet Things Secur., pp. 1–4, 2019, doi: 10.1109/SysCoBIoTS48768.2019.9028035.
- [44] S. Ullah et al., "A comprehensive survey of wireless body area networks on PHY, MAC, and network layers solutions," *J. Med. Syst.*, vol. 36, no. 3, pp. 1065–1094, 2012, doi: 10.1007/s10916-010-9571-3.
- [45] F. Di Franco, C. Tachtatzis, R. C. Atkinson, I. Timmiello, and I. A. Glover, "Channel estimation and transmit power control in wireless body area networks," no. May 2013, pp. 11–19, 2015, doi: 10.1049/iet-wss.2013.0070.
- [46] K. Takizawa, Channel model for wearable WBAN, no. July. 2008. [online]. Available: <https://mentor.ieee.org/802.15/documents>
- [47] M. Boumaiz, M. E. L. Ghazi, S. Mazer, M. E. L. Bekkali, A. Bouayad, and M. Fattah, "Performance analysis of DQPSK and DBPSK modulation schemes for a scheduled access phase based Wireless Body Area Network," 2018 9th Int. Symp. Signal, Image, Video Commun., pp. 163–167, 2018, doi: 10.1109/ISIVC.2018.8709232
- [48] Ningdalli, Ashwini and Dr. B. K. Sujatha. "FPGA Implementation of PI/2 BPSK and PI/4 QPSK for IEEE 802.15.6 WBAN Standard." (2015),
- [49] P. S. Hall, "Antennas challenges for body centric communications," *Conf. Proc. - 2007 IEEE Int. Work. Antenna Technol. Small Smart Antennas Metamaterials Appl. iWAT 2007*, no. 2, pp. 41–44, 2007, doi: 10.1109/IWAT.2007.370076.
- [50] S. L. Cotton, R. D. Errico, and C. Oestges, "A review of radio channel models for body centric communications," pp. 371–388, 2014, doi: 10.1002/2013RS005319.Received.
- [51] K. Y. Yazdandoost, Channel Modeling Subcommittee Report. 2010. [online]. Available: <https://mentor.ieee.org/802.15/documents>
- [52] A. Chi et al., "A 1 V 5 mA Multimode IEEE 802.15.6/Bluetooth Low-Energy WBAN Transceiver for Biotelemetry Applications," vol. 48, no. 1, pp. 186–198, 2013. doi: <https://doi.org/10.1109/JSSC.2012.2221215>
- [53] A. Boulis, "Castalia A simulator for Wireless Sensor Networks and Body Area Networks - User's Manual," Version 3.0, no. March, p. 79, 2010.

# Ensemble Machine Learning Model for Higher Learning Scholarship Award Decisions

Wirawati Dewi Ahmad<sup>1</sup>, Azuraliza Abu Bakar<sup>2</sup>

Centre for Artificial Intelligence Technology, Faculty of Information Science and Technology  
University Kebangsaan Malaysia, 43600 Bangi, Selangor Darul Ehsan, MALAYSIA

**Abstract**—The role of higher learning in Malaysia is to ensure high quality educational ecosystems in developing individual potentials to fulfill the national aspiration. To implement this role with success, scholarship offer is an important part of strategic plan. Since the increasing number of undergraduates' student every year, the government must consider to apply a systematic strategy to manage the scholarship offering to ensure the scholarship recipient must be selected in effective way. The use of predictive model has shown effective can be made. In this paper, an ensemble knowledge model is proposed to support the scholarship award decision made by the organization. It generates list of eligible candidates to reduce human error and time taken to select the eligible candidate manually. Two approached of ensemble are presented. Firstly, ensembles of model and secondly ensembles of rule-based knowledge. The ensemble learning techniques, namely, boosting, bagging, voting and rules-based ensemble technique and five base learners' algorithm, namely, J48, Support Vector Machine (SVM), Artificial Neuron Network (ANN), Naïve Bayes (NB) and Random Tree (RT) are used to develop the model. Total of 87,000 scholarship application data are used in modelling process. The result on accuracy, precision, recall and F-measure measurement shows that the ensemble voting techniques gives the best accuracy of 86.9% compare to others techniques. This study also explores the rules obtained from the rules-based model J48 and Apriori and managed to select the best rules to develop an ensemble rules-based models which is improved the study for classification model for scholarship award.

**Keywords**—Scholarship classification; ensemble learning; rules-based classification; rules-based ensemble

## I. INTRODUCTION

Since the year 2013, the number of Higher Education scholarship applications has increased in line with the increasing number of graduates at under graduate level every year. Therefore, the scholarship provider institution need to use the best approach or technique to select the best candidate for awarding education scholarship in appropriate time. Some of the scholarship provider institution have begun to adopt analytical data approach with appropriate classification techniques in managing and selecting the best candidates for scholarships award. By implementing this approach, the best eligible candidate can be selected automatically and provide better decision than manual methodology.

Through preliminary studies, the scholarship application processing method is using three key methods [9], namely, (1) pre-applicant checking that need to meet the general requirements using the online application system; (2) manual

checking by responsible officer to verify applicant information and (3) final selection by a voting committee to select eligible candidates based on the information or characteristics of the applicant manually. These methods have two (2) major weaknesses, namely, (1) requiring skilled system programmers to maintain the system code and (2) required a long time and a large number of employees to process and do the checking for each application for decision within a specified time frame. Hence, this study aims to propose an analytical data approach using ensemble learning to select the best eligible candidate based on existing scholarship recipient's data. This approach helps to provide a solution for time and people issues in the current method by providing eligible candidate based on classification model of ensemble learning technique.

This study conducted a research on the best ensemble classification techniques for awarding scholarships to graduate students in Malaysia who applied for scholarships provided by the Ministry of Higher Education Malaysia. A set of real data contains a collection of application including the successful and unsuccessful candidate that were collected from the online scholarship application system are processed and transformed according to each technique used. In this process, there are variety of filters to narrow down the important features. It is also important that any preparation of the data prior to fitting with the algorithm used. In order to develop the model, we carried out two types of experiments, namely, common ensemble learning model and rules-based ensemble model. These model are been trained and tested using two data distribution techniques namely strata distribution and k-folds distribution to ensure best learning and performance model.

Moreover, in the model evaluation task, we used two different evaluation, namely technical evaluation and expert evaluation to ensure that the assessment of performance performance of each technique is evaluated. For technical evaluation, common performance measurement namely accuracy, precision, recall and F-measure are used to evaluate model performance in terms of accuracy and error for models produced using the selected technique. In addition, evaluations by domain specialists are also done to ensure better results are made. After evaluation of the model developed with the best performing techniques is the recommended technique.

It has been observed that most of the studies in ensemble learning especially in education domain is focusing in several main issues but not in scholarship award decision model. There also been observed that in scholarship award decision model, only single classification was applied.

Therefore, this paper is focusing in propose the best ensemble learning model for scholarship award in higher learning to help the scholarship provider turn to new and systematic way on selecting the best candidate to receive a scholarship while in their program of study.

## II. RELATED WORK

### A. Data Analytic Approach in the Education Sector

The study of data analytic approach in the Education sector includes higher learning programs has expanded extensively. These studies contributed in providing solutions to addressing problems in the education sector in a country using classification and prediction methods [1][32]. The useful information obtained through this method may provide a solution to help the decision-making process in education insight in the future. The rapid growth in the size of the educational data also shows that the classification and prediction method applied requires a more efficient set of algorithms and techniques in giving the best results [1]. From the literature studies found that, there are two main aspects of the education which are often being focused in the research namely student group issues and higher learning institution issue as shown in Fig. 1.

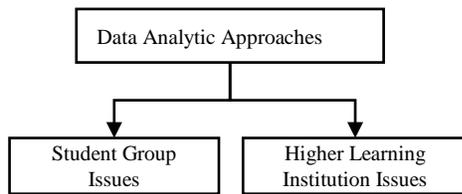


Fig. 1. Data Analytic Applied Studies in Educational Sector.

For the student's group issues, there are many studies involved such as a classification and prediction model in monitoring student academic performance based on existing student achievement records [2][3][4], selecting best student who are eligible to be offered certain university program and selecting the best study programme or courses for student to register.

While for the higher learning organization many studies involved such as establishing a list of programme and courses to be offered by higher learning institution, selecting the best recipients for research grant award[3] and determine the best student intake for certain program ensuring a high pass rate [5]. Beside that, studies on a scholarship offers also been done using this approach to support decision making process in selecting best recipients.

### B. Ensemble Learning Method

There are studies on producing the scholarship award decisions using data analytic approach which used a single classification method in producing the classification and prediction model [6][7][8][9]. However, the disadvantages of a single classification method may result in improperly constructed models and the quality of the decision made will be affected. A single classification method is found to reduce the chances of exploring knowledge using other techniques that cause the loss of information that is potentially important for the prediction process [10]. Additionally, the increasingly

large data size factors required more effective technique to be used rather than single algorithm approach.

Therefore, the technique of ensemble learning is uses to combine more than two algorithms to produce the best learning model. Many studies as shown in [11][10][12][31] have proven that using this method may overcome the weaknesses of single classification method and a more robust final model can be developed. This is because using this method, the final model is capable to combine the characteristics of a single classifier used with either the same or different function [10] and gives a better result than a single classifier [10][13][12][11].

This technique has two main objectives: 1) to increase the accuracy of overall forecasts compared to a single classifier; and 2) improve the rate of generalization better because of its specific measurement measurements. As a result, the final classifier can solve unresolved issues with a single predictive model [13]. The model performance on examples not seen during training demonstrate the actual capabilities of the model [14].

- Ensemble Techniques

Boosting is the basic technique used in combined classification to enhance the capability of a classification model. This technique focuses on weak classifier by re-testing classified items inaccurate so as to be accurately classed in the next iteration. It was introduced by Schapire in 1990 and has been widely used to date. The welding process combines this technique as in Packing Techniques by sampling data on each iteration using the majority draw technique. One of the commonly used algorithms is AdaBoost.[15]

The AdaBoost algorithm was introduced by Freund and Schapire in 1997 [15]. It generates a set of hypotheses, and combines them through parliamentary majority parallel to the class achieved by each hypothesis. Hypotheses are generated by training weak classifier, using the records or samples obtained during the iteration process using regular training data. This cycle ensures that the classification of the wrong records classified by the previous classifier is more likely to be included in the subsequent iteration training data. Hence, the training data being constructed is increasingly difficult to classify but the accuracy of Success is classified accordingly is highly dependent on the previous welding pattern. It also seeks to improve welding accuracy and is commonly used in many fields. It is easy to implement and is not exposed to over-fitting problems [16]. It also produces a basic model sequence with different weight distribution over a set of exercises [10].

Bagging is an easy technique used in ensemble learning. It may combine base learners call classifiers and produce final ensemble model or use an algorithm with multiple set of testing data as ensemble basis. Based on the review by [17] the single-core tree algorithm and decision tree will produce different tree outputs. Different trees are formed when there is a change in the starting point of the training data that results in a decrease in stability. It can be used for any other classification method in the data mining approach. This technique is also suitable to be used in finding suitable models

for large-sized data as the classification becomes easier as the training data sets are broken down according to certain techniques [15] (Deng 2007). The output of this technique is the final output with the majority draw technique which selects the best weld results among good models.

Voting is also a popular joining welding technique [18] [19][20][20]. It combines many classmates and conducts a lottery process during welding. Many methods are used to select the best classifier vote among existing classmates. Average, minimum and maximum techniques are among those used. Classification using the majority technique is often chosen because it produces the output of the best model that is fair and balanced. For example, a set of data that has two class labels with the best three classifiers will determine the majority of classmates who label the most.

Table I shows the comparison on advantages and disadvantages between these three-ensemble techniques that leads to the experimental design.

In conclusion, the disadvantages of ensemble learning model is depends on some factors such as bias, noise and variants [21][22][12]. These weaknesses can be overcome by using the rules-based ensemble model. This technique also known as association classification (AC) whereby it combine rules generated from different rules-based model to perform as final ensemble model. Many studies have shown that AC is capable to enhance the ensemble classification capabilities with higher accuracy rather than common ensemble techniques. [23][24][25]. Therefore, in this study use two rules-based classification algorithm, namely, Apriori algorithm and J48 from decision tree algorithm to develop an ensemble rules-based model to determine higher learning scholarship award decision.

TABLE I. COMPARISON ON ENSEMBLE LEARNING TECHNIQUE

Technique	Advantages	Disadvantages
boosting	Improve the ability of a weak model by repeatedly conducting exercises using different data samples. Focusing on exemplary examples in the model iteration further results in higher model accuracy	Longer and higher cost modeling time Classification model becomes more complex and difficult to implement using real-time platforms Less suitable to use if the data has high noise levels and outliers
Bagging	Improve accuracy and reduce variance, thereby reducing overfitting problems Take samples of different data samples to improve model accuracy The easiest method compared to the upgrade technique Suitable for data sets with even high noise levels	Improve the computational complexity of the model Effectively only for examples of records with high frequency
Voting	Performance improvements for a large number of classes. Can be used on all types of data	Needing more good classifier is better because the classification is based on the majority.

The association rules is a very popular and extensive method that can be used to discover interesting relationships between variables in a very large database [26]. This technique seeks to find interesting and frequent patterns, knowing the correlation between data sets in a data repository that will provide useful information in support of future decision-making. In this study, the use of apriori algorithm techniques used to look at the attributes of attributes in Malaysia's higher education scholarship packages. This attribute represents the frequency of candidate's pattern of success or is not offered a higher education scholarship offer.

Apriori is a simple algorithm used to find frequent set of records in a data set. The output of this algorithm is a set of sequential tips with the frequency set of the data item. The generated tips are dependent on several filtering metrics such as minimum support and confidence level [21]. This means that with this algorithm and using existing data sets, it is able to provide frequent data set patterns and help in identifying useful information to support future decision-making processes. There are two main processes in this technique, namely the process of obtaining the maximum set of items by relying on the minimum support parameters and the rules of generating rule rules based on the minimum confidence. Both of these processes are interconnected in determining the characteristics and number of rule rules that are generated [27].

In this study, a base learners under decision tree category used was J48. This algorithm was chosen because the decision tree is compatible with both types of data whether numerical or nominal. In addition to whatever data size, higher accuracy in the decision tree classification technique illustrates that the technique successfully classified well. The decision tree can handle and handle large quantities of input data such as text with numerical or nominal numerical data [26]. It is a supervised learning approach that has the ability to extract information from a large amount of data based on rules or decision rules [26][28].

In this paper, we present the comparison of ensemble learning model for scholarship award decision based on the real scholarship recipient's database using five different ensemble learning techniques in order to propose the best technique in scholarship award decision making. Eventhough we used the existing techniques, however, the same research strategy has not been implemented on the real scholarship data in Malaysia which requires specific and systematic reviews. Hence, the results of this study will proposed an ensemble model for shcolarship award decisions which can improve the scholarship decision making process in the future.

### III. RESEARCH METHODOLOGY

This section explains the process that involved in the proposed ensemble learning framework for scholarship award decision. The study of classification model for scholarship recipients has been done widely by implementing the data mining approaches using a single classification technique. Therefore, in this study we present a better technique in a classification model using ensemble learning classification approaches to determine scholarship award decision by using common ensemble classification method and a rules-based

ensemble technique. The actual scholarship data that been used was obtained from the government scholarship provider. Fig. 2 shows the proposed ensemble learning framework based on the literature review conducted.

**A. Scholarship Data Collection and Pre-Processing**

After a preliminary study of the data, the data preparation process was conducted to provide a complete set of data for modeling purposes. The two main processes undertaken in this stage are data integration and cleaning and features selection and transformation.

1) *Data intergration and cleaning:* The raw data obtained was analyzed to see the type and size of the attributes and their importance. A total of 87, 000 data was retrieved with more than 40 attributes recorded. After a preliminary study of the data on several steps has been taken to ensure clean and complete data is provided. Data noise exists due to a poorly managed database. Records that do not store the correct values for the attribute type and size are deleted. Likely, actual databases are sometimes used to carry out system entry tests that receive meaningful test data.

For a record that has no value for the final decision of a committee it is defined as incomplete data and eliminated from the net data set. Similarly, a set of data that is of no value to the upload document which indicates the applicant is not committed to applying for a scholarship.

An important attribute that determines the candidate's decision is a decent value of an office that stores a value other than 1 or 2 is also eliminated. This odd value may be due to a system testing process or an applicant who does not resubmit an incomplete document during the application process. This data record is eliminated because it is considered not important in the welding modeling process.

Inconsistent data records are also found during the data analysis process. For example, the applicant's age exceeds 64 years, the applicant's age is less than 22 years, the non-existent identity card number, or a record exceeding one application is also eliminated.

As a result of the above process, a total of 57,000 net records were successfully provided. Attributes can be categorized into three categories, attributes that are eliminated i.e. over 29 attributes, attributes are retained by 11 attributes, new attributes created to replace attributes that are deleted by 4 attributes. Table II shows an example of a list that has been deleted after the data clearing and integration process.

2) *Features selection and transformation:* In this study, the initial set of data was processed into a new data set with attributes of the correct type and value for the purpose of modeling. This new set of data is assessed using specific feature selection techniques and transformed into appropriate forms to meet modeling algorithm requirements. Table III depicts the selected attributes for ensemble knowledge model.

**B. Ensemble Learning Modelling**

The preprocessed scholarship data set was spilt into two set of training and testing data. This distribution used two techniques namely strata distribution and k-folds distribution. While using these two techniques, five base learner algorithm were chosen as a base learner, namely, J48, Support Vector Machine (SVM), Naïve Bayes (NB), Artificial Neuron Network (ANN) and Random Tree (RT). These five bases learner will be used in three common ensemble techniques namely Boosting, Bagging and Voting. Each model produced by these technique will be evaluated using common evaluation metrics. To develop the ensemble models, there are two type of ensemble learning modelling framework used namely Common ensemble learning model and Rules-based ensemble model.

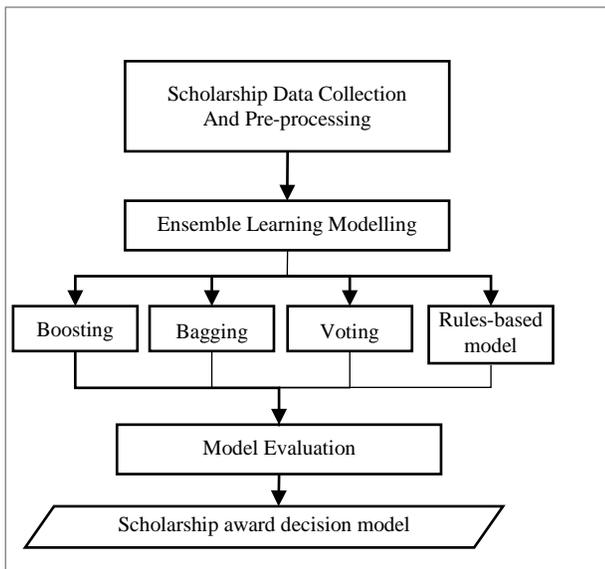


Fig. 2. Proposed Ensemble Learning Framework to Determine Scholarship Award Decision.

TABLE II. TABLE SCHOLARSHIP DATASET FINAL ATTRIBUTES

Attribute name	Type
Age	Varchar
Gender	Varchar
marital status	Varchar
disabilities	Varchar
sponsorship status	Varchar
work status	Varchar
approval status	Integer
field of study	Integer
program structure	Integer
graduated years of study	Varchar
Level of study	Integer
University	varchar

TABLE III. ATTRIBUTES TRANSFORMATION CODING

Attribute name	Definition
age range	1 = 20 to 25 years 2 = 25 to 29 years 3 = 30 years to 34 years 4 = 35 years to 39 years 5 = 40 years to 49 years 6 = 50 years to 64 years
Gender	1 = male 2 = female
marital status	1 = single 2 = married 3 = Others
disabilities	1=Yes 2=No
sponsorship status	0 = Yes 1=No
work status	0=Employed 1=Unemployed
approval status	1=Eligible 2=Not Eligible
field of study	1=Social science 2=Science & Technology
program structure	1=Course work 2=Mix mode 3=Research
graduated years of study	1=2014 and 2015 2=2013 3=2012 4=2011 5=between 2006 and 2010 6=between 1990 and 1999 7=Others
Level of study	8=Master 13=PhD
University codes	Special codes for each university

For common ensemble learning model, we used three common ensemble techniques namely boosting, bagging and voting. From these techniques, three ensemble learning experiment were designed in order to obtain the best ensemble model for higher learning scholarship award decision using ensemble learning technique. For boosting and bagging experiment design, the five base learners were used as an individual classifier while in voting experimental design, the five base learners will be taken as one ensemble classifiers as shown in Fig. 3.

For the last experiment, the rules-based ensemble model was developed using the set of rules generated previously by the decision tree rules based model, J48 and an Apriori rules based model as shown in Fig. 4. Each rules generated from this experiment will be analyze and measured in order to select the best rules. The ensemble rules then combined in an ensemble rules-based model and tested with using testing data prepared earlier.

The rules-based ensemble method is the techniques that can improve ensemble learning capabilities with high accuracy rather than single ensemble model. In this study, the rules obtained from the single-layer rules-based model were excavated and evaluated to be combined with rules obtained from the rules-based classification model using the Apriori algorithm. The combination of these rules will determine whether the model's ability to classify Eligible or Not qualified candidates is better than common ensemble techniques used.

In this framework, two sets of training data are used as training data used in previous experiment and new training data sets only sample candidates are not eligible. This is aimed at getting more tips for "Not Eligible" candidate cases. These rules are then combined with the selected rules from the J48 model in a new set of rules that will be used in the rules-based ensemble approach experiment. To determine the correct classification and incorrect classification rate, the majority algorithm is shown as Fig. 5.

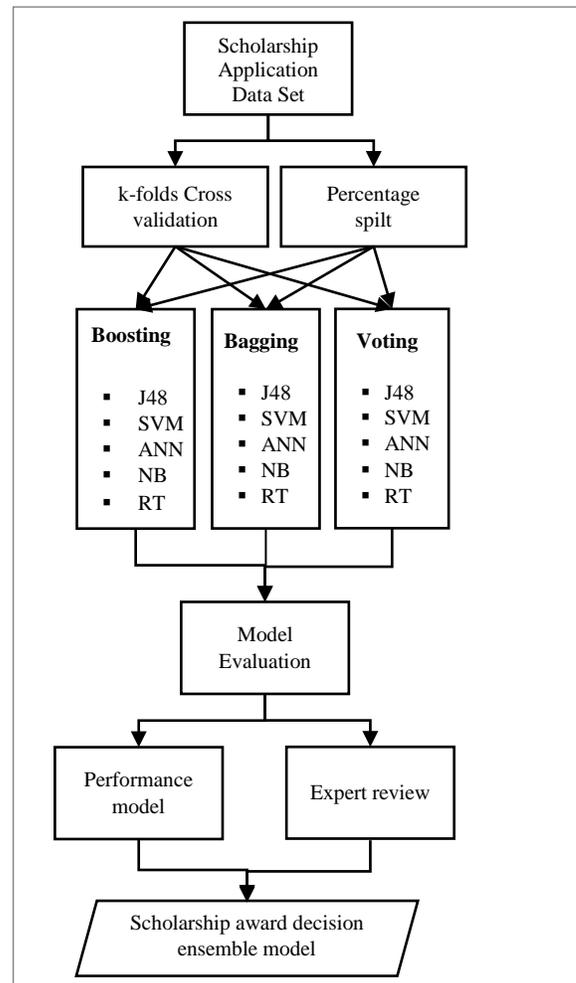


Fig. 3. Common Ensemble Learning Model to Determine Scholarship Award Decision.

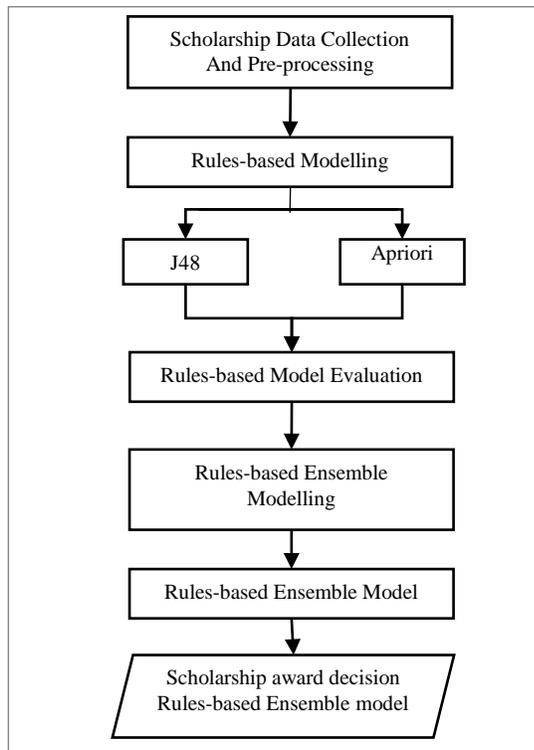


Fig. 4. Rules-based Ensemble Model to Determine Scholarship Award.

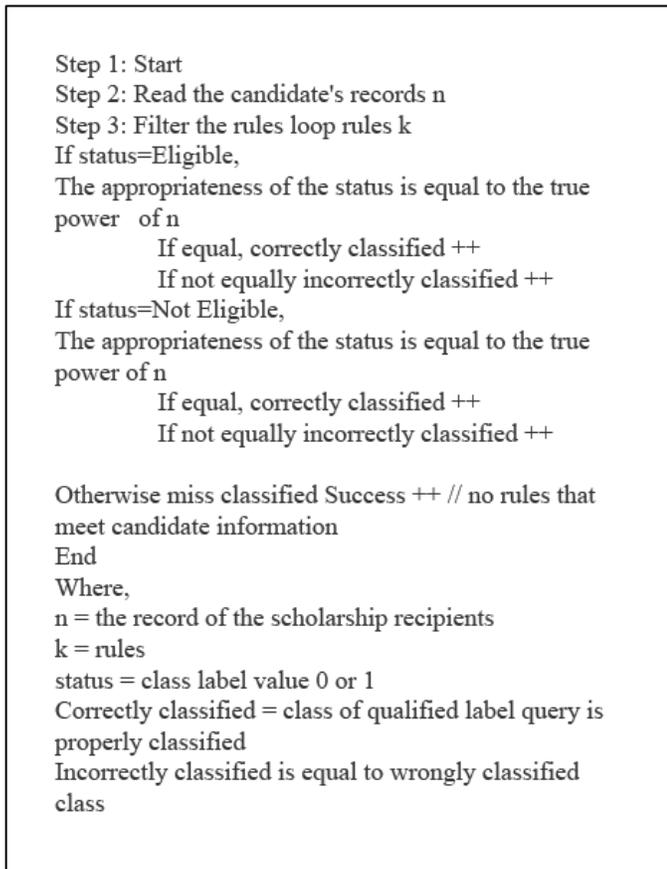


Fig. 5. Majority Algorithm.

### C. Model Evaluation

In this study, model evaluation is based on two approaches, namely technical evaluation and expert evaluation. The performance on classification model were measured using common performance measurement namely accuracy, precision, recall and F-measure. The formula to calculate these measure are shown below (1) (2) (3) and (4).

$$Accuracy = \frac{\text{correctly classified scholarship result}}{\text{total scholarship data}} \quad (1)$$

$$Precision = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \quad (2)$$

$$recall = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad (3)$$

$$F - \text{measure} = \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

Where,

1) True positive(TP) = eligible candidate correctly classified

2) False positive(FP) = not eligible candidate incorrectly labeled as eligible candidate

3) False negative(FP) = eligible candidate incorrectly labeled as not eligible

In this study, the expert evaluation also used to review and validate the acquisition of rules from the rules-based ensemble modelling. The scholarship domain expert who has been involved in selecting and verifying the best rules to be used in rules-based ensemble modelling. The results from all experiment, were then compared to propose the best ensemble model generated by this study.

## IV. RESULT AND DISCUSSION

### A. Ensemble Learning Modelling

The accuracy of a classification model for scholarship award decision can be measured by calculating the percentage of correctly classified number of eligible scholarship candidate and divide by the total of candidate number. In this study, the voting technique showed the highest accuracy of 86.9% compared to bagging and boosting techniques. This is followed by bagging ANN model with accuracy of 86.5% while the best model for boosting technique is the ensemble classification model using boosting SVM which achieved 86.3% of accuracy rate. While the weakest model is ensemble classification model using boosting RT algorithm which is only 81.6% of accuracy. However, overall, the ensemble technique with five base learners produced accuracy more than 80% and can be categorized as good with the best technique of voting techniques as shown in Fig. 6.

Beside accuracy, precision rate is used to show the rate of predicted positive by dividing the number of eligible candidate correctly classified and the total of the number of eligible candidate correctly classified and the number of not eligible candidate incorrectly labeled as eligible candidate. By observed this measure, we can see the ability of a model to predict the eligible candidate to be awarded. In this study, the bagging SVM technique show the highest precision rate of

0.868 followed by the model of bagging NB technique with precision rate of 0.863. While the weakest model was boosting SVM model with precision rate at 0.7712.

In this study, recall measurement indicate the ability of a model to predict the eligible candidates by dividing the number of eligible candidate correctly classified and the total of the number of eligible candidate correctly classified and the number of eligible candidate incorrectly labeled as not eligible candidate. In this study, the ensemble classification model using bagging SVM technique achieves the highest recall rate of 0.869. The overall result show that recall measurement for overall model shows good value which is more than 0.800. This demonstrates the ability of each model to classify the eligible candidate class is good.

For F-measure rates, both measurements precision and recall was combined to see the optimal blend of prediction ability for an ensemble learning model. In this study, the best F-measure rate is 0.808, using bagging SVM technique. Overall result for precision, recall and F-measure are shown in Fig. 7.

### B. Rules-Based Ensemble Modelling

1) *J48 model rules extraction:* In the classification process besides being able to specify the class labels for data records, the algorithm is used to generate rules that determine the class labels. One of the common rules-based algorithm widely used is J48 algorithm. The rules pattern generated from J48 classification modeling process can be used to obtain hidden information that may be useful to the data owner and combine with other rules-based model to generate new rules-based model.

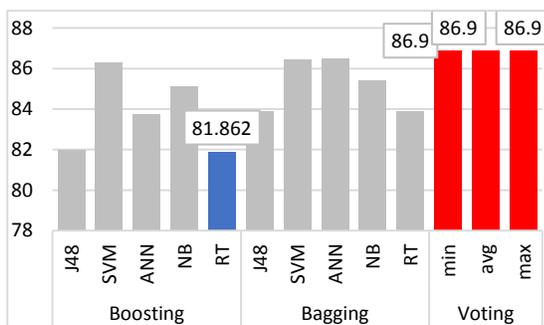


Fig. 6. Rules-based Ensemble Model to Determine Scholarship Award.

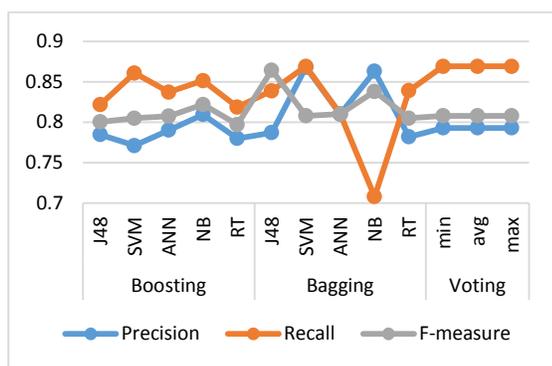


Fig. 7. Others Measures Result for Common Ensemble Learning Model.

Extracting useful information and knowledge to support decision making is one of the main goals of data mining. A knowledge analysis process enables hidden information to be obtained and verified to determine its importance to the stakeholders. In this study, knowledge analysis was conducted on the results of the rules generated by the J48 model to develop an ensemble rules-based model in determine a higher learning scholarship eligible candidate. Rules with high frequencies are taken and been verified to ensure that the rules used in ensemble rules-based model are valid and useful in selecting the most qualified candidates for awarding Higher Education scholarships based on records of successful and failed candidates.

The rules generated by this algorithm is depended by two class labels, which are rules that give the label class 1 the 'Eligible' and class 2 'Not Eligible'. The rule rules generated by this model are 2771 rules with the size of the rules 3265. These rules are verified with the help of expert domains to see their importance and validity. The process of validation this rules is important to ensure that rules are not significant can be filtered and remove. The method of filtering useful rules is run by the number of frequencies each rules and expert assistance [29]. The result of this process manages to identify two categories of rules namely the rule of determining the classification of eligible candidates and the rule of determining the classification of not eligible candidates. For the classification of qualified candidates, the discussion concurred that the rationing tips of more than 15 candidates were classified as important tips of 70 of the rules while the rules for unqualified cases considered important were 67 rules. Table IV lists examples of rules that define the Eligible label class and the knowledge acquisition gathered when the rules transformed based on the original data set.

Based on the analysis of these rules, it can be concluded that the rules generated by the J48 algorithm are more focused on the case of 'Eligible' candidates more than the 'Not eligible' candidates. This is because the number of rules for determining the Ineligible candidate case is small and the number of records is small compared to the Eligible Candidate.

In order to improve the model's capabilities, the improvement in the classification of negative records, which is not feasible, should be improved. In order to address this issue, the study will then conduct experiments on one of the techniques to improve the accuracy of a classifier model, i.e. the rules-based ensemble approach. To derive more useful rules, this study developed a model from a common rules-based algorithm namely Apriori and extracted the rules generated by this model.

Furthermore, the identified rules will be combined with the rules from the J48 model as discussed earlier into a new rules-based ensemble model using the simple majority technique which will be discussed in next section. This data mining approach is solution that can improve classification and prediction task with higher accuracy. According to the study (Mittal et al., 2017), this technique provide a solution to extract a classifier that contains the simple If-Then's rule and produces high accuracy.

TABLE IV. RULES WITH HIGH FREQUENCY FOR ELIGIBLE CANDIDATE

Rules	Information Gathered
age_range = 2 AND level_of_study = 8 AND university_code = M0101028 AND field_of_study = 2 AND gender = 2	Women candidates aged 25 to 29 who apply to continue their studies at the University 'M0101028' and continue their studies in Science and Technology
university_code = M0100612 AND work_status = 1 AND sponsorship_status = 0 AND field_of_study = 2 AND graduated_years_of_study = 5 AND program_structure = 1 AND marital_status = 2 AND gender = 2	Candidates applied to continue their studies at University 'M0100612' and continue their studies in Science and Technology, have been sponsored at undergraduate, graduating years between 2006 and 2010 and married status
age_range = 2 AND level_of_study = 8 AND university_code = M0100612 AND work_status = 1 AND sponsorship_status = 0 AND field_of_study = 2 AND graduated_years_of_study = 5 AND program_structure = 1 AND marital_status = 2 AND gender = 2	Women candidates aged 25 to 29 who apply to continue their studies at the University 'M0100612' at Masters level and continue their studies in Science and Technology, have been sponsored at undergraduate and marital status
age_range = 2 AND level_of_study = 8 AND university_code = M0100612 AND work_status = 1 AND sponsorship_status = 0 AND field_of_study = 2 AND julat_thn_ijz = 5 AND program_structure = 1 AND gender = 2	Women candidates aged 25 to 29 who apply to continue their studies at the University 'M0100612' at Masters level and continue their studies in Science and Technology

2) *Apriori model*: In this experiment, the input data is a set of scholarship data of 10,000 candidates as D1. This data set has gone through the process of cleansing and transformation as described in previous chapters. This data set is also the same set of data used in the process of ensemble learning modeling before. While the output of this algorithm is a set of rules with the frequency. The minimum support value used in this experiment is 0.1 with a confidence level of 0.9. The purpose of selecting these values is to produce a number of important rules.

As a result, a total of 500 rules were observed, and only 221 rules related to the class label were identified. The rules generated by this model are includes of 2, 3, 4 and 5 sequences. Examples of rules with the frequency values obtained are as in Tables V and VI. Furthermore, the rules generated by this model is evaluated by technical measurement and expert domains knowledge to find best rules to be used in rules-based ensemble model.

In the rules evaluation process, only rules with minimum confident level and minimum support was reviewed and selected. During this process, the rules was taken and analyzed using SQL to determine whether it is useful or not. Expert review also been performed to identify the best rules. From this process, we found that only 18 important rules were identified from a total of 221 rules on the corresponding label class. These rules includes rules for eligible candidate cases. As no suitable rules are in place for Not eligible cases, another rules-based model was developed using the only Not eligible candidate data set (D2) using the same value for confident level and minimum support.

The results of this experiment produced estimated a total of 200 rules. From the evaluation process, only 69 rules were associated with the class label and from that only 8 important rules identified from this experiment. The total of 26 rules were obtained from Apriori modelling will be used as rules condition for rules-based ensemble model to determine scholarship award decision. Table VII shows the result of rules selected from these two experiments.

3) *Rules-based ensemble model*: In this study, the available rules from a single classifier model (J48) experiment and the model of associative rules are combined using simple majority technique as described in previous chapters. The rules-based ensemble model was develop using PHP programming language and the same training data in previous experiment. As shown in Table VIII. A total of 137 rules were taken from the J48 model while 26 rules from the Apriori model then ensembled in the rules-based model. This model then tested using the same set of data and the accuracy of the model was observed.

As the result, Table IX shows that the ability of this model to weld properly for qualified candidates is high and this gives a good value for the accuracy and accuracy of this model of 85.94% and 0.8609.

TABLE V. 2 AND 3 SEQUENCE RULES EXAMPLE

2 sequences	3 sequences
age_range = 2 5174 ==> gender = 1 5169	age_range = 2 class_label = 1 4665 ==> gender = 1 4661
marital_status = 1 4984 ==> gender = 1 4979	class_label = 1 field_of_study = 1 4582 ==> gender = 1 4578

TABLE VI. 4 AND 5 SEQUENCE RULES EXAMPLE

4 sequences	5 sequences
age_range = 2 disabilities = 1 class_label = 1 4653 ==> gender = 1 4649	disabilities = 1 sponsorship_status = 0 level_of_study = 8 program_structure = 1 4620 ==> gender = 1 4614
disabilities = 1 class_label = 1 field_of_study = 1 4573 ==> gender = 1 4569	gender = 1 disabilities = 1 sponsorship_status = 0 class_label = 1 4937 ==> program_structure = 1 4930
disabilities = 1 sponsorship_status = 0 work_status = 1 class_label = 1 5716 ==> gender = 1 5708	gender = 1 sponsorship_status = 0 level_of_study = 8 program_structure = 1 4621 ==> disabilities = 1 4614

TABLE VII. NUMBER OF RULES OBTAINED FROM APRIORI MODEL

	Number of rules gathered	Number of rules includes class label	Number of selected rules
Apriori D1	500	221	18
Apriori D2	200	69	8
Total	700	290	26

TABLE VIII. NUMBER OF RULES FOR RULES-BASED ENSEMBLE MODELLING

	Number of rules
J48	137
Apriori	26
Total	163

TABLE IX. RULES-BASED ENSEMBLE MODEL RESULT

Measures	Value
TP	8585
FP	1387
TN	2
FN	0
Accuracy	85.94%
Precision	0.8609
Recall	1
F-measure	0.9252

From the experiment result shows that True positive (TP) value is height and this gave high accuracy rate (85.95%). For the FP value, there are 1,387 record of not eligible candidate were incorrectly labeled as eligible candidate. However, compared to other ensemble model experiment, the F-measure for rules-based ensemble model is higher 0.9252.

## V. CONCLUSION

In this study we present a better technique for a classification model for higher learning scholarship award decision using ensemble learning approach by using common

ensemble classification method and a rules-based ensemble technique. All the model, then evaluated by technical and domain experts to obtain the best model.

Additionally, the preprocessing data managed to provide a complete dataset to be used as training and testing dataset. The preparation of this dataset also managed to propose the suitable type and transformation codes for some attributes which can be applied at the beginning of input data process in scholarship application form in the future.

Overall result shows that each of ensemble learning techniques used, perform a good result on the classification and prediction performance as shown Table X with the best model was voting technique with highest accuracy of 86.9% which is identical to the experimental results of [30]. In addition, this study also explores the study of rules-based ensemble model for higher learning scholarship award. This study found that the accuracy of rules-based ensemble are slightly different than the best ensemble model (85.94%) but acceptable. This study also explores the rules obtained from the rules-based model J48 and Apriori and managed to select the best rules to enhance to develop an ensemble rules-based models which is improved the study for classification model for scholarship award in [8][6][9].

This means that when we ensemble five base learners J48, SVM, ANN, NB and RT, the ensemble model performs significantly better than ensemble it alone using boosting and bagging techniques.

The experiments also suggest the experiment using different dataset with different scholarship program ensemble techniques can be adopted to build better learning and prediction model to predict scholarship award for higher learning students.

TABLE X. OVERALL RESULT SHOWS THAT EACH OF ENSEMBLE LEARNING TECHNIQUES USED

Ensemble Techniques	Boosting					Bagging					Voting		
	J48	SVM	ANN	NB	RT	J48	SVM	ANN	NB	RT	min	avg	max
Accuracy (%)	81.97	86.299	83.7325	85.11034	81.862	83.9	86.45	86.5	85.4	83.9	86.9	86.9	86.9
Precision	0.7848	0.7712	0.79025	0.8094	0.78	0.869	0.868	0.81	0.863	0.782	0.793	0.793	0.793
Recall	0.8216	0.8608	0.83725	0.8514	0.8188	0.869	0.869	0.81	0.708	0.839	0.869	0.869	0.869
F-measure	0.8004	0.805	0.8075	0.8218	0.797	0.869	0.808	0.81	0.838	0.805	0.808	0.808	0.808

#### ACKNOWLEDGMENT

We would like to thank Ministry of Higher Education for their support and encouragement. This project is supported by Research University Grand Challenge Grant CP-2017-15/1.

#### REFERENCES

- [1] Dutt A, Ismail MA, Herawan T. A Systematic Review on Educational Data Mining. *IEEE Access*. 2017;5:15991–6005.
- [2] Adejo OW, Connolly T. Predicting student academic performance using multi-model heterogeneous ensemble approach. *J Appl Res High Educ*. 2018;10:61–75.
- [3] Mesarić J, Šebalj D. Decision trees for predicting the academic success of students. *Croat Oper Res Rev* [Internet]. 2016;7:367–88. Available from: <http://hrcak.srce.hr/174215?lang=en>.
- [4] Shahiri AM, Husain W, Rashid NA. A Review on Predicting Student's Performance Using Data Mining Techniques. *Procedia Comput Sci* [Internet]. Elsevier Masson SAS; 2015;72:414–22. Available from: <http://dx.doi.org/10.1016/j.procs.2015.12.157>.
- [5] Yu CH, Digangi S, Jannasch-Pennell A, Kaprolet C. A Data Mining Approach for Identifying Predictors of Student Retention from Sophomore to Junior Year. *J Data Sci*. 2010;8:307–25.
- [6] Raharja YP. Rancang Bangun Sistem Rekomendasi Beasiswa Menggunakan Algoritma Klasifikasi C4.5 pada Universitas Dian Nuswantoro. *Undinus* [Internet]. 2014;1–4. Available from: <http://eprints.dinus.ac.id/13408/>.
- [7] Tun KT, Aye AM. Selection of Appropriate Candidates for Scholarship Application Form using KNN Algorithm. *Int J Sci Eng Technol Res*. 2014;3:1019–26.
- [8] Alhassan JK, Lawal SA. Using Data Mining Technique for Scholarship Disbursement. 2015;9:1511–4.
- [9] Azuraliza, Arshad A. Rough Set and Decision Tree Model for Determining Scholarship Award Qualification. 2013;12:65–70.
- [10] Oza NC, Tumer K. Classifier ensembles: Select real-world applications. *Inf Fusion*. 2008;9:4–20.
- [11] Dietterich TG. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees. *Mach Learn* [Internet]. 2000;40:139–57. Available from: <http://en.scientificcommons.org/42637098%5Cnuuid/7906280C-AEF8-405A-9A94-6BAA1DDAED1E>
- [12] Rokach L. Ensemble-based classifiers. *Artif Intell Rev*. 2010;33:1–39.
- [13] Stapel M, Zheng Z, Pinkwart N. An Ensemble Method to Predict Student Performance in an Online Math Learning Environment. *Proc 9th Int Conf Educ Data Min*. 2016;231–8.
- [14] Schapire RE. The boosting approach to machine learning: an overview. *Nonlinear Estim Classif* [Internet]. 2003;171:149–71. Available from: <http://www.ams.org/mathscinet/search/publications.html?pg1=MR&s1=MR2005788%5Cnpapers2://publication/uuid/13F19159-186B-4FAA-9433-AE2BE3D153D8>.
- [15] Deng H. A Brief Introduction to Adaboost. 2007;1–35.
- [16] Palaniappan S, Rajinikanth T V, Govardhan A. Emerging Trends in Electrical, Communications and Information Technologies. 2017;394:165–74. Available from: <http://link.springer.com/10.1007/978-981-10-1540-3>.
- [17] Goyal A, Thakur S, Chowdhury R. Using Ensemble Learning and Association Rules to Help Car Buyers Make Informed Choices. *Proc Int Conf Big Data Adv Wirel Technol - BDAW '16* [Internet]. 2016;1–5. Available from: <http://dl.acm.org/citation.cfm?doid=3010089.3010093>
- [18] Ahmed M, Rasool AG, Afzal H, Siddiqi I. Improving handwriting based gender classification using ensemble classifiers. *Expert Syst Appl* [Internet]. Elsevier Ltd; 2017;85:158–68. Available from: <http://dx.doi.org/10.1016/j.eswa.2017.05.033>.
- [19] Hamsagayathri P, Sampath P. Decision Tree Classifiers for Classification of Breast Cancer. *Int J Curr Pharm Res* [Internet]. 2017;9:31. Available from: <http://innovareacademics.in/journals/index.php/ijcpr/article/view/17377>.
- [20] Pham BT, Tien Bui D, Prakash I. Landslide Susceptibility Assessment Using Bagging Ensemble Based Alternating Decision Trees, Logistic Regression and J48 Decision Trees Methods: A Comparative Study. *Geotech Geol Eng*. Springer International Publishing; 2017;35:2597–611.
- [21] Maniar H. A Predictive Student Performance Analytics Scheme using Auto-Adjust Apriori Algorithm. 2017;157:4–7.
- [22] Polikar R. Ensemble based systems in decision making. *Circuits Syst Mag IEEE*. 2006;6:21–45.
- [23] Friedman JH, Popescu BE. Predictive learning via rule ensembles. *Ann Appl Stat*. 2008;2:916–54.
- [24] Trandafilii E, Alkocı A, Kajo E, Xhuvani A. Discovery and evaluation of student's profiles with machine learning. *Proc Fifth Balk Conf Informatics - BCI '12* [Internet]. 2012;174. Available from: <http://dl.acm.org/citation.cfm?doid=2371316.2371350>.
- [25] Wook M, Yahaya YH, Wahab N, Isa MRM, Awang NF, Seong HY. Predicting NDUM Student's Academic Performance Using Data Mining Techniques. 2009 Second Int Conf Comput Electr Eng [Internet]. 2009;357–61. Available from: <http://ieeexplore.ieee.org/document/5380417/>.
- [26] Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK. Validation of QSAR Models - Strategies and Importance. *Int J Drug Des Discov*. 2011;2:511–9.
- [27] Mittal K, Aggarwal G, Mahajan P. a Comparative Study of Association Rule Mining Techniques and Predictive Mining Approaches for Association Classification. *Int J Adv Res Comput Sci* [Internet]. 2017;8:365–72. Available from: <http://ijarcs.info/index.php/Ijarcs/article/viewFile/4984/4331>.
- [28] Fatima M, Pasha M. Survey of Machine Learning Algorithms for Disease Diagnostic. *J Intell Learn Syst Appl* [Internet]. 2017;9:1–16. Available from: <http://www.scirp.org/journal/doi.aspx?DOI=10.4236/jilsa.2017.91001>.
- [29] Azuraliza, Idris N, Hamdan AR, Othman Z, Ahmad Nazari MZ, Zainudin S. Classification models for outbreak detection in oil and gas pollution area. *Proc 2011 Int Conf Electr Eng Informatics, ICEEI 2011*. 2011;0–5.
- [30] Manaf SA, Mustapha N, Sulaiman N, Husin NA, Shah Zainuddin MN, Mohd Shafri HZ. Majority voting of ensemble classifiers to improve shoreline extraction of medium resolution satellite images. *J Theor Appl Inf Technol*. 2017;95:4394–405.
- [31] Hamood Ali Alshalabi, Sabrina Tiun, Nazlia Omar. A Comparative Study Of The Ensemble And BaseClassifiers Performance In Malay Text Categorization. *Asia-Pacific Journal of Information Technology and Multimedia*. Vol. 6 No. 2, December 2017; 53 - 64.
- [32] Azhar Mohd Khairy, Afzan Adam, Mohd Ridzwan Yaakub. Data Analytics In Malaysian Education System: Revealing The Success Of Sijil Pelajaran Malaysia From Ujian Aptitud Sekolah Rendah. *Asia-Pacific Journal of Information Technology and Multimedia*. Vol. 7 No. 2, December 2018; 29 - 45.

# Where is the Highest Rate of Children with Anemia in Peru? An Answer using Grey Systems

Alexi Delgado<sup>1</sup>

Department of Engineering, Mining Engineering Section  
Pontificia Universidad Católica del Perú  
Lima, Peru

Nicolly Perez Ccancce<sup>2</sup>

Systems Engineering Program  
Universidad de Ciencias y Humanidades  
Lima, Peru

**Abstract**—Anemia, or chronic malnutrition, in children under 5 is a social problem that affects the infant crucially in his or her growth and cognitive development. However, this problem presents itself in different ways in each department of Peru. Thus, in this work, the methodology of grey clustering, which is based on grey systems theory, was applied. The case study was conducted in the 25 departments of Peru, to analyze the departments most affected by chronic malnutrition in children under 5 years of age. The results of the study were that the departments of Cajamarca, Huancavelica, Loreto and Pasco have the highest rate of children with anemia; this may be due to the fact that these departments have greater poverty or negligence with respect to proper food handling. The results of this study could help local authorities such as the Ministry of Health to combat malnutrition, and also serve as a basis for future studies to evaluate the social impact of other conditions on health from a mathematical perspective.

**Keywords**—Anemia; CTWF; grey systems; grey clustering

## I. INTRODUCTION

One of the public health problems that most afflicts the global population is anemia; which is not only the most frequent and most widespread, but also affects particularly the most vulnerable populations such as infants and women of reproductive age as stated by [1] according to [2]. Anemia according to [3] is defined as a decrease in hemoglobin concentration. It is a complex condition that occurs when there is a shortage of red blood cells in the body to adequately meet physiological needs [4]. In this regard, WHO stated that 528.7 million women and 273.2 million children under 5 years of age were anemic by 2011 worldwide [5]. This is alarming given that, in the case of preschool children, it affects their growth and destabilizes their nutrients [6]. However, specifically in the case of Peru, according to a technical report published by the Ministry of Health, [7] it is estimated that 6 out of 10 children between the ages of 6 and 12 months present anemia, a worrisome reality that requires greater monitoring.

For the present research, Grey clustering methodology, which is based on grey systems theory, was used. Initially developed by Deng in 1985 [8], it is a mathematical method

clustering analysis [9], which is based on the application of Center-Point Triangular Whitenization Weight Functions (CTWF) [10], with the objective of classifying observation groups into particular categories [11]. This methodology was selected since it has been applied to different challenges, which show a high level of uncertainty or limited information [10], such as social impact assessment [12], crime management [10], security management [13], level of learning [8], among others. Therefore, it has a high degree of adaptability in the case study of this work.

Likewise, in order to achieve the most precise monitoring possible, data provided by the National Institute of Statistics and Informatics (INEI by its Spanish acronym) [14] will be used to facilitate the classification of a central point in the class interval [15].

The purpose of this study is to identify the departments that have the highest percentage of children under 5 years of age who have anemia in Peru, in order to identify the populations that have the greatest need for prevention programs by the Ministry of Health.

This paper is organized as follows, Section II is an explanation of the Grey Clustering method. Section III describes the application in the case study. This is followed by the results and discussion in Section IV. Finally, the conclusions are presented in Section V.

## II. METHODOLOGY

The methodology to be used for the development of this research will be Grey Clustering; this is developed as follows: first, assume that there is a set of  $m$  objects,  $n$  criteria,  $s$  classes of grey and sample value  $x_{ij}$  ( $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ ). Then, the steps of the Grey Clustering method can be described as below [16]:

Step 1: The intervals of the criteria are divided into  $s$  grey classes, then the central points  $\lambda_1, \lambda_2, \dots$  and  $\lambda_s$  are determined.

Step 2: Grey classes are distributed in two directions, adding the grey classes 0 and  $(s + 1)$  with their central points  $\lambda_0$  and  $\lambda_{s+1}$ ; as shown in Fig. 1.

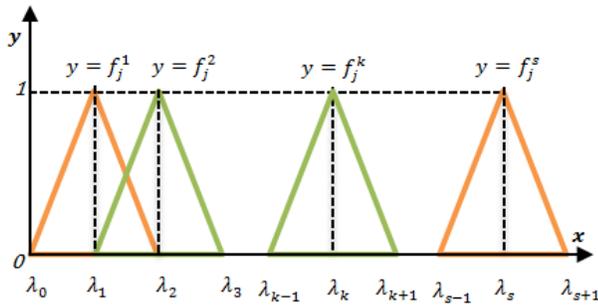


Fig. 1. Center-Point Triangular Whitenization Weight [17].

Step 3: The clustering coefficient, that indicates the weight of the criteria, for the group  $i, i = 1, 2, \dots, m$ , in the grey class  $k, k = 1, 2, \dots, s$ , is determined using (1).

$$f_j^k(x_{ij}) = \begin{cases} 0, & x \notin [\lambda_{k-1}, \lambda_{k+1}] \\ \frac{x - \lambda_{k-1}}{\lambda_k - \lambda_{k-1}}, & x \in [\lambda_{k-1}, \lambda_k] \\ \frac{\lambda_{k+1} - x}{\lambda_{k+1} - \lambda_k}, & x \in [\lambda_k, \lambda_{k+1}] \end{cases} \quad (1)$$

Step 4: If  $\max_{1 \leq k \leq s} (f_j^k) = f_j^{k^*}$  it is decided that object  $i$  belongs to grey class  $k^*$ , the objects can be ordered according to the magnitudes of their coefficients [10].

### III. CASE STUDY

For the application of the analysis, the data of children under 5 years old with chronic malnutrition rate (also known as anemia) was observed in the metrics of the twenty-five Peruvian departments. Fig. 2 shows the values given to each department for the present study.



Fig. 2. Values given to the Peruvian Departments in the Current Research.

### A. Context Overview

Currently, anemia in children under five years of age can be identified as a major risk since the growth and proper development of the infant is affected [18]. One approach to solving this problem could be to analyze the prevalence of children under five years of age with malnutrition in order to assess the management of this issue in each department. To conduct this research, data for 2017 was obtained from the website of the National Institute of Statistics and Information (INEI) [14]. The data obtained is presented in Table I.

TABLE I. CHILDREN UNDER 5 WITH ANEMIA BY DEPARTMENT

Department	Children affected	Department	Children affected
0	17.1	13	10.5
1	16.1	14	5.1
2	20.9	15	9.6
3	4.9	16	23.8
4	20.0	17	7.3
5	26.6	18	3.4
6	5.2	19	22.8
7	13.4	20	15.9
8	31.2	21	16.1
9	19.6	22	12.1
10	8.3	23	3.2
11	17.3	24	8.2
12	15.6	25	19.4

### B. Assessment Criteria

Calculations based on Grey Clustering method are preceded as follows:

Step 1: The center points  $\lambda_1, \lambda_2, \lambda_3, \lambda_4,$  and  $\lambda_5$ , of five grey classes were determined as shown in Table II.

Step 2: To the four grey classes determined above, classes  $s_0$  and  $s_5$  were added; as shown in Table III and Fig. 3.

Step 3: The values presented in Table III were replaced in (1), to obtain the CTWF for the four grey classes. The results are shown in (2)-(5):

$$f_j^1(x) = \begin{cases} 0, & x \notin [0, 13.7] \\ \frac{x-0}{6.7}, & x \in [0, 6.7] \\ \frac{13.7-x}{7}, & x \in [6.7, 13.7] \end{cases} \quad (2)$$

$$f_j^2(x) = \begin{cases} 0, & x \notin [6.7, 20.7] \\ \frac{x-6.7}{7}, & x \in [6.7, 13.7] \\ \frac{20.7-x}{7}, & x \in [13.7, 20.7] \end{cases} \quad (3)$$

$$f_j^3(x) = \begin{cases} 0, & x \notin [13.7, 27.7] \\ \frac{x-13.7}{7}, & x \in [13.7, 20.7] \\ \frac{27.7-x}{7}, & x \in [20.7, 27.7] \end{cases} \quad (4)$$

$$f_j^4(x) = \begin{cases} 0, & x \notin [20.7, 34.4] \\ \frac{x-20.7}{7}, & x \in [20.7, 27.7] \\ \frac{34.4-x}{6.7}, & x \in [27.7, 34.4] \end{cases} \quad (5)$$

Following this, the values in Table I were replaced in (2)-(5). The results are shown in Table IV.

Step 4: Finally, the requirement  $\max_{1 \leq k \leq 5} (f_i^k) = f_i^{k^*}$  was applied; thus, it was decided that  $i$  belongs to the grey class  $k^*$ , for each department. The results are shown in Table V.

TABLE II. CASE STUDY CLASSES

Very low class ( $\lambda_1$ )	Low class ( $\lambda_2$ )	Medium class ( $\lambda_3$ )	High class ( $\lambda_4$ )
[3.2 – 10.2)	[10.2–17.2)	[17.2–24.2)	[24.2–31.2)
6.7	13.7	20.7	27.7

TABLE III. CENTRAL POINTS OF THE EXTENDED GREY CLASSES

Central points of the extended grey classes					
$\lambda_0$	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$
0	6.7	13.7	20.7	27.7	34.4

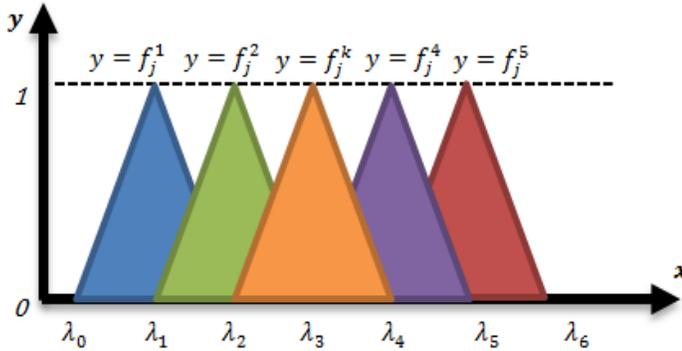


Fig. 3. CTWF for the Case Study.

TABLE IV. CTWF VALUE FOR EACH PERUVIAN DEPARTMENT

Departments	0	1	2	3
$f_j^1(x)$	0.00	0.00	0.0	0.73
$f_j^2(x)$	0.51	0.65	0.00	0.00
$f_j^3(x)$	0.48	0.34	0.97	0.00
$f_j^4(x)$	0.0	0.00	0.02	0.00
Departments	4	5	6	7
$f_j^1(x)$	0.00	0.00	0.77	0.04
$f_j^2(x)$	0.1	0.00	0.00	0.95
$f_j^3(x)$	0.9	0.15	0.00	0.00
$f_j^4(x)$	0.00	0.84	0.00	0.00
Departments	8	9	10	11
$f_j^1(x)$	0.00	0.00	0.77	0.00
$f_j^2(x)$	0.00	0.15	0.22	0.78
$f_j^3(x)$	0.00	0.84	0.00	0.51
$f_j^4(x)$	0.47	0.00	0.00	0.00
Departments	12	13	14	15
$f_j^1(x)$	0.00	0.46	0.76	0.59
$f_j^2(x)$	0.73	0.54	0.00	0.41
$f_j^3(x)$	0.27	0.00	0.00	0.00
$f_j^4(x)$	0.00	0.00	0.00	0.00
Departments	16	17	18	19
$f_j^1(x)$	0.00	0.91	0.50	0.00
$f_j^2(x)$	0.56	0.09	0.00	0.00
$f_j^3(x)$	0.44	0.00	0.00	0.70
$f_j^4(x)$	0.00	0.00	0.00	0.30
Departments	20	21	22	23
$f_j^1(x)$	0.00	0.00	0.22	0.47
$f_j^2(x)$	0.69	0.66	0.77	0.00
$f_j^3(x)$	0.31	0.34	0.00	0.00
$f_j^4(x)$	0.00	0.00	0.00	0.00
Departments	24	25		
$f_j^1(x)$	0.79	0.00		
$f_j^2(x)$	0.21	0.19		
$f_j^3(x)$	0.00	0.81		
$f_j^4(x)$	0.00	0.00		

TABLE V. FINAL VALUES FOR EACH DEPARTMENT

Department 0	Department 1	Department 2	Department 3	Department 4	Department 5	Department 6
0.51 Medium class	0.65 Medium class	0.97 Medium class	0.73 Medium class	0.9 Very low class	0.84 Medium class	0.77 High class
Department 7	Department 8	Department 9	Department 10	Department 11	Department 12	Department 13
0.95 Low class	0.47 High class	0.84 Medium class	0.77 Low class	0.78 Medium class	0.73 Medium class	0.54 Low class
Department 14	Department 15	Department 16	Department 17	Department 18	Department 19	Department 20
0.76 Very low class	0.59 Low class	0.56 High class	0.91 Very low class	0.50 Very low class	0.70 High class	0.6 Medium class
Department 21	Department 22	Department 23	Department 24	Department 25		
0.66 Medium class	0.77 Low class	0.47 Very low class	0.79 Low class	0.81 Medium class		

#### IV. RESULTS AND DISCUSSION

##### A. Results on the Case Study

Fig. 4 was constructed from Table V. The result is shown below according to the levels of anemia obtained according to the Grey clustering methodology.

As shown in Fig. 4, the province of Huancavelica shows a high rate of children suffering from anemia at national level, followed by Cajamarca, Loreto and Pasco. This may be due to the fact that, as mentioned by the Peruvian Ministry of Health (MINSA in Spanish), the main factors that contribute to the existence of malnutrition in children are: a bad nutritional combination, poor hygiene in food preparation and avoiding healthy eating. Factors that are present in the department of Huancavelica; therefore, this entity seeks to implement collective activities to raise awareness among the population [7].

Another possible cause of anemia (or chronic malnutrition) to be considered is the lack of education and economic deprivation of mothers, which affects the nutrition of their children, since they acquire their knowledge and life skills through informal education [19].

On the other hand, provinces such as Arequipa, Madre de Dios, Moquegua, Lima and Tacna have the lowest levels of anemia in children. This may be due to the prevention of anemia in these provinces, as is the case in the department of Arequipa, where a growth and development monitoring program (CRED) organized by the Peruvian Ministry of Health has been implemented [4].

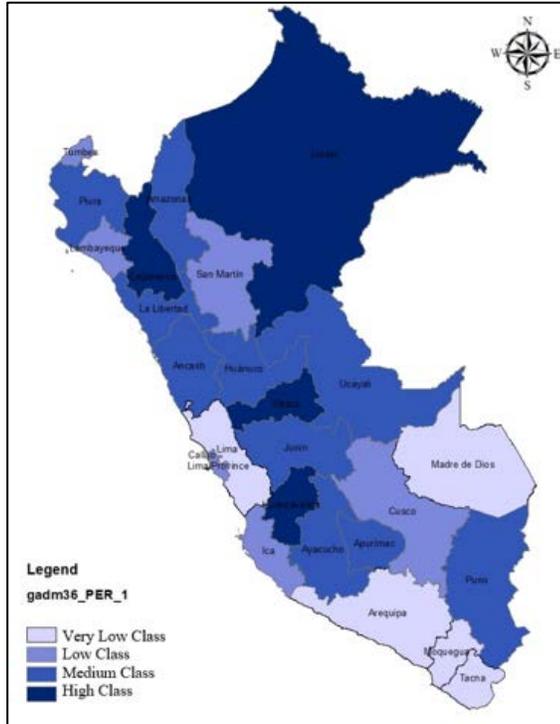


Fig. 4. Anemia Level in Children by Department.

##### B. Discussion on the Methodology

The Grey clustering method has helped to classify the data within levels optimally, based on previously collected data. The methodology has also proven to have advantages over other methods, such as Delphi [20], or the analytical hierarchy process (AHP) [21]; since the methods outlined do not consider uncertainty within the analysis [13]. Which is an important benefit, especially when dealing with social aspects.

#### V. CONCLUSIONS

The results obtained for the case study on the prevalence of anemia in children under 5 years of age are of interest, since it has been possible to identify the departments with the highest rate of the condition (Huancavelica, Loreto, Pasco and Cajamarca). This will be useful for decision-making by departmental authorities in Peru, such as MINSA and/or the central government, since it will allow them to concentrate their efforts on the inclusion of healthy food and greater hygiene programs in order to combat the current situation.

In addition, it has been demonstrated that the Grey Clustering method is an approach with a great capacity that allows us to apply it to social problems that present a high level of insecurity, given that this methodology allows us to analyze uncertainty.

Finally, in future research, this method could be applied to problems of social topics, such as the mathematical analysis of various diseases as diabetes, obesity, etc. as the social problem has been performed in the present research.

#### REFERENCES

- [1] J. L. Miller, "Iron deficiency anemia: A common and curable disease," *Cold Spring Harb. Perspect. Med.*, vol. 3, no. 7, Jul. 2013. DOI: 10.1101/cshperspect.a011866.
- [2] N. Zavaleta and L. Astete-Robilliard, "Effect of anemia on child development: Long-term consequences," *Rev. Peru. Med. Exp. Salud Publica*, vol. 34, no. 4, pp. 716–722, Oct. 2017. DOI: dx.doi.org/10.17843/rpmpesp.2017.344.3251.
- [3] Hospital Provincial de Castro, "Clinical Practice Guidelines: Prevention, Diagnosis and Treatment of Iron Deficiency Anemia in Children and Adults - Guía de Práctica Clínica: Prevención, Diagnóstico y Tratamiento de la Anemia por Deficiencia de Hierro en Niños y Adultos," *Cons. Salubr. Gen.*, pp. 1–57, 2009.
- [4] C. Sotomayor-Beltran and G. Z. Segura, "A spatial assessment of anemia among Peruvian children aged 6 months to 5 years between 2016 and 2017," *Congr. Argentino Ciencias la Inform. y Desarro. Investig. CACIDI 2018*, pp. 1–5, 2018. DOI: <https://doi.org/10.1109/CACIDI.2018.8584359>.
- [5] OPS and OMS, "Iron deficiency anemia: Research for efficient and viable solutions - Anemia ferropénica: Investigación para soluciones eficientes y viables."
- [6] J. A. M. M. Torres Lagrava, "Nutritional Chispitas and its administration to children from months to less than 2 years old Municipality of Icla Chuquisaca 2015 - Chispitas Nutricionales y su administración a niños de meses a menores de 2 años Municipio de Icla Chuquisaca 2015."
- [7] Ministerio de Salud del Perú, "National Plan for the Reduction and Control of Maternal and Child Anaemia and Chronic Malnutrition in Peru: 2017 - 2021 - Plan Nacional para la REDUCCIÓN Y CONTROL DE LA ANEMIA Materno Infantil y la Desnutrición Crónica Infantil en el Perú: 2017 - 2021," Lima, 2017.
- [8] A. Delgado, "Why do any secondary students prefer the mathematics? A response using grey systems," *Proc. 2017 Int. Symp. Eng. Accreditation, ICACIT 2017*, pp. 1–4, 2018. DOI: <https://doi.org/10.1109/ICACIT.2017.8358082>.

- [9] Y. Li, Y. Niu, W. Wang, and B. Li, "Grey-incidence clustering decision-making method with three-parameter interval grey number based on regret theory," in 2017 IEEE International Conference on Grey Systems and Intelligent Services, GSIS 2017, 2017, pp. 211–218. DOI: <https://doi.org/10.1109/GSIS.2017.8077706>.
- [10] A. Delgado, P. Montellanos, and J. Llave, "Air quality level assessment in Lima city using the grey clustering method," in IEEE ICA-ACCA 2018 - IEEE International Conference on Automation/23rd Congress of the Chilean Association of Automatic Control: Towards an Industry 4.0 – Proceedings, 2019, 8609699.
- [11] S. Liu and Y. Yang, "Explanation of terms of grey clustering evaluation models," *Grey Syst. Theory Appl.*, vol. 7, no. 1, pp. 129–135, Feb. 2017. DOI: <https://doi.org/10.1108/GS-11-2016-0046>.
- [12] A. Delgado and I. Romero, "Applying the Grey Systems Theory to Assess Social Impact from an Energy Project," *Proc. 2018 IEEE 25th Int. Conf. Electron. Electr. Eng. Comput. INTERCON 2018*, no. 2, pp. 1–4, 2018. DOI: <https://doi.org/10.1109/INTERCON.2018.8526372>.
- [13] C. Li, K. Chen, and X. Xiang, "An integrated framework for effective safety management evaluation: Application of an improved grey clustering measurement," *Expert Syst. Appl.*, vol. 42, no. 13, pp. 5541–5553, 2015. DOI: <https://doi.org/10.1016/j.eswa.2015.02.053>.
- [14] INEI, "PERU Instituto Nacional de Estadística e Informática INEI." [Online]. Available: <https://www.inei.gob.pe/estadisticas/indice-tematico/sociales/>. [Accessed: 22-Oct-2019].
- [15] S. Liu and Y. Lin, "Interval Analysis and Grey Systems Theory," in Springer Science & Business Media, 2010.
- [16] M. Sacasqui, J. Luyo, and A. Delgado, "A Unified Index for Power Quality Assessment in Distributed Generation Systems Using Grey Clustering and Entropy Weight," 2018 IEEE ANDESCON, ANDESCON 2018 - Conference Proceedings, 2018, 8564631.
- [17] S. Liu and Y. Lin, *Grey Systems Theory and Applications*, vol. 68. Springer Berlin Heidelberg, 2011.
- [18] Ministerio de Salud del Peru (MINSA), "Technical health standard for monitoring the growth and development of children under five - Norma tecnica de salud para el control del crecimiento y desarrollo de la niña y el niño menor de cinco años," *Bibl. Cent. del Minist. Salud*, p. 149, 2011.
- [19] H. B. Urke, T. Bull, and M. B. Mittelmark, "Socioeconomic status and chronic child malnutrition: Wealth and maternal education matter more in the Peruvian Andes than nationally," *Nutr. Res.*, vol. 31, no. 10, pp. 741–747, 2011. <https://doi.org/10.1016/j.nutres.2011.09.007>.
- [20] M. R. Álvarez M. Torrado-fonseca, "Delphi method - El método Delphi," *REIRE. Rev. d'Innovación i Recer. en Educ.*, vol. 9, no. 9 (1), pp. 0–2, 2016. <https://doi.org/10.1344/reire2016.9.1916/>.
- [21] A. Priyadi, M. H. P, and M. Pujiantara, "Wind farm site selection base on fuzzy analytic hierarchy process method; Case study area Nganjuk," no. Mcdm, pp. 545–550, 2016. <https://doi.org/10.1109/ISITIA.2016.7828718>.

# Mapping UML Sequence Diagram into the Web Ontology Language OWL

Mo'men Elsayed<sup>1</sup>, Nermeen Elkashef<sup>2</sup>  
Department of Mathematics and Computer Science  
Faculty of Science, Alexandria University  
Alexandria, Egypt

Yasser F.Hassan<sup>3</sup>  
Professor of Computer Science  
Faculty of Computer Science and Artificial  
Intelligence, Pharos University, Alexandria, Egypt

**Abstract**—In this paper, we propose a new mapping technique from the OMG's UML modeling language into the Web Ontology Language (OWL) to serve the Semantic Web. UML (Unified Modeling Language) is widely accepted and used as a standardized modeling language in Object-Oriented Analysis (OOA) and Design (OOD) approach by domain experts to model real-world objects in software development. On the other hand, the conceptualization, which is represented in OWL, is designed to process the content of information rather than just present the information. Therefore, the matter of migrating UML to OWL is becoming an energetic research domain. OWL (Web Ontology Language) is a Semantic Web language designed for defining ontologies on the Web. An ontology is a formal specification naming and definition of shared data. This technique describes how to map UML Models into OWL and allows us to keep semantic of UML sequence diagrams such as messages, the sequence of messages, guard invariant, etc. to make data of UML sequence diagrams machine-readable.

**Keywords**—Mapping; Unified Modeling Language; UML; sequence diagram; ontology; Web Ontology Language; OWL

## I. INTRODUCTION

Nowadays we are observing a growing effort for supporting semantics of data that is stored on the web that makes the web more intelligent that procreate a promising technology is called "Semantic Web". The Semantic Web is developed by W3C for providing the knowledge data interchange over the web that available standard formats, reachable, manageable, and understand which will be used by machines for overcoming the original web limitation of only interchanging data through documents [1]. The semantic web is also called "Web of data". The web will be able to process and explicate information to be better to meet the human requirement and able to provide full and immediate answers for natural language queries. Ontology and also the various languages designed for sharing data: Extensible Markup Language (XML), Resource Description Framework (RDF) and Web Ontology Language (OWL) are the stilts of this technology [2].

The ontology lies at the core of semantic data integration. The ontology is the knowledge data domain that will be shared and explored. The Semantic Web does not break away from the web but rather an extension of the current one, where data is given distinct meaning, better-enabling computers and people to work in cooperation [3]. Therefore, the issue of

converting UML to OWL is becoming an active research point.

UML, the Unified Modeling Language, is the most utilized language to the requirements specification [4]. UML is a standardized modeling graphical language that includes an integrated set of diagrams. Every diagram depicts the modeling system in different portions, but together they can provide a full map of the modeling system. We focus on one diagram that is a UML sequence diagram. A UML sequence diagram is a type of interaction diagram where it illustrates how and in what order a set of objects works together. A sequence diagram is employed for dealing with the dynamic view of a system, whereas OWL is developed to form a semantic web to represent the explicit specification of a conceptualization, not just a document web. A picture is worth a thousand words, this idiom definitely fits describing UML. UML is a standard notation language that can be used for specifying, visualizing, constructing, and documenting the phases of software systems.

Ontology means an ontology may be a characterization (like a suitable specification of a program) of concepts and relationships which is able to exist among them through a community. In other words – an ontology illustrates a part of the globe.

This paper is organized as follows: Section 2 offers the background of our work. In Section 3 discusses our technique for mapping UML sequence diagrams into OWL 2 DL in detail. Section 4 presents an overview of our technique with a running case study. Section 5 concludes and points out the fields of future work.

## II. BACKGROUND

### A. Unified Modeling Language

UML is a standardized blueprint representation to design and analyze a model of a system. These blueprints provide more than 10 diagrams in which every diagram supports an aspect to characterize every part of a system.

UML is developed to make available communication among the software developers by specifying, visualizing, constructing, and documenting the aspects of software systems. UML includes things, relationships, and diagrams, as shown in Table I. One exceptional diagram is a sequence diagram which is categorized as a behavioral diagram. Behavior diagrams depict a dynamic aspect of the objects in a

system, a chain of changes to the system through time. A sequence diagram is one of the interaction diagrams that illustrate the interactions between object instances and the ordering of messages according to time. A sequence diagram illustrates the interactions among objects, and emphasizes a sequential order of messages.

**B. Ontology Foundations**

An ontology is a specification of a conceptualization [3]. The term is borrowed from philosophy, where an ontology can be a methodical account of existence. The meaning depends on our understanding what "exists" is that which can be represented wherever the terms "specification" and "conceptualization" is a description (like a formal specification of a program) of concepts and describable relationships between them that can reflect in the representational vocabulary with which an abstraction of a program. Typically by using UML, we can represent the abstract of a program. For example, A UML object o is the object of class C is drawn by using UML, as given in Fig. 1.

In ontologies, the concepts of the program are represented in a set of axioms that depict the specification of a conceptualization [4], and the relationship among concepts as properties. Therefore, the UML Class C is mapped to OWL 2 as Declaration ( Class( C ) )

In the OWL the UML object o of class C is called an individual and is expressed as a class assertion:

ClassAssertion( C o )

TABLE I. UML THINGS, RELATIONSHIPS AND DIAGRAMS

UML Category	UML Elements	UML Elements
Things	Structural Things	Classes Interfaces Collaborations Use Cases Active Classes Components Nodes
		Behavioral Things
	Grouping Things	Packages
	Annotational	Things Notes
Relationships	Structural Relationships	Dependencies Aggregations Associations Generalizations
	Behavioral Relationships	Communicates Includes Extends Generalizes
Diagrams	Structural Diagrams	Class Diagrams Component Diagrams Deployment Diagrams Use Case Diagrams
	Behavioral Diagrams	Sequence Diagrams Communication Diagrams Statechart Diagrams Activity Diagrams

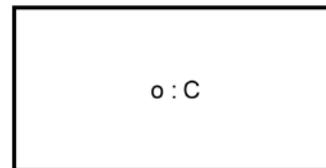


Fig. 1. A UML Object o of Class C.

**C. OWL and UML**

The Ontology Definition Metamodel (ODM) [5] is derived from Meta-Object Facility (MOF) [5,6] and based on the Object Management Group (OMG) [7] specifications that permit integrating ontology engineering into concepts of OMG modeling. The ODM follows an identical hierarchy just like the one mentioned in a four-layer OMG modeling hierarchy [8].

The ODM was primarily developed to support the ontology structure [5]. It contains classes, associations, and constraints.

MOF defines an abstract framework and language for constructing, managing, and specifying technology-neutral metamodels. It is the foundation for defining any modeling language like UML. Consequently, The UML is also derived from the MOF and fundamental form of ODM. Therefore, UML notations are also used for ontology modeling [6, chapter 7].

The ODM provides metamodels for several knowledge representation languages such as OWL and RDF [8]. In our technique, ODM metamodel OWL is used to represent the MOF / UML based models. In our technique, we use a decidable fragment of OWL 2 DL.

Whereas the ODM and UML are derived from MOF [5,6], therefore there exist common features, as well as there, are also different features. The common features are shown in Table II. The comparison between UML and ODM aforementioned is given in terms of UML and OWL 2 DL. Table III shows the features in UML which do not have equivalent OWL 2 elements.

TABLE II. UML ELEMENTS THAT HAVE THE DL EQUIVALENT OWL 2 DL ELEMENTS

UML Elements	OWL Elements
Class	Class
Instance	Individual
Enumeration	Oneof
Multiplicity	Min/max/exact cardinality
Datatype	Datatype

TABLE III. UML ELEMENTS THAT HAVE THE DL EQUIVALENT OWL 2 DL ELEMENTS

UML Elements	OWL Elements
Ordering	Not available
Messages	Not available
Operations	Not available
Guards	Not available
Fragment	Not available
Operands	Not available

#### D. Reasoners

Reasoning is a critically important capability for the Semantic Web application development. A reasoner plays a vital role in developing that automatically infers logical consequences from a collection of logical facts or axioms such as Pellet, FaCT++, HerMIT, etc.

Pellet [9] provides functionality to check consistency and infer subsumption of ontologies and Semantic Web Rule Language. Pellet is a complete and capable OWL-DL reasoner, which is written in Java and is open source. Based on these criteria, we have chosen Pellet [9] which satisfies our requirements to reasoner processes the ontology and generates a validation report.

### III. UML SEQUENCE DIAGRAMS INTO OWL 2 DL

In order to create better OWL 2 ontologies from a UML sequence model, we explain in this section our understanding of UML sequence diagram concepts and show how to migrate these concepts into OWL 2.

The UML sequence diagrams capture the dynamic behavior of a system. The sequence diagram provides the behavioral interface of object instances and the sequence of messages that they send to each other over time by using the vertical axis of the diagram to show time what messages are sent and what.

#### A. Classes and Objects

A concept that grouped multiple objects that have the same features and share the same behaviors is commonly known as a class in UML. The concept of class in UML is equivalent to the concept of class in OWL 2 because both concepts are similar. A UML Class C is mapped to OWL 2 as:

Declaration( Class( C ))

An object is an instance (or element) of a class. In UML, objects have the behaviors of their class. Each object in a sequence diagram belongs to a specific class in a class diagram. A UML object o of class C is mapped into ClassAssertion axiom in OWL 2 called an individual:

ClassAssertion( C o )

Furthermore, By default in a UML sequence diagram, every object is different from another. However, OWL 2 follows the open-world assumption, so we must mention that all individuals are different from each other. For example, objects o1, ..., on in a sequence diagram, we use the DifferentIndividuals axiom in OWL 2:

DifferentIndividuals(o1 ... on)

#### B. Sequence of Messages

Messages that depict the call of operations belong to a specific class are shown horizontally in Fig. 2. They are sent from a source object that is defined as a caller and received by a target object is defined as a receiver in a sequence diagram. A vertical position indicates the sequence of the messages of the sequence diagram, wherever the first message is always shown at the top in the diagram. Next, subsequent messages are added to the sequence diagram little down from the earlier message.

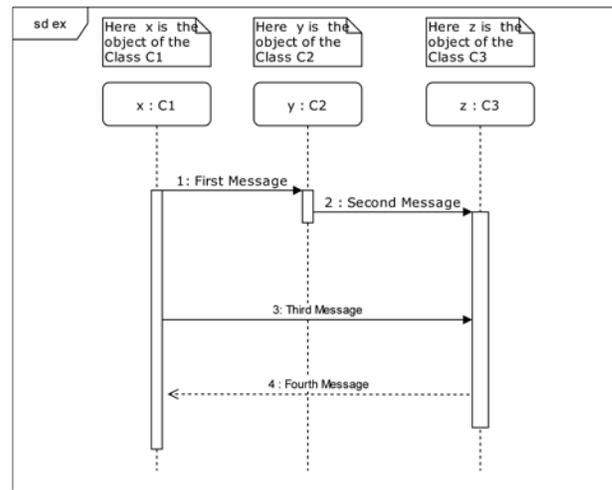


Fig. 2. A Simple UML Sequence Diagram.

A Sequence of messages is a number of messages that come one after another in a particular order as shown in Fig. 2. OWL 2 contains no axioms specifically for defining sequence or ordering [10]. However, OWL 2 has axioms that can be used to model sequence. Our technique describes a design pattern for modeling a sequence of messages using OWL 2 axioms, as summarized in the diagram in Fig. 3.

Before starting the conversion, we create a class OWL called “Fragment” to represent the sequence of messages in the UML sequence diagram. The class “Fragment” represents the kind of fragment in the UML sequence diagram. The object property “hasOperand” connects operands to the class “Fragment” which contains the operands of it. The class “Operand” represents the sequence of elements in the UML sequence diagram that can have multiple elements that is executed by the class “Element” and the object property “hasMessage”/“hasFragment”. The class “Operand” specifies a particular size refers to the number of messages in the operand. As shown in Fig. 3, the object property “hasNext” connects an individual of the class “Element” to exactly another one. Consequently, the object property “hasNext” supports the sequence of elements. Furthermore, “hasNext” is accompanied by its related transitive properties and inverse. Two object properties are defined, “firstElement” and “lastElement”, to determine which are the first and the last elements in the class “Operand”, as sub-properties of “element”. The elements connected with these two properties cannot be respectively preceded or followed by another element. Moreover, the class “Element” represents the type of elements in the operand this is executed by the class “Message” and “Fragment” respectively with the object property “hasMessage” and “hasFragment”. The class “Message” specifies a particular type that refers to the type of a message in the UML sequence diagram. In order to identify which are the caller and the receiver objects of a message, two object properties are defined caller and receiver. The object property “next message” connects an individual of the class “Message” to exactly another one. Two objects, properties are defined, “firstMessage” and “lastMessage”, to determine which are the first and the last messages in an operand and a fragment, as sub-properties of “hasMessage”. Consequently,

the object property “nextOperand” supports the sequence of operands. Two object properties are defined, “firstOperand” and “lastOperand”, in order to determine which are the first and the last operand in the class “Fragment”.

This class and the properties are related to this class are defined as follows:

Class: Fragment  
SubClassOf:  
    kind exactly 1,  
    firstOperand exactly 1 Operand,  
    lastOperand exactly 1 Operand  
DisjointWith: Operand, Element, Message  
ObjectProperty: hasOperand  
    Domain: Fragment  
    Range: Operand  
    SubPropertyChain: hasOperand o nextOperand  
    InverseOf: operandOf  
ObjectProperty: followedBy  
    Characteristics: Irreflexive  
ObjectProperty: precedeBy  
    Characteristics: Irreflexive  
    InverseOf: followedBy  
ObjectProperty: firstOperand  
    Characteristics: Functional  
    SubPropertyOf: hasOperand  
    Domain: Fragment  
    Range: precedeOperand exactly 0 Operand  
ObjectProperty: lastOperand  
Characteristics: Functional  
SubPropertyOf: hasOperand  
Domain: Fragment  
Range: nextOperand exactly 0 Operand  
Class: Operand  
SubClassOf:  
    size exactly 1 xsd:nonNegativeInteger,  
    guard max 1 xsd:string,  
    firstElement max 1 Element,  
    lastElement max 1 Element  
DisjointWith: Fragment, Element, Message  
DataProperty: guard  
    Domain: Operand  
    Range: xsd:string  
DataProperty: size  
    Characteristics: Functional  
    Domain: Operand  
    Range: xsd:nonNegativeInteger  
ObjectProperty: nextOperand  
    Characteristics: Functional  
    SubPropertyOf: followedBy  
    Domain: Operand  
    Range: Operand  
ObjectProperty: precedeOperand  
    Characteristics: Functional  
    SubPropertyOf: precedeBy  
    InverseOf: nextOperand  
ObjectProperty: hasElement  
    Domain: Operand  
    Range: Element  
    SubPropertyChain: hasElement o hasNext  
    InverseOf: elementOf  
ObjectProperty: firstElement  
    Characteristics: Functional

SubPropertyOf: hasElement  
Domain: Operand  
Range: hasPrecede exactly 0 Element  
InverseOf: firstElementOf  
ObjectProperty: lastElement  
    Characteristics: Functional  
    SubPropertyOf: hasElement  
    Domain: Operand  
    Range: hasNext exactly 0 Element  
    InverseOf: lastElementOf  
Class: Element  
SubClassOf: inverse hasElement some Operand  
DisjointWith: Fragment, Operand, Message  
ObjectProperty: hasNext  
    Characteristics: Functional  
    SubPropertyOf: followedBy  
    Domain: Element  
    Range: Element  
ObjectProperty: hasPrecede  
    Characteristics: Functional  
    SubPropertyOf: precedeBy  
    InverseOf: hasNext  
ObjectProperty: hasFragment  
    Domain: Element  
    Range: Fragment  
ObjectProperty: hasMessage  
    Domain: not Message  
    Range: Message  
    InverseOf: messageOf  
ObjectProperty: firstMessage  
    Characteristics: Functional  
    SubPropertyOf: hasMessage  
ObjectProperty: lastMessage  
    Characteristics: Functional  
    SubPropertyOf: hasMessage  
Class: Message  
SubClassOf:  
    index exactly 1 xsd:positiveInteger,  
    type exactly 1,  
    caller max 1 not Message  
    receiver max 1 not Message  
DisjointWith: Fragment, Operand, Element  
DataProperty: index  
    Characteristics: Functional  
    Domain: Message  
    Range: xsd:positiveInteger  
ObjectProperty: caller  
    Characteristics: Asymmetric  
    Domain: Message  
    Range: not Message  
ObjectProperty: receiver  
    Characteristics: Asymmetric  
    Domain: Message  
    Range: not Message  
DisjointWith: caller  
ObjectProperty: nextMessage  
    Characteristics: Functional  
    SubPropertyOf: followedBy  
    Domain: Message  
    Range: Message  
ObjectProperty: precedeMessage  
    Characteristics: Functional  
    SubPropertyOf: precedeBy  
    InverseOf: nextMessage

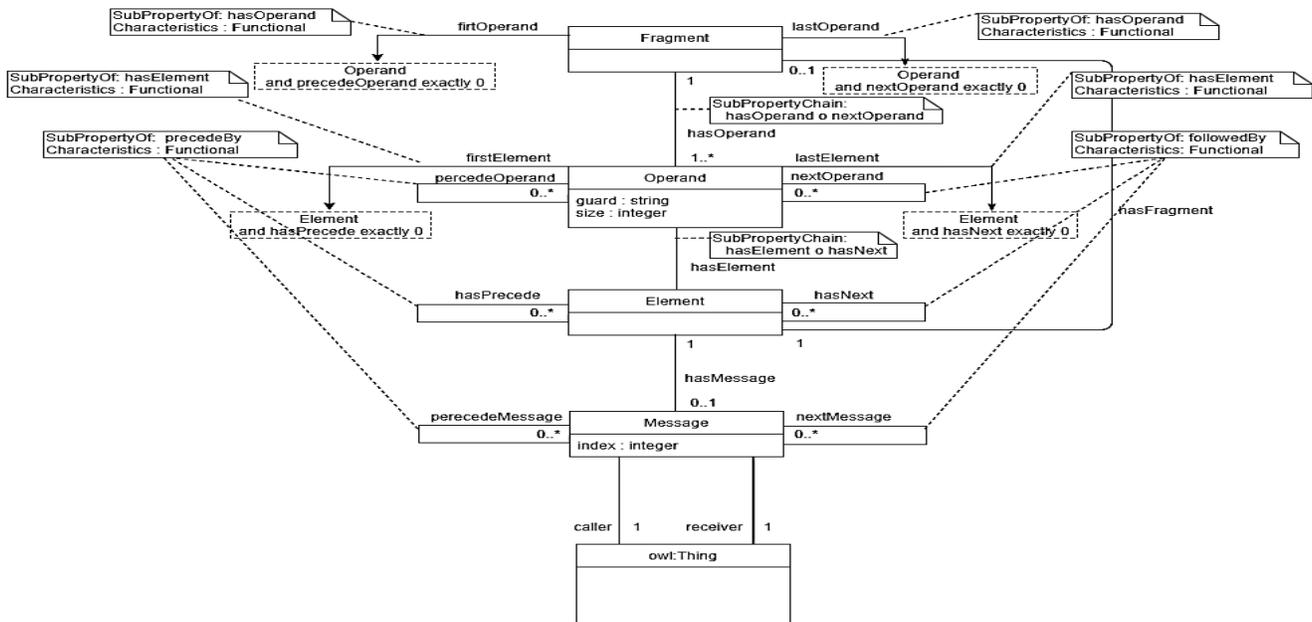


Fig. 3. Diagram Summarizing the Class “Fragment” and Its Related Classes and Properties.

The reflexive object properties “followedBy” and “precededBy” (respectively super-“hasNext”, “nextMessage” and “nextOperand”, and “hasPrecede”, “precedeMessage” and “precedeOperand”) refers to all the elements that follow/precede a particular element. In OWL 2, no cycles are permitted. Acyclicity means that an element cannot follow or precede itself, i.e., an element e1 either follows or precedes e2, e2 follows or precedes e3 and e3 follows or precedes e1 is disallowed. A sufficient and necessary condition for acyclicity of followedBy/precededBy is irreflexive closure of the property is transitive. We make followedBy/precededBy irreflexive and transitive at the same time. However, to set those two properties as transitive is not possible [8] since the logic system would no longer be decidable, and we would keep the ontology in a DL framework. We map transitivity in Semantic Web Rule Language (SWRL). The transitivity of the properties “followedBy” and “precededBy” is written in SWRL as

- followedBy(?e1, ?e2) ^ followedBy(?e2, ?e3) -> followedBy(?e1, ?e3)
- precedeBy(?e1, ?e2) ^ precedeBy(?e2, ?e3) -> precedeBy(?e1, ?e3)

The UML sequence diagram is mapped in a way that is possible to infer some implicit data. Leaving to a reasoner. For example, it is not necessary to specify all the operands of the fragment and all the elements of the operand. In fact, through the properties chain axiom defined in for the properties.

- hasOperand : hasOperand o nextOperand
- hasElement : hasElement o hasNext

It can specify the first (properties firstOperand and firstElement) and the last (properties lastOperand and lastElement) elements. In this technique, the reasoner will be able to infer all the remaining hasOperand and hasElement.

Moreover, the mixture of the above property chain can be very useful when combined with the subsequent SWRL rules:

- firstMessage(?o, ?m1) ^ index(?m1, 1) ^ size(?o, ?v) ^ index(?m2, ?v) -> lastMessage(?o, ?m2)
- firstOperand(?f, ?o) ^ firstMessage(?o, ?m) -> firstMessage(?f, ?m)
- elementOf(?e, ?o) ^ firstElementOf(?e, ?o) ^ hasMessage(?e, ?m) -> firstMessage(?o, ?m)
- elementOf(?e, ?o) ^ firstElementOf(?e, ?o) ^ hasFragment(?e, ?f) ^ firstMessage(?f, ?m) -> firstMessage(?o, ?m)
- lastOperand(?f, ?o) ^ lastMessage(?o, ?m) -> lastMessage(?f, ?m)
- elementOf(?e, ?o) ^ lastElementOf(?e, ?o) ^ hasMessage(?e, ?m) -> lastMessage(?o, ?m)
- elementOf(?e, ?o) ^ lastElementOf(?e, ?o) ^ hasFragment(?e, ?f) ^ lastMessage(?f, ?m) -> lastMessage(?o, ?m)
- elementOf(?e, ?o) ^ hasMessage(?e, ?m) -> hasMessage(?o, ?m)
- hasOperand(?f, ?o) ^ hasMessage(?o, ?m) -> hasMessage(?f, ?m)
- elementOf(?e, ?o) ^ hasFragment(?e, ?f) ^ hasMessage(?f, ?m) -> hasMessage(?o, ?m)
- hasNext(?e1, ?e2) ^ hasMessage(?e1, ?m1) ^ hasMessage(?e2, ?m2) -> nextMessage(?m1, ?m2)
- hasNext(?e1, ?e2) ^ hasMessage(?e1, ?m1) ^ hasFragment(?e2, ?f) ^ firstMessage(?f, ?m2) -> nextMessage(?m1, ?m2)

- $\text{nextOperand}(?o1, ?o2) \wedge \text{lastMessage}(?o1, ?m1) \wedge \text{firstMessage}(?o2, ?m2) \rightarrow \text{nextMessage}(?m1, ?m2)$
- $\text{hasNext}(?e1, ?e2) \wedge \text{hasFragment}(?e1, ?f1) \wedge \text{lastMessage}(?f1, ?m1) \wedge \text{hasFragment}(?e2, ?f2) \wedge \text{firstMessage}(?f2, ?m2) \rightarrow \text{nextMessage}(?m1, ?m2)$
- $\text{hasNext}(?e1, ?e2) \wedge \text{hasFragment}(?e1, ?f) \wedge \text{lastMessage}(?f, ?m1) \wedge \text{hasMessage}(?e2, ?m2) \rightarrow \text{nextMessage}(?m1, ?m2)$
- $\text{messageOf}(?m1, ?o) \wedge \text{messageOf}(?m2, ?o) \wedge \text{index}(?m1, ?v1) \wedge \text{add}(?v2, ?v1, 1) \wedge \text{index}(?m2, ?v2) \rightarrow \text{nextMessage}(?m1, ?m2)$
- $\text{nextMessage}(?m1, ?m2) \wedge \text{index}(?m1, ?v1) \wedge \text{add}(?v2, ?v1, 1) \rightarrow \text{index}(?m2, ?v2)$
- $\text{Operand}(?o) \wedge \text{firstMessage}(?o, ?m) \wedge \text{lastMessage}(?o, ?m) \rightarrow \text{size}(?o, 1)$
- $\text{Operand}(?o) \wedge \text{firstMessage}(?o, ?m1) \wedge \text{index}(?m1, 1) \wedge \text{lastMessage}(?o, ?m2) \wedge \text{index}(?m2, ?v) \rightarrow \text{size}(?o, ?v)$
- $\text{nextOperand}(?o1, ?o2) \wedge \text{lastMessage}(?o1, ?m1) \wedge \text{index}(?m1, ?v1) \wedge \text{lastMessage}(?o2, ?m2) \wedge \text{index}(?m2, ?v2) \wedge \text{subtract}(?v3, ?v2, ?v1) \rightarrow \text{size}(?o2, ?v3)$
- $\text{hasNext}(?e1, ?e2) \wedge \text{hasMessage}(?e1, ?m1) \wedge \text{index}(?m1, ?v1) \wedge \text{hasFragment}(?e2, ?f) \wedge \text{firstOperand}(?f, ?o) \wedge \text{lastMessage}(?o, ?m2) \wedge \text{index}(?m2, ?v2) \wedge \text{subtract}(?v3, ?v2, ?v1) \rightarrow \text{size}(?o, ?v3)$
- $\text{hasNext}(?e1, ?e2) \wedge \text{hasFragment}(?e1, ?f1) \wedge \text{lastOperand}(?f1, ?o1) \wedge \text{lastMessage}(?o1, ?m1) \wedge \text{index}(?m1, ?v1) \wedge \text{hasFragment}(?e2, ?f2) \wedge \text{firstOperand}(?f2, ?o2) \wedge \text{lastMessage}(?o2, ?m2) \wedge \text{index}(?m2, ?v2) \wedge \text{subtract}(?v3, ?v2, ?v1) \rightarrow \text{size}(?o2, ?v3)$

Let us introduce an example to show how to use our technique for describing a UML sequence diagram. Suppose one wants to describe the example, in Fig. 2. It is possible to model this scenario straightforwardly using our technique:

```
:example a :Fragment
; :kind "sd"
; :firstOperand :op1
; :lastOperand :op1 .
:op1 a :Operand
; :size "4"^^xsd:nonNegative
; :firstElement :el1
; :lastElement :el2 .
:el1 a :Element
; :hasMessage :first_message
; :hasNext :el2 .
:el2 a :Element
; :hasMessage :second_message
; :hasNext :el3 .
:el3 a :Element
; :hasMessage :third_message
; :hasNext :el4 .
:el4 a :Element
; :hasMessage :fourth_message .
```

```
:first_message a :Message
; :index "1"^^xsd:positiveInteger
; :type "synchronous"
; :caller :x
; :receiver :z .
:second_message a :Message
; :index "2"^^xsd:positiveInteger
; :type "synchronous"
; :caller :y
; :receiver :z .
:third_message a :Message
; :index "3"^^xsd:positiveInteger
; :type "asynchronous"
; :caller :x
; :receiver :z .
:fourth_message a :Message
; :index "4"^^xsd:positiveInteger
; :type "return"
; :caller :z
; :receiver :x .
```

### C. Messages

A UML sequence diagram shows interactions with which messages are exchanged among a set of objects participate. It concentrates to determine the behavioral view of a system. There are two dimensions in the UML sequence diagram, a vertical dimension, and a horizontal dimension respectively representing time and objects participating in the interaction. The horizontal dimension also captures the message which can be a signal or a class operation call between two vertical dashed lines which are called lifetimes. Each lifetime indicates an individual participant over a time in the interaction.

A message is an abstract element that has a name. It specifies the kind of communication between two lifelines of an interaction. It does not specify only the sort of communication but also the caller and therefore the receiver. Caller and receiver are normally two occurrence specifications (points at the ends of messages). The message is shown from the caller message end to the receiver message end.

The types of communication of the message as defined in the UML are listed below.

- Synchronous Message: represents a class operation call. All other calls of the caller are blocked waiting for the receiver to have processed the message and returned.
- Asynchronous Message: the caller of message does not need to wait for a replay to continue. Like synchronous messages.
- Reply Message: it is also defined return message, used to refer to the receiver that has processed the message and returned a result to the message caller.
- Self Message: when a caller and a receiver are the same it means a caller sent a message to itself. is represented as a U shaped arrow.
- Create Message: the receiver of this message is created during the interaction by the message that is being sent.

- Delete Message: it destroys its receiver. Targets can be destroyed during the interaction by the message that is being sent.
- Found Message: is a message from an unknown caller to a known receiver.
- Lost Message: is a message from a known caller to an unknown receiver.

In OWL 2 DataOneOf axiom is suitable for defining a type of the message. "DataOneOf" defines a datatype with a set predefined value space.

```
DatatypeDefinition( type DataOneOf( "synchronous"  
"asynchronous"... ) )
```

The message is a signal or a class operation. In a UML class diagram an operation op in class C can be one of the following:

```
op = { SimpleOp, ParOp }
```

- SimpleOp (simple operation): is an operation without parameters.
- ParOp (parameter operation): is an operation with parameters.

For each SimpleOp, we create a data property by respectively its domain and range associating with the class corresponding to the operation and the XSD type corresponding to the type of the operation in the UML class diagram.

Simple operations have a multiplicity restriction to max one value. The simple operations of a UML class in a class diagram are mapped in OWL 2 as a "DataProperty". The range of the data property is datatype. The datatype can be xsd:string, xsd:int and other datatypes [11].

Simple operations that use basic types are mapped by declaring a data property with the operation's name. An operation is a required component of its class. Consequently, the data properties describing operations have a max cardinality of one. The Simple operation SimpleOp of the UML class C having any of the above-mentioned DataType is translated in OWL 2 as

```
Declaration(Class(C))  
Declaration(DataProperty(SimpleOp))  
SubClassOf(C DataMaxCardinality(1 SimpleOp))  
DataPropertyDomain(SimpleOp C)  
DataPropertyRange(SimpleOp DataType)
```

An operation with parameters is mapped to:

- an OWL class (named className) with data property for every additional parameter and one of its datatype, the data properties have an exact cardinality of one.
- and object property between the new class and the class contains this operation, the object property has an exact cardinality of one.

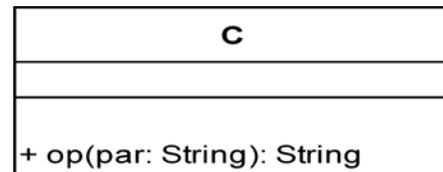


Fig. 4. A UML Operation op with Parameter par of Class C.

Fig. 4 shows an example of an operation op with parameters of class C is drawn by using UML.

The example of an operation op with parameters of class C is mapped into OWL 2 as

```
Declaration(Class(C))  
Declaration(Class(Op))  
Declaration(DataProperty(par))  
SubClassOf(Op DataExactCardinality(1 par))  
DataPropertyDomain(par Op)  
DataPropertyRange(par String)  
Declaration(DataProperty(datatype))  
SubClassOf(Op DataExactCardinality(1 datatype))  
DataPropertyDomain(datatype Op)  
DataPropertyRange(datatype String)  
Declaration(ObjectProperty(Op_C))  
SubClassOf(Op ObjectExactCardinality(1 Op_C))  
ObjectPropertyDomain(Op_C Op)  
ObjectPropertyRange(Op_C C)  
DisjointClasses(Op C)
```

#### D. Combined Fragment

A Combined fragment is an interaction fragment using an interaction operator to define the semantics of the combined fragment, such as alternative, option, and loop. Each combined fragment contains at least one interaction operand that is like a UML sequence diagram that can contain interaction fragments and messages together to model conditional behavior in a UML sequence diagram. An interaction operand illustrates the interactions between classes or object instances and the ordering of messages according to time. An interaction operand may have interaction constraints also called guards, which is a boolean conditional expression. A guard is a semantic condition that specifies the condition under which the interaction fragments and messages will be performed inside the interaction operand.

Interaction Operators as defined in the UML sequence diagram [12] are listed below.

- sd: abbreviation for sequence diagram, has one operand used for framing an entire sequence diagram.
- alt: abbreviation for alternatives, means that the combined fragment represents alternative or choice

paths of execution. Only the one whose guard is true will be chosen to execute.

- opt: abbreviation for option, equivalent to an alt only with one path with a guard, and the option operand is executed if the guard is true.
- loop: has one operand with a guard, means that the combined fragment represents a loop. The operand will be repeated at least the minimum count and no more than the maximum count as long as the guard is true.
- break: has one operand with or without a guard that is performed instead of the remainder of the enclosing interaction fragment.
- par: abbreviation for alternatives, means that the combined fragment represents more than one operands which can be executed parallel. Operands can be freely interleaved. In any order, but must be according to the ordering imposed by each operand separately.
- critical: abbreviation for critical region, is a region cannot be interleaved by other occurrence specifications.
- ref: abbreviation for critical reference, refers to an interaction defined on another diagram.

There are other interaction operators, such as Strict Sequencing, Weak Sequencing, Negative, Ignore, Consider, and Assertion which are also defined in [12, 13].

In OWL 2 DataOneOf axiom is suitable for defining a type of a combined fragment. "DataOneOf" defines a datatype with a fixed predefined value space.

DatatypeDefinition( kind DataOneOf( "alt" "opt" ... ) )

We can use a number of value restriction infix operators with the guard constraint of the class "Operand", such as =, >=, <=, > and <. The guard constraint is mapped in OWL 2 as a DataProperty. The value constraint of the guard is written in the UML sequence diagram as [Guard Op Value], where OP is the infix operator and Value represents the guard value. The map of the value constraint in OWL 2 is based on the infix operator used with the guard operator, such as ">=" is mapped using the OWL 2 axiom DatatypeRestriction and xsd:minInclusive, ">" is mapped using the OWL 2 axiom DatatypeRestriction and xsd:minExclusive, "<=" is mapped using the OWL 2 axiom DatatypeRestriction and xsd:maxInclusive, "<" is mapped using the OWL 2 axiom DatatypeRestriction and xsd:maxExclusive, "!=" is mapped using the OWL 2 axiom complementOf and the axiom DataHasValue, and "=" mapped using the OWL 2 axiom DataHasValue. For example, [Guard = Value] is a guard constraint, is mapped to OWL 2 as:

DataHasValue( Guard }Value}^^ datatype)

In this translation, Guard is the name of the guard, Value is the value of the guard, and datatype is the datatype of the guard Value.

#### IV. CASE STUDY

There has been substantially related work on mapping UML diagrams into Ontology has been discussed by several authors in the past. For instance, B. Bouchra. Author in [2] discusses the conversion method by building an e-learning ontology from its UML class diagram. In their approach. They have recourse to the Collection Ontology to map the composition relationship and the Value Partitions Design Pattern to map the inheritance. Moreover, the approach presented in [4] discusses the migrating UML class diagrams to Ontology. In their approach, the model information is stored in the XMI document by using a Power Designer tool then creating an ontology by passing this XMI document as the input of their mapping algorithms. The approach presented in [8] describes conversion rules from UML diagrams to Ontology containing multiple class, object and statechart diagrams. However, the goal of his work is analyzing the consistency and satisfiability of models. Moreover, they do not discuss the mapping UML sequence diagrams into Ontology.

We present a summary of our technique that we expound with a running example. Fig. 5 shows a UML sequence diagram that describes the withdrawal cash scenario of an ATM system, where Messages are numbered top-down. It exposes the object of each class and messages that can be invoked on them. It consists of four classes, namely, "User", "ATM", "Bank" and "Account". To evaluate our technique we need to first map this diagram into OWL 2 by following the mapping discussed in the previous sections, after mapping the diagram we pass the OWL 2 ontology to the OWL 2 reasoner. Fig. 6 and 7 illustrate the diagram after is mapped.

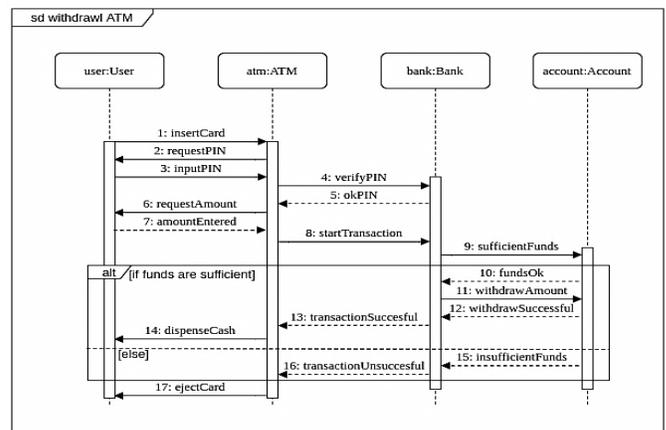


Fig. 5. Withdrawal Cash Scenario of an ATM System.

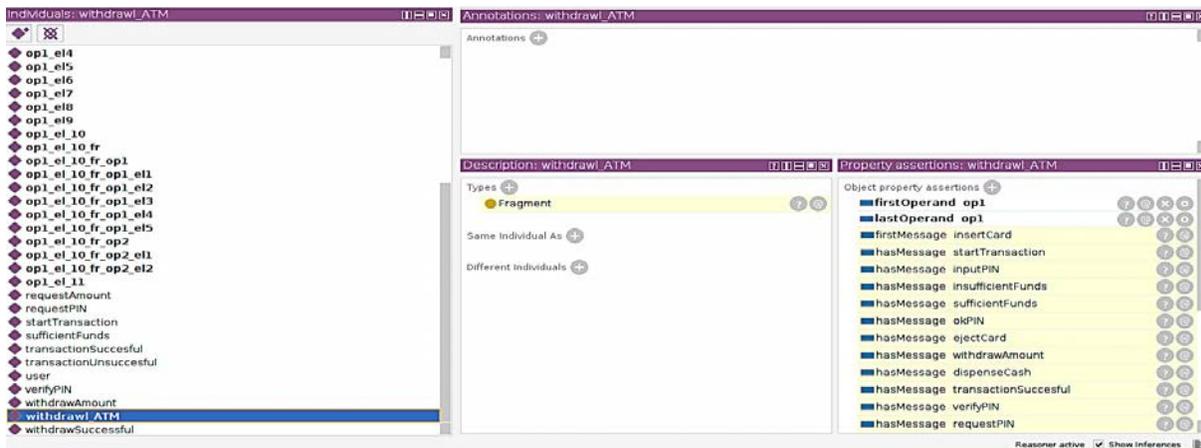


Fig. 6. The Withdrawal ATM Diagram after is Mapped.

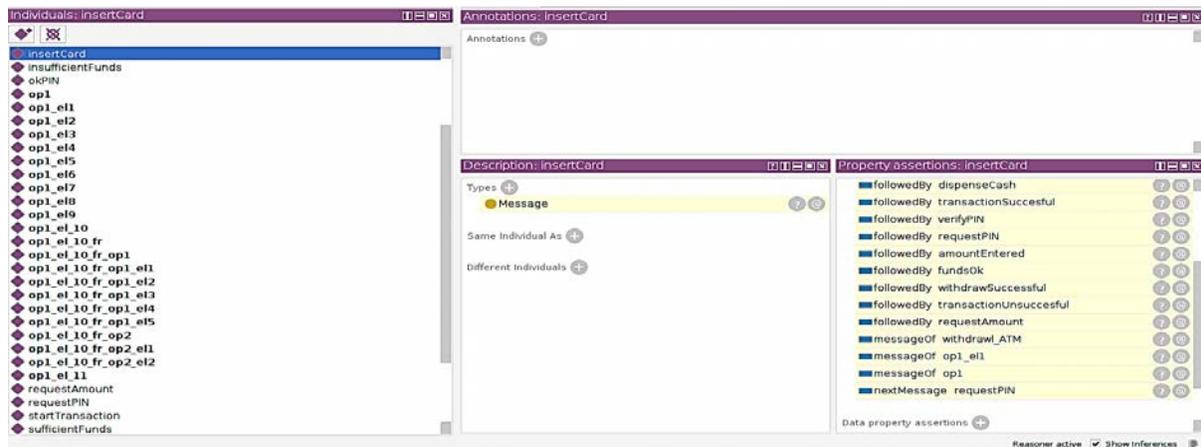


Fig. 7. The First Message insertCard after is Mapped.

## V. CONCLUSION

UML sequence diagrams are used to describe the behavior of systems. In this paper, we have demonstrated a technique for mapping behavioral knowledge expressed in the UML sequence diagrams as an OWL ontology. The OWL DL and SWRL rules are used to formalize the semantics of the sequence diagrams model. We have analyzed similarities and differences among UML and OWL elements in-depth. With this knowledge, we have developed rules for addressing the issue of defining sequence in OWL. Furthermore, we formalized a suitable way to handle the fragment operator with an arbitrary number of operands, which is crucial when specifying transformations of sequence diagrams.

As our future work, we will continue to study additional cases in order to complete the set of rules. We plan to provide the support of other interaction constraints.

## REFERENCES

[1] P. Kumar K N, R. Kumar V, K. Raghuvver, "A Survey on Semantic Web Technologies for the Internet of Things," In 2017 IEEE International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), 2017, doi: 10.1109/CTCEEC.2017.8454974.  
[2] B. Bouchra, B. Mohamed, "Building an e-learning system's OWL ontology by exploring the UML model", Journal of Theoretical and Applied Information Technology (JATIT), Vol.87, No.3, 2016, pp.380-387.

[3] H. Martina, V. Bureš , "Formal Ontologies in Information Systems Development: A Systematic Review", Information, Vol. 11, No. 2, 2020.  
[4] E. Oussama, A. Khadija, A. Larbi, B. Mohamed, "Mapping UML to OWL2 Ontology", Journal of Theoretical and Applied Information Technology (JATIT), Vol. 90, No. 1, 2016, pp. 126-143.  
[5] OMG "Ontology Definition Metamodel Superstructure Version 1.1," 2014, <https://www.omg.org/spec/ODM/About-ODM/>.  
[6] K. John, "Model-Based Development and Evolution of Information Systems – A Quality Approach", Springer, London, UK, 2012.  
[7] Object Management Group Website (OMG), <https://www.omg.org/>.  
[8] H.K. Ali, P. Ivan, "Consistency of UML class, object and statechart diagrams using ontology reasoners", Journal of Visual Languages & Computing, Vol. 26, 2015, pp. 42-65.  
[9] S. Evern, P. Bijan, C.G. Bernardo, K. Aditya, K.Yarden, "Pellet: a practical OWL-DL reasoner" , Journal of Web Semantics, Vol. 5, No. 2, 2007, pp.51-53.  
[10] C. Paolo, P. Silvio, "The Collections Ontology: creating and handling collections in OWL 2 DL frameworks", Semantic Web Journal (SWJ), Vol. 5, No. 6, 2014, pp. 515- 529.  
[11] W3C, OWL Working Group, "OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax. W3C Recommendation 11 December 2012," <https://www.w3.org/TR/owl2-syntax/>.  
[12] OMG, "Unified Modeling Language Superstructure Version 2.5.1," 2017, <https://www.omg.org/spec/UML/About-UML/>.  
[13] R. James, J. Ivar, and B. Grady, "The Unified Modeling Language Reference Manual 2nd Edition. Addison Wesley", 2004.

# Intelligent Fleet Management System for Open Pit Mine

Hajar BNOUACHIR<sup>1</sup>, Atae SEMMAR<sup>7</sup>  
Hicham MEDROMI<sup>8</sup>

Engineering Research Laboratory (LRI), System  
Architecture Team (EAS), National and High School of  
Electricity and Mechanic (ENSEM), Hassan II University of  
Casablanca, Research Foundation for Development and  
Innovation in Science and Engineering, Casablanca,  
Morocco

Meryiem CHERGUI<sup>2</sup>, Aziza CHAKIR<sup>5</sup>

Engineering Research Laboratory (LRI), System  
Architecture (EAS), National and High School of Electricity  
and Team Mechanic (ENSEM), Hassan II University of  
Casablanca, Casablanca, Morocco

Nadia MACHKOUR<sup>3</sup>, Mourad ZEGRARI<sup>4</sup>

Structural Engineering, Intelligent Systems and Electrical  
Energy Laboratory, Ecole Nationale Supérieure des Arts et  
Métiers, Hassan II University of Casablanca, Innovation  
Lab for Operations, Mohammed VI Polytechnic University  
Benguerir, Morocco, Casablanca  
Morocco

Laurent DESHAYES<sup>6</sup>

Innovation Lab for Operations  
Mohammed VI Polytechnic University  
Benguerir, Morocco

**Abstract**—Fleet management systems are currently used to coordinate mobility and delivery services in a wide range of areas. However, their traditional control architecture becomes a critical bottleneck in open and dynamic environments, where scalability and autonomy are key factors in their success. In this article, we propose an intelligent distributed Fleet Management System architecture for an open pit mine that allows mining vehicles control in a real time context, according to users' requirements. Enriched by an intelligence layer made possible by the use of high-performance artificial intelligence algorithms and a reliable and efficient perception mechanism based on Internet of Things technologies and governed by an smart and integrated decision system that allows the fleet management system to improve its agility and its response to user requirements, our architecture presents numerous contributions to the domain. These contributions enable the fleet management system to meet the interoperability and autonomy requirements of the most widely used standards in the field, such as ISA 95.

**Keywords**—Fleet management system; open pit mine; monitoring; architectures; artificial intelligence; real time system

## I. INTRODUCTION

Monitoring the company's activity in real time has many assets and advantages not to be ignored. This instant management allows us to visualize and know the real situation of the company as well as to monitor its objectives, key figures and performance indicators; it facilitates decision-making, manages its daily life more effectively and improves its reactivity and competitiveness.

Regardless the structure size, any good manager must be equipped with the appropriate resources to manage his activity effectively. This means that the positioning of carried out actions and committed budgets should be monitored, for both, quantitative and qualitative objectives.

The Fleet Management System (FMS) [1] is a modern technological solution to manage the fleet's vehicles in terms of automation and business process optimization and to link all business sectors such as dispatching, purchasing, communications, navigation, accounting and finance in a single system. FMS is easy to control and optimize, its foundations date back to the 1980s, when the computer was integrated in the vehicle and connected to different wireless networks by satellite and terrestrial [2].

In the mining sector, the main objective of the fleet management system is to optimize mine production and efficiency based on real-time data. More specifically, the FMS aims to maximize mine production, minimize stock handling, feed the pro-cessing plant at the planned rate and meet quality mixing constraints [3].

In this article, we will present first, fleet management systems' state of the art as well the use of artificial intelligence in it; we will also initiate a fleet management system evaluation model based on ISA 95. We will discuss after that, open pit mine FMS problems before proposing a general intelligent FMS architecture. We will end this article with conclusion and perspectives.

## II. STATE OF THE ART

### A. Fleet Management Systems

Generally, Fleet management system (FMS) is a term used to refer to a wide range of solutions for different fleet applications in the fields of transport, distribution and logistics [4] [5]. It includes targeted planning, supervision and control of fleet operation according to available transport resources and application constraints. The FMS aims to reduce risks, increase service quality and improve a fleet's operational efficiency while minimizing costs [6].

The implementation of operational research techniques as the main means of addressing the high cost of transporting equipment to open-pit mines, has attracted re-searchers and laid the foundation for fleet management systems (FMS) used in open-pit mines.

They considered the FMS as a real time decision-making system [7] [8] for materials handling in an open pit mine. FMS obtains the required information about mining operations from the database and makes decisions accordingly. These decisions are then implemented in the operation and the FMS is recalled whenever a new decision is required.

Fig. 1 shows the communication between open-pit mining and the FMS.

The modular mining system revealed models [10] [11] and algorithms, namely:

- **Shortest path:** It aims at optimizing travel time using the Dijkstra method [12] [3], its advantages are that the algorithm does not need to study all contours and it has an order of  $n^2$  so that it is sufficiently efficient to use for relatively large applications. But the model takes a long time, it has failures in case of negative edges.
- **Allocation planning[13]:** Minimizes the total number of trucks required using the simplex method, the model is in real time, the flow rate of each road is based on the volume of material rather than the number of trucks. However, it is only appropriate when some variables are involved and it is non-negative constraints for all variables.
- **Real-time Dispatching[14]:** Minimizes button waste by applying the dynamic programming method, it propose a progressive time horizon when the assignment order is required, Undercarriage and over trucking conditions are taken into account, but the definition of a progressive time horizon for an assignment order is indispensable, taking into account sub-campaign and over-trucking conditions.

There are many limits of the current algorithms, namely

- The link between the strategic plan and operations: the proposed fleet management systems do not allow both short-term and long-term objectives to be achieved,
- Most models are deterministic: they assume a constant average grade for each size front and do not cover the entire life of the mine,
- The current systems do not take into account the loss of tones associated with the movement of excavators from one level to another, they present one of the main factors of deviation from the production rate,

We can identify two aspects of the fleet management system as shown in Fig. 2:

- A real-time control system using the appropriate dashboards.
- Truck loader assignment system based on algorithms & methods.

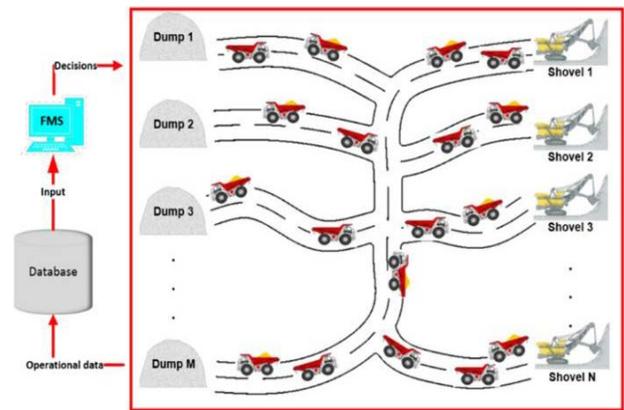


Fig. 1. Schematic of the Surface mining Operation and how it Communicates with the FMS [9].



Fig. 2. Fleet Management System.

In this article, we will focus on real-time control system, with decision making Dashboards.

### B. Fleet Management System Evaluation

Based on the ISA-95 standard [15], which is an international standard for control systems integration, we identify the criteria to evaluate fleet management systems as shown in Fig. 3.

ISA 95 consists of 11 production control functions[16] [17]:

- The control and resource allocation function;
- The dispatching function of production;
- The data collection and acquisition function;
- The quality management function;
- The management function of the manufacturing process;
- The planning and monitoring function;
- The performance analysis function;
- The operation management and fine-tuning scheduling function;
- The documentation management function;
- The workforce management function;
- The maintenance management function.

Nowadays, there are many companies around the world providing mining fleet management systems. Some of the most

popular are: Modular Mining Systems, Jig-saw Software and Wenco, Dynamine TATA services Consulting [3]. However, Mi-cromine Pitram system, Viste and CAT® MineStar™ FLEET are the most common.

In our article entitled Fleet management system in real time “Stat of Art”, we evaluate these industrial solutions according to ISA 95 evaluation criteria and we conclude that the majority of them do not afford:

- Management of skills and authorizations
- Dynamic resource management
- Information transmission to the different work stations  
Management of non-production control operations
- Observation statistical quality control and statistical process control
- Steering by automatic correction of the drifts
- Comparison with objectives
- Compliance check of files and forms
- Information Transfer between resources

C. Artificial Intelligence for Fleet Management Systems

Artificial intelligence (AI) is omnipresent nowadays in many technological applications (apps, websites, gaming predictions, services, etc.). It’s a way to give accurate user recommendations and change its experience in many fields [18].

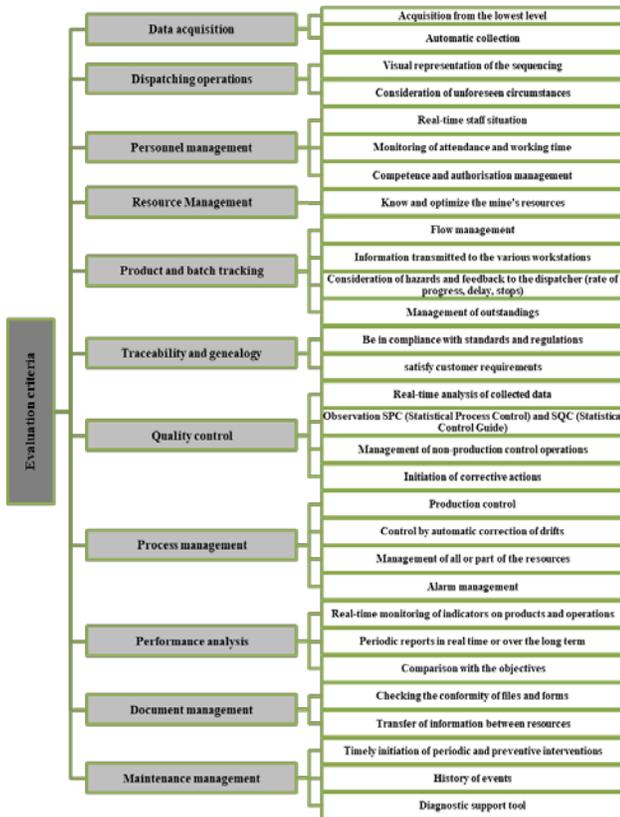


Fig. 3. The Criteria for Evaluating the Fleet Management System.

Fleet management has revolutionized AI. The priority given to commercial results and driver safety preserving cost and efficiency introduce intelligent fleet management systems.

The use of smartphones and devices recommending the best way to take is for the average driver synonym of AI presence. Drivers’ lives are easier thanks to complex traffic apps that combine GPS and artificial intelligence [19]. There are many benefits of AI-based applications for fleet management such as route recommendations, road risk data analysis and driver coaching. In fact, earlier technology failed to provide the accuracy, efficiency, convenience as AI apps do. As a result, it is becoming the optimal and safer solution to transport goods and services in general and in mining field as well.

We can define AI fleet management as the use of artificial intelligence to manage fleet operations, it rationalizes fleet manager tasks by eliminating human error in transportation processes [20].

With AI recommendations in FMS, drivers, managers, and mechanics can make correct and optimal decisions to improve the fleet long-term performance. It also al-lows drivers’ autonomy during each transport cycle. Table I presents some key aspects of fleet management that AI can optimize:

TABLE I. IA FOR FMS OPTIMIZATION

IA ASPECT	USES	RESULTS
Real-Time Fleet Analytics	Collect data for predictive analytics: <ul style="list-style-type: none"> <li>• traffic</li> <li>• road conditions,</li> <li>• environmental hazards,</li> <li>• real-time weather,</li> <li>• mechanical faults</li> </ul>	<ul style="list-style-type: none"> <li>• Prioritization opportunities and risks</li> <li>• Make better routes</li> <li>• Schedules, maintenance delivery</li> <li>• Dispatch fleet outcomes and activities arrangements</li> <li>• Stay prepared for any unexpected events.</li> </ul>
Better Repair & Maintenance Decisions	Predictive maintenance Fleets state at all times	<ul style="list-style-type: none"> <li>• Accurate self-diagnostics</li> <li>• Saving a routine maintenance cost</li> </ul>
Fleet Integration	<ul style="list-style-type: none"> <li>• Planning, maintenance and monitoring operations</li> </ul>	<ul style="list-style-type: none"> <li>• Integrating every department on a single platform</li> <li>• Save time and costs</li> <li>• Cohesive fleet</li> </ul>
Simpler Recruitment Process	<ul style="list-style-type: none"> <li>• Knowledge Management</li> <li>• The most qualified drivers’ recommendation</li> </ul>	<ul style="list-style-type: none"> <li>• Simplify the on boarding process</li> </ul>

III. PROBLEMATIC

The proposed models must be as close as possible to reality, taking into account the heterogeneity of trucks and the complete modeling of an open-pit mining operation. FMS should also determine the shortest dynamic routes taking into account the current location of the truck, its next destination and the time required to reach the objective in any current traffic congestion.

Otherwise, to identify FMS users' needs and to define its main attributes in real time, a process was initiated by Zeimpekis and Vasileios [21]. The results revealed that the mining fleet management systems must meet logistics and distribution requirements such as:

- Real-time tracking of vehicle location
- Generation of vehicle and distribution system performance reports.
- Generation of proof of delivery at the end of loading and shipment.

On the other hand, we carried out within the framework of this project, a needs study within the OCP mining company (Benguerir mine), and we identified the following specifications:

- Integrated management of the production chain
- Optimization of equipment performance and elimination of losses
- Improved maintenance efficiency
- Ensure equipment security

According to the previous specifications and the evaluation of the FMS conducted in Section 2.3 and given the benefits of AI in this context, we propose a generic architecture of an Intelligent Fleet Management System for an open pit mine.

#### IV. PROPOSED ARCHITECTURE

The real-time aspect in the FMS is necessary to deal with unforeseen events. It makes it possible to detect deviations from the initial allocation plan and adjust the schedule [6] by offering immediate and efficient re-routing.

In this section, we present our fleet management proposal with a description of an FMS intelligent architecture (see Fig. 4).

The intelligent FMS architecture includes three basic layers.

IoT layer contains intelligent sensors for both autonomous and manual vehicles in the open pit mine. Every sensor is connected to an intelligent agent as software and these agents are linked to control device agent to manage them and to announce the real time status of the mine fleet for different stakeholders.

In fact, control device agent is responsible for detecting new device (new vehicle IoT sensor), update its status in FMS mapping, suspend a device and detect maintenance needs, dangers and special behaviours.).

Indeed, on each vehicle, IoT sensor is implemented and connected to open pit mine network. These sensors allow the acquisition and transmission of vehicles' information's such as the position of the machines, their speed and their conditions. They are also aware of each other's existence .One more important thing some vehicles in the mine need more intelligent sensors for specific mission or critical situations and each vehicle in the fleet is aware and reacts to the entire fleet's

needs .A disruption engine determines if a fleet vehicle will be critically late for a scheduled trip and allocate alternative vehicles, as required.

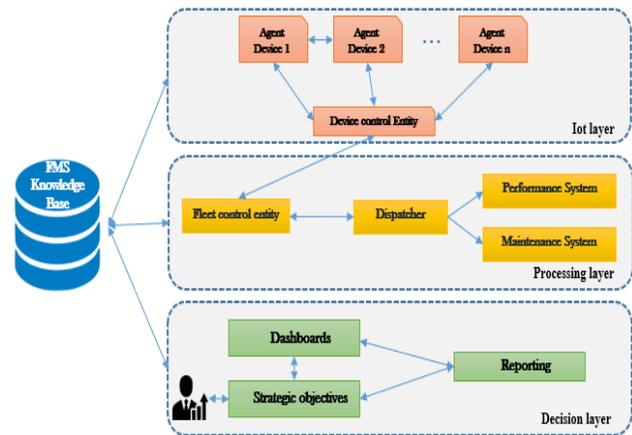


Fig. 4. Fleet Management System Architecture.

Processing layer represents the fleet coordination entities. In fact, in this level, we have:

Fleet control entity that monitors the operating states and positions of the vehicles through the communication with the device control agent. It informs stakeholders and the other agents as well of any changes in the fleet that would require an adjustment of the assignment of tasks. It has fleet reference states to be able to draw up a real-time, vigilant and continuous comparison.

Dispatcher Agent: once executed, this entity recalculates the overall allocation of tasks for mining vehicles. It has two dispatching algorithms based on a set of assignment criteria: Short-term allocation algorithm and a long-term allocation algorithm. The dispatcher is one of the most intelligent entities in our FMS. It is up to it to de-fine the mining missions by taking into consideration the strategic decisions and the real state of the fleet. Dispatcher agent also executes optimized dispatching in collaboration with performance and maintenance systems to research fleet vehicle that should go next and in which itinerary.

Maintenance system manages vehicles under maintenance. It allows updating their status and gives information about what is wrong with them. It also helps drivers and open pit mine responsible (s) to have real time informations about vehicles maintenance status. It is the necessary data to make dispatching allocation strategies (long term and short term).

Performance system: It is the intelligent system that ensures FMS performance management, through machine learning models. The datasets are made from real and successful open pit mine missions in OCP. This system y is connected to the knowledge base just like all the entities of the processing layer.

Decision layer it is the layer which mainly interfaces the FMS with its potential users (directors and executive managers) via:

- Collecting the strategic objectives of the open pit mine from top management and sending them in

synchronous or asynchronous mode to the dispatcher to develop the mining missions.

- Display of performance indicators and real-time and instantaneous graphs representing the state of the mine (production and fleet).

Decision layer is also able to generate stats and administrative documents in order to facilitate mining regulations respect.

## V. IMPLEMENTATION

As for expected tools for the proposed intelligent fleet management system, we use mainly Internet of Things (IoT) technology.

IoT refers to a network of actuators and sensors continuously collecting data from IoT is a network of sensors continuously collecting environmental data. In the open pit fleet management system, the IoT collects data for analysis. It ensures the transparent sharing of information between all stakeholders in the supply chain, such as drivers, production managers and dispatchers.

For our intelligent fleet management system, we use the three main technologies of IoT:

- Wireless communication (4G, Bluetooth, WiFi) to transmit the relevant information between the different vehicles of the mine.
- Global positioning system (GPS) for precise location tracking in real time.
- Integrated diagnostic analyzers (such as OBDII and J1939) for self-diagnosis and reporting

The smart FMS also has cameras to ensure that video data can be captured, analyzed and accessible at all times, allowing better study of driver behaviour, road conditions or hazards. All the more, machine learning technology allows fleets to learn data collected over time and make managed adjustments based on that data in the performance management entity.

The proposed intelligent FMS with all the components and technologies used is able to perform the following tasks:

- Collection of precise mining transport data and its transmission to other devices.
- Transmission of information at all levels of the supply chain from top management to implementers.
- Analyze data in real time and advice drivers on the best course of action.
- Performing self-diagnostics of mining vehicles and recommending solutions via predictive maintenance.

This is important because it creates a new era of mining fleet management system in which human errors are reduced and open pit mine production is optimized in different aspects of the transport cycle. This led to better results and cost optimization. Fig. 5 presents the first version of open pit mine platform dashboards. The platform covers ISA95 main domains.



Fig. 5. The First Version of Open Pit Mine Platform Dashboards.

Primary results were positively validated and appreciated by experiment mine actors.

## VI. CONCLUSION AND PERSPECTIVES

In this article, we proposed an intelligent fleet management system focusing on re-al-time management of mining equipment by applying a multi-agent system. First, we presented a state of the art of fleet management systems as well as some definitions of the multi-agent system. In addition, we identified the criteria for evaluating a fleet management system based on the ISA95 standard and finally we proposed architecture for an intelligent fleet management system composed of three layers: the vehicle layer, the processing layer and the decision layer. We implemented the first version of this architecture and test it in Benguerir experimental mine in order to improve it and extend its application scope.

In perspective, we will evaluate the proposed system using the criteria of the ISA95 standard and improve dispatching algorithms; we will also detail and implement every part of the smart fleet management system architecture.

## ACKNOWLEDGMENT

This research was supported by Mohammed VI Polytechnic University de Benguerir, Morocco and OCP group. We thank our colleagues from university and from OCP group who provided insight and expertise that greatly assisted the research.

## REFERENCES

- [1] Ch. Achillas, D. Bochtis, D. Aidonis, V. Marinoudi, and D. Foinas, "Voice-driven fleet management system for agricultural operations," *Inf. Process. Agric.*, vol. 6, no. 4, pp. 471–478, Dec. 2019, doi: 10.1016/j.inpa.2019.03.001.
- [2] E. Krelja Kurelović and B. Škabić, "Comparison of fleet management systems," *Zb. Veleuč. U Rijeci*, vol. 6, no. 1, pp. 357–370, 2018, doi: 10.31784/zvr.6.1.27.
- [3] A. M. Afrapoli and H. Askari-Nasab, "Mining fleet management systems: a review of models and algorithms," *Int. J. Min. Reclam. Environ.*, vol. 33, no. 1, pp. 42–60, Jan. 2019, doi: 10.1080/17480930.2017.1336607.
- [4] "Wellman - (54) FLEET MANAGEMENT SYSTEM.pdf."
- [5] P. Chaowasakoo, H. Seppälä, H. Koivo, and Q. Zhou, "Improving fleet management in mines: The benefit of heterogeneous match factor," *Eur. J. Oper. Res.*, vol. 261, no. 3, pp. 1052–1065, Sep. 2017, doi: 10.1016/j.ejor.2017.02.039.

- [6] H. Billhardt et al., "Dynamic Coordination in Fleet Management Systems: Toward Smart Cyber Fleets," *IEEE Intell. Syst.*, vol. 29, no. 3, pp. 70–76, May 2014, doi: 10.1109/MIS.2014.41.
- [7] H. Yamamoto and R. B. Ramli, "Real-time Decision Making of Agents to Realize Decentralized Autonomous FMS by Anticipation," p. 11, 2006.
- [8] S. Parhi and S. C. Srivastava, "A Highly Efficient and Optimised Simulation Based Multi objective Decision-Making for FMS Control," *Int. J. Emerg. Res. Manag. Technol.*, vol. 6, no. 6, p. 98, Jun. 2018, doi: 10.23956/ijermt.v6i6.254.
- [9] A. Moradi Afrapoli, M. Tabesh, and H. Askari-Nasab, "A multiple objective transportation problem approach to dynamic truck dispatching in surface mines," *Eur. J. Oper. Res.*, vol. 276, no. 1, pp. 331–342, Jul. 2019, doi: 10.1016/j.ejor.2019.01.008.
- [10] A. W. I. S. | [www.anchorwave.com](http://www.anchorwave.com), "DISPATCH Fleet Management System | Modular Mining Systems, Inc." <https://www.modularmining.com/product/dispatch/> (accessed Mar. 16, 2019).
- [11] "Modular Mining | MineSmarterTM," Modular Mining. <https://www.modularmining.com/> (accessed Jun. 13, 2019).
- [12] M. Najafi, R. Rafiee, and S. M. E. Jalali, "Open pit limit optimization using dijkstra's algorithm," p. 5, 2020.
- [13] Shiv PHoomari-Nasaban Ask and Shiv Prakash Upadhyay, "Dynamic shovel allocation approach to short-term production planning in open-pit mines," *International Journal of Mining, Reclamation and Environment*, 2017.
- [14] R. B. Dannenberg, "Time-Flow Concepts and Architectures For Music and Media Synchronization," p. 6.
- [15] L. Yue, P. Niu, and Y. Wang, "Guidelines for defining user requirement specifications (URS) of manufacturing execution system (MES) based on ISA-95 standard," *J. Phys. Conf. Ser.*, vol. 1168, p. 032065, Feb. 2019, doi: 10.1088/1742-6596/1168/3/032065.
- [16] "Le standard ISA 95 - Smart Manufacturing," *TLG Pro*, Feb. 05, 2018. <https://www.tlgpro.fr/2018/02/standard-isa-95/> (accessed Jan. 02, 2019).
- [17] B. Scholten, *The road to integration: a guide to applying the ISA-95 standard in manufacturing*. Research Triangle Park NC: ISA, 2007.
- [18] M. Chergui, H. Medromi, and A. Sayouti, "Inter-organizational Workflow for Intelligent Audit of Information Technologies in terms of Enterprise Business Processes," *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, no. 5, 2014, doi: 10.14569/IJACSA.2014.050515.
- [19] A. M. Nascimento et al., "A Systematic Literature Review About the Impact of Artificial Intelligence on Autonomous Vehicle Safety," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–19, 2019, doi: 10.1109/TITS.2019.2949915.
- [20] R. Salazar-Cabrera, A. P. De La Cruz, and J. M. Madrid Molina, "Fleet Management and Control System from Intelligent Transportation Systems perspective," in *2019 2nd Latin American Conference on Intelligent Transportation Systems (ITS LATAM)*, Bogota, Colombia, Mar. 2019, pp. 1–7, doi: 10.1109/ITSLATAM.2019.8721347.
- [21] V. Zeimpekis, G. M. Giaglis, and I. Minis, "Development and Evaluation of an Intelligent Fleet Management System for City Logistics," in *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, Waikoloa, HI, USA, Jan. 2008, pp. 72–72, doi: 10.1109/HICSS.2008.121.

# A Human Gait Recognition Against Information Theft in Smartphone using Residual Convolutional Neural Network

Gogineni Krishna Chaitanya<sup>1</sup>, Dr. Krovi.Raja Sekhar<sup>2</sup>

Department of Computer Science and Engineering  
Koneru Lakshmaiah Education Foundation  
Vaddeswaram, 522502, Andhra Pradesh, India

**Abstract**—The genuine user of the smartphone is identified and information theft is prevented by continuous authentication, which is one of the most emerging features in biometrics application. A person is recognized by analysing the physiological or behavioural attributes is defined as biometrics. The physiological qualities include iris acknowledgment, impression of finger, palm and face geometry are used in the biometric validation frameworks. In the existing entry-point authentication techniques, a confidential information is lost because of internal attacks, while identifying the genuine user of the smartphone. Therefore, a biometric validation framework is designed in this research study to differentiate an authorized user by recognizing the gait. In order to identify the unauthorized smartphone access, a human gait recognition is carried out by implementing a Residual Convolutional Neural Network (RCNN) approach. A personal information of end user in smartphone is secured and presented a better solution from unauthorized access by proposed architecture. The performance of RCNN method is compared with the existing Deep Neural Network (DNN) in terms of classification accuracy. The simulation results showed that the RCNN method achieved 98.15% accuracy, where DNN achieved 95.67% accuracy on OU-ISIR dataset.

**Keywords**—Authentication; biometric analysis; genuine user; information loss; residual convolutional neural network; smartphone

## I. INTRODUCTION

In the past few years, the researchers considered reliable authentication of users as most significant technique in various applications namely smart phone unlocking or online banking system, because the verification provides security to credential information of end users [1]. According to every access control framework, identification has three different strategies to verify the possessions. A person's keys, badges and cards are considered as first strategy, where secret password, Personal Identification Number (PIN), client id are presented in second strategy and the final strategy is biometrics such as ear, face, fingerprint, etc. [2]. The individual's identification is verified by biometric verification techniques straightforwardly, when compared with first and second techniques [3]. The existing algorithms uses the physiological biometrics such as retina, unique mark, and palm prints for characterizing the human personality on the basis of information collected from the human body [4]. A particular activity of a man is characterized by using behavioural biometrics such as keystroke elements,

voice ID, signature elements and movements of human [5]. In biometrics frameworks, the existing techniques widely used the human's face, fingerprint impression and iris [6,7]. But, there are some major drawbacks presented in the existing techniques, where the health conditions, age, external appearances influence the human face and ink, drugs or burns on fingers influences the finger impression. Therefore, gait behavioural analysis is introduced by the researchers as new innovation, because a novel element of an individual is described by human gait [8,9].

The personal information of user can be accessed by unlocking the smartphone device, where entry-point authentication plays an important role for unlocking. A direct contact, fixed emotions or specific postures are mostly used by many existing biometrics for pattern classifications, but gait recognition doesn't require any specific notifications to identify the persons that works in a natural way [10-11]. In general, gait images or videos are recorded/captured from a far distance. In order to identify a person, vision techniques are used by gait recognition with a low resolution images/videos. The normal cameras in a mobile phone is used to collect the gait images and videos, therefore, recognition of gait is attractive, simple and effective method. Moreover, gait recognition algorithms are effective even with low-quality images [12,13]. In this research study, there are three major steps presented in the proposed architecture, where the steps include pre-processing the data, extracting the features and matching process. A personal information of smartphone is secured by developing the gait recognition of smartphone architecture. The research study designed the Residual Convolutional Neural Network for verifying the user behaviour in gait recognition. The steps for training and prediction of user behavioural patterns are explained through continuous authentication in the proposed algorithm. The validations are conducted on CASIA-A, B and OU-SIR datasets for validating the performance of RCNN in terms of accuracy, precision, F1-score and sensitivity.

The remaining research paper is consisting of: Section 2 presents the discussion of various existing techniques, where problem statement of the research study is provided in Section 3. The explanation of the system design with proposed algorithm is illustrated in Section 4 and the validation of proposed RCNN method with traditional techniques is depicted in Section 5. Finally, the conclusion of the research study is presented in Section 6.

## II. LITERATURE REVIEW

In this section, the existing techniques includes different neural networks and fuzzy logic techniques are discussed that are used for predicting the gait behaviour of a person. Moreover, the key benefits of the existing techniques with its limitations are presented from the year 2017 to 2020. The parameters such as recognition accuracy, precision, recall and f-measure are used to validate the effectiveness of existing techniques.

N. Takemura et al. [14] designed an input and output architecture for CNN for recognizing the cross-view gait. Two major aspects were considered in this developed method, where the first aspect was identification versus verification and the final aspect was the trade-off between the displacements of spatial information that was caused by view difference and subject difference. The experiments were conducted on OU-ISIR large population dataset, where the results proved that Siamese network and triplet network were suitable for verification and identification. Here, the developed method identified the cross-view gait only by considering the spatial displacement that were caused by view angle difference. But, the displacement of spatial was caused by clothing difference, carrying status difference, walking speed difference, etc., which were not focused by the developed CNN method.

S. Tong et al. [15] addressed the multi-view gait recognition problem by implementing the Spatial-Temporal Deep Neural Network (STDNN) that consists of Spatial Feature Network (SFN) and Temporal Feature Network (TFN). The low-level edge features were extracted by adopting the sub-network of TFN and given as an input to the Spatial-Temporal Gradient (STG) network. A multilayer CNN was used to extract the spatial features, where a long short-term memory unit was used to extract the STG features. When comparing with intra-class variations, inter-class variations were larger by optimizing the SFN using verification loss and classification loss. The simulations were conducted on three datasets includes CASIA-B, CMU MoBo and OU-SIR datasets. The results stated that the STDNN method achieved better performance than other existing techniques, however the developed method was insufficient to handle the over-fitting problems.

I. Huitzil et al. [16] studied the recognition of gait system that depends on fuzzy ontologies and Microsoft Kinect. When comparing with existing techniques, the study proposed the novel recognition algorithm based on fuzzy logic achieved better performance for straight line walks. The issues of identification of unknown individuals were solved by the developed method, which were not presented in the system knowledge base. In order to improve the performance of the method, a new dataset with 91 individuals were developed and the results proved that the developed algorithm was robust against small changes in the biometrical values across different steps. But, the reflective clothing and footwear affected the recordings quality, when building the new datasets, which was the major drawback of the method.

X. Wang et al. [17] achieved better gait recognition by using Trituple gait silhouettes (TTGS) feature representation

and Multichannel CNN (MCNN). The essential features of human gait were extracted by using MCNN approach, where gait data was pre-processed by TTGS that contained local information and didn't required the segmentation of strict gait cycle. The experiments were conducted on popular datasets included CASIA Dataset B and OU-ISIR Large Population Dataset in terms of cumulative match characteristics (CMCs). The results proved that the developed method achieved better performance than other existing techniques. However, the method failed to process the original gait videos and worked only on the silhouette images.

H. Arshad et al. [18] recognized the human gait by integrating framework of DNN with Fuzzy Entropy controlled Skewness (FECS) approach. The pre-trained CNN model was used to extract the DCNN features and their information were mixed using parallel fusion approach. The best subsets of features were selected by using the FECS approach with the calculation of skewness and entropy vectors from the fused vectors. The four datasets include CASIA-A, B, C and AVAMVG gait was used to test the efficiency of the developed method and the results proved that the FECS approach achieved better performance. However, the developed method neglected some useful features and the system accuracy was affected due to low-resolution video sequences.

In order to address the issues of existing techniques, the research study designed a RCNN approach for recognizing the human gait in order to preserve the credential information of end users.

## III. PROBLEM STATEMENT

In this section, the major problem of this research study that are presented in smartphone is explained. In general, a vast amount of personal information is presented in the user's smartphone. The entry-point authentication is generally used to secure the privacy of user's information, where knowledge-based passwords in authentication is used for unlocking the smartphones that are visible to user's surrounding people. Then, the unauthorized users (maybe friends/relatives) access the smartphone by stealing the knowledge-based passwords using internal attacks. Therefore, the confidential information of users are easily accessed that may cause mischiefs. Moreover, the major existing techniques such as neural networks may suffers from overfitting problems.

## IV. PROPOSED METHODOLOGY

The explanation of the proposed method is presented in this section. In order to address the theft of user's information, the existing techniques designed the two-phase authentication of the smartphone. However, this technique needs extra time for verification and it suffers from an internal attack, when the password space of the second phase is visible. Therefore, behavioural biometric of smartphone is introduced in this research study by developing a Residual Convolutional Neural Network (RCNN), because behavioural patterns of the individual user are identical. The proposed architecture is consisting of two phases namely enrollment phase and verification phase, which is shown in Fig. 1.

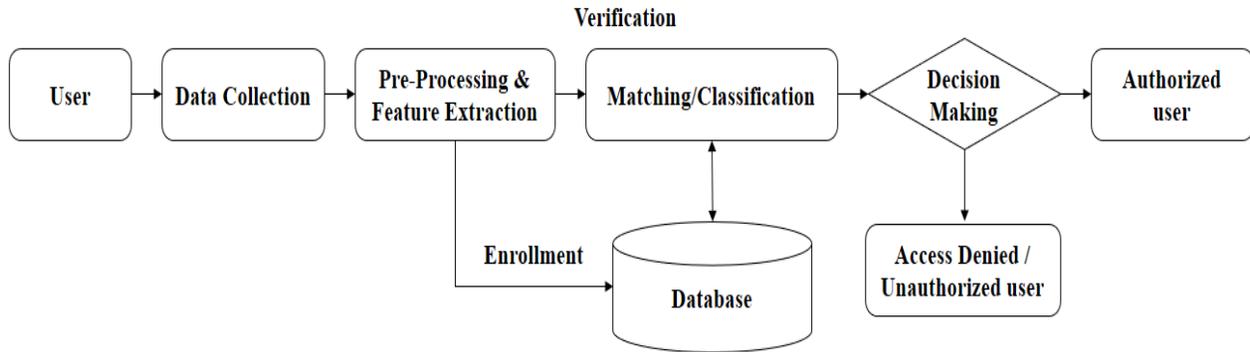


Fig. 1. Proposed Architecture.

There are three stages presented in the enrollment phase namely data collection, pre-processing with extracting the important data and database. In the verification phase, there are four stages are presented such as collecting the data, pre-processing with feature extraction, database and matching the data for final decision (i.e. decision making). In the below section, every stages in enrollment and verification phases are illustrated as follows:

#### A. Data Collection

The behaviour activity of user is collected by using tiny sensors in the smartphones, where the gyroscope sensors as  $G$  and accelerometer sensors as  $A$  of the smart phone are selected for collecting the biometric patterns of users. Consider  $G = [G_x, G_y, G_z]$  and  $A = [A_x, A_y, A_z]^T$ , where  $x, y, z$  are defined as vectors and those vectors are stored in two dimensional arrays called  $A$  and  $G$ . The transpose of two dimensional arrays are horizontally concatenated and stored in a single two-dimensional array, where the Eq. (1 and 2) shows the transpose of  $A$  and  $G$  and Eq. (3) shows the storage of concatenation of Eq. (1 and 2).

$$A = [A_x, A_y, A_z]^T \quad (1)$$

$$G = [G_x, G_y, G_z] \quad (2)$$

$$D = [A, G] \quad (3)$$

Where, a two-dimensional array is illustrated as  $D$ . After collecting the data, it will give as an input for pre-processing stage.

#### B. Pre-Processing and Feature Extraction

In order to improve the behavioural pattern quality, the missing values are removed in the pre-processing stage from the raw behavioural data. Then, the feature extraction stage occurred by obtaining the pre-processed behavioural data. In the process of extraction, the RCNN method is used to extract both spatial and temporal features from the pre-processed data.

#### C. Database

The behavioural templates are stored in the database that are labelled with user ID, where two phases are connected with

database logically as depicted in Fig. 1. The user templates for verification is used in database for enrollment phase and matching templates are stored for final classification for verification phase.

#### D. Matching and Decision Making

In this section, the developed method concludes that whether the end user is either genuine/authorized users or non-genuine/unauthorized users. The user identity's matching score is calculated by comparing templates using matching algorithm. In order to produce a match score, the research study uses the RCNN classifier, which is explained in below section. Therefore, the user as whether authorized user or unauthorized user is decided by using the produced match score. The user can access the resources in smartphones only if the user is authorized, otherwise, the access is denied.

#### E. Design of Classifier

In visual imagery, the most commonly used technique is Residual Convolutional Neural Network or ResNet [19], which is a class in DNNs. The identity of user is classified by using RCNN with learning models for behavioural analysis, which is the major objective of the proposed classifier. The RCNN classifier obtained the training and testing for analysing the human gait behaviour in this work. The feature maps are produced by converting the images with filters in the RCNN approach. In order to receive high-level features from the input data, the next convolution layer received this kind of feature maps. Fig. 2 illustrates the basic architecture of RCNN [20].

The dimensionality of image is reduced and non-linearity is added by using the down sampling operation and non-linearity function between the convolution layers. When the dominant features are predicted in the feature maps, the down sampling operation used the max pooling layer for reducing the dimensionality. The feature maps are vectorized by using the flattening layers, once the NN is initialized. At every output neurons, a number is produced by forwarding flatten input vector into network in the NN that shows how much a certain activity is classified as input vector. In the proposed RCNN techniques, the parameters for mapping function, pooling process and loss function of softmax classifier are used. The following equations will explains the parameters for each function. Fig. 3 shows the working procedure of RCNN classifier.

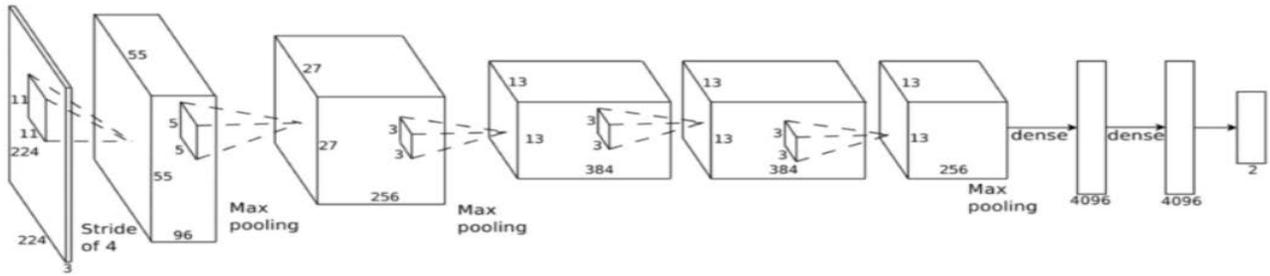


Fig. 2. Basic Architecture of CNN.

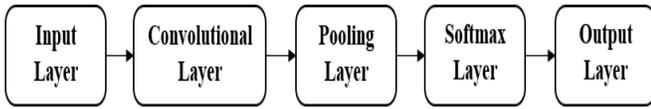


Fig. 3. Working Procedure of RCNN Classifier.

The local information of data is extracted by using convolution layer in RCNN classifier. In addition, the input features are improved by the convolutional operations that minimized the interference of noise. The mathematical equation (4) shows the mapping operation in the process of convolution.

$$x_j^l = f_c \left( \sum_{i \in M_j} x_i^{l-1} \times k_{i,j}^{l-1} \times k + \theta_j^l \right) \quad (4)$$

where, convolutional layer's mapping set is denoted as  $x_j^l$ , feature set in  $(l-1)$  convolutional layer is illustrated as  $x_j^{l-1}$  and convolutional kernel between mapping set and feature set is represented as  $k_{i,j}^l$  in the convolutional layer  $l$ . The bias is described as  $\theta_j^l$  and activation function is depicted as  $f_c$ . During training process, over-fitting is highly reduced by using the pooling process. The mathematical expression (5) describes the pooling process as:

$$x_j^l = f_p \left( \beta_j^l \text{down}(x_j^{l-1}) + \theta_j^l \right) \quad (5)$$

where, downsampling method is denoted as  $\text{down}(\cdot)$  from the layer of  $(l-1)$  to  $l^{\text{th}}$  layer, additive bias is described as  $\theta_j^l$ , multiplicative bias is denoted as  $\beta_j^l$  and the activation function is illustrated as  $f_p(\cdot)$ . There are two types presented in the pooling process, where types include maximum and average pooling. A rasterization layer is formed by arranging the final pooling layer (i.e. matrix features), where fully connected layer is further connected with the matrix features. Eq. (6) stated the node output as:

$$h_j = f_h \left( \sum_{i=0}^{n-1} w_{i,j} x_i - \theta_j \right) \quad (6)$$

where, the input vector connection weight is illustrated as  $w_{i,j}$ , node threshold is denoted as  $\theta_j$  and activation function is described as  $f_h(\cdot)$ . The mathematical equation 7 represents the loss function of softmax classifier.

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=1}^k l \{ y^{(i)} = j \} \log \frac{e^{\theta_j}}{\sum_k e^{\theta_k}} \right] \quad (7)$$

where,  $j^{\text{th}}$  input neuron is denoted as  $e^{\theta_j}$  in the  $l$  layer, all neuron input is described as  $\sum_k e^{\theta_k}$ ,  $j^{\text{th}}$  neuron output is stated as  $\frac{e^{\theta_j}}{\sum_k e^{\theta_k}}$  constant is illustrated as  $e$  and indicator function is depicted as  $l(\cdot)$ . The indicator function's result is one, when the brace value is true and the result of indicator function is zero, when the brace value is false. The rule items are added in  $J(\theta)$  for preventing the local optimum falling. After the rule items are added, the equation (8) shows the softmax classifier's loss function as  $J(\theta)$ .

$$J(\theta) = - \left[ \sum_{i=1}^m \sum_{j=1}^k l \{ y^{(i)} = j \} \log \frac{e^{\theta_j}}{\sum_k e^{\theta_k}} \right] + \frac{\rho}{2} \sum_{i=1}^k \sum_{j=0}^n \theta_{ij}^2 \quad (8)$$

where, weighted term is denoted as  $\frac{\rho}{2} \sum_{i=1}^k \sum_{j=0}^n \theta_{ij}^2$  and the excessive parameters are stabilized by using this weighted term in the training set. The output sum is normalized at the output layer by employing the softmax activation function, where one is added to all numbers at output neuron. The weight at the NN is updated by utilizing the learning function in the training phase and filters at the convolution layers of the RCNN method. In order to update the weights and filters, the activities loss is considered as input by the learning algorithm and propagates the error into the network. Finally, the output of the RCNN classifier predicted the user as either authorized user or unauthorized user from the input data. The next section will explain the validation of the RCNN with other techniques in terms of various parameters.

## V. RESULTS AND DISCUSSION

The dataset description, performance analysis of RCNN and comparative analysis of developed method is explained in this section, where the RCNN method is implemented in 3.0 GHz Intel i5 processor, with 4 GB RAM and 1 TB hard disk. The algorithm is designed by using Python language, where the performance metrics namely sensitivity, precision rate, False Negative Rate (FNR), F1-score and accuracy are used to validate the RCNN with different existing techniques. Initially, the collection of data in this research study is presented as follows.

### A. Database Description

In this research study, three datasets such as CASIA A, B and OU-ISIR dataset are used to validate the performance of RCNN method. The gait video of CASIA-A dataset is collected in an outdoor environment with two alternate days. Four gait sequences are performed by involving 20 subjects in three distinct views as frontal, obliquely and laterally. Therefore, the dataset consists of total 240 gait sequences (*i.e.*  $20 \times 4 \times 3 \times 240$ ) the image resolutions of  $352 \times 240$  that are recorded in 25 frames per second (fps). The average length of every sequence is about 90 frames, where 168 video sequences are used for training process and others for testing in this research study.

A 31 females and 93 males (total of 124) are used in this dataset, where USB cameras are used to record the gait videos for 11 various views in an indoor environment. Hence, this dataset is broadly used for multi-view gait recognition. The frame size of the recorded video is  $320 \times 240$  at 25fps rate, where the difference of angle view direction for each video is 180 that are arranged as 0, 18, ..., 72, 90, ..., 162 and 1800. For multi-view video, gait sequences are recorded with three variations, which includes six video sequences for walk and finally, four video sequences by carrying bags and wearing a coat. The total video sequences is  $10 \times 11 \times 124 = 13,640$  for CASIA-B dataset. The ratio of 70:30 is used for validation process and only 900 views are considered in this research study that consists of 1240 video sequences.

In order to collect the gait videos, 2135 males and 1872 females of total 4007 subjects with 1 to 94 years old ages are used in the OU-ISIR dataset. A gallery and probe are considered as two gait sequences for every subject. Four observation angles includes 550, 650, 750, 850 are used to capture the gait sequences for each and every subject. In this dataset, variations are not provided for walking conditions and when comparing with CASIA-B dataset, the cross-view angle between gallery and probe is small. The performance of various gait recognition techniques are evaluated by using this dataset, because it consists of vast amount of subjects with different ages.

### B. Parameter Evaluation

In this section, the performance metrics are discussed, which is used to test the efficiency of the proposed architecture system and also used to justify the practical developments of proposed RCNN method. There are five metrics namely accuracy, precision, F1-score, sensitivity and FNR are used to

validate the performance of RCNN classifiers. The mathematical expression for this metrics are depicted in the following equations (9), (10), (11) and (12).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (10)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \quad (11)$$

$$F1-Score = \frac{2TP}{2TP + FP + FN} \quad (12)$$

where, True Positive is denoted as TP, True Negative is illustrated as TN, False Positive is described as FP and False Negative is depicted as FN in the above three equations.

### C. Performance Analysis of Proposed Method for CASIA-A Database

In this section, the performance analysis of RCNN is validated with different classifiers includes Linear Support Vector Machine (LSVM), Quadratic-SVM (QSVM), Weighted K-Nearest Neighbour (WKNN) and FEcS-DNN [18] in terms of various parameters for CASIA-A database. Table I provides the validated results of RCNN in terms of accuracy, sensitivity, FNR, precision, and F1-score. Fig. 4 shows the graphical representation for precision and sensitivity, where Fig. 5 presents the validation results of FNR.

TABLE I. EXPERIMENTAL ANALYSIS OF PROPOSED METHOD FOR CASIA-A DATABASE

Methodology	Parameters				
	Sensitivity(%)	Precision(%)	FNR	F1-score(%)	Accuracy(%)
LSVM	97.67	98.0	1.9	97.83	98.1
QSVM	99.00	99.33	0.7	99.16	99.3
WKNN	96.67	96.67	3.6	96.67	96.4
FEcS-DNN	99.33	99.00	0.7	99.16	99.3
RCNN	99.59	99.47	0.2	99.50	99.9

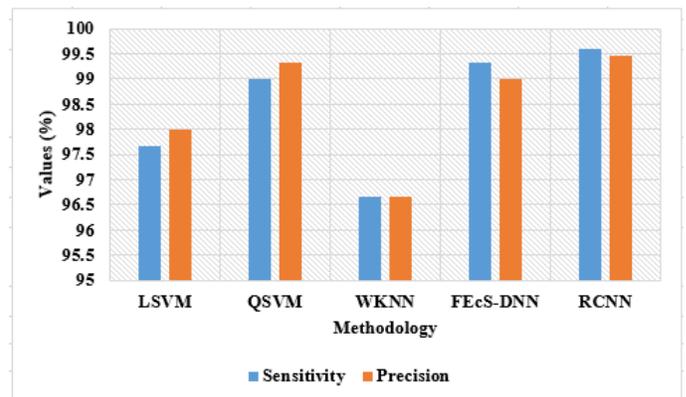


Fig. 4. Graphical Representation of Proposed RCNN in Terms of Sensitivity and Precision.

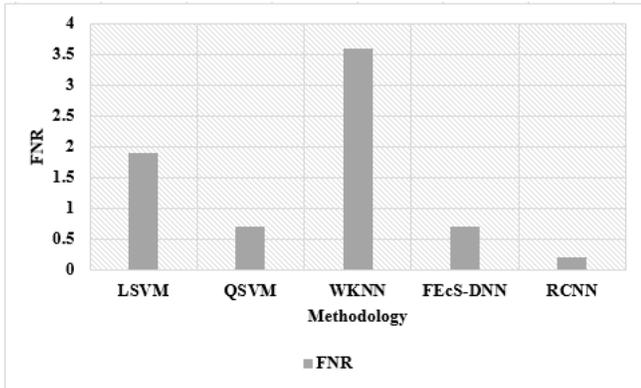


Fig. 5. Validation of Proposed RCNN by means of FNR.

The existing techniques WKNN and LSVM provides poor performance than other techniques, for instance, the WKNN achieved nearly 96% of sensitivity and precision, where LSVM achieved nearly 97.5% of sensitivity and precision. This is because, the WKNN works based on the weight calculation and LSVM's parameter are not effectively assigned. Therefore, FEcS-DNN method extracted spatial and temporal features and increased the performance such as it achieved nearly 99% of sensitivity and precision. However, the DNN method neglect the useful information for gait recognition, where the proposed RCNN method solved the issues of existing techniques. Hence, the proposed RCNN method achieved 99.59% of sensitivity and 99.47% of precision rate.

In the experiments of FNR on CASIA-A dataset, the WKNN method provided poor performance i.e. it obtained 3.6 FNR than other techniques. The existing LSVM obtained 1.9 FNR, where QSVM and FEcS-DNN obtained 0.7 FNR, because the existing techniques are affected by overfitting issues. But, the proposed RCNN method solved the overfitting problems by using the speed learning of residuals and obtained only 0.2 FNR. Fig. 6 illustrates the graphical results of proposed RCNN on the basis of F1-score and accuracy.

From the experimental analysis, it is proved that the proposed method achieved better performance than various classifiers for CASIA-A dataset. Among other techniques, the existing WKNN achieved poor performance, because the method classifies the extracted input based on weights calculations. When compared with FEcS-DNN technique (99.16% F1-score and 99.30% accuracy), the proposed method slightly increased the F1-score (i.e. 99.50%) and accuracy (i.e. 99.90%) values, this is because the Residual in the proposed method reduces the impact of vanishing gradients. The next section will discuss the performance analysis of proposed method for CASIA-B dataset.

#### D. Performance Analysis of Proposed Method for CASIA-B Database

Table II presents the performance analysis of RCNN method with LSVM, QSVM, WKNN and FEcS-DNN [18] for CASIA-B database. Fig. 7 presents the graphical results of

RCNN in terms of precision and sensitivity, where Fig. 8 shows the FNR results of proposed method.

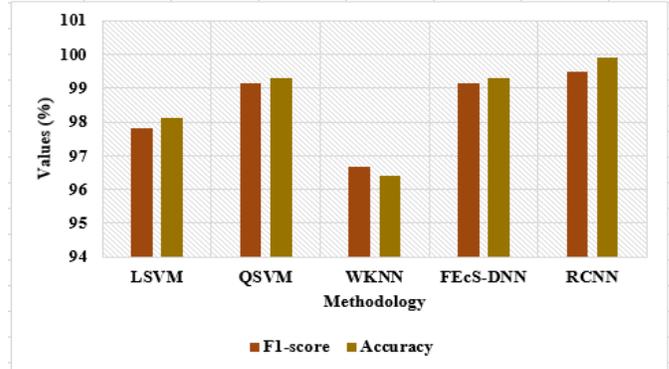


Fig. 6. Analysis of RCNN by means of F1-Score and Accuracy.

TABLE II. EXPERIMENTAL ANALYSIS OF PROPOSED METHOD FOR CASIA-B DATABASE

Methodology	Parameters				
	Sensitivity(%)	Precision(%)	FN R	F1-score(%)	Accuracy(%)
LSVM	71.75	73.00	28.2	72.37	71.8
QSVM	86.25	86.75	13.6	86.50	86.4
WKNN	66.25	66.75	33.7	66.50	66.13
FEcS-DNN	92.00	92.40	7.8	92.20	92.2
RCNN	94.16	95.47	4.1	94.68	97.9

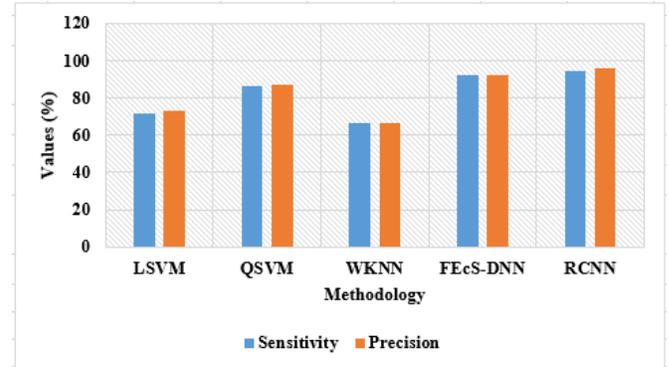


Fig. 7. Graphical Representation of RCNN in terms of Sensitivity and Precision

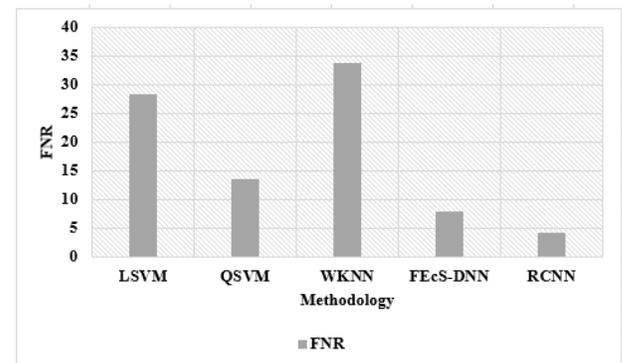


Fig. 8. Analysis of Proposed RCNN by means of FNR.

From the Table II, the experimental analysis shown that the proposed RCNN achieved better performance in terms of sensitivity and precision rate. The existing WKNN technique only provided poor performance than all other techniques, i.e. it achieved only 66% of sensitivity and precision rate. The QSVM and FECS-DNN method achieved nearly 86-92% of sensitivity and precision rate, where the proposed RCNN achieved 94% of sensitivity and 95% of precision rate on CASIA-B dataset.

The WKNN method obtained high FNR rate (i.e. 33.7%), where LSVM and QSVM obtained 28.2 FNR and 13.6 FNR on CASIA-B dataset. When compared with existing FECS-DNN technique, the proposed method obtained less FNR on CASIA-B dataset. However, the results showed that the RCNN method obtained 4.1 FNR due to practical usability issues are occurred during authentication. This leads to unauthorized users as genuine users and access the sensitive information of users in the smartphones. Fig. 9 represents the validated results of RCNN method by means of F1-score and accuracy.

The proposed RCNN method achieved less accuracy (i.e. 97.9%) due to high FNR values on this dataset. However, the proposed method achieved better performance than other existing techniques. For instance, the WKNN and LSVM achieved only nearly 66-72% of accuracy, where FECS-DNN achieved only 92.2% accuracy. The reason behind the less accuracy is that the existing techniques are failed to represent the accurate temporal features and negation of some useful information. In order to achieve better gait recognition, the research study developed the RCNN method and extract the temporal and spatial features. The next section will describe the comparative analysis of RCNN on OU-ISIR dataset.

E. Comparative Analysis of Residual Convolutional Neural Network

In this section, the proposed method is compared with CNN with Chrono-gait image (CNN-CGI), SFN, TFN and STDNN [15] in terms of accuracy for CASIA-B and OU-ISIR dataset. Table III shows the comparative analysis of proposed RCNN by means of accuracy. The graphical representation of accuracy is illustrated in Fig. 10.

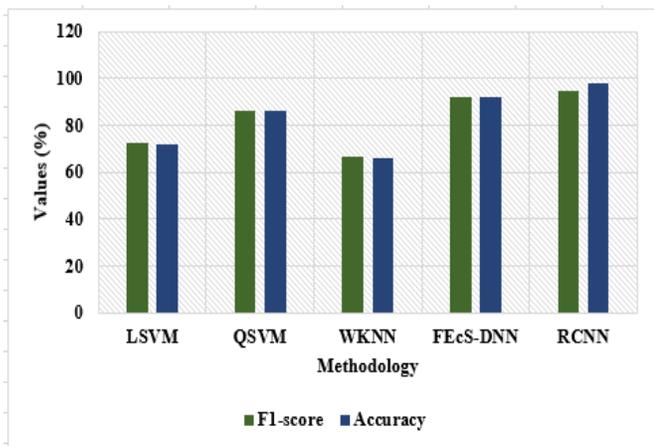


Fig. 9. Performance Analysis of Proposed RCNN in Terms of F1-score and Accuracy.

TABLE III. COMPARATIVE ANALYSIS OF PROPOSED RCNN IN TERMS OF ACCURACY (%)

Methodology	Database	
	CASIA-B	OU-ISIR
CNN-CGI	89.83	91.42
SFN	92.17	92.64
TFN	93.26	93.87
STDNN	94.24	95.67
Proposed RCNN	97.90	98.15

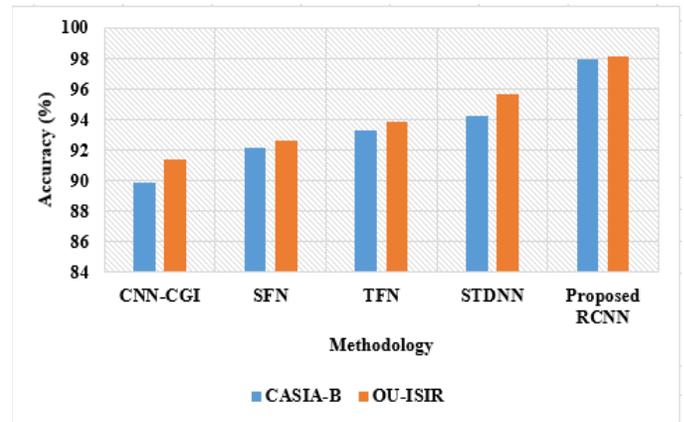


Fig. 10. Comparative Analysis of Proposed RCNN by means of Accuracy.

From the Table III, it is clearly stated that the proposed RCNN method achieved better performance for two datasets such as CASIA-B and OU-ISIR. Initially, the CNN-CGI achieved 91.42% accuracy for OU-ISIR than CASIA-B dataset, because a certain amount of temporal information is carried by CNN. However, the CNN-CGI method failed to represent the temporal features accurately, so the performance is degraded. The existing techniques includes SFN, STDNN and TFN achieved nearly 92%-95% accuracy in both datasets, because low-level edge features and spatial features were extracted. But, those existing techniques are suffered from over-fitting problems that leads to poor accuracy. Therefore, the research study developed the RCNN method, which achieved 97.90% accuracy on CASIA-B dataset and achieved 98.15% accuracy on OU-ISIR dataset. The developed method extract both spatial and temporal features and the overfitting problems are solved by RCNN method.

VI. CONCLUSION

A smartphone can be accessed by either authorized user or unauthorized user, therefore a human gait recognition is developed in this research study. A continuous authentication is important for accessing the smartphone, hence the developed architecture is suitable for applying on any operating system. The RCNN classifier is implemented in this study for improving the security of credential information from the unauthorized user access in smartphones. The experiments are conducted on three datasets such as CASIS-A, B and OU-ISIR dataset in terms of accuracy, sensitivity, precision, FNR and F1-score. The results proved that the RCNN method achieved 99.9% accuracy for CASIA-A, 97.9% accuracy for CASIA-B

and 98.15% for OU-ISIR dataset, where the existing FECS-DNN approach achieved only 99.3% accuracy for CASIA-A and 92.2% for CASIA-B dataset. The experimental results showed that the RCNN technique achieved 4.1% FNR for CASIA-B dataset, because the environmental conditions provide the practical usability issues at the time of authentication. Therefore, the genuine user can't able to access the smartphone at certain instance, because those users are misclassified due to usability issues. In future work, some efficient methodologies must be developed to address the practical usability issues during authentication.

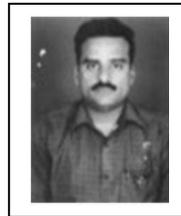
#### REFERENCES

- [1] M. Gomez-Barrero, C. Rathgeb, G. Li, R. Ramachandra, J. Galbally, and C. Busch, "Multi-biometric template protection based on bloom filters", *Information Fusion*, vol.42, pp.37-50, 2018.
- [2] M. M. Al Rahhal, M. L. Mekhalafi, M. Guermoui, E. Othman, B. Lei, and A. Mahmood, "A Dense Phase Descriptor for Human Ear Recognition", *IEEE Access*, vol.6, pp.11883-1187, 2018.
- [3] S. A. Bargal, A. Welles, C. R. Chan, S. Howes, S. Sclaroff, E. Ragan, and C. Gill, "Image-based Ear Biometric Smartphone App for Patient Identification in Field Settings", In: *Proc. of 10th International Conf. On VISAPP*, vol.3, pp.171-179, 2015.
- [4] C. Cintas, M. Quinto-Sánchez, V. Acuña, C. Paschetta, S. De Azevedo, C. C. S. de Cerqueira, and S. Canizales-Quinteros, "Automatic ear detection and feature extraction using geometric morphometrics and convolutional neural networks", *IET Biometrics*, vol.6, pp.211-223, 2016.
- [5] M. A. Murillo-Escobar, C. Cruz-Hernández, F. Abundiz-Pérez, and R. M. López-Gutiérrez, "A robust embedded biometric authentication system based on fingerprint and chaotic encryption", *Expert Systems with Applications*, vol.42, pp.8198-8211, 2015.
- [6] R. Venkatesh, N. U. Maheswari, and S. Jeyanthi, "Multiple Criteria Decision Analysis Based Overlapped Latent Fingerprint Recognition System Using fuzzy Sets", *International Journal of Fuzzy Systems*, vol.20, pp.2016-2042, 2018.
- [7] E. J. Ragan, C. Johnson, J. N. Milton, and C. J. Gill, "Ear biometrics for patient identification in global health: a cross-sectional study to test the feasibility of a simplified algorithm", *BMC research notes*, vol.9, pp.484, 2016.
- [8] N. Geethanjali, and K. Thamaraiselvi, "Feature Level Fusion of Multimodal Biometrics and Two Tier Security in ATM System", *International Journal of Computer Applications*, vol. 70, pp. 14 2013.
- [9] X. Li, Y. Yin, Y. Ning, G. Yang, and L. Pan, "A hybrid biometric identification framework for high security applications", *Frontiers of Computer Science*, vol.9, pp.392-401, 2015.
- [10] X. Wang, and W. Q. Yan, "Cross-view gait recognition through ensemble learning", *Neural Computing and Applications*, pp. 1-13, 2019.
- [11] U. Martinez-Hernandez, and A. A. Dehghani-Sanij, "Adaptive Bayesian inference system for recognition of walking activities and prediction of gait events using wearable sensors". *Neural Networks*, vol. 102, pp. 107-119, 2018.
- [12] J. P. Singh, S. Jain, S. Arora, and U. P. Singh, "Vision-based gait recognition: a survey". *IEEE Access*, vol. 6, pp. 70497-70527, 2018.
- [13] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, "Joint Intensity Transformer Network for Gait Recognition Robust Against Clothing and Carrying Status", *IEEE Transactions on Information Forensics and Security*, vol. 14, pp. 3102-3115, 2019.
- [14] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "On input/output architectures for convolutional neural network-based cross-view gait recognition", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, pp. 2708 – 2719, 2017.
- [15] S. Tong, Y. Fu, X. Yue, and H. Ling, "Multi-view gait recognition based on a spatial-temporal deep neural network". *IEEE Access*, vol. 6, pp. 57583-57596, 2018.
- [16] I. Huitzil, L. Dranca, J. Bernad, and F. Bobillo, "Gait recognition using fuzzy ontologies and Kinect sensor data", *International Journal of Approximate Reasoning*, vol. 113, pp. 354-371, 2019.
- [17] X. Wang, J. Zhang, and W. Q. Yan, "Gait recognition using multichannel convolution neural networks", *Neural Computing and Applications*, pp. 1-11, 2019.
- [18] H. Arshad, M. A. Khan, M. I. Sharif, M. Yasmin, J. M. R. Tavares, Y. D. Zhang, and S. C. Satapathy, "A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition", *Expert Systems*, pp. e12541, 2020.
- [19] H. Gao, B. Cheng, J. Wang, K. Li, J. Zhao, and D. Li, "Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment". *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 4224-4231, 2018.
- [20] A. Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks", *Advances in neural information processing systems*, pp. 1-9, 2012.

#### AUTHORS' PROFILE



Gogineni Krishna Chaitanya research scholar received his Bachelors Degree in Computer Science from Acharya Nagarjuna University and Masters Degree from JNTUK. He is currently pursuing Ph.D degree with Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation, Vaddeswaram, 522502 Andhra Pradesh, India. His research interests include digital forensics, Biometrics, Authentication and Machine Learning.



Dr. K Raja Sekhar received his Ph.D Degree in Computer Science and Engineering from Acharya Nagarjuna University. He is currently a Professor with the of Computer Science and Engineering Koneru Lakshmaiah Education Foundation, Vaddeswaram, 522502 Andhra Pradesh, India. He has published more than 50 articles in journals and Conference proceedings. His research interests include Digital forensics, Biometrics, Network Security and Usable security. He received several

Excellence awards and several best paper awards. He has been on the editorial boards of several journals.

# The Cuckoo Feature Filtration Method for Intrusion Detection (Cuckoo-ID)

Wafa Alsharafat  
Al al-Bayt University, Jordan

**Abstract**—Intrusion Detection Systems (IDSs) play a crucial role in keeping online systems secure from attacks. However, these systems usually face the challenge of needing to handle and analyze a vast volume of data in order to achieve intrusion detection. Feature filtration is a solution that overcomes this challenge by focusing on the characteristic network features that play a significant role in enabling these systems to achieve high detection rates. This paper presents an intelligent cuckoo feature filtration method that is intended to prune away insignificant network features. Then, an IDS (the Cuckoo-ID) is designed in which an eXtended Classifier System (XCS) uses the filtered features for improving the rate of detection of network intrusions. Thus, the main objective of Cuckoo-ID is to maximize the detection rate (DR) and minimize the false alarm rate (FAR). Experiments were then run on the KDDcup'99 dataset to test the intrusion detection (ID) efficiency of the proposed system. The results showed that cuckoo filtration does profoundly raise the ID rate of the entire system. Finally, the DR and FAR of Cuckoo-ID were compared with those of intrusion detection methods that depend on network feature filtration.

**Keywords**—Cuckoo algorithm; feature filtration; intrusion detection; XCS; detection rate

## I. INTRODUCTION

More lights have been spotted recently than before on the networks because of the sensitivity of such an environment. The networks have a vast number of resources and information. Therefore, protective solutions against possible threats and vulnerabilities of mechanisms must be implemented.

Now, an Intrusion Detection System (IDS) is one of the popular mechanisms for protecting network resources against attacks [1]. James Anderson, in 1980, published the first study of IDSs for improving the security levels of networks [2]. The IDS has been defined by Teodoro [3] as being 'a security tool, like other measures such as antivirus software and firewalls proposed for security, to enable that security to become more powerful, in terms of protecting an information and communication system'. Therefore, intrusion detection (ID) aims at achieving intelligent control of the events occurring in a. Researchers in [1,4,5] have explained that the IDSs have two major taxonomies, depending on the source of the data to collect. These are:

- Host-based Intrusion Detection System (HIDS) which are targeted to collect information about activity in a specific host [5].
- Network-based Intrusion Detection System (NIDS) which collect information from the network [5].

Different Artificial Intelligence (AI) techniques including Deep Learning, Cuckoo Algorithm, Fuzzy Logic, Genetic Algorithm, and the Artificial Neural Network have been applied and enhanced to automatically detect possible intrusions [6, 7, 8]. Furthermore, different AI techniques have been adapted to identify the key and most relevant network features that guide and lead the IDS to achieve accurate detection as in Jamali and Jafarzadeh [9]; Lee, Park, and Lee [10]; Abd Eldayem [11]; and Alsharafat [12]. Within this context, feature filtration has the crucial advantage of overcoming the conflict between network features and attacks.

Bearing this in mind, this study was initiated to assess the potential for the Cuckoo Algorithm to serve as a feature filtration method that will reduce the number of the network features needed to detect intrusion; Determine the impact of feature filtration on ID and Compare the level of ID performance of the proposed feature filtration method (Cuckoo-ID) with those of other ID methods (both the ones involving feature filtration and those which do not).

In short, the key contributions of this work are enumerated as follows.

- 1) Summarize a set of interrelated works related to IDS and those involving feature filtration.
- 2) Propose a new feature filtration method for intrusion detection system based on Cuckoo Algorithm.
- 3) Implement an intrusion detection model that applied the Cuckoo algorithm for feature filtration.
- 4) Evaluate the performance of the proposed feature filtration method and compare the result with methods involving feature filtration and those which do not.

The remainder of this paper is organized as follows. Section 2 reviews related to previous works. Section 3 describes the benchmark dataset that is widely used in testing the levels of performance of IDSs. Section 4 clarifies the proposed method for improving ID. Section 5 discusses the evaluation of the performance of the proposed ID method. Lastly, Section 6 contains a discussion of the experimental results and concludes them.

## II. RELATED WORKS

A review of the published literature reveals that different researchers have employed various AI techniques for ID [6 – 20]. These techniques include Artificial Neural Networks (ANNs), the Cuckoo Algorithm, Fuzzy Logic, metaheuristic algorithm, Random forest, cuttlefish algorithm, and the Genetic Algorithm (GA), besides other new trends to be

implemented in detecting network attacks, especially in the environment of the Cloud, which is regarded by intruders as being a preferred target with regard to exploiting its weak points.

Some researchers keenly delved into a search for key factors that have a positive influence on DR. In this regard, the search was guided (in some studies) by the investigation of the effect of network feature filtration on the proposed IDS [6–20]. For feature filtration, Alsharafat in 2010 [12] has developed an ID model called ANN-XCS that proceeds in two phases. The first phase concerned about feature filtration by applying ANN. Then, the filtered features will be considered in the entire work of the second phase, which applied an Extended Classifier System (XCS) for the purpose of intrusion detection. For enhancement, XCS applies a set of modifications to GA for the breeding classifier pool. As a result, the DR of ANN-XCS was 98.01%, and the concomitant FAR was 0.9%.

Alzboon, Alkhalid, and Alsharafat [13] proposed an IDS in 2017 that was based on using the Cuckoo search method integrated with GA as a classifier generator within XCS. In addition, the network features were filtered in this system using Fuzzy Clustering by the Local Approximation Membership (FLAME) method, FLAME-XCS as an abbreviation, which reduced the number of network features to track from 41 to 20 features. As a result, the DR reached 99.9% while the FAR reached 0.005%, corresponding to an outstanding ID efficiency.

A Learning Automata Intrusion Detection System (LA-IDS) was developed in 2017 by Jamali and Jafarzadeh [9]. This model applies a seven-level hierarchical structure in which each level is responsible for processing one of the network features. By reducing the number of features to trace from 41 to 7, this system (LA-IDS) proved to have a DR of 98.9% and a FAR of 1.3%. In the same year, Lee, Park, and Lee [10] focused on feature selection, which represents the first phase of the proposed work, which aimed at constructing a subset of features using a sequential, forward-floating search (SFFS) instead of the method used by Jamali and Jafarzadeh [9]. The second phase of the proposed system, however, entailed the construction of a classification model based on using a random forest classifier (RFC) to select a feature subset. Lee in his work achieved 99.9 as DR and 0.1 as FAR. In another study, an optimization technique, the Naïve Bayes (NB) classifier, was applied by Abd-Eldayem in 2014. Experimentation revealed that this optimization technique had a DR of about 99% and a FAR of nearly 1%.

In 2018, Shone and his colleagues [6] proposed a novel deep-learning (DL) classification model that was constructed using stacked, non-symmetric, deep auto-encoder a multiple, hidden-layer, unsupervised, neural network-based, feature extraction algorithm. This model had a DR of 97.9% and a FAR of 2.1%.

Also, Yan and Han [8] proposed an IDS which would use the stacked sparse auto-encoder (SSAE) for extracting the

features that have a significant influence on intrusion behavior. Then, a different classifier was employed by using low-dimensional sparse features. Performance evaluation revealed that this model had a DR of 99.01% and a FAR of 0.13%.

In 2019, Sara and her colleagues [18] proposes a wrapper method for feature selection based on linear correlation coefficient (FGLCC) algorithm and cuttlefish algorithm (CFA), and the Decision tree act as a classifier in IDS. The results obtained DR equals 95.23% with a low FAR of 1.65%. Also, Boonyopakornin 2019 has applied Fuzzy logic and association with genetic network programming (GNP), FL-GNP as an abbreviation, for feature selection to create an associated rule to detect attacks [19]. Thus 24 features used instead of 41 features which produced an IDS with a 94.8 detection rate.

Convolutional neural network (CNN) has been applied by Xiao in 2019 [20] in detecting network attacks which called (CNN-IDS). The auto-encoder (AE) as a nonlinear dimension reduction technique has been used to reduce features. The overall DR and FAR were 93.0 and 0.005, respectively.

The pigeon inspired optimizer algorithm was proposed [16] by utilizing the selection process of network features. A new methodology used to binarize a continuously metaheuristic algorithm. Also, the proposed work, externally, compared with the sigmoid function. As a result, the accuracy of detection was 0.947. While Aishwarya and his colleagues [17] examine the efficiency of J48, Naive Bayes (NB) and Random forest (RF) as a classification models. The RF reached 99.9 and 0.004 as a DR and FAR, respectively.

### III. THE KDD'99 DATASET

For assessing the performance of the proposed Cuckoo-ID system, there was a need to run it on actual data. For this purpose, the researcher utilized the KDD'99[21] dataset, which is (so far as the author knows) the most appropriate dataset for this purpose. KDD'99 is a benchmark dataset that is frequently used by different researchers to evaluate the levels of performance of their proposed IDSs.

The origin of KDD'99 is DARPA, which produced it in MIT Lincoln Labs. This dataset is a standard dataset that is commonly used to evaluate results in this line of research. A standard set of data needs to be audited, since it includes a huge number of intrusion records that have been simulated in a military network Laboratories [21]. In this respect, the analysis of the KDD'99 provides useful information for the expansion of IDSs. The data in this dataset are classified into normal records and attack records, and they involve mining rules. This dataset contains 41 features. Hence, one of the objectives of the feature filtration method is its ability to extract the most relevant set of features. In this context, different algorithms can be applied to select the relevant features from the KDD'99 dataset, as in Jamali and Jafarzadeh [9]; Lee, Park, and Lee [10]; Abd-Eldayem[11]; and Alsharafat [12].

#### IV. PROPOSED WORK

Here, a new IDS will be presented by adapting a Cuckoo Algorithm to include feature filtration, which is due to working overall the XCS steps. The Cuckoo Algorithm is used in the present study as a feature filtration method to enhance the detection of the network attacks. The motivation for performing this work emerged from the high potential of this particular algorithm for feature filtration as had been confirmed in various previous studies such as Gandomi, Yang, and Alavi [22]; Yang and Deb [23]; and Yang, Deb, Karamanoglu, and Xingshi [24]. In 2009, Yang and Deb [23] presented this meta-heuristic algorithm—namely, the Cuckoo Algorithm—which they based on a parasitic-like nesting behavior for the brooding of some cuckoo species. This algorithm works coherently and effectively with the Lévy flights (LFs). Thus, the general framework of the proposed work consists of two key processes; feature filtration using a cuckoo algorithm which specifies the key features that are considered next process which concerns about intrusion detection using XCS as illustrated in Fig. 1.

##### A. Cuckoo Feature Filtration

One of the critical decisions in any IDS is the selection of the critical network features. Thus, for enhancing the detection of the proposed method, the cuckoo algorithm was employed as it plays an important role in distinguishing the features that will seriously affect the performance of the system. This algorithm increases the DR and reduces the associated FAR. In addition to this, the Cuckoo Algorithm works effectively with Lévy Flight (LF) and generates new solutions around the most suitable solution obtained so far, which will speed up the process of local search [22–26].

The solution-generation process of the cuckoo search algorithm depends on three rules:

- Each cuckoo randomly chooses a nest in which to lay one egg at a time.
- The elitism nests with high-quality eggs will be transferred to the next generation.
- The number of host nests is fixed, where each host can detect a strange egg with a probability of  $P_a \in [0,1]$ . As

well, the host bird can discard the egg or leave the nest to construct a new nest in another place [23]. For such enhancements, certain characteristics of the current solution must be modified in order to breed a new solution.

The Cuckoo Algorithm for feature filtration can be summarized in the follow Pseudo-code as in Fig. 2 [23,26]:

By incorporating the LF in the Cuckoo Algorithm, the LF will guarantee to find new solutions in the region of the best-stated solution, which will boost the speed of the local search space [23,25,26] and guarantee that the system will not be trapped in a local optimum [22,23].

##### B. Extended Classifier System (XCS)

In 1995, Wilson has introduced The XCS[27], which is a real (integer or binary) code one of the Learning Classifier System categories. It is a fitness classifier system in which every rule depends on a fitness value for the evaluation of this system. As such, the XCSs can be classified as a rule-based classifier system. Each rule, or classifier, is constructed from two parts:

- 1) The Condition Part: This part is encoded using the step size parameter ( $\alpha$ ). So, a binary code is applied by using the notion of  $\{0, 1, \#\}$ , where the symbol # refers to a non-significant value of a feature.
- 2) The Action Part: This part represents the result of the selected rule that will be fired at the environment. The result can be either a normal record or an attack.

Thus, in all XCS components, the filtered features produced from applying the cuckoo algorithm will be considered in condition part in every rule(classifier) through population set [P], matching set [M], prediction set and action set [A].

The XCS concentrated on two main operations: Reinforcement Learning (RL) and GA breeding. RL is concerned with gaining criticism from the external environment about the fired result. It is also concerned with taking advantage of the selected rule in order to be useful in similar situations [28,29]. On the other hand, the GA is used for breeding new rules.

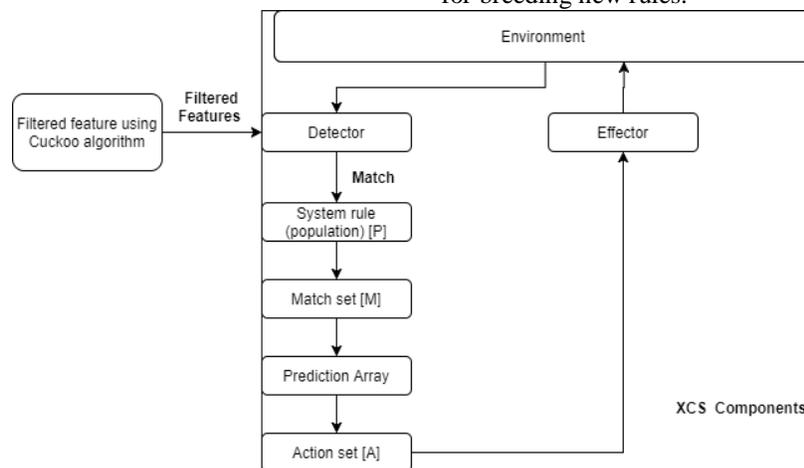


Fig. 1. The Proposed Work; Cuckoo Algorithm and XCS.

```
Objective function: F = DR; // for maximization detection rate
Generate n host nests; // nest contains a set of network features.
Initialize generation numbers(g), Discovery probability (Pa)and Step size(α)

While ((not Maximum Generation) or termination criterion)
    New cuckoo generated randomly and replace it in a nest (i) by performing
        Lévy Flights; // cuckoo egg represent a network feature
    Evaluate cuckoo fitness; F (i) //
    Choose a nest among n (j) randomly;
        if ( Fi <Fj),
            Replace j by the new solution;

        end if
    Abandon a strange cuckoo egg with a probability (Pa) and new nest is built;
    Keep the elitism nests;
    Rank the nests and find the current best nest;
    Pass the elitism nest to the next generation;
end while
```

Fig. 2. Cuckoo Algorithm Pseudo-Code for a Feature Filtration.

The main components of the XCSs are as follows [27–30]:

- **Detector:** The detector is an input gateway that considers the environment as being a traffic record to be converted into a binary code. Here, the filtered feature will be encoded using binary code while other features replaced by sign (#) to represent irrelevant features in detection.
- **Match set [M]:** The match set is a repository of the rules that are compiled after performing matching between system rules and records (population of rules [P]) via the detector according to the conditional part, regardless of the action result.
- **Prediction Array:** The prediction array is an array created from [M] according to the average weighted fitness of the prediction rules employed in an action set [A].
- **Action set [A]:** All the rules in [M] that support a specific action will be placed in an [A].
- **Effectors:** The effectors are output gateways for a rule-selected form [A]. They are intended to determine whether the traffic record is a normal record or an attack.

## V. PERFORMANCE EVALUATION

For evaluating the performance of the proposed IDS and comparing it with levels of performance of comparable IDSs, two performance evaluation criteria were employed; the DR and the FAR.

- **Detection rate (DR)**

The detection rate (DR) has been defined as ‘the ratio between the number of correctly detected attacks and the total number of attacks’ [31]. This ratio is calculated as a percentage using Equation (1) [13,14]:

$$DR = ((TP)/(TP + FP)) * 100 \% \quad (1)$$

- **False alarm rate (FAR)**

The false alarm rate (FAR) has been defined as being ‘the number of “normal” patterns classified as attacks (False Positive) divided by the total number of “normal” patterns’ [31]. It is usually expressed as percentage and can be estimated using Equation (2) [13,14]:

$$FAR = ( (FP) / (FP + TN)) * 100 \% \quad (2)$$

The definition of the entire parameters in equation (1) and (2) are listed in Table I.

TABLE I. DEFINITION OF THE PARAMETERS OF EQUATIONS (1) AND (2)

Parameter	Definition
True Positive (TP)	Attack records correctly classified as an attack.
False Positive (FP)	Normal record that is inaccurately classified as an attack.
True Negative (TN)	Normal record that is correctly classified as normal.
False Negative (FN)	Attack record that is inaccurately classified as normal.

VI. EXPERIMENTAL RESULTS AND CONCLUSION

One of the aims of the IDS proposed in this study is the enhancement of feature filtration in the ID process. Feature filtration aims at selecting the features that have a crucial role in the detection of each type of attack. Thus, the Cuckoo Algorithm was employed to achieve this purpose.

A comparison was held between different IDSs employing feature selection methods enhancing the efficiency of ID in terms of the DR and FAR. Comparisons in terms of the DR and FAR are shown in Table II, whereas comparisons in terms of the DR alone are illustrated by Fig. 3. It should be underscored first that the number of features identified by the different filtration methods as crucial to ID was different, varying from 7 to 41 features (Table II and Fig. 1).

In terms of the DR, it can be noticed that the IDS proposed in this study (Cuckoo-ID) has an equal performance to that LA-ID developed by Jamali [9] and a somewhat better performance (i.e. higher DR) than the ANN-XCS suggested by Alsharafat [12]. In the meantime, SFFS[10], BN in Abd-Eldayem [11], and FLAME-XCS [13] slightly outperform the system proposed in the current study (Table II and Fig. 1).

In terms of the FAR, it is found (Table II) that the IDS suggested in the present study ranks third in performance next to the system developed by Xiao[20], which had FARs of 0.09% and 0.005%, respectively (Table II and Fig. 3). The other systems included in the comparison had lower performance (that is, higher FARs) than the IDS proposed by the present study (Table II).

An issue that is worth highlighting in Fig. 3 is the effect of the number of selected features on the DR, which tends to decrease as the number of features is increased (though decreasing inconsistently). As far as the IDSs compared in Fig. 3 are concerned, it can be assumed that several features in the range of 13–20 can yield the best results. Lower DRs are likely to be obtained if the number of the employed features is lower than 13 or is higher than 20. However, this issue merits further investigation, especially since the present study used 19 features but still obtained a slightly lower DR than SFFS [10], who used 10 features.

TABLE II. COMPARISON BETWEEN DIFFERENT FEATURE SELECTION METHODS IN FILTRATION EFFICIENCY IN TERMS OF THE DR AND FAR

Feature Selection method	No. of Features	DR (%)	FAR (%)
CNN-ID [20]	< 41	93	0.005
FL-GNP [19]	24	94.8	NA
FGLCC [18]	10	95.23	1.63
ANN-XCS [12]	< 41	98.01	0.09
Cuckoo-ID	19	98.9	0.09
LA-IDS [ 9 ]	7	98.9	1.3
SSAE [ 7 , 8 ]	<41	99.01	0.13
NB [11]	13	99.03	1
FLAME-XCS [13]	20	99.9	0.005
SFFS [10]	10	99.9	0.1

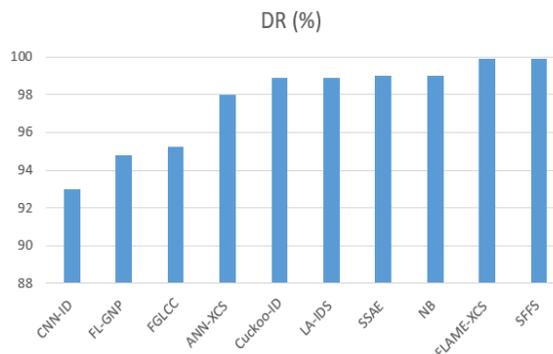


Fig. 3. A Comparison in Detection Rates (DRs) between Cuckoo-ID and IDSs that Filtered Features of the KDD'99 Dataset.

Based on her reading of Fig. 3, as discussed in the preceding paragraph, the researcher suggests using three or four sets of features in future research. These sets contain varying numbers of features in fixed combinations like the first 10 features, the first 20 features, the first 30 features, and then the whole number of features (41). Other combinations of feature sets with almost the same number of features but which involve overlaps in features are also suggested. These include features 1–12, 10–22, 20–32, and 30–41. This is due, on the one hand, to standardize the ID and performance testing processes and the comparisons between the various proposed IDSs; and, on the other hand, to specify the effect of the number of features included in performance testing on the performance of the proposed IDS itself. The researcher thinks that doing so will generate new, valuable knowledge that will contribute to the calibration of the IDS performance evaluation process and to the refinement of the ID process itself.

Given these findings (Table II and Fig. 3), the researcher concludes that her proposed IDS marks a slight improvement over one of the previously-developed IDSs called FLAME-XCS[13] in terms of the DR and varying improvements over multiple previously-developed IDSs in terms of the FAR as SFFS, LA-ID, NB, FGLCC and [8–11,18]. However, these results should be interpreted with caution for two reasons. First, as can be seen in Table II, the different researchers developed their IDSs using datasets with different numbers of features, varying from 7 to 41. Second, there is no optimal set of feature that can be applied in all these studies. Viewed from another angle, the differences between studies in the numbers of features employed in developing their suggested IDSs and evaluating their levels of performance highlights the need for standardization of the number and the type of features to use in such evaluations.

A review of the literature uncovered that there are cases when IDSs were developed using all the features (41) in the KDD'99 dataset. A list of this group of studies which the researcher knew about is given in Table III. The focus of these studies was the development of IDSs that will ensure the enhancement of the ID process in terms of the DR and FAR rather than feature filtration. Besides, a graphical comparison between these studies in the DRs associated with each of them is provided by Fig. 4. Both in Table III and Fig. 4, the researcher included her own proposed IDS (Cuckoo-ID) for comparison.

TABLE III. A COMPARISON IN PERFORMANCE BETWEEN THE PROPOSED CUCKOO-IDS AND IDSs WITH NO FEATURE SELECTION IN TERMS OF DR AND FAR

IDS	DR (%)	FAR (%)
SCDNN[33]	92.23	7.9
TLMD 4 [34]	93.11	0.761
DENDRON [35]	95.97	1.08
CNN+LSTM[36]	96.96	0.2
DEEP [6]	97.9	2.1
RF [17]	98.7	NA
Cuckoo-ID	98.9	0.09
LA-GRU[7]	98.9	0.134
ECOC[32]	99	NA
Stacked [8]	99.01	0.13

Table III and Fig. 4 points out that, in terms of the DR, Cuckoo-ID and LA-GRU[7], which both have the same DR (98.9%), rank third in performance next to the Stacked method [9] whose DR was 99.01% and ECOC method [32] whose DR was 99.0%. Again, this result supports that the researcher’s own proposed IDS has noticeably good performance and is promising, owing to the fact that it performs better in terms of the DR than several previous IDSs have done (Table III).

In other respects, the comparison between the IDSs listed in Table III in terms of the FAR brings to notice that Cuckoo-ID placed in the second-lowest FAR (0.09%). Thus, the results of the performance comparison (summarized by Table III, and depicted in Fig. 5) reinforce the researcher’s former conclusion (drawn from Table II and Fig. 3, Fig. 4 and Fig. 5) that the IDS proposed in this study has profoundly good performance and is quite promising as an IDS that could replace the IDSs included in the comparisons in this paper (Table II and Table III).

In other respects, the comparison between the IDSs listed in Table II and those listed in Table III underlines that, in general, the IDSs employing feature selection methods have superior performance, in terms of the DR, than do those systems which use all of the features in the KDD’99 dataset without performing feature selection. The same holds almost as true with regard to the FAR in Fig. 4, where the IDSs employing feature selection methods (Table II) generally have much lower FARs than do those IDSs which do not apply feature selection (Table III). Therefore, the researcher recommends the use of IDSs with feature selection.

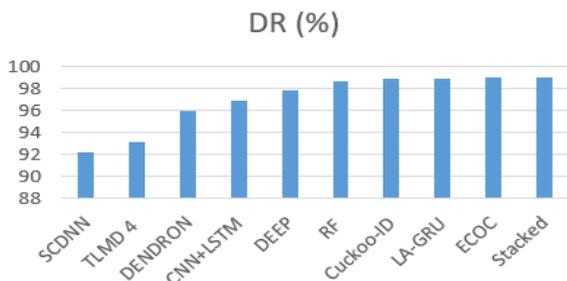


Fig. 4. A Comparison in Detection Rates (DRs) between Cuckoo-ID and IDSs that used All Features of the KDD’99 Dataset.

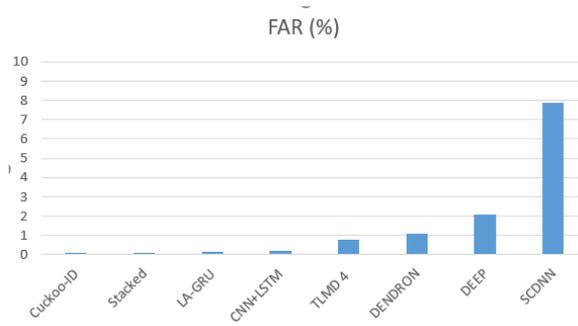


Fig. 5. A Comparison of FAR between Cuckoo-ID and IDSs that used All Features of the KDD’99 Dataset.

Fig. 3 and Fig. 4 lead the researcher to the conclusion that there is no optimal set of features for detecting network intrusions. This is because the various IDSs cited and compared in this paper (Table II and Table III, as well as Fig. 3 and Fig. 4) used different numbers of features. Accordingly, the features themselves used in these tests must be different, though there maybe common features. This is an issue that, though it cannot be negated, cannot be settled out.

The Cuckoo Algorithm was employed for feature filtration in the IDS proposed in this study. It reduced the number of features from the original 41 features down to 19 features and, in consequence, contributed to the development of a highly efficient IDS that has a high DR and a low FAR.

In view of the study’s findings, the researcher maintains that it will be quite interesting for future research to develop IDSs that will produce still higher DRs for all types of attacks. Moreover, different evolutionary algorithms are already known. The use of such algorithms can improve the DR of the entire system even if sometimes using all 41 features of the KDD’99 dataset without filtration. Furthermore, another type of dataset will be examined as UNSW-NB 15 to explore new types of attacks and to test the capability of Cuckoo-ID to detect new type of attacks that not included in KDD’99.

REFERENCES

- [1] Kumar DA, Venugopalan SR. Intrusion detection system: A review. *Int J Adv Res Comput Sci.* 2017;8(8):356–370.
- [2] Bruneau G. The history and evolution of intrusion detection. SANS Institute Reading Room whitepaper [updated 2001 Oct 13]. Available from: <https://www.sans.org/reading-room/whitepapers/detection/paper/344>
- [3] García-Teodoro P, Díaz-Verdejo J, Maciá-Fernández G, et al. Anomaly-based network intrusion detection: Techniques, systems and challenge. *Sci Direct* 2009;28(1–2):18–28.
- [4] Wikimedia.Foundation. Intrusion detection system [cited 2009]. Available from: <http://en.wikipedia.org/wiki/Intrusion-detection>
- [5] McHugh, J. Intrusion and intrusion detection. *Int J Inf Secur.* 2001;1:14–35.
- [6] Shone N, Ngoc TN, Phai VD, et al. A deep learning approach to network intrusion detection. *IEEE Trans Emerg Topics Comput Intell.* 2018;2(1):41–50.
- [7] Yan B, Han, G. LA-GRU: Building combined intrusion detection model based on imbalanced learning and gated recurrent unit neural network. *Secur Commun Netw.* 2018a;1:1–13.
- [8] Yan B, Han G. Effective feature extraction via stacked sparse auto encoder to improve intrusion detection system. *IEEE Access* 2018 b;6:41238–41248.

- [9] Jamali S, Jafarzadeh P. An intelligent intrusion detection system by using hierarchically structured learning automata. *Neural Comput Appl*. 2017;28:1001–1008.
- [10] Lee J, Park D, Lee C. Feature selection algorithm for intrusions detection system using sequential forward search and random forest classifier. *KSII Trans Internet Inf Syst*. 2017;11(10):5132–5148.
- [11] Abd Eldayem MM. A proposed HTTP service based IDS. *Egypt Inf J*. 2014;15(1):13–24.
- [12] Alsharafat W. Applying artificial neural network and extended classifier system for network intrusion detection. *Int Arab J Inf Technol*. 2013;10(3):230–238.
- [13] Alzboon K, Alkhaldy J, Alsharafat W. Intrusion detection model based on clustering algorithm FLAME and cuckoo search selection in genetic algorithm in XCS [master's thesis]. Jordan: Al al-Bayt University; 2017.
- [14] Saeed K, Karim F. A novel classification method using hybridization of fuzzy clustering and neural networks for intrusion detection. *Int J Modern Educ Comput Sci*. 2014;11:11–24.
- [15] Biswajit P, Olugbenga O, Priyanka M. Training of intelligent intrusion detection system using neuro fuzzy. 15th Institute of Electrical and Electronics Engineers (IEEE)/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). 2014 Jun 30–Jul 2; Las Vegas, NV, USA.
- [16] Alzazzam H, Sharieh A, Sabri Kh, A Feature Selection Algorithm for Intrusion Detection System Based on Pigeon Inspired Optimizer. *Expert Systems With Applications*, <https://doi.org/10.1016/j.eswa.2020.113249>, 2020.
- [17] Aishwarya CH, Venkateswaran N, Supriya T, et al. Intrusion detection system using KDD Cup 99 dataset. *Int J Innov Technol Explor Eng*. 2020;9(4):3169–3171.
- [18] Mohammadi S, Mirvaziri H, Ghazizadeh-Ahsaee M, et al. Cyber intrusion detection by combined feature selection algorithm. *J Inf Secur Appl*. 2019;44:80–88.
- [19] Boonyopakorn P. The optimization and enhancement of network intrusion detection through fuzzy association rules. 2019 6th International Conference on Technical Education (ICTechEd6). 2019 March 19–20; Bangkok, Thailand.
- [20] Xiao Y, Xing Ch, Zhang T. An intrusion detection model based on feature reduction and convolutional neural networks. *IEEE Access*. 2019;7:42210–42219.
- [21] University of California, Irvine. KDD Cup 1999 Data [data set]. Third International Knowledge Discovery and Data Mining Tools Competition, held in conjunction with KDD-99 The Fifth International Conference on Knowledge Discovery and Data Mining [updated 1999 October 29]. Available from: <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>
- [22] Gandomi AH, Yang XS, Alavi AH. Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems. *Eng Comput*. 2013;29:17–35.
- [23] Yang XS, Deb S. Cuckoo search via Lévy flights. *Proceedings of the World Congress on Nature & Biologically Inspired Computing*. 2009 Dec 9–11; Coimbatore, India.
- [24] Yang XS, Deb S, Karamanoglu M, et al. (2012). Cuckoo search for business optimization applications. *National Conference on Computing and Communication Systems*. 2012 Nov 21–22; West Bengal, India.
- [25] Yildiz AR. Cuckoo search algorithm for the selection of optimal machining parameters in milling operations. *Int J Adv Manuf Technol*. 2013;64:55–61.
- [26] Santillan JH, Tapucar S, Manliguez C, et al. (2018). Cuckoo search via Lévy flights for the capacitated vehicle routing problem. *J Ind Eng Int*. 2018;14:293–304.
- [27] Wilson SW. Classifier fitness based on accuracy. *Evol Comput*. 1995;3(2):149–175.
- [28] Lanzi PL. Learning classifier systems: Then and now. *Evol Intell*. 2008;1(1):63–82.
- [29] Holmes JH, Lanzi PL, Stolzmann W, et al. Learning classifier systems: New models, successful applications. *Inf Process Lett*. 2002;82(1):23–30.
- [30] Bull L, Kovacs T, editors. *Foundations of learning classifier systems*. Berlin: Springer Science & Business Media; 2005.
- [31] Elhamahmy ME, Elmahdy HN, Saroit IA. A new approach for evaluating intrusion detection system. *CiiT International Journal of Artificial Intelligent Systems and Machine Learning*. 2010;2(11):290–298.
- [32] Zare, P. Intrusion detection system based on combination of optimized genetic and firefly algorithms in cloud computing structure. *Comput Eng Intell Syst*. 2019;10(4):6–11.
- [33] Ma T, Wang F, Cheng J, et al. A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks. *Sensors*. 2016;16(10):1701.
- [34] Yuan Y, Huo L, Hogrefe D. (2017). Two layers multi-class detection method for network intrusion detection system. *IEEE Symposium on Computers and Communications*. 2017 Jul 3–6; Heraklion, Greece.
- [35] Papamartzivanos D, Mármol FG, Kambourakis G. Dendron: Genetic trees driven rule induction for network intrusion detection systems. *Future Gener Comput Syst*. 2018;79(2):558–574.
- [36] Wang W, Sheng Y, Wang J, et al. (2017). HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection. *IEEE Access*. 2017;6:1792–1806.

# Segmentation of Fuzzy Enhanced Mammogram Mass Images by using K-Mean Clustering and Region Growing

Nidhi Singh<sup>1</sup>, S. Veenadhari<sup>2</sup>  
Dept. Name of CSE  
Rabindranath Tagore University  
Bhopal, India

**Abstract**—Providing intention to encourage radiologist's appraisal for distinguishing proof or order of mammogram images, different methods were suggested by specialists since past two decades. By means of this technical paper, we propose segmentation on advanced mammogram imaging with k-means clustering and locale developing systems tending to support specialists or radiologists to figure out cancerous areas with computer-aided techniques. The suggested task is further classified within two stages: Applied/implemented pre-processing, at primary stage. With the pre-processing stage, we carried a median filter to expel undesirable salt and pepper clamor. Further, we apply fuzzy intensification operator (INT) to upgrade the distinction of intake images. During subsequent stage, improved fuzzy imaging conduces as input for k-mean clustering. Secondly, the locale developing technique is employed with previously generated clustered imagery to partition mammogram into homogeneous areas indicated through force from pixels. With the end goal of the experiment, we utilized the smaller than normal MAIS dataset. The experiment's end result shows that proposed strategy accomplishes higher precision.

**Keywords**—INT operator; feature extraction; k-mean clustering; mammogram; median filter; segmentation

## I. INTRODUCTION

In today's time, Cancer (tumor) is major fatal diseases precisely made of widely different related ailments. In every single category about cancer expansion, body somatic (cell) starts to separate repeatedly additionally disseminate amongst encompassing tissue. Similarly, a breast carcinoma emerges from breast organ tissues. Breast comprises of billions of infinitesimal cells. These somatic would begin expanding compulsively which causes cancer of breasts. Cancer of breast growth may be partitioned into two sorts: Ductal Carcinomas and Lobular Carcinomas. Ductal carcinomas are highly prominent cancers that commences in mammary duct whereas lobular carcinomas are particularly ailments that grows around lobes [1]. Prior, the analysis and treatment end up with being ruinous without productive methods. At every stage of carcinoma, the demise rate including dynamism from this disease raises. To decrease the deaths rate and limit the dynamism elicits need for earlier breast cancer identification strategies. So, automated computerized detection is unavoidable. There exists none absolute reasons available for breast carcinoma, and we might see these as causative factors only. They were going to be hereditary or natural. Hereditary

factors incorporate family ancestry, individual wellbeing history, menstrual and reproductive history, dense breast tissues, certain genome changes, age, sex etc. The ecological factors incorporate corpulence, under stellar eating ruts, liquor utilization, radiation, less physical activity, etc. [2]. The elementary factors of breast cancers are forming lumps. Reason for this is tiny sedimentation of calcium called micro calcification and tumors called circumscribed mass such tumors are often benign and not malignant. The benign tumors are generally non-aggressive and non-harmful. It does not disseminate to another body parts [3]. There are several distinctive imaging systems for earlier prediction of breast cancer. These incorporate MRI, X-Ray imaging, ultrasound imaging, computerized mammography, screening etc. The computerized mammograms are widely utilizes currently on account including favorable circumstances over others. X-Ray imaging is typically used for the benefit of discovering indications of carcinoma while utilizing mammograms that generally examines the problem. A mammogram uses X-rays to make breasts imaging [4]. Earlier there exist film mammograms with which images being stored on films, today computerized mammography are generally used that captures and stores straightforwardly on digital computer and every single corners and niches are visible for simple detection. In breasts ultrasound [5], the sound waves are used to form the images. However, currently this is not used since that's done with handheld devices. It would facilitate false positives as well as false negatives, if in this specific instance the person operating is not good grasp or skilled and thereby the quality about the images will differ. Point wise identification of the images gets done by applying feature extracting and texture extraction methods [6]. This alone open other research area, as broad assortment of strategies employed for segmentations, feature extraction, enhancements, done mostly by wavelet techniques [7], clustering using GLCM matrix [8] etc. specifically depicted plainly during relevant works. Therefore a definitive point from this survey is to furnish distinctive improvements, detection and classification approaches for quick recognizing of breast carcinoma.

Several authors have published research papers for breast regions segmentation based on difference in density. For instance, Saidin et al. [9] proposed technique to perform breast segments into four regions: backgrounds, skin-air boundaries, fatty boundaries and pectoral muscles boundaries.

Karssemeijer [10] used approach for mammograms subdivisions towards three separate regions: breast tissues, pectoral muscles and backgrounds. Adel et al. [11] discuss way for breast regions segmentation with three distinct areas: pectoral muscles, fatty and fibro glandular regions by using Bayesian techniques with the adaption of Markov random field for region detection of various tissues on mammograms. El-Zaart et al. [12] provides mammograms segmented images with three regions, which are fibro glandular disc, breast regions and backgrounds. Camilus et al. [13] intend a technique for graph cuts computations seeking the pectoral muscles automatically.

The intended work is divided within two phases. During its primary phase, we applied pre-processing. To be a part of pre-processing step, we had utilized median filter to expel undesirable salt and pepper commotion. We apply fuzzy intensification operator (INT) to strengthen the differentiation of information picture. In the subsequent level, enhanced effective fuzzy images used as input for k-means clustering. Besides, the algorithms for region growing should apply to earlier produce clustered images to divide mammograms within homogeneous parts (regions) in accordance with the relevant intensity of the pixels. Intending to the experiment we used the mini MAIS datasets. The trial result indicates this intended strategy accomplishes higher precision.

The setup of this proposed paper is arranged as follows: Introduction is provided in Section I; Section II discusses substances/material and methods; Section III portrays the proposing (suggested) method; Section IV include outcomes (results) and discussion; conclusion and future work are provided with Section V; we end by references in Section VI.

## II. MATERIAL AND METHOD

### A. Database

To test out proposed methodology we have use mini-MIAS dataset [14]. This dataset consist of 322 mammogram image. These images are kind of three separate classes. Among 322 images, 106 images belong to Fatty (F) class, another 104 images belongs to Fatty-Glandular (G) class and rest 112 images belong to Dense-Glandular (D) class images. All images have size of 1024 × 1024 pixels in PGM Format. Each pixel in the images is corresponded to the 8-bit word, where the images are in grayscale format with a pixel intensity of range [0, 255] [15].

### B. Median Filter

Median filter [16] recurrently used to overcome salt and paper noise while trimming down noise. It also preserves edges in image.

For applying the median filter we have to select a square window of size 2\*k+1, where k lies between 1 and N, around the considered pixel. After then arrange and sort all pixel values belongs into the square window and then calculate median value and replace median value with value considered pixel value. This process is repeated for every pixel in the image from left to right and top to bottom manner.

Fig. 1 illustrates an example calculation. In this example we have select k=1; so window size will be 3\*3.

105	100	106	97	98	100	102
108	107	98	97	99	98	100
100	104	99	120	101	102	101
100	101	99	99	99	99	97
97	99	100	101	103	98	101

Fig. 1. Matrix of 5×7 where Every Element Represents Gray Value of Pixels.

### Neighborhood pixel values:

98,97,99,99,120,101,99,99,99

### Sorted pixel values:

97,98,99,99,99,99,99,101,120

Median value: 99

It can be seen that the central pixel value of 120 is rather unrepresentative of the surrounding pixels and is replaced with the median value: 99. A 3×3 square neighborhood is used here and larger neighborhoods will produce more severe smoothing. Image Enhancement using fuzzy intensification operator.

### C. Image Enhancement using Fuzzy Intensification Operator

Image enhancement using fuzzy logic [17] is done by using Fuzzification which converts image form spatial domain to fuzzy domain, enhancement if fuzzy membership value and Defuzzification which convert back original image from fuzzy domain to spatial domain.

1) *Fuzzification*: An image X of size M×N can be viewed as an array of fuzzy singleton by converting into fuzzy set notation using following formula:

$$\mu_{ij} = (X_{ij} - X_{\min}) / (X_{\max} - X_{\min}) \quad (1)$$

Here, where  $\mu_{i,j}$  is a membership value that represents the amount of brightness acquired by the pixel intensity value  $x_{i,j}$  at  $i^{th}$  row and  $j^{th}$  column in image [18].

2) *Modification of the membership function*: The goal of our proposed method is to take care of the fuzzy nature of an image and formulate the contrast improvement more adjustable and valuable and to prevent from over-enhancement/under-enhancement. So, we employ Adjustment of memberships function  $\mu_{i,j} \rightarrow \mu_{i,j}$  by applying following PAL and KING transformation [19] (or the intensification operator (INT)).

$$\mu_{ij} = T(\mu_{ij})$$

$$\mu_{ij} = \begin{cases} 2 * [\mu_{ij}]^2 & 0 \leq \mu_{ij} \leq \mu_c \\ 1 - 2 * [1 - \mu_{ij}]^2 & \mu_c < \mu_{ij} \leq 1 \end{cases} \quad (2)$$

It transforms the membership values that are above threshold value to much larger values and membership values that are below than threshold to much smaller values in a nonlinear manner to obtain an enhanced image. Otherwise, display the unenhanced image.

3) *Defuzzification*: Defuzzification procedure is performed by applying reverse operation of fuzzification.

$$X'_{ij} = G^{-1}(\mu'_{ij}) \quad (3)$$

$$X'_{ij} = X_{\min} + \mu'_{ij} * (X_{\max} - X_{\min}) \quad (4)$$

Thus, the ultimate image will be contrast improved image [18].

#### D. Histogram based Self Initializing K-Means Clustering

K-Mean clustering [20] is among the most applicable unsupervised learning for classification that needs manual initialization clusters count and primary centroids of each cluster.

The automatic initialization of initial centroids can be achieved with uniform distribution of gray values of the normalized histogram.

##### 1) Algorithm 1: K-Means clustering

a) Input the image X and Initiate the number of clusters K.

b) Calculate the histogram of X. Find out the smallest and largest histogram gray level respectively, say p and q. Based on the value of p and q, apply discrete uniform distribution to initiate the initial centroids  $c_j$  where  $j=1, 2 \dots K$ .

The general formula for the probability density function (pdf) for the uniform distribution is:

$$f(x) = 1/b - a \quad \forall a \leq x \leq b$$

c) Allocate each pixel to the nearest class. This process is completed by minimizing the function J as given below.

$$J = \sum_{i=1}^N \sum_{j=1}^K \|x_i^j - c_j\|^2$$

Where,  $\|x_i^j - c_j\|^2$  is an absolute distance from a data point  $x_i^j$  to the cluster center  $C_j$ .

d) Evaluate mean value for each cluster and substitute this mean value as new centroids.

e) Do again steps and 4 until no more new centroids are created.

#### E. Region-based Segmentation

RBS [21-22] is a method for finding out region on the basis of some region membership criteria directly. The basic steps for RBS are:

a) Pick a set of seed points. Number of seed points depends on number of segments. Seed point selection is based on user choice or any random process.

b) The process starts from the selected seed points and then regions are expanded from these points to its neighboring points depending on region relationship norms. The norms may be pixel intensity, pixel color or texture.

c) Repeat step 2 for each of the recently added pixels; stop if no more pixels can be added.

### III. PROPOSED METHOD

In this Article ornate (elaborate) the intended method working sensible to easily identify and segment the boundary of breast tissues area in mammograms image. Our suggested method deals via a pair of phases:

- 1) Preprocessing.
- 2) K-means Clustering based Segmentation.

The first step, preprocessing comprises of two sub section i.e. noise Removal and improvement in image contradictions. Median filter utilizes in removal of salt and paper noise and contrast of images strengthened from assistance of Fuzzy Pal and King Method.

Second phase, the image splits in multiple clusters by applying K-mean clustering and appertaining to Seed Based Region Growing (SBRG) Technique, required appropriate cluster is segmented by selecting the seed/start point.

##### a) Proposed Algorithm

Step 1: Input the mammogram images.

Step 2: Median filter is imposed thus removes various noises from imaging.

Step 3: Change images (getting from step 2) from spatial domain to fuzzy area and subsequently applying the pal king function, as demanded INT operator on fuzzy domain that modifies fuzzy area for low contrast area to upper contrast area and moreover lastly(modified) fuzzy area transformed into intensified image using Defuzzification process.

Step 4: Perform the Histogram based Self Initializing K-Means Clustering algorithm that partition the images in different number of clusters. The count of clusters is mentioned before processing.

Step 5: Apply SBRG methodology which segments the appropriate region by performing region growing from seed/start points to adjoining points in concordance with region membership criterion.

Fig. 2 show the Block diagram of proposed work.

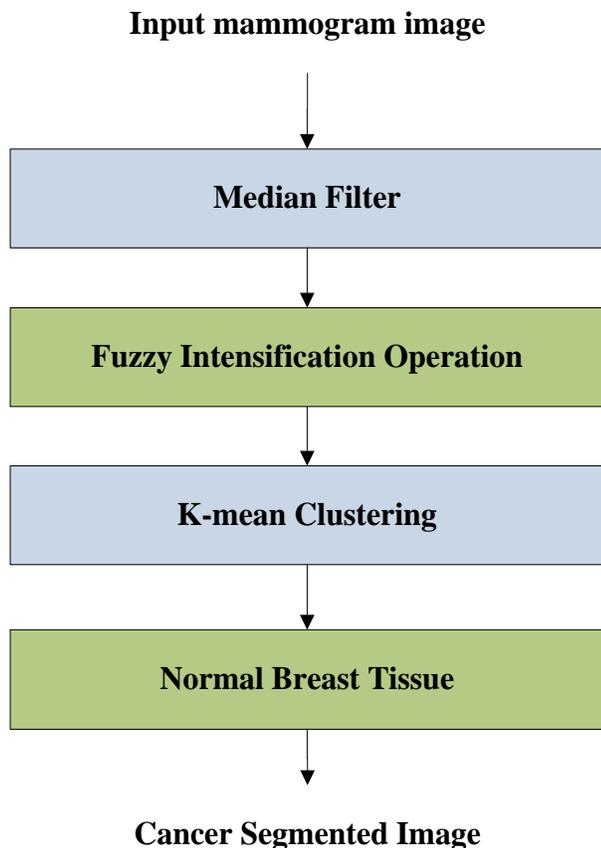


Fig. 2. Flow Chart of the Proposed Method.

#### IV. EXPERIMENTS AND RESULT

Thirty-one MIAS images brought into play the experiment. 15 together were cancerous images and 16 were normal images. Regarding this the proposed algorithm, referring to 15 cancerous images, 12 patients detected correctly and remaining 3 images, the cancerous region including tumor found incorrect. When the proposed methodology applied to remaining 16 patients on normal images, the significant cancer tissues was determined in 2 images, and in case of 14 images, the algorithm determines that these images does not contain carcinoma.

Fig. 2 shows the result of proposed method against the various mammogram images. Fig. 3(a) depicts Original image, Fig. 3(b) shows Mammography image after preprocessing step, Enhanced Mammography image after fuzzy INT operation is shown in Fig. 3(c). Fig. 3(d) depicts segmented image after K-mean clustering and finally Fig. 3(e) shows segment extracted by using region growing technique.

Table I depicts confusion matrix which describes relation between actual and predicted class by the performed classifier in our case we have two classes i.e. benign and malignant.

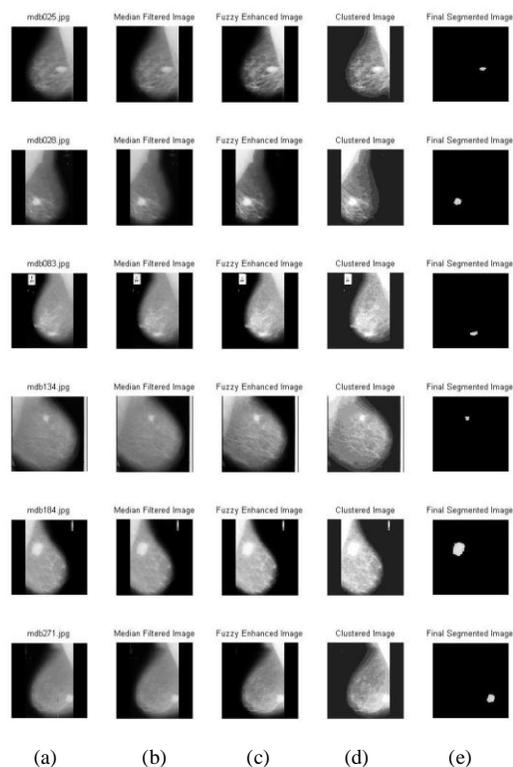


Fig. 3. Segmentation of Images mdb025, mdb028, mdb083, mdb134, mdb184 and mdb271:(a) Original Image, (b) Mammography Image after Preprocessing Step, (c) Enhanced Mammography Image after Fuzzy INT Operation (d) Segmented Image by K-mean Clustering (e) Segment Extraction by using Region Growing.

Table II shows the performance measure calculated from the Table I. From the Table II, it is clear that our proposed method has good precision value while error is low.

Two vital evaluation parameters for calculating effectiveness of our suggested technique are True Positive Rate (TPR) and False Positive Rate (FPR). TPR determines the correctly classified cancerous ROIs among all cancerous ROIs available whereas FPR determines incorrectly classified cancerous ROIs among all non-cancerous ROIs.

TABLE I. CONFUSION MATRIX

Actual Class	Predicted class	
	Cancerous image	Normal Image
Cancerous image	12	03
Normal Image	02	14

TABLE II. PERFORMANCE MEASURE

Measure	Definition
Error	0.1612
FPR	0.8235
precision	0.8235
ACC	83.87
F1_Score	0.8484
MCC	0.6778

Depending of evaluating the binary classification, F-measure and MCC also contribute a very significant role. F-measure examines harmonic mean of precision and recall and expounds as:

$$F - measure = \frac{2 \times recall \times precision}{recall + precision} \quad (5)$$

The MCC anatomizes correlation coefficient within observed and predicted classification and described as:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{(TP + FN)(TN + FP)(TP + FP)(TN + FN)} \quad (6)$$

The F-measure value separate from 0 to + 1 and MCC value lies from -1 to +1. Higher value measurement of both F-measure and MCC specify higher classification quality.

## V. CONCLUSION

For radiologist's consideration to easily identification and categorization of concerning mammography images, different methods by researchers sustained bearing in mind since past decades. By means of this research report, we aims mammography image segmentation with k-means clustering and region growing techniques benefit to experts or radiologists to detect cancerous regions with computer-aids. The intended work carve up within two stages: At start stage, pre-processing employed, we retained a median filter that expel undesirable salt and pepper noise. Further, taken pal king approach that reinforce input images contrast. In subsequent stage, more effective image is exerted as an input for k-mean clustering. Image is divided in order to get multiple clusters and with Seed Based Region Growing (SBRG) Technique, required appropriate cluster is segmented by selecting the seed point. For achieving rationale behind this experiment, we are using mini MAIS dataset. The end outcome of experiments demonstrates that proposed strategy accomplishes higher precision.

## REFERENCES

- [1] Saurabh PrasadLori Mann BruceJohn E Ball," A Multi-classifier and Decision Fusion Framework for Robust Classification of Mammographic Masses" International Conference of the IEEE Engineering in Medicine and Biology , 2008.
- [2] K.Subashini, K.Jeyanthi, 'Masses detection and classification in ultrasound images', IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS), Volume 9, Issue 3 Ver. II (May -Jun. 2014), PP 48-51.
- [3] Ojo J. A., Adepoju T. M., Omdiora E. O., Olabiyisi O. S. and Bello O. T, 'Pre-Processing Method for Extraction of Pectoral Muscle and Removal of Artefacts in Mammogram,' IOSR Journal of Computer Engineering (IOSR-JCE) e-Volume 16, Issue 3, Ver. V (May-Jun. 2014).
- [4] Huanping Zhao, Weidong Xu, Lihua Li, Juan Zhang "Classification of Breast Masses Based on Multi-View Information Fusion Using Multi-Agent Method", 5th International Conference on Bioinformatics and Biomedical Engineering, (iCBBE) 2011.
- [5] K.Subashini, K.Jeyanthi, 'Masses detection and classification in ultrasound images', IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS), Volume 9, Issue 3 Ver. II (May -Jun. 2014), PP 48-51.
- [6] S. Deepa, Dr.V.Subbiah Bharathi, 'Textural Feature Extraction and Classification of Mammogram Images using CCCM and PNN', IOSR Journal of Computer Engineering (IOSR-JCE) , Volume 10, Issue 6 (May. - Jun. 2013).
- [7] Mencamttini, A.et al., "Mammographies images enhancement and denoising for breast cancer detection using dyadic wavelet processing", IEEE Trans. Instrumentation Measure., 57: 1422-1430. DOI: 10.1109/TIM.2007.915470.
- [8] Abdelali ELMOUFIDI et al., "Detection of Regions of Interests in Mammograms by Using Local Binary Pattern, Dynamic K-Means Algorithm and Gray Level Co-occurrence Matrix", 2014 Fifth International Conference on Next Generation Networks and Services (NGNS'14) 28- 30 May 2014, Casablanca, Morocco.
- [9] N. Saidin et al., "Density based breast segmentation for mammograms using graph cut and seed based region growing techniques", 2nd International Conference on Computer Research and Development (ICCRD '10), pp. 246-250, Kuala Lumpur, Malaysia, May 2010.
- [10] N. Karssemeijer, "Automated classification of parenchymal patterns in mammograms", Phys. Med. Biol., vol. 43, 1998, pp. 365-378.
- [11] A.El-Zaart et al., "Expectation Cmaximization technique for fibroglandular discs detection in mammography images", Comput. Biol. Med. 40(4), 392-401 (2010).
- [12] A.El-Zaart et al., "Expectation Cmaximization technique for fibroglandular discs detection in mammography images", Comput. Biol. Med. 40(4), 392-401 (2010).
- [13] K. S. Camilus et al., "Computer- Aided Identification of the Pectoral Muscle in Digitized Mammograms", Journal of digital imaging, October 2009, pp. 1-19.
- [14] [Online] Available: <http://www.wiau.man.ac.uk.services/MIAS/MIASweb.htm>.
- [15] Suckling, J. et al., "The Mammographic Image Analysis Society digital mammogram database", Exerpta Medica, International Congress Series 1069 pp. 375-378., 1994.
- [16] Makandar, Aziz and Bhagirathi Halalli. "Breast Cancer Image Enhancement using Median Filter and CLAHE." (2015).
- [17] M. Hanmandlu and D. Jha, "An Optimal Fuzzy System for Color Image Enhancement", Image Processing, IEEE Transactions on, Vol. 15, pp. 2956-2966, 2006.
- [18] Khairunnisa Hasikin and Nor Ashidi Mat Isa, "Enhancement of the low contrast image using fuzzy set theory" 14th International Conference on Modelling and Simulation, IEEE, 2012.
- [19] Pal S K, King R A. Image enhancement using smoothing with fuzzy sets. IEEE Trans. Systems, Man & Cybernetics, 1981, 11(7): 494-501.
- [20] S.Julian Savari Antony ,Dr.S.Ravi, 'A New Approach to Determine the Classification of Mammographic Image Using K-Means Clustering Algorithm', International Journal of Advancements in Research & Technology, Volume 4, Issue 2, February -2015.
- [21] R. Adams and L. Bischof. "Seeded region growing", IEEE Trans, Pattern Analysis Machine Intelligence, 16(6):641-647, Jun 1994.
- [22] Harikrishna Rai G.N et al., "Gradient Based Seeded Region Grow method for CT Angiographic Image Segmentation", InterJRI Computer Science and Networking, Vol. 1, Issue 1, July 2009.

# IoT Technology for Facilities Management: Understanding End user Perception of the Smart Toilet

Venushini Raendran<sup>1</sup>, R. Kanesaraj Ramasamy<sup>2</sup>  
Faculty of Computing and Informatics  
Multimedia University  
Cyberjaya, Malaysia

Intan Soraya Rosdi<sup>3</sup>, Ruzanna Ab Razak<sup>4</sup>  
Nurazlin Mohd Fauzi<sup>5</sup>  
Faculty of Management, Multimedia University  
Cyberjaya, Malaysia

**Abstract**—The Internet of Things (IoT) plays an important role as an emerging technology. IoT platforms enable electronic devices to collect, process, and monitor various types of data. The Smart Toilet system featured in this paper is based on IoT technology, and it is designed for resource optimization in facility management services and for bringing convenience to individual end users. This paper presents a study conducted on individual end user perception of the Smart Toilet. A total of 124 respondents had participated in the study's online survey and statistical data analysis methods were used to analyse the data. Results indicate user perception of the proposed Smart Toilet and the Smart Toilet app.

**Keywords**—IoT system; facilities management; smart toilet; resource optimization; user acceptance; theory of planned behaviour

## I. INTRODUCTION

Information and communication technology (ICT) has been dominating the world in all aspects with the foundation of rapid internet technology growth in the last few decades. The increased growth in cities and dependencies on ICT and efficient management of critical infrastructure address the challenges of development and sustainability of technology [1]. For the last few years, the development of technology and research on ICT have been devoted to the Internet of Things (IoT) [2]. IoT platforms offer provisions for most electronic devices to collect, process, and monitor various types of data. Evolution of IoT in the smart cities, smart homes, wearable technologies etc. enables us to monitor real-time data from electronic devices and use the data for critical operations such as rescue missions during emergencies [3]. Its popularity is powered by the fact that many smart city solutions promise to alleviate the real pains of people living in cities. IoT technology is generally used to save energy and resources.

With the current technology of IoT in buildings and residential areas, the environment can be monitored in real time. Retrieving real-time data for rescue operations may well result in survivors during disasters and emergencies [3]. IoT may also be used to enhance security surveillance and improve the deployment of wireless communication. According to [4], with urbanization breaking the 50% barrier, it is of importance to understand the need for increased efficiency in city management. The author also highlights that

highly developed sensing technology allows the gathering and evaluation of real-time data, which is then converted into useful information. However, disadvantages of the technology include precision demands, cost, and intrusiveness and privacy infringement risks. Deploying additional devices would increase the accuracy but at the same time it will increase intrusiveness as well as the cost.

This paper features the proposed Smart Toilet system, which is based on IoT technology and is designed to capture real-time data from sensors placed in toilet facilities towards the optimization of toilet cleaning resources and the sourcing of toilet peripherals. References made to existing studies on toilet management technologies reveal works done such as by [5], who proposes a smart automatic urinal flushing system. Water management system in the form of an automated flushing system is implemented in many public toilets. It helps control the amount of water running in the toilet as well as reduce the chances of cross-infection from toilet users pressing the flush button. However, the drawback of this system is that it is only able to manage the frequency of the cleaning the toilet bowl. Comparatively, a smart toilet system should be a system that manages the entire environment inside the toilet and not just focused on any one toilet part or feature.

Based on the challenges in toilet facility maintenance, the Smart Toilet system architecture is designed to optimize toilet-cleaning resources by way of monitoring toilet usage rates using motion sensors, automated janitor scheduling system and predictive maintenance of toilet peripherals through a web dashboard, and providing toilet occupancy data to users in the general public using the Smart Toilet app. However, besides the technology issues, there is a need to understand individual end users' perception of the Smart Toilet. Hence, the research question is: 'How do individual end users perceive the proposed Smart Toilet?' The research objective is to analyze individual end user perception of the proposed Smart Toilet and the Smart Toilet app.

The significance of this paper is its highlighting of the importance of IoT technology, the application of IoT technology in the Smart Toilet System for more effective facilities management, and finally, the perception of targeted end users of the proposed Smart Toilet and the Smart Toilet app.

This paper is organized such that Section II presents the related works on IoT technology in the context of facilities management and describes the Smart Toilet system architecture. Section III explains the user perception study conducted for the proposed Smart Toilet and covers the discussion of findings and finally Section IV the conclusion of the paper.

## II. RELATED WORKS

### A. IoT for Facility Management

In the context of this study, the focus on IoT technology is its applications in the facilities management sector as management of toilet facilities falls under this business sector. The field of facility management (FM) is a combination of various disciplines that ensure workplace environment functionality by implementing integration between people, place, process and technology [6]. Facilities management is responsible for the identification, design, specification, procurement and delivery of premises and service requirements that support a company's core business [7]. Facilities management has a strategic nature and it plays a vital role in implementing company policies that are used in strategic costing in service management control [8]. Through facilities management, a company will be able to ensure that the space surrounding their work environment can be enhanced without interfering with the daily movements of the staff and the core business operations of the company [9].

In Malaysia, facilities management standards and regulations are not making as fast a progress as they should, making the country quite left behind in terms of artificial intelligence and IoT technology to manage and integrate facilities management [10]. In 2019, Sirim QAS International Sdn. Bhd, Malaysia's leading certification, inspection and testing body introduced the Facility Management Systems Certification Scheme to pave the way for forward-looking industry players to position themselves for the fourth industrial revolution. Dubbed (ISO) 41001:2018, among its benefits include improved safety, health and productivity of the workforce, better efficiency and effectiveness in facilities management and improved organizational costs [11]. The Smart Toilet system, which will be described in the following section, is a good step in the right direction towards meeting the standards of the ISO certification.

### B. The Smart Toilet

The proposed Smart Toilet system can be implemented in public toilets at common places such as universities, shopping malls, and any other types of low- or high-rise buildings. Its implementation is intended to provide automated janitor scheduling system through a web dashboard, to develop awareness of the importance of clean toilet culture with technology-based tools in line with ISO 17966:2016, and to provide information on toilet occupancy status to the public. By identifying the toilet occupancy status and the location of the nearest toilet in the any given building, end users will able to avoid queuing especially during peak hours. Implementation involves a complete IoT system from sensing capability to data management system with cloud-based Integration. Fig. 1 shows the overall architecture of Smart Toilet implementation.

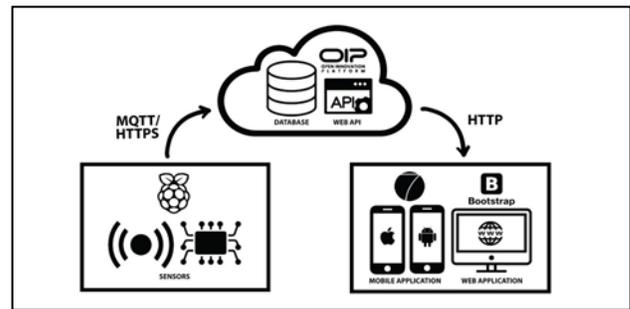


Fig. 1. Overall Architecture of Smart Toilet Implementation.

Raw data collected from the sensors are also known as “real-time” which will be stored in the cloud database for further processing. Cloud technology is used not only as a medium to store the data, but it is also required to process the stored data before it is made available to the end user. When data are processed and organized or presented as meaningful information, the data will be extracted to the mobile application via web Application Programming Interface (API) for the end user in a presentable and understandable way. Hybrid type mobile application will be developed. A hybrid application can give freedom for developers to use their web development skills like HTML5, CSS and JavaScript to develop a mobile application. It is cost effective as a single development process handles both Android and IOS application. There are three main phases in this Smart Toilet as shown in Fig. 1 above. Firstly, sensory phase (1), then, cloud management (2) and lastly, mobile development phase for monitoring purposes. (3). Table I shows the description and tools used in the three phases.

Fig. 2 shows the mobile application user interface for the end user and administrative respectively. The Smart Toilet mobile application will help end users to check the occupancy status of toilet cubicles. The green color indicates the toilet cubicles are available while the red color indicates their unavailability. The users would also be able to check on other available toilets in the specific building through the floor directory as shown in Fig. 2. End users can also give feedback in terms of the cleanliness rating of the toilet through mobile application. If the toilet rated less poorly on cleanliness, the janitors on duty will receive notification from their supervisors to clean the respective toilet.

More importantly, the administrative view is as shown in Fig. 3 and 4 respectively, on maintenance prediction and automated janitor scheduling system. There will be three different colors used to ease the readability for the management personnel with green indicating all peripherals are working in good condition. Next, yellow indicates that the peripheral may be due for repairs or servicing within the next three months, and finally, red indicates faulty toilet peripherals.

Fig. 5 shows the activities that have been carried out during the implementation of the Smart Toilet. There are four major phases of activities, namely, 1) Analysis, 2) Modelling, 3) Development and 4) Testing.

TABLE I. PHASES IN SMART TOILET, THEIR DESCRIPTION AND TOOLS USED

Phase	Tools
Sensory phase	<p>In this phase, the sensors are installed in every toilet cubicles as shown in the figure. This sensor will monitor the real-time environment. The retrieved data will send to the database to process in the next stage. The hardware used in this level as below:</p> <p>A. Raspberry Pi B+.[12] It is a credit card-sized computer, which can plug into any HDMI input device. This device is used as a microcontroller. The main technical specifications of the latest model of Raspberry Pi also known as Model B+ (2018) have the following features:</p> <ol style="list-style-type: none"> <li>2.4GHz and 5GHz IEEE 802.11.b/g/n/ac wireless LAN, Bluetooth 4.2, BLE</li> <li>Full-size HDMI</li> <li>4 USB 2.0 ports</li> <li>5V/2.5A DC power input</li> </ol> <p>B. Infrared (IR) Sensor This sensor will be used to detect the obstacles. In ST, this sensor will be used to for detect the person in the toilet in and out. Specification of IR sensors are:</p> <ol style="list-style-type: none"> <li>Input Power: 3.3V or 5VDC</li> <li>3 pin interface which is OUT, GND and VCC</li> <li>Two LED indicators, one (Red) as a power indicator, another (green) as object detection indicator</li> <li>adjustable sensitivity with an onboard potentiometer, this translate to the adjustable detection range</li> </ol> <p>C. Push-to-On Button This button is for janitors to indicate the cleaning process of the toilet is completed after the dirty notification is alerted. Specification as below:</p> <ol style="list-style-type: none"> <li>Contact rating: 125V 3A</li> <li>Thread diameter: 15.5mm</li> <li>Button diameter: 13mm</li> <li>Material: Metal (body) and plastic (button)</li> </ol> <p>D. Light-emitting diode (LED) If the user entered the toilet, this LED is responsible for alerting other users that the specific toilet is occupied.</p>
Cloud Management Phase	<p>We used the cloud database to store the data and the system process as well. All the data will be accumulated from the sensors will be stored in the cloud database. We use Telekom Malaysia Research and Development (TMRnD), Open Innovation Platform (OIP). Besides this server, another server such as Microsoft Azure or Amazon Web Service (AWS) IoT server which provides almost the same functionality for IoT implementation.</p>
Mobile Development Phase	<p>Mobile development is the last stage of completing Smart Toilet. In this phase, the mobile application is developed for end users and administrative respectively where the data from the cloud database will be revealed with presentable User Interface (UI). We used Framework 7 as our development framework which supports Model-View-Controller (MVC) architecture. It is a free and open source mobile HTML framework to develop hybrid mobile apps or web apps with iOS &amp; Android native look.</p>

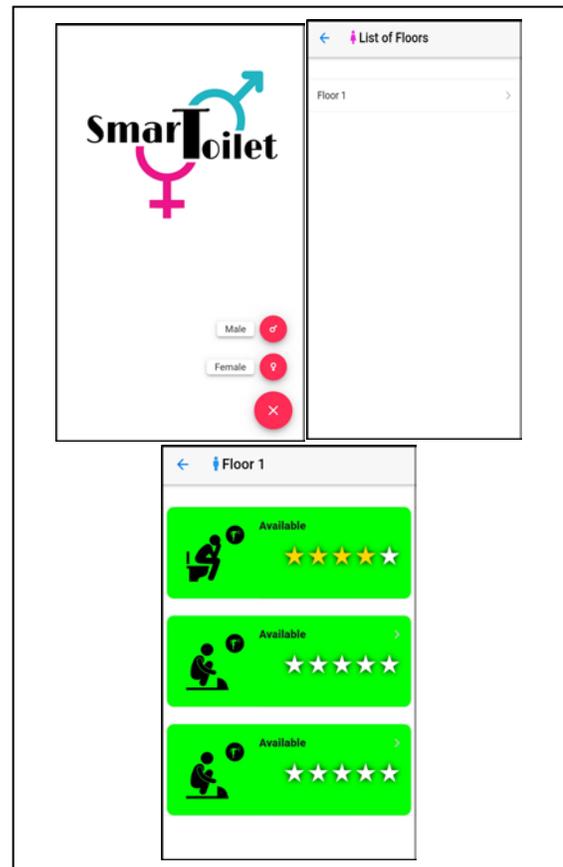


Fig. 2. The Mobile Application user Interface for end user.



Fig. 3. The web Dashboard view for Administrative on Maintenance Prediction.

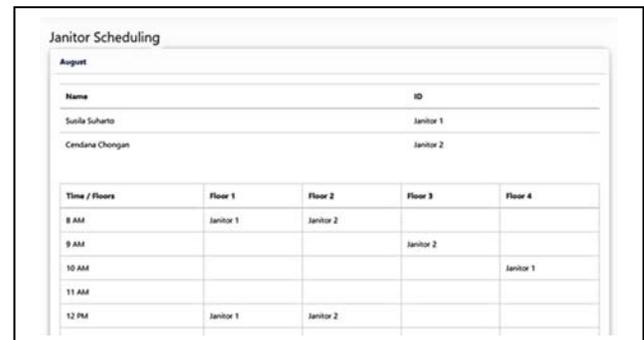


Fig. 4. The Web Dashboard view for Administrative on Janitor Scheduling.

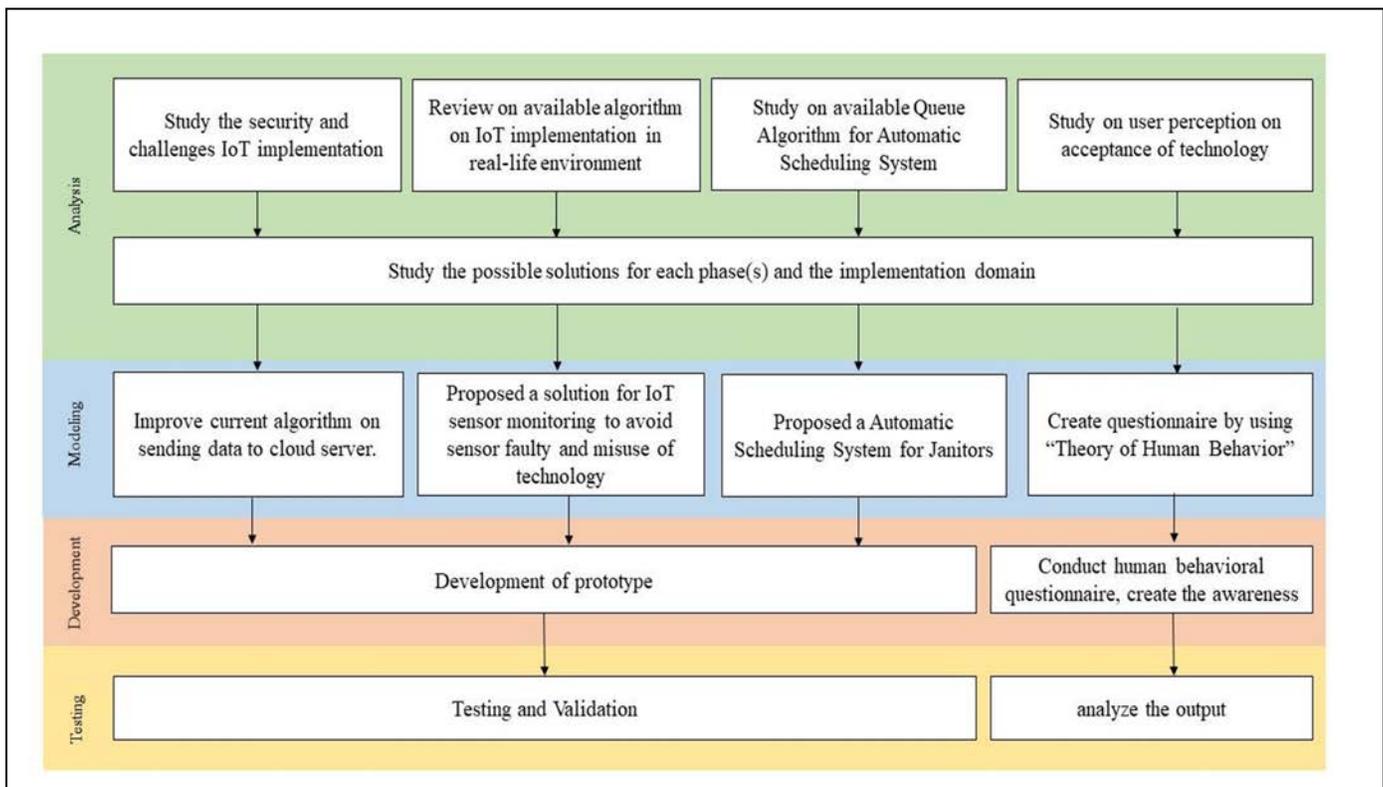


Fig. 5. Activities Carried Out During IoT Implementation.

### C. Analysis

Review of domain-specific related studies (in this case, toilets) in defining user requirements. At the same time, studies related to algorithms on IoT implementation on the real environment and automated scheduling is considered.

### D. Modelling

The modelling phase involved designing the hardware and software.

### E. Development

In this phase, we have implemented the prototype by building the miniature toilets.

### F. Testing

In this phase, the miniature Smart Toilet is tested via the performance test and stability test.

## III. USER PERCEPTION STUDIES ON THE SMART TOILET

Researchers in the field of human behaviour define 'perception' as the process that individuals go through in organizing and interpreting information detected by their senses in order to assign meaning to their environmental elements. Informational inputs are then processed into individual decisions and actions. Hence, individual behaviour is based on people's perception of their reality and not necessarily on reality itself [13][14]. To study end-user perception and individual decision making in a systematic manner, reference was made to the theory of planned behaviour (TPB) in the field of social sciences [15].

### A. Theory of Planned behavior

The TPB is an extension of the theory of reasoned action (TRA) developed by [16]. The TPB explains individual perception and decision making by establishing that a person's actual behavior is directly influenced by his or her behavioral intention [17]. In turn, one's behavioral intention is shaped by three antecedents: attitudes, subjective norm, and perceived behavioral control. While a person's attitude refers to his or her favorable or unfavorable assessment of the behavior in question, the subjective norm represents the perceived social pressure to engage in the behavior in question. The third antecedent, which is perceived behavioral control, signals a person's perceived control over his or her ability to perform the behavior. The TPB extends the earlier TRA by adding perceived behavioral control as an additional variable when it was found that there are situations where one may not have complete control over one's behavior. It is established that the level of importance of the three antecedents in predicting behavioral intention changes across different situations and behaviors [16][18].

The TPB has been extensively applied to studies on various issues and across many different disciplines such as health [19], dietetics [20], human resource management [21], entrepreneurship [22], sports science [23], logistics [24], and many more. It has also been applied in studies relating to information technology and smart systems, i.e. Internet of Things (IoT) such as works by [25], [26].

**B. Research Design, Instrument and Variables**

In developing the Smart Toilet system, information was sought on end-user perception to ensure that the end product catered to the real needs of users, thus resulting in successful user adoption. The study on end-user perception had involved administering a survey questionnaire as the data collection instrument. The survey was conducted online and sent via social networking platforms such as WhatsApp and Facebook. As the product is developed for the Malaysian market, target survey respondents were Malaysians from the general public across different age groups and locations.

Based on the original TPB model, three constructs that determine a person’s intention which subsequently influences his or her actual behaviour are attitude, subjective norm, and perceived behavioural control (PBC) [18].

On attitude, many previous studies on technology adoption reported that attitude is an important factor influencing the intention to use particular technology tools, for example, on mobile health technology [27] and on smart home services [28]. Previous studies also found that subjective norm is positively associated with the intention to adopt mobile health service [29], intention to purchase smartphones and intention to use smart home services [30]. Subjective norm is defined as a person’s perception of his or her important people’s opinion on his or her intention to use [17]. With regards to perceived behavioural control (PBC), past studies support the positive impact of PBC on the intention to use mobile data services [31], [32]. Following the definition by [28], PBC refers to potential users’ perception of the level of easiness to use the technological service.

The work by [28] is used as the main reference for this study as it concerns a similar topic, i.e. end user acceptance of IoT in the context of smart home technology. Another reason is that [28] extended the original TPB model to include five other variables that are relevant predictors in the IoT context, which are: perceived risk, automation, mobility, interoperability, and trust in service providers. Some of the variables studied by [28] are also highlighted in other technology-related studies such as perceived risk in [33] and [34], automation and mobility in [35], and trust in service provider in [36]. However, there has yet to be existing studies found on the concept of interoperability introduced by [28].

As such, furthering the work of [28], the Smart Toilet user perception study had examined public perception on the following nine variables: attitude towards the Smart Toilet app, subjective norm, perceived behavioral control, perceived risk, automation, mobility, interoperability, trust in service provider, and intention to use the Smart Toilet app. In addition, the study also collected feedback on the respondents’ frequency in using public toilets, preferred toilet types, and preferred washing amenities. Information from the public survey was to serve as feedback to the technical development team to make adjustments in providing the required necessities to future product end users.

**C. Statistical Data Analysis and Results**

A total of 124 respondents had participated in the survey. Since this initial survey gauges the public's perception towards

embedded technology in toilets and their intention to use the Smart Toilet app, the statistical methods are limited to descriptive statistics only.

Participants of the survey consisted of 67.6% females and 32.4% males. While participants come from across all the different age groups, i.e. Below 30 years old, between 30-39 years, between 40-49 years, and 50 years old and above, the majority of them belong to the 30-39 years old category. Fig. 6 presents the breakdown of age groups.

Cumulatively, 73.5% of the respondents frequently visit public places, and more than half of them frequently use the available public toilets. In terms of respondents’ preferred type of toilet and washing amenities, the majority of them chose sitting toilets (55.9%) as opposed to squatting ones, and they preferred toilets with water hoses (73.5%) rather than dry toilets or ones with built-in bidets.

The online survey respondents were presented with brief statements relating to their perception of the nine variables, which are: attitude towards the Smart Toilet app, namely subjective norm, perceived behavioral control, perceived risk, automation, mobility, interoperability, trust in service provider, and intention to use the Smart Toilet app. Their responses were measured using a Likert Scale of 1 (Strongly Disagree) to 5 (Strongly Agree). Tables II and III outlines the attributes in the study, the corresponding variables measured, and the descriptive statistics for each variable.

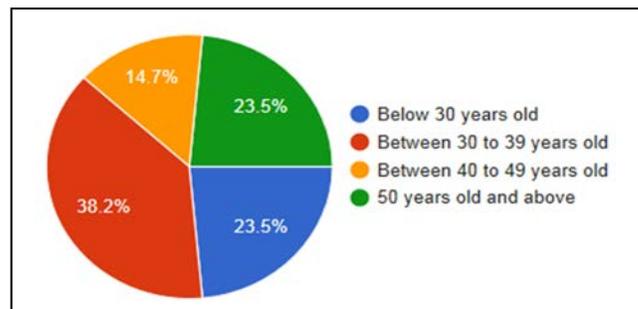


Fig. 6. Percentage of Respondents by Age.

TABLE II. ATTRIBUTES IN THE STUDY

Item	Statement
A	I have positive feelings about using the smart toilet app.
B	People whose opinions are important to me would support my use of the smart toilet app.
C	I have the resources, knowledge, and ability to use the smart toilet app.
D	I do not have any privacy concerns from the human motion detection sensor installed in toilet cubicles to support the smart toilet app.
E	It is convenient that the smart toilet app helps users proactively without human intervention.
F	It is convenient to access the smart toilet app anywhere at any time.
G	It is convenient that the smart toilet app can run on existing hardware (smartphone, tablet, etc.)
H	I think the smart toilet app service provider has customers' best interests in mind (convenience, user-friendliness, etc.).
I	I predict I would use the smart toilet app once it becomes available in the future.

TABLE III. DESCRIPTIVE STATISTICS OF EACH VARIABLE

Item	Variable	Mean	Std Dev	Mode	Cum. % for Level 4 & 5
A	Attitude	4.03	0.92	5	73.5%
B	Subjective norm	3.79	0.96	4	70.6%
C	Perceived Behavioral Control	3.97	0.95	4	73.5%
D	Perceived Risk	3.94	0.66	4	76.5%
E	Automation	3.97	0.98	4	73.5%
F	Mobility	4.15	0.87	5	76.5%
G	Interoperability	4.12	0.82	4	79.4%
H	Trust	3.97	0.88	5	67.7%
I	Intention to use	4.18	0.77	4	85.3%

Table III contains the mean agreement level and its standard deviation, mode and cumulative percentage of respondents who agree and strongly agree to each item statement.

#### D. Discussion of Findings

Based on Table I, the mode for each item indicates the opinion of the majority of respondents (e.g. Level 4 = Agree; Level 5 = Strongly Agree). The majority of the respondents strongly agreed on the mobility of the app, had positive attitudes towards using the app, and trustful of the app service provider. The mean level of agreement is close to 4, which suggests that respondents mainly agree with all statements. The last column in the table shows the cumulative percentage of respondents who agreed and strongly agreed with the statements. In brief, the results show that many of the respondents have positive feelings about using the smart toilet app and intend to use it in the future. In addition, the respondents do not have any privacy concerns from the sensor installed in toilet cubicles.

The user perception study presented in this paper is conducted to provide preliminary insights into the market potential of the proposed Smart Toilet and the Smart Toilet app and the public reaction to the idea. Further in-depth empirical investigation of user perception of the technology and how the variables influence user intention in using the app and related services will be included in future research.

#### IV. CONCLUSION

This paper describes how IoT technology is used to address challenges in toilet facility maintenance. It describes the Smart Toilet system, which utilizes a complete IoT system from sensory level to data management system with cloud-based integration. The Smart Toilet system allows the collection of real-time data on toilet usage and cleanliness ratings. Big data analytics (BDA) on these data will enable the optimization of resources and predictive maintenance. This paper also answers the question of how individual end users perceive the proposed Smart Toilet and the Smart Toilet app. Findings from the user perception survey shows that respondents have positive feelings about the Smart Toilet and they have the intention to use the Smart Toilet and the Smart

Toilet app in the future. Its implementation in facilities management has the potential to improve the quality of public facilities and services, reduce asset owners' maintenance cost, prolong asset and facilities lifespan, and increase asset value. It can be concluded that the Smart Toilet addresses the needs of the public as well as organizations involved in facilities management.

#### ACKNOWLEDGMENT

This work was supported by the Telekom Malaysia Research & Development grant number [RDTC/180969].

#### REFERENCES

- [1] A. K. Sikder, A. Acar, H. Aksu, A. S. Uluagac, K. Akkaya, and M. Conti, "IoT-enabled smart lighting systems for smart cities," 2018 IEEE 8<sup>th</sup> Annual Computing and Communication Workshop and Conference (CCWC), pp. 639-645, January 2018.
- [2] E. Theodoridis, G. Mylonas, and I. Chatziagiannakis, "Developing an IoT smart city framework," IISA 2013, Piraeus, pp. 1-6, July 2013.
- [3] K. Akkaya, I. Guvenc, R. Aygun, N. Pala, and A. Kadri, "IoT-based occupancy monitoring techniques for energy-efficient smart buildings," 2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), New Orleans, LA, pp. 58-63, March 2015.
- [4] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An information framework for creating a smart city through internet of things," IEEE Internet of Things Journal, vol. 1, no. 2, pp. 112-121, 2014.
- [5] K. Osathanukul, K. Hantarkul, P. Pramokchon, P. Khoenkaw and N. Tantitharanukul, "Design and implementation of an automatic smart urinal flusher," 2016 International Computer Science and Engineering Conference (ICSEC), Chiang Mai, pp. 1-4, December 2016.
- [6] J. Lai, K. Tu, J. Lian, and J. Kim, "Facilities management education in the Four Asian Dragons: a review," Facilities, vol. 37, no. 11/12, pp. 723-742, 2019.
- [7] B. Williams, Facilities Economics in the United Kingdom, 2nd ed. International Facilities and Property Information Limited: Kent, 2003.
- [8] H. Koleoso, M. Omirin, and F. Adejumo, "Comparison of strategic content of facilities managers functions with other building support practitioners in Lagos, Nigeria," Property Management, vol. 36, no. 2, pp. 137-155, 2018.
- [9] J. M. Wiggins, Facilities Managers Desk Reference. Wiley Blackwell Publishers: 2010.
- [10] N. Elyna Myeda, and M. Pitt, "Facilities management in Malaysia: Understanding the development and practiced," Facilities, vol. 32, no. 9/10, pp. 490-508, 2014.
- [11] The Star, Retrieved from <https://www.thestar.com.my/business/business-news/2019/08/26/setting-the-standards-for-facilities-management>, 26 August 2019.
- [12] Raspberry Pi 3 Model B. Retrieved September 3, 2018, from <https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus/>
- [13] S. P. Robbins, and T. A. Judge, Organizational Behavior, 15th ed. Pearson: London, England, 2013.
- [14] J. Wood, R. Zeffane, M. Fromholtz, R. Wiesner, R. Morrison, and P.-S. Seet, Organizational Behavior - Core Concepts and Applications, 3rd ed. Wiley: Australia, 2013.
- [15] I. Ajzen, "From intentions to actions: A theory of planned behavior," In J. Kuhl and J. Beckmann (eds), Action Control: From cognition to behavior. Springer: Heidelberg, 1985.
- [16] M. Fishbein, and I. Ajzen, Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison-Wesley: Reading, MA, 1975.
- [17] I. Ajzen, "The theory of planned behavior," Organizational Behavior and Human Decision Processes, vol. 50, pp. 179-211, 1991.
- [18] I. Ajzen, and M. Fishbein, Understanding Attitudes and Predicting Social Behaviour. Prentice-Hall : Englewood Cliffs, NJ, 1980.

- [19] A. Grønhoj, T. Bech-Larsen, K. Chan, and L. Tsang, "Using theory of planned behavior to predict healthy eating among Danish adolescents," *Health Education*, vol. 113, no. 1, pp. 4-7, 2013.
- [20] A. Silva, I. Figueiredo, T. Hogg, and M. Sottomayor, "Young adults and wine consumption a qualitative application of the theory of planned behavior," *British Food Journal*, vol. 116, no. 5, pp. 832-848, 2014.
- [21] M. C. H. Ang, T. Ramayah, and H. Amin, "A theory of planned behavior perspective on hiring Malaysians with disabilities," *Equality, Diversity and Inclusion*, vol. 34, no. 3, pp. 186-200, 2015.
- [22] W. J. Aloulou, "Predicting entrepreneurial intentions of final year Saudi university business students by applying the theory of planned behavior," *Journal of Small Business and Enterprise Development*, vol. 23, no. 4, pp. 1142-1164, 2016.
- [23] C. M. Veraldo, and B. J. Ruihley, "Theory of planned behavior and women in senior-level athletic administration". *Sport, Business and Management*, vol. 7, no. 1, pp. 21-37, 2017.
- [24] C. Busse, A. Regelmann, H. Chithambaram, and S. M. Wagner, "Managerial perceptions of energy in logistics: An integration of the theory of planned behavior and stakeholder theory," *International Journal of Physical Distribution & Logistics Management*, vol. 47, no. 6, pp. 447-471, 2017.
- [25] T. Sommestad, H. Karlzén, and J. Hallberg, "The sufficiency of the theory of planned behavior for explaining information security policy compliance," *Information and Computer Security*, vol. 23, no. 2, pp. 200-217, 2015.
- [26] R. Yadav, V. Chauhan, and G. S. Pathak, "Intention to adopt internet banking in an emerging economy: a perspective of Indian youth," *International Journal of Bank Marketing*, vol. 33, no. 4, pp. 530-544, 2015.
- [27] Z. Hussein, S. W. Oon, and A. Fikry, "Consumer attitude: Does it influence the intention to use mHealth?" *Procedia Computer Science*, vol. 105, pp. 340-344, 2015.
- [28] H. Yang, H. Lee, and H. Zo, "User acceptance of smart home services: an extension of the theory of planned behavior," *Industrial Management & Data Systems*, vol. 117, no. 1, pp. 68-89, 2017.
- [29] Y. Sun, N. Wang, X. Guo, and Z. Peng, "Understanding the acceptance of mobile health services: a comparison and integration of alternative models," *Journal of Electronic Commerce Research*, vol. 14, no. 2, pp. 183-200, 2013.
- [30] H. Chi, H. R. Yeh, and Y. Yang, "Applying theory of reasoned action and technology acceptance model to investigate purchase behavior on smartphone," *Journal of International Management Studies*, vol. 6, no. 3, pp. 1-11, 2011.
- [31] Z. Deng, X. Mo, and S. Lui, "Comparison of the middle-aged and older users' adoption of mobile health services in China," *International Journal of Medical Informatics*, vol. 83, no. 3, pp. 210-224, 2014.
- [32] M. T. Lu, G. H. Tzeng, H. Cheng, and C. C. Hsu, "Exploring mobile banking services for user behavior in intention adoption: using new hybrid MADM model," *Service Business*, vol. 9, no. 3, pp. 541-565, 2014.
- [33] M. C. Lee, "Factors influencing the adoption of internet banking: an integration of TAM and TPB with perceived risk and perceived benefit," *Electronic Commerce Research and Application*, vol. 8, no. 3, pp. 130-141, 2009.
- [34] J. S. Chou, and I. G. A. N. Yutami, "Smart meter adoption and deployment strategy for residential buildings in Indonesia," *Applied Energy*, vol. 128, no. 1, pp. 336-349, 2014.
- [35] M. E. Porter, and J. E. Heppelman, "How Smart, Connected Products Are Transforming Competition," *Harvard Business Review*, 2014. Access date: 24s August 2018. <https://hbr.org/2014/11/how-smart-connected-products-are-transforming>
- [36] M. Noh, and K. Lee, "An analysis of the relationship between quality and user acceptance in smartphone apps," *Information System E-Business Management*, vol. 14, pp. 273-291, 2016.

# New Learning Approach for Unsupervised Neural Networks Model with Application to Agriculture Field

Belattar Sara<sup>1\*</sup>, Abdoun Otman<sup>2</sup>, El khatir Haimoudi<sup>3</sup>  
Computer Science Department, Advanced Science and  
Technologies Laboratory, Polydisciplinary Faculty  
University of Abdelmalek Essaadi (UAE), Larache, Morocco

**Abstract**—An accurate and lower cost hybrid machine learning algorithm based on a combination of Kohonen-Self Organizing Map (SOM) and Gram-Schmidt (GSHM) algorithm was proposed, to enhance the crop yield prediction and to increase the agricultural production. The combination of GSHM and SOM allows to withdraw the most informative components about our data, by overcoming correlation issues between input data prior to the training process. The improved hybrid algorithm was trained firstly on data that have a correlation problem, and it was compared with another hybrid model based on SOM and Principal Component Analysis (PCA), secondly, it was trained using selected soil parameters related to the atmosphere (e.g. pH, nitrogen, phosphate, potassium, depth, temperature, and rainfall). A comparative study with the standard SOM was conducted. The improved Kohonen-Self Organizing Map when applied to correlated data, demonstrated better results in terms of classification accuracy (8/8), and rapidity = 0.015s compared to a classification accuracy (7/8) and a rapidity = 97,828 s using SOM combined with PCA. Moreover, the proposed algorithm resulted in better results for crop prediction in terms of maximum iteration number of 675, mean error  $\leq 0.00022$ , and rapidity = 18.422s versus an iteration number of 729, mean error  $\leq 0.000916$  and rapidity= 23.707s with the standard SOM. The proposed algorithm allowed us to overcome correlation issues, and to improve the classification, learning process, and rapidity, with the potential to apply for predicting crop yield in the agricultural field.

**Keywords**—Kohonen-self organizing map; gram-schmidt algorithm; principal component analysis; agriculture field; crop yield prediction

## I. INTRODUCTION

Preserving balanced agriculture is of paramount importance as it is considered the main source of nutrition for maintaining human life. Thus, the consumption of fruits and vegetables increases on a daily basis with the increase of population [1]. Recently, climate changes hamper the accurate prediction of the crop. However, the environmental changes influence the agricultural production leading to the minimization or a decrease of crop yields which is highly dependent on climate and rainfall conditions this decrease lead also to migration problem from developing countries because the agriculture field represent the source life of the most population [2]. For that, many researchers in agriculture fields started to develop machine learning (ML) algorithms to

improve agricultural production. Such as; bias-corrected random forest (BRF), multi-layer perceptron neural network (MLP), and support vector machines (SVM), were proposed to improve the estimation of agricultural drought in South-Eastern Australia [3]. Developed an estimation model of crop yield based on stepwise linear regression (SLR) and vegetation indices in order to predict crop yields [4]. Used a modified Self Organizing Map SOM based on Learning Vector Quantization (LVQ) for weather and crop prediction [5]. Another, study applied the boosted tree regression and artificial neural networks to forecast upland rice yield under climate change in a region (Sahel) with vulnerable weather and with very little capacity to adapt [6]. Earlier research [4, 5, 7] that targeted crop prediction using machine learning tools don't account for correlated data and dependence problems, in the preprocessing step which reduces classification rate and leads to an increased cost. Pushpa Mohan et al. [5], developed a coupling of SOM and LVQ to enhance the accuracy of prediction rate, in the case of using the correlated data, a decrease of accuracy and prediction rate will be noticed. The same limitations issues were observed in the cited literature. Therefore, developing an efficient new hybrid model for solving this issue and increase the accuracy of the prediction rate. The reduction of agricultural yield is a major common problem in Africa [8], especially in Morocco. For this purpose, in this work, we decide to create an intelligent system based on a new hybrid of neural networks and the Gram-Schmidt algorithm to combater this phenomenon and also to improve the classification, prediction accuracy and rapidity of our intelligent system. Hence, this intelligent system capable of predicting the crop yield for specific soil and atmosphere based selected parameters such as (pH, nitrogen, phosphate, potassium, depth, temperature, and rainfall), which might help to overcome the ecological changes, and to improve the agricultural production. Each vegetable or fruit can be characterized by certain soil parameters to ensure good production. Various models or architectures of artificial neural networks can be found in literature this includes, Kohonen-Self-Organizing Map, Conventional neural network (CNN), and Multilayer perceptron (MLP), etc. Each model or architecture can give very good results effective in a type of task where it was applied. The model proposed in this work, is the SOM this paradigm uses the unsupervised learning algorithm and it is one of the most popular artificial neural networks models, it applied in many industries from which [9,

\*Corresponding Author

10, 11, 12, 13], this model is intelligent to learn how to classify quantitative multi-parameter data using an unsupervised learning algorithm which allowed us to make a good prediction system. The goal of combining the SOM and GSHM algorithms is to improve the standard SOM. The GSHM algorithm is one of the fundamental procedures in linear algebra, the purpose of using it is to filter or to withdraw the most informative components for the sake more precise, eliminate inter-data correlation from the pattern, identify each object, and to obtain the new matrix contain novel factorizations or components. The improved SOM was trained by this new matrix, for making this training we used the correlated data from [14] and the crop prediction data collected and preprocessed from [7], the objective of using the correlated data is to test the robustness of our improved SOM in classification with the data that have a correlation problem between inputs. The empirical results showed that our new hybrid model or improved SOM proposed has a significantly better than standard SOM and hybrid model based on SOM and Principal component analysis (PCA) using the data that have a correlation problem because the PCA method can also resolve the problem of correlation and dependence between inputs, and for the crop prediction data, the outcomes showed that the improved SOM gave the best performance in terms of maximum iteration number, rapidity, and with high accuracy of mean error versus standard SOM.

This paper is organized as follows. In Section II, some techniques and methods intelligence applied in the agriculture field to solve the various issues about this domain are portrayed in the Literature review section. All methods and algorithms used in this work are presented and described in Section III. In Section IV, the experimental results of standard SOM and new hybrid or improved SOM. In Section V, we make the discussion about our work in terms of results and limitations. This followed by a section of the conclusion, we conclude the results obtained and give our perspectives for future research.

## II. LITERATURE REVIEW

In this section, we present the most current works in the agriculture field with intelligent methods by researchers in crop and weather prediction, cited in our introduction. In this study, some brief important works found in the literature are presented.

Feng, P., Wang, B., Liu, D. L., & Yu, Q, [3] examined three advanced methods of machine learning, i.e. Bias-Corrected Random Forest, Support Vector Machine, and Multi-layer Perceptron Neural Network for remote sensing drought indices, their objective were to determine numerous remotely-sensed drought factors might be used for evaluation agricultural drought in south-eastern Australia, they supported on thirty remotely-sensed drought factors from the Tropical Rainfall Measuring Mission (TRMM) and therefore Moderate Resolution Imaging Spectroradiometer (MODIS) to breed a ground-based drought index of Standardized Precipitation Evapotranspiration Index (SPEI) at some stage in 2001-2017. The results showed that the proposed methodology of bias-corrected random forest gave a far better performance than SVM and MLP for SPEI prediction.

Pushpa Mohan, Kiran Kumari Patil, [5] have made a combination between Self-Organizing Map (SOM) and Learning Vector Quantization (LVQ), this combination named Weighted-Self Organizing Map (W-SOM). In this work, the purpose of using the Weighted-Self Organizing Map is to predict the weather and crop for the Mysore region and to increase the prediction accuracy for rice production. In their experimental results, they found that the proposed approach of W-SOM improved the accuracy term in weather and crop prediction up to 0.5% than standard Self-Organizing Map, Ensemble Neural Network (ENN), and Kernel-Nearest Neighbors (KNN).

Zhang, L., Traore, S., Ge, J., Li, Y., Wang, S., Zhu, G. ... Fipps, G, [6] employed tree regression and artificial neural networks to modeled upland rice yield under climate change within the Sahel, for artificial neural networks, they used multilayer perceptron, probabilistic neural network, generalized feedforward, and linear regression, so the probabilistic neural network followed by boosted tree regression gives the best performance of the calibrated rice yield models.

Snehal S. Dahikar, Sandeep V. Rode, Pramod Deshmukh, [7] used artificial neural networks and in particular the paradigm of feed-forward backpropagation network to combat the climate change because it has a direct effect on crop production. In this paper, in order to make an intelligent prediction system, they used soil parameters and also the parameters related to the atmosphere (e.g. pH, nitrogen, phosphate, potassium, depth, temperature, and rainfall), for predicting a decent crop.

Han, J.-C., Huang, Y., Li, Z., Zhao, C., Cheng, G., & Huang, P, [9] proposed the Self-Organizing Map in groundwater level prediction (GWL), this prediction can contribute to maintaining reliable water supply in the various field (e.g. agriculture, domestic and especially in arid and semi-arid regions). The goal of this work was to apply the SOM methodology to determine spatially homogeneous clusters of GWL piezometers in order to make a good prediction. In this literature, the proposed modeling system has the capacity to take decision-making in order to inform the use of groundwater resources control, particularly in arid regions.

## III. METHODS

### A. Data Pre-Processing / Source

This work uses data from the literature [7] since our region from loukkos of Morocco does not have for this moment a data of crop-specific. Especially, this data are collected from Shri Shivaji Agriculture College, Amravati of Vidarbha region, India. This region is agricultural, and it is rich in a forest, among the important crop yields of this region are, cotton, oranges, and soybeans. Each crop has specific parameters of soil and weather to give a good agricultural production (e.g. ph, nitrogen, phosphate, potassium, depth, temperature, and rainfall) (see parameters in Table I). Another data contained the correlation problem from [14] is used in this work for testing the capability of our improved model in

classification. Hence, preprocessing it utilized in order to analyze the collected data.

### B. Kohonen-Self Organizing Map

The proposed methodology in this work named as standard Self-Organizing Map or Kohonen network is based on the artificial neural network. This method consisting of two layers, an input layer, and an output layer. An input layer or any pattern to be classified is represented by a multidimensional vector qualified as input vectors, and each pattern has an input neuron. An output layer is also called a competition layer where neurons compete. In SOM, the values from input neurons are passed on to all neurons in the competition layer, at the same time to approximate values passed by input neurons, and can imagine the output layer as a grid, (usually one or two-dimensional). In each node in the grid there is a "neuron" and each neuron is bound to a weight vector of the same dimension as input data vectors responsible for an area in the data space (again called input space).

In a Kohonen-Self-Organizing Map, weight vectors provide a discrete representation of the input space. They are positioned in such a way that they maintain the topological shape of the input space. By keeping neighborhood relations in the grid, they allow easy indexing (via coordinates in the grid). It can be seen from (Fig. 1) that each entry neuron has a link to all competition neurons.

The most famous learning types found in the artificial neural network are, unsupervised learning and supervised learning. SOM is a type of unsupervised learning. This type of learning is based on competitive and/or cooperative [15]. The algorithms based on unsupervised learning, are represented in particular to clustering algorithms, these latter used as an important data mining technique to group information set into a group of a similar object [16]. For that in this scenario, we propose it in order to improve it in terms of classification accuracy, learning process results, and make a good intelligent prediction system.

Before starting the learning procedures of SOM, the input data is stored as a matrix of (N-rows X M-columns) (see Fig. 2).

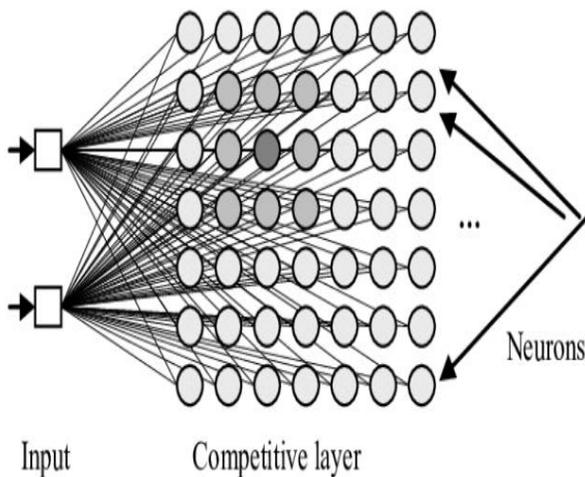


Fig. 1. Kohonen-Self Organizing Map [17].

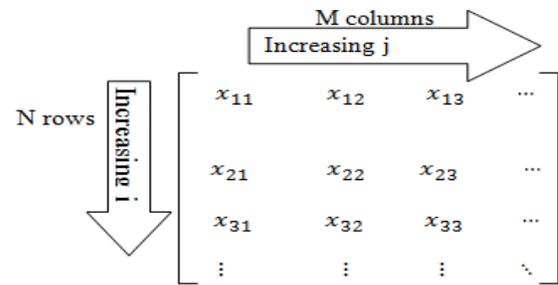


Fig. 2. Input Data Matrix.

Where, the rows are the objects and the columns are the components.

The inputs have to be average not too far from zero and variance not too far from 1. The input values should also not have too much weight. We can make some monotonous non-linear transformations that reduce the large values and minimize the number of training epochs of the neural networks, which leads to a better predictor and speed up learning. For that, in the first step, we add the algebraic formula of normalization [18] depicted in (1),

$$v_{ij} = x_{ij} / \sqrt{\sum_{i=1}^N x_{ij}^2} \quad (1)$$

Where,  $x_{ij}$  is the input objects and  $N$  is the number of variables inside the vector  $x$ .

Input vectors are connected with all nodes or neurons presented in the grid or map by weight vectors, the outputs of SOM methodology named the winning node. There are numerous functions for finding this winning node  $j(x)$ , (e.g. Euclidean distance, Product scalar, and Manhattan distance). In this work, we used the Euclidean distance function. The distance between the inputs vector and the weight vector for each node by applying the Euclidean distance is calculated in (2),

$$j(x) = \arg \min_j \| x(n) - w_j(n) \| \quad (2)$$

Where,  $j$  is the winner unit and  $j=1, 2, \dots, N$ ,  $n$  represent the number of node,  $w$  is the synaptic weight vector of the winner unit  $j$  and  $x$  is the input vector.

Hence, for updating the weights of winning node and neighborhood region, we have to apply the formula presented in (3) [19],

$$w_j(n+1) = w_j(n) + \alpha_j(n+1) h_{j,i}(n+1) (x(n) - w_j(n)) \quad (3)$$

Where, index  $i$  stands for the winning node and index  $j$  stands for other nodes presented in the grid,  $w_j(n+1)$  and  $w_j(n)$  are the synaptic weight vectors after and before updating in  $n$  and  $n+1$  iteration, the learning rate  $\alpha_j = \alpha_0 \exp(-t/\theta)$  in iteration  $t$ ,  $\alpha_0$  is the initial learning rate and neighborhood function  $h_{j,i}(n) = \exp(-(j-i)^2/\sigma_i^2(n))$ . This neighborhood function is centered at the wining node  $i$ .

C. Improved Kohonen-Self Organizing Map based on Gram-Schmidt Algorithm

To achieve the improvement of Kohonen-Self Organizing Map, we make the hybridization of two well-known algorithms of Gram-Schmidt and standard SOM. Thus, the development steps of the new hybrid model are described in the flowchart (Fig. 3).

Firstly, our data are passed through the GSHM processing block before it is processed with the SOM, in order to eliminate the correlation, dependence between inputs or objects, and to withdraw the most informative data. These capabilities of this new implementation will affect classification accuracy in the data that have the problem of correlation from the pattern and will affect on iteration number, mean error, and rapidity, for crop prediction data.

The Gram-Schmidt algorithm is an orthogonal space projection algorithm for finding orthogonal and orthogonal families in a Hilbert space, and it most widely used representative of a broad class of orthogonalization techniques and strategies we refer to [20-22]. In this work, we used a modified GSHM algorithm to identify each input object by  $A$ ,  $B$  factorizations or components.

To illustrate the GSHM algorithm, we consider the matrix (M) which includes the learning data as follows (see Fig. 4).

Each row of the original matrix (M) (Fig. 4) corresponds to a point in an n-dimensional space (the following Fig. 5- for n = 3). In this figure, points (1, 2, and m) in space correspond to the 1st, 2nd, and mth lines of the matrix (see Fig. 5).

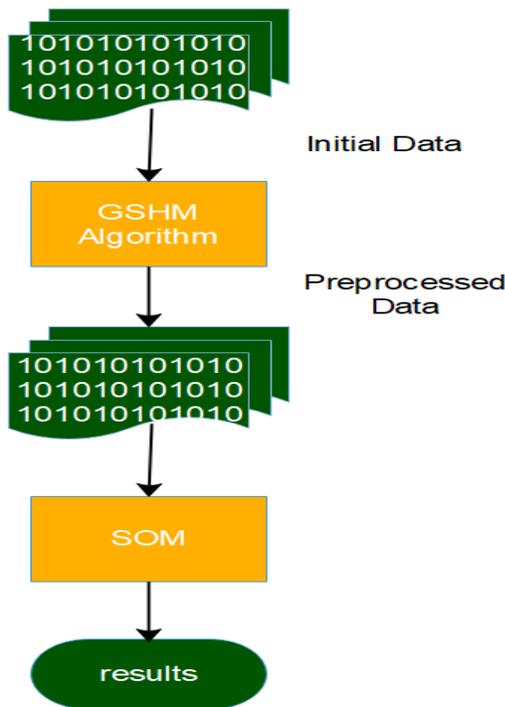


Fig. 3. Flowchart Design of our New Hybrid.

$$M = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{k1} & x_{k2} & x_{k3} \\ \dots & \dots & \dots \\ x_{m1} & x_{m2} & x_{m3} \end{bmatrix}$$

Fig. 4. Organization of our Original Data in a Matrix.

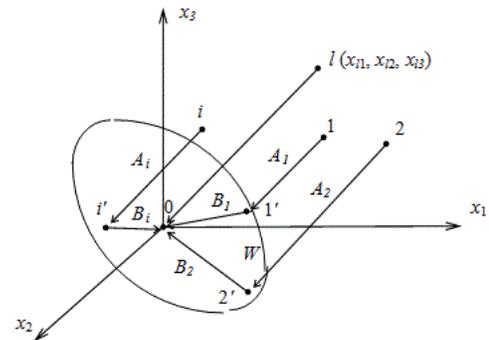


Fig. 5. Representation of Matrix Lines in 3D Space.

The procedure for calculating additional components begins with the choice of the point furthest from the initial coordinates (Fig. 5), that is, the row (S) of the matrix where the square sum of its elements is maximum. We draw a vector from the selected point (1) to the initial coordinates, and perpendicular to it passing through the origin of the coordinates. We draw a plan whose equation is as follows:

$$x_{11} * x_1 + x_{12} * x_2 + x_{13} * x_3 = 0 \tag{4}$$

The distances of an arbitrary point in a given space relative to the drawn plan are equal to the absolute value of (d) see the formula below:

$$d = \frac{x_{11} * x_1 + x_{12} * x_2 + \dots + x_{13} * x_3}{\sqrt{x_{11}^2 + x_{12}^2 + \dots + x_{13}^2}} \tag{5}$$

The values of the resulting expression for (d) correspond to the value of the component ( $A_l$ ) calculated from the following function:

$$A_l^{(j)} = \sum_{p=1}^m (M_{l,p}^{(j)} \cdot M_{S,p}^{(j)}) / \sum_{p=1}^m (M_{S,p}^{(j)})^2 \tag{6}$$

As  $|d|$  it is the distances between the bases points (1) and the initial coordinates, it can be demonstrated that the component ( $A_l$ ), is the relative distance between each point and the drawn plan see formula below:

$$M_{l,i}^{(2)} = M_{l,i}^{(1)} - A_l^{(1)} \cdot M_{s,i}^{(1)} \tag{7}$$

After calculating the component ( $A_l$ ), we will apply the previous formula only once to calculate the lines of the matrix that make up the projected points of the original matrix, the Gram-Schmidt orthogonalization algorithm agrees with the constructed plan.

The second component ( $B_l$ ) that will be calculated is the distance on the plan between the points projected on the plan

and the point of the initial coordinates. The formula for calculating the component ( $B_l$ ) is depicted in (8).

$$B_l = \sqrt{\sum_{i=1}^m (M_{li}^{(2)})^2} \quad (8)$$

Hence, in order to obtain the most informative components ( $A_l$  and  $B_l$ ), we have to develop all the previous steps above. For more details, following the steps presented in Fig. 6.

So, the new matrix  $M'$  obtained by GSHM algorithm is portrayed in Fig. 7.

// The procedure steps for generating the new matrix based on the Gram-Schmidt algorithm are,

1. Read the original data (matrix  $M$ );
2. Divide the number of columns / 2;
3. With the GSHM algorithm, calculate the value of the component ( $A$ ) and calculate the distances on the plan, between the points projected on the plan and the point of the initial coordinates ( $B$ );
4. Add the components ( $A$ ) and ( $B$ ) in original data and also keep them in a new matrix  $M'$ ;
5. Use the finally original data contained the first components ( $A$ ) and ( $B$ ) for calculating other components.
6. Repeat the step 3, 4 and 5 until to get new matrix  $M'$  contained the same columns numbers of matrix  $M$ ;
7. If the matrix  $M$  contained an odd number of columns, in this case, we can take only component ( $A$ ).

Fig. 6. The Steps for Generating the New Matrix by Gram-Schmidt Algorithm.

$$M' = \begin{pmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \dots & \dots & \dots \\ y_{k1} & y_{k2} & y_{k3} \\ \dots & \dots & \dots \\ y_{m1} & y_{m2} & y_{m3} \end{pmatrix}$$

Fig. 7. The New Matrix Obtained by GSHM Algorithm.

#### D. Principal Component Analysis

The principal component analysis (PCA) is a well-known and useful technique among the data analysis methods; it often applied in many applications [23, 24, 25]. The ultimate goal of PCA is to reduce and extract relevant information from original data, these proprieties allow resolving the correlation problem between inputs in data. For that PCA represent a robust tool especially in data mining and machine learning [26] in order to make the best analyze and taking the decision based upon the data.

### IV. RESULTS

#### A. Experimental Results with Standard Kohonen-Self Organizing Map

##### 1) Simulation using correlated data

In this section, we make our experimental test by using the data that have a correlation and regularities problem between

the pattern and also the crop prediction data. Firstly, the correlated data are collected from [14], in this work they developed an improved SOM based on PCA for solving the problem of correlation. Their data consist of 8 inputs objects or vectors contained (similar objects, objects have regularities between input and normal objects). So, we going, to training our standard SOM from these data. The results of the classification SOM between us and the last cited paper, are depicted in Fig. 8.

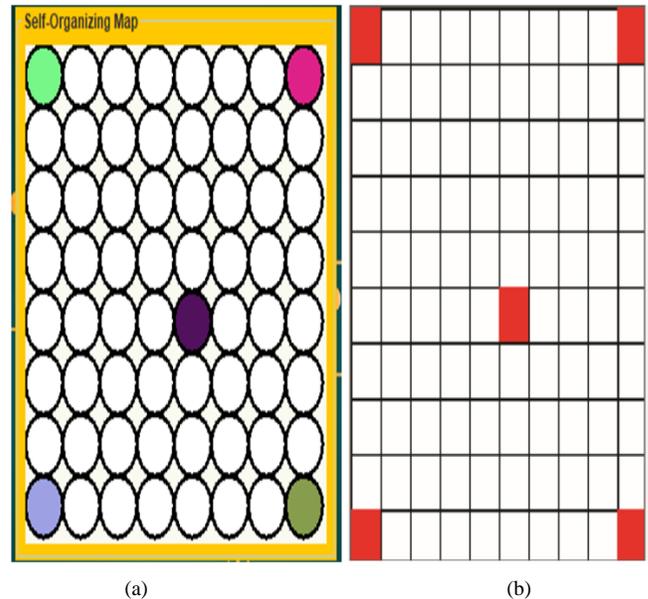


Fig. 8. (a) Our Classification Results, (b) Their Classification Results.

The figure above (Fig. 8) shows the results of the classification between our standard SOM and their standard SOM. The experimental results show that the standard Kohonen-Self Organizing Map can't perform a better classification of the data, it gave just five winning nodes in map for both of us, so our wining nodes are colored by a different color. We can conclude, that if we use data that have a problem of correlation in prediction, we can't get the best predictor system due to bad classification, this is among the goals for which we proposed our new hybrid or improved SOM for solving the classification problem in this type of data and to make the good prediction and to improve its accuracy.

##### 2) Simulation using crop prediction data (non-correlated inputs).

Secondly, we make another test using the crop prediction data, these data consist of 9 crops (cotton, sugarcane, jowar, bajra, soybeans, corn, rice, wheat, groundnut), and each crop has specific parameters to give a good prediction (e.g. pH, nitrogen, phosphate, potassium, depth, temperature, and rainfall). The essential parameters of the crop are presented in Table I.

In order to train our standard Kohonen-Self Organizing Map, we based on the training parameters (Table I). Now, we give the dimension for creating the map of SOM (see Fig. 9).

TABLE I. THE PARAMETERS RELATED TO THE ATMOSPHERE AND SOIL FOR CROPS (INPUT VECTORS)

Crops	Essential Parameters of Crop						
	PH	N	P	K	Depth	Temp	Rainfall
Cotton	8	100	50	50	30	33	900
Sugarcane	7	175	100	100	60	30	750
Jowar	8	80	40	40	50	30	1000
Bajra	7	40	20	25	15	28	600
Soybeans	7	30	75	15	19	27	800
Corn	8	100	25	0	40	20	500
Rice	7	100	50	50	15	22	90
Wheat	5	100	50	50	30	24	1400
Groundnut	6	25	50	30	20	24	1250

In the figure (Fig. 9) we gave (8X8) for map dimension, in next step we will insert the iteration rate (first stop condition), learning error (second stop condition), learning rate, and input data file (S\_matrice) for initializing and starting the learning step of SOM (see Fig. 10).

After finishing the learning, the proposed methodology of the standard Kohonen-Self Organizing Map gave nine winning nodes in the map (each input vector has a winning node). The outputs of the standard SOM or winning nodes are presented in Fig. 11.

The learning is completed when the two conditions mean error and the maximum iteration number is realized. Then, the winning neurons or nodes are colored with different colors. Thus, the coordinates of these winning nodes in the map are depicted in Fig. 12.

We noticed that the standard SOM classified correctly our data, which allowed us to make a good prediction system in this situation. So, the learning process results of the standard SOM are presented in Table II.

In the table II, we can conclude that the standard SOM is finished the learning by maximum iteration number = 729, mean error = 0.000916, learning number = 0.72900.

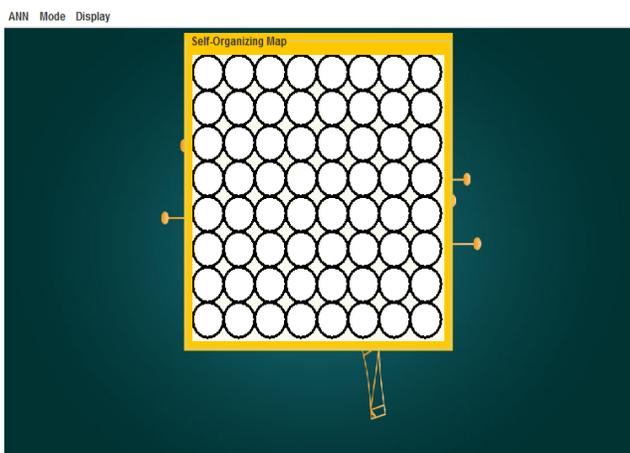


Fig. 9. Standard SOM Creation.

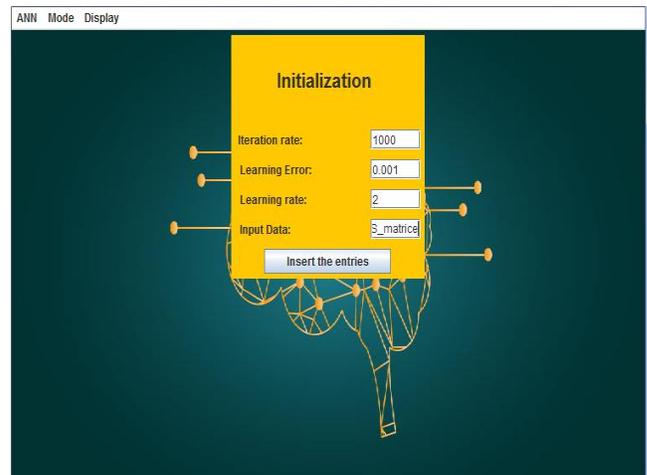


Fig. 10. Standard SOM Initialization.

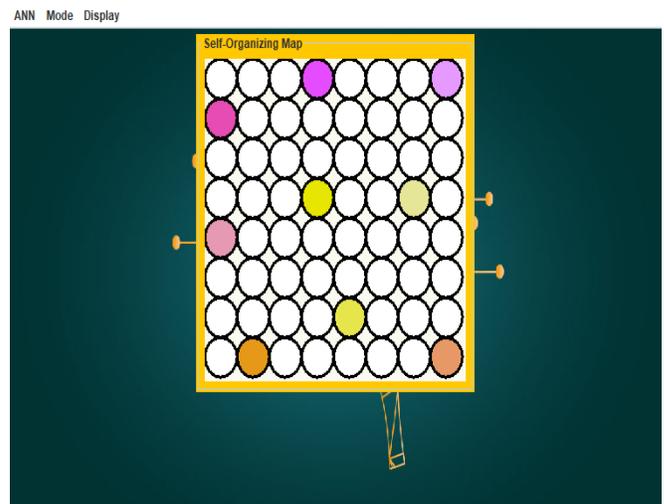


Fig. 11. Winning Nodes of Standard SOM.

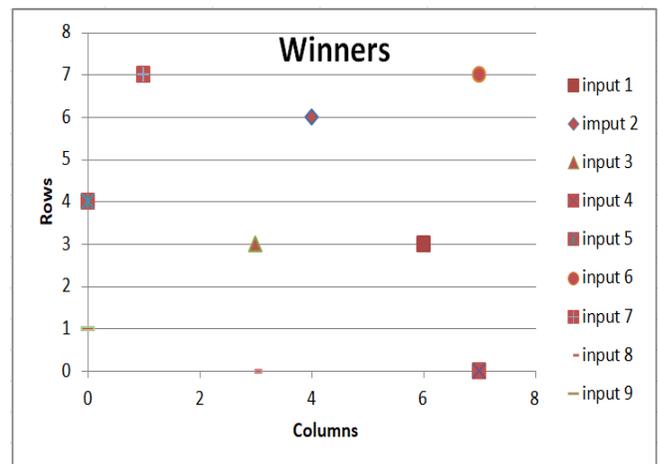


Fig. 12. Winning Nodes Coordinates of Standard SOM.

TABLE II. LEARNING PROCESS RESULTS OF STANDARD SOM

Iteration rate	1000
Maximum iteration number	729
Learning error	0.001
Mean error	0.000916
Learning rate	2
Learning number	0.72900
Rapidity	23.707s

Hence, our application is ready to predict the suitable crop associated with various parameters of soil and parameters related to the atmosphere (e.g. pH, nitrogen, phosphate, potassium, depth, temperature, and rainfall), now we will insert these seven parameters for predicting the decent crop. The results of the prediction are presented in Fig. 13.

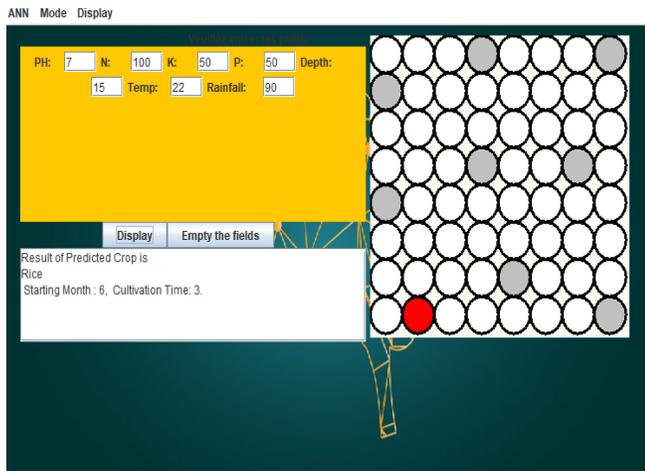


Fig. 13. The Predicted Suitable Crop is Rice.

The predicted crop based on these parameters pH=7, N=100, K=50, P=50, Depth=15, Temperature=22, Rainfall=90 is rice and the winning node of this input is colored by red color in the map. Additionally, our intelligent system gave a suitable starting month and cultivation time for this predicted crop.

In this situation, the crop data used in this test by SOM gave the best results in classification and prediction but with a big max iteration number and max error. Therefore, the objective of our improved SOM is to improve the standard SOM by minimizing iteration number, mean error, and also the rapidity.

**B. Experimental Results with Improved Kohonen-Self Organizing Map**

**1) Simulation using correlated data obtained by PCA and GSHM**

Now, we use correlated data obtained by the Gram-Schmidt algorithm and PCA for training the improved SOM based on PCA and GSHM, in this test we use the correlated date transformed by PCA from [14], in order to make the comparison in term of classification accuracy. The results of classification between SOM-PCA and GSHM-PCA as shown in Fig. 14.

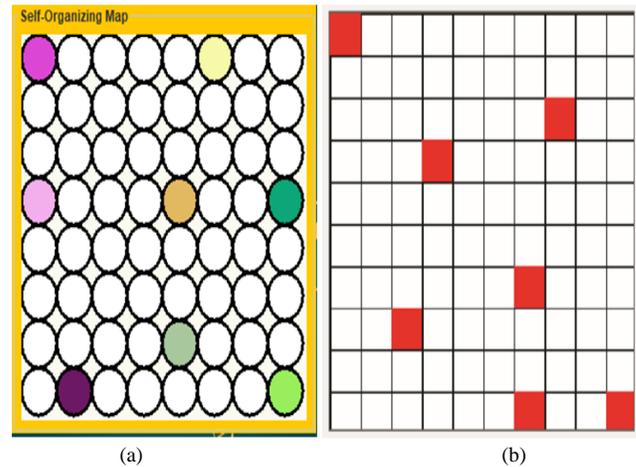


Fig. 14. (a) Improved SOM by Gram-Schmidt algorithm (classification results), (b) Improved SOM by Principal Component Analysis (classification Results).

In Fig. 14, our new hybrid of GSHM-SOM and their proposed model of SOM-PCA gave the best performance in classification accuracy, they can classify correctly the data and resolve the correlation problem by eliminating the dependence between inputs, but our new improved SOM based on the GSHM algorithm gave a high accuracy of classification (8/8) than SOM-PCA (7/8). Moreover, the rapidity of the Gram-Schmidt algorithm is 0.015s compared to Principal component analysis is 97,828 s, because the PCA is an iterative algorithm, for that it takes more time for giving the results against the GSHM is a linear algorithm that allows us to obtain the results in fast time.

**2) Simulation using crop prediction data obtained by GSHM**

The Gram-Schmidt algorithm is powerful than principal component analysis in classification and rapidity. For that, we proposed it in this work. Now, we will test the crop prediction data obtained by the GSHM algorithm for analyzing the experimental results. Although these data don't contain the correlation issues between inputs, our improved SOM will have an influence on the maximum iteration number, mean error, and rapidity. The data obtained by the GSHM algorithm are presented in Table III.

TABLE III. INPUT DATA OBTAINED BY GSHM ALGORITHM

Input Vector N <sup>o</sup>	Input Vectors Obtained by Gram-Schmidt Algorithm
1	[0, 907.51, 0, 914.97, 0, 922.37, 0]
2	[0.848, 107.85, 0.144, 188.58, 0.230, 178.65, 0.217]
3	[2.410, 6.869, 0.162, 18.222, 0.173, 18.799, 0.183]
4	[2.044, 9.666, 0.692, 7.801, 0.624, 7.830, 0.627]
5	[4.042, 3.558 E-12, 1.196, 1.663, 1.193, 1.860, 1.196]
6	[3.192, 57.992, 5.300, 5.641E-12, 4.520, 8.266, 4.615]
7	[0.841, 34.294, 6.944, 2.480, 6.582, 2.462E-12, 6.547]
8	[17.438, 52.244, 8.380, 26.950, 7.966, 13.357, 7.822]
9	[16.055, 30.509, 3.756, 44.583, 3.808, 28.923, 3.621]

After GSHM we got the most informative components about original data of 9 crops (Table III). Thus, we going to training our improved Kohonen-Self Organizing Map, through the new data that we got by the Gram-Schmidt algorithm. Hence, for starting the learning procedure of SOM, we used the same conditions of the first test (i.e. iteration rate, learning error, learning rate, and map dimension number) and input file data (M\_matrice) (see Fig. 15).

As a result, the improved SOM gave also nine winning nodes in the map (each input vector has a winning node), so the wining nodes of improved SOM are depicted in Fig. 16.

The winning nodes of improved SOM are colored by different colors. The coordinates of these winning nodes in the map are shown in Fig. 17.

As a conclusion, we noticed that the improved SOM classified correctly our data transformed by the Gram-Schmidt algorithm and gave a good classification. The learning process results of the improved SOM are displayed in Table IV.

Table IV shows that the new hybrid of improved SOM is stopped by 675 for maximum iteration number and by 0.00022 for means error versus standard SOM it stopped by 729 for maximum iteration number and by 0.000916 for mean error. For that, we can conclude that the coupling of the Gram-Schmidt algorithm and standard SOM has been applied successfully, in terms of maximum iteration number, mean error, rapidity, and precision of classification. These capabilities allow us to make a good prediction system in a fast time.

Our improved SOM is ready to predict the suitable crop by using various soil parameters and parameters related to the atmosphere obtained by the GSHM algorithm, now we will insert these parameters for predicting the decent or suitable crop. The operation for predicting the crop yield is shown in Fig. 18.

The predicted crop based on these parameters pH=0.841, N= 34.294, K= 6.944, P= 2.840, Depth= 6.582, Temperature = 2.462.  $10^{-12}$ , Rainfall= 6.547 is rice and the winning node of this input object is colored by red color. In addition, our intelligent system gave a suitable starting month and cultivation time for this predicted crop (Fig. 18).

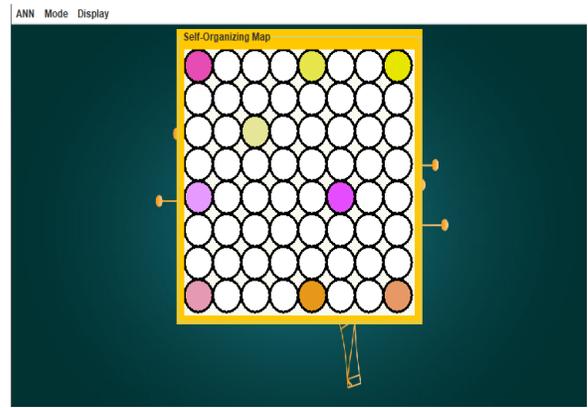


Fig. 16. Winning Nodes of Improved SOM.

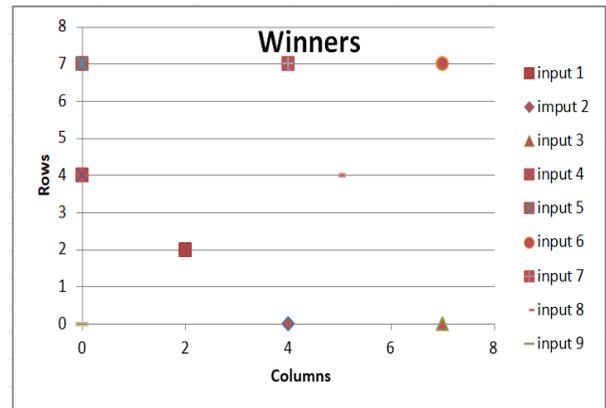


Fig. 17. Winning Nodes Coordinates of Improved SOM.

TABLE IV. LEARNING PROCESS RESULTS OF IMPROVED SOM

Iteration rate	1000
Maximum iteration number	675
Learning error	0.001
Mean error	0.00022
Learning rate	2
Learning number	0.78567
Rapidity	18.422s

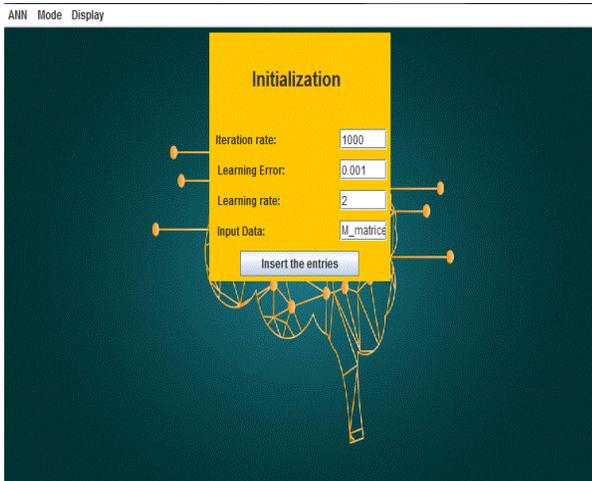


Fig. 15. Improved SOM Initialization.

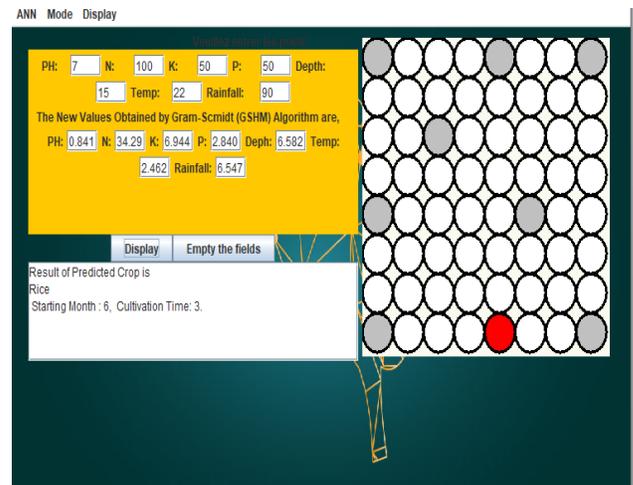


Fig. 18. The Predicted Suitable Crop is Rice.

### C. A Comparison between Standard SOM and Improved SOM

In this section, the maximum iteration number and mean error of standard SOM are compared with the maximum iteration number and mean error of improved SOM.

The comparison between standard SOM and improved based on crop prediction data in terms of maximum iteration number and mean error for every two columns from the original matrix  $M$  and new matrix  $M'$ , is displayed in Fig. 19 and Fig. 20.

Fig. 19 determines the maximum iteration numbers graph comparison through our original matrix  $M$  of standard SOM and new Matrix  $M'$  obtained by the Gram-Schmidt algorithm of the improved SOM in all different two columns. The result shows, that we got an increasing and decreasing in iteration number between columns 2, 4, 6, 7 for original matrix  $M$  and new matrix  $M'$ , but generally, the new matrix  $M'$  of 7 columns gave the best maximum iteration number by 675 versus 729 of maximum iteration number in original matrix of 7 columns. In conclusion, we can say that our proposed new hybrid is not only based on iteration numbers for columns but is based to solve the correlation problem in the multitude of learning like the linear dependence pattern components. Knowing that our new hybrid, allowed us to develop a linear algorithm in contrary to the algorithm published in [27].

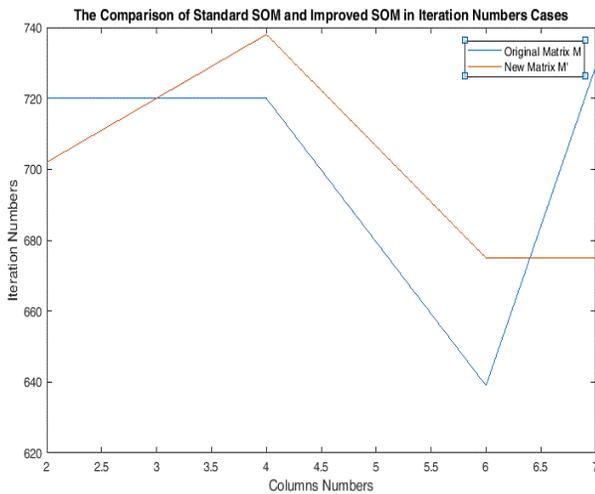


Fig. 19. Comparison in Terms of Iteration Numbers.

Fig. 20 specifies that the improved SOM has better in terms of mean error compared to standard SOM because we have significantly a decreasing of mean error in all two columns for new matrix  $M'$  of improved SOM than original matrix  $M$  of standard SOM, this means demonstrated that the proposed new hybrid has high accuracy.

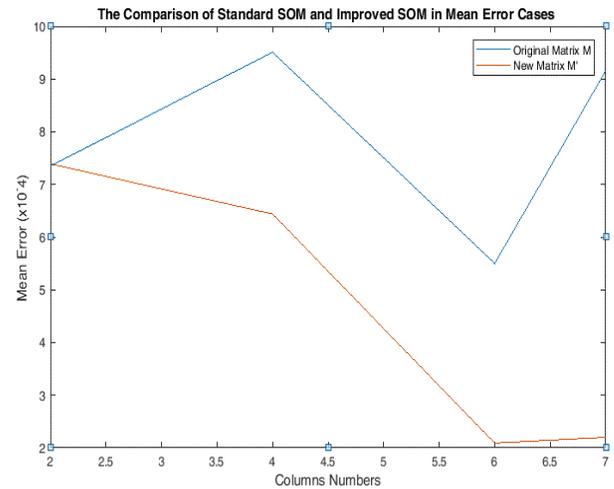


Fig. 20. Comparison in Terms of Mean Error.

### V. DISCUSSION

Artificial neural networks offer very powerful and effective methods, capable of performing complicated tasks, besides, it has the properties not found in the classical methods (e.g. learning ability, adaptation, classification, and generalization of results). Thus, these capabilities allow us to solve various issues and tasks requiring intelligence. For that we proposed the new approach of ANN to improve it, in order to be more powerful to apply it in numerous problems facing several fields, especially in the prediction system.

This work proposed a new hybrid model of Kohonen-Self Organizing Map and Gram-Schmidt algorithm in order to improve it at the level of classification accuracy, maximum iteration number, mean error, and rapidity, this study showed that the improved model gave the high accuracy of classification when we used the correlated data and also it minimizes the number of iteration, mean error with high accuracy, and speed up our intelligent system when we use the crop prediction data compared to standard SOM. These powerful of improved SOM can solve the most of problems that we may encounter in the prediction field, which allow us to make a good predictor system for predicting the suitable crop by using various parameters of soil and parameters related to the atmosphere. We can say that the GSHM algorithm represents a powerful tool for machine learning, especially for SOM.

The results of this study couldn't be directly compared to previous literature, as we use different metrics to evaluate the quality of the developed model and very limited studies focused on correlations issues. However, some of the results such as classification and rapidity could be indirectly compared to a previous study by our group [14]. The author targeted correlation problems in the input data using SOM combined with PCA. The proposed model presented here demonstrated better results in terms of classification and computational cost.

However, they are some limitations that need to be acknowledged, this includes the problem of collecting real data in morocco related to crop prediction. In addition, it is difficult to find big data with correlated inputs.

## VI. CONCLUSION

This paper illustrated a prediction of the crop by various parameters of soil for the Vidarbha region, this prediction system is realized by our new proposed methodology of Kohonen-Self Organizing Map. In this study, the improved SOM based on the Gram-Schmidt algorithm was implemented to avoid problems related to the correlation issues and speed up the learning process results with high accuracy. The improved SOM gave the best performance than standard SOM in terms of classification, maximum iteration number, mean error, rapidity and it was efficient in prediction. In future work, for further improving the results obtained by improved SOM and also we will use the data of our region loukkos, Morocco; this region is rich in fruits and vegetables (e.g red fruits, potatoes, tomatoes, etc.). And to solve several issues that face farmers.

## REFERENCES

- [1] Tanha Talaviya, Dhara Shah, Nivedita Patel, Hiteshri Yagnik, Manan Shah, Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides, Artificial Intelligence in Agriculture, 2020, ISSN 2589-7217, <https://doi.org/10.1016/j.aiia.2020.04.002>.
- [2] Falco, C., Galeotti, M., & Olper, A. (2019). Climate change and migration: Is agriculture the main channel? *Global Environmental Change*, 59, 101995. doi:10.1016/j.gloenvcha.2019.101995.
- [3] Feng, P., Wang, B., Liu, D. L., & Yu, Q. (2019). Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. *Agricultural Systems*, 173, 303–316.
- [4] O. Satir and S. Berberoglu, "Crop yield prediction under soil salinity using satellit derived vegetation indices", *Field Crops Research*, Vol.192 , pp.134-143, 2016.
- [5] Pushpa Mohan, Kiran Kumari Patil, Weather and Crop Prediction Using Modified Self Organizing Map for Mysore Region, REVA University, Computer Engineering, India, December 23-2017, *International Journal of Engineering and Intelligent Systems*, Vol.11, No .2, 2018.
- [6] Zhang, L., Traore, S., Ge, J., Li, Y., Wang, S., Zhu, G., ... Fipps, G. (2019). Using boosted tree regression and artificial neural networks to forecast upland rice yield under climate change in Sahel. *Computers and Electronics in Agriculture*, 166, 105031.
- [7] Snehal S. Dahikar, Sandeep V. Rode, Pramod Deshmukh An Artificial Neural Network Approach for Agricultural Crop Yield Prediction Based on Various Parameters, *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)* Volume 4, Issue 1, January 2015.
- [8] Sultan, B., Defrance, D. & Iizumi, T. Evidence of crop production losses in West Africa due to historical global warming in two crop models. *Sci Rep* 9, 12834 (2019). <https://doi.org/10.1038/s41598-019-49167-0>.
- [9] Han, J.-C., Huang, Y., Li, Z., Zhao, C., Cheng, G., & Huang, P. (2016). Groundwater level prediction using a SOM-aided stepwise cluster inference model. *Journal of Environmental Management*, 182, 308–321. doi:10.1016/j.jenvman.2016.07.069.
- [10] Li, Y., Wright, A., Liu, H., Wang, J., Wang, G., Wu, Y., & Dai, L. (2019). Land use pattern, irrigation, and fertilization effects of rice-wheat rotation on water quality of ponds by using self-organizing map in agricultural watersheds. *Agriculture, Ecosystems & Environment*, 272, 155–164.
- [11] Wang Pan, Yimin Sun, Michela Turrin, Christian Louter, Sevil Sariyildiz, Design exploration of quantitative performance and geometry typology for indoor arena based on self-organizing map and multi-layered perceptron neural network, *Automation in Construction*, Volume 144, 2020, 103163, ISSN 0926-5805, <https://doi.org/10.1016/j.autcon.2020.103163>.
- [12] Mehmet Ozcalici, Mete Bumin, An integrated multi-criteria decision making model with Self-Organizing Maps for the assessment of the performance of publicly traded banks in Borsa Istanbul, *Applied Soft Computing*, Volume 90, 2020, 106166, ISSN 1568-4946, <https://doi.org/10.1016/j.asoc.2020.106166>.
- [13] Shin Nagai, Hiroshi Morimoto, Taku M. Saitoh, A simpler way to predict flowering and full bloom dates of cherry blossoms by self-organizing maps, *Ecological Informatics*, Volume 56, 2020, 101040, ISSN 1574-9541, <https://doi.org/10.1016/j.ecoinf.2019.101040>.
- [14] El Khatir HAIMOUDI, Loubna CHERRAT, Hanane FAKHOURI & Mostafa Ezziyani. (2016). Towards a New Approach to Improve the Classification Accuracy of the Kohonen's Self Organizing Map During Learning Process. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 3.
- [15] Lee, M., Song, T.-G., & Lee, J.-H. (2020). Heartbeat classification using local transform pattern feature and hybrid neural fuzzy-logic system based on self-organizing map. *Biomedical Signal Processing and Control*, 57, 101690. doi:10.1016/j.bspc.2019.101690.
- [16] Wang, B., Kong, Y., Zhang, Y., Liu, D., & Ning, L. (2019). Integration of Unsupervised and Supervised Machine Learning Algorithms for Credit Risk Assessment. *Expert Systems with Applications*.
- [17] [https://www.researchgate.net/figure/Kohonen-network-architecture\\_fig1\\_224355225](https://www.researchgate.net/figure/Kohonen-network-architecture_fig1_224355225)
- [18] Bazrafkan, A., & Pakravan, M. R. (2017). An MADM network selection approach for next generation heterogeneous networks. 2017 Iranian Conference on Electrical Engineering (ICEE). doi:10/1109/iranianee.2017.7985361.
- [19] S. Hosseini, "Binary Tree Time Adaptive Self-Organizing Map", *Neurocomputing*, Vol.74, No.11, pp.1823–1839, May 2011.
- [20] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, PA, (1996)
- [21] G.H. Golub and C.F. VanLoan, *Matrix Computations*, Third Edition, John Hopkins University Press, Baltimore, MD, (1996).
- [22] N. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, PA, (1996).
- [23] Christine Borgen Linander, Rune Haubo Bojesen Christensen, Graham Cleaver, Per Bruun Brockhoff, Principal component analysis of d-prime values from sensory discrimination tests using binary paired comparisons, *Food Quality and Preference*, Volume 81, 2020, 103864, ISSN 0950-3293, <https://doi.org/10.1016/j.foodqual.2019.103864>
- [24] Seyed Ali Ghorashi, Farhang Honarvar, Morteza Tabatabaeipour, Automated extraction of local defect resonance using the principal component analysis in lock-in ultrasonic vibrothermography, *Infrared Physics & Technology*, Volume 105, 2020, 103204, ISSN 1350-4495, <https://doi.org/10.1016/j.infrared.2020.103204>
- [25] Yu Zhang, Xiaobo Tian, Rongguang Liang, Phase extraction from two randomly phase shifted interferograms by combining advanced principal component analysis and Lissajous ellipse fitting, *Optics and Lasers in Engineering*, Volume 132, 2020, 106134, ISSN 0143- 8166, <https://doi.org/10.1016/j.optlaseng.2020.106134>.
- [26] Qianqian Wang, QuanXue Gao, Gan Sun, Chris Ding, Double robust principal component analysis, *Neurocomputing*, Volume 391, 2020, Pages 119-128, ISSN 0925-2312, <https://doi.org/10.1016/j.neucom.2020.01.097>
- [27] Haimoudi El Khatir, Tsymbal Yu.V. Method of preprocessing of realization set for counter propagation neural networks, *vidbir-zm-2003\_19e*, p 154-158.

# A Secured Large Heterogeneous HPC Cluster System using Massive Parallel Programming Model with Accelerated GPUs

Khalid Alsubhi

Faculty of Computing and Information Technology  
King Abdulaziz University, Saudi Arabia

**Abstract**—High Performance Computing (HPC) architectures are expected to develop first ExaFlops computer. This Exascale processing framework will be proficient to register ExaFlops estimation every subsequent that is thousands-overly increment in current Petascale framework. Current advancements are confronting a few difficulties to move toward such outrageous registering framework. It has been anticipated that billion-way of parallelism will be exploited to discover Exascale level secured system that provide massive performance under predefined limitations such as processing cores and power consumption. However, the key elements of the strategies are required to develop a secured ExaFlops level energy efficient system. This study proposes a non-blocking, overlapping and GPU computation based tri-hybrid model (OpenMP, CUDA and MPI) model that provide a massive parallelism through different granularity levels. We implemented the three different message passing strategies including and performed the experiments on Aziz-Fujitsu PRIMERGY CX400 supercomputer. It was observed that a comprehensive experimental study has been conducted to validate the performance and energy efficiency of our model. Experimental investigation shows that the EPC could be considered as an initiative and leading model to achieve massive performance through efficient scheme for Exascale computing systems.

**Keywords**—High Performance Computing HPC; MPI; OpenMP; CUDA; Supercomputing Systems

## I. INTRODUCTION

Since last three decades, High performance computing (HPC), played a fundamental role in scientific endeavour where vendors emphasized to improve system performance by dramatic increasing through on-chip parallelism. According to Top-500 supercomputers list, an improvement of 10x in system performance is discovered after every 3.6 years [1]. A supercomputer in 2012, Titan Cray XK7 was capable to achieve 18 PFLOPs under the 8.3 MW power consumption [2]. Moving on the vision to enhance system performance to solve the complex problems, Tianhe-II the current supercomputer manufactured by NUDT is capable to deliver 55.2 PFLOPs with 17MW power consumption [3]. The demand of computation for solving complex problem envisioned to develop new supercomputer [4]. This extraordinary scale processing framework will be proficient to compute 1018 FLOPS activities for each subsequent that is thousand-crease increment in current Petascale framework. As per expectations, Exascale figuring framework will be involved countless

heterogeneous process hubs connected by complex systems [5]. The essential issue for HPC frameworks is that such Extreme (Exascale) processing framework doesn't exist yet, anyway everything toward Exascale is simply expectations and contemplations. To improve the system throughput, the trend has been changed from traditional way of doubling clock speeds by doubling number of cores, threads or other parallelizing mechanisms [4]. However, it is predicted that millions of cores of heterogeneous devices including CPUs and GPUs will be comprised by the Exascale computing system.

## A. Exascale Computing Limitations and Challenges

As indicated by the innovation and programming approaches that are being utilized in existing Petascale registering framework, the power consumption is about 25 to 60 MW by utilizing 30 M number of centres. The interest of intensity utilization for Exascale registering framework will be more than 130 Megawatts [6]. United State Department of Energy characterized some essential limitations such as Power Consumption roughly 20-30 MW, Development Cost (D.C) up to 200 M US dollars, Delivery Time (DT) till 2020, and Cores about 100 Million [7]. However, development of targeted Exascale Supercomputer under the delimitation of these constraints is the tremendous challenge for vendors and development communities.

Leading to the massive powerful computing system, there are several challenges which are still the blockage for development toward emerging HPC systems. In [7], some primary Exascale computing challenges discussed are presented in Table I. For 21st century, these imperative difficulties are the basic way to create innovatory answers for Exascale figuring framework. Nonetheless, an emotional reformulation at both equipment and programming levels, programming models, vitality proficient strategies, investigating apparatuses and overhaul calculations are requested to accomplish the calculation in ExaFlops [8]. Since last few years the development process for Exascale computing system is being rapidly fast. Under these listed challenges, many new approaches have been proposed.

## B. Software Technology Navigation

In current study, our contribution is related to challenges 1, 2, and 5 from Table I to improve the system performance through efficient and massive parallelism under minimum power consumption. From software perspectives, still it has not

been determined that at what level [9], the software framework is adoptable to achieve massive parallelism for Exascale computing systems. The recent energy efficient GPU technology introduced by NVIDIA outperforms the traditional processing on CPU cores [34, 35, 36]. Therefore, involving GPU accelerated computation in system, the hierarchy level of programming model navigation is shown in Fig. 1.

According to this navigational model, the anticipation is going to be that Tri-level model outperforms much better than traditional single or dual models [10]. It provides massive parallelism where energy efficient accelerated devices (GPGPU) collaborate with other models that deal with fine-grain and coarse-grain parallelism.

The rest of paper is organized as follows. Section II related work describes the existing state-of-art-approaches at Single, Dual, and Tri levels. Further Section III depicts a comprehensive overview of proposed EPC model, its features and components. Section IV, presents the experimental platform and applications used to evaluate EPC model. Last Section V concludes and explains the results in term of summary.

TABLE I. EXASCALE COMPUTING CHALLENGES

Challenges	Description
Power Consumption Management	Power consumed by the system and its management
Novel Architectures	New non-conventional architectures to support Exascale frameworks
Memory Technology	Memory management and storing systems to support massive storage.
Scalable System Software	Scalable and resilience system are needed to support sudden power fluctuation
Programming Systems	Novel programming techniques to support parallel programming libraries and frameworks.
Data Management	Efficient data management approaches are demanded.
Exascale Algorithms	New algorithms should be proposed to manage massive parallelism and advance programming.
Discovery and Design Algorithms	Discovery should be facilitated by mathematical models.
Resilience & Correctness	Faults and verification challenges should be addressed.
Scientific Productivity	Scientific productivity is necessary to through novel software tools.
Power Consumption Management	Power consumed by the system and its management

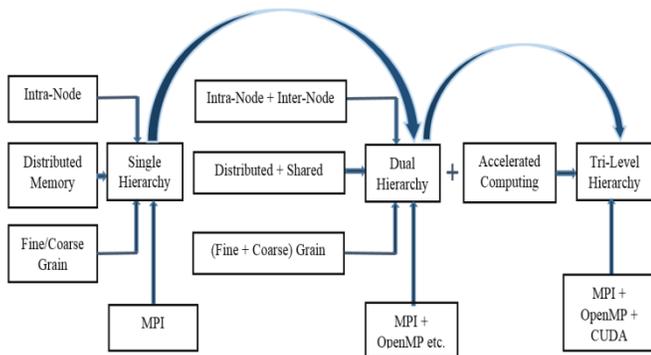


Fig. 1. Hierarchy Level of Programming Model Navigation.

## II. RELATED WORK

Pushing toward HPC (High Performance Computing), equipment and programming rising advances have been examined toward Petascale registering framework in [11]. Prompting Petascale figuring framework, numerous equipment point of view methods where studied such as Conventional innovation, Preparing In-Memory structures (PIM), Digital superconductor advances, Computation Fluid Dynamics (CFD), Special-reason equipment, Web-based Petascale Computing, atomic nanotechnology and insightful planetary rocket and so on [12]. An information parallel programming language with respect to procedures for Petascale framework were proposed [13]. These models where capable to gain parallelism for both course grain and fine grain level using traditional homogenous system on multicore CPU devices [14].

In the end of recent decade, to bring scalability in system, technology trend was changed from traditional homogenous to heterogeneous cluster system where many-core devices were introduced such as General Purpose Graphics Processing Unit (GPGPU), Graphics Processing Unit (GPU) by NVIDIA [15] and MIC (Many Integrated cores) by Intel [16]. These accelerated devices are based on Single Instruction Multiple Data (SIMD) from Flynn’s classification. Beyond these accelerated devices, many parallel programming models have been proposed such as CUDA, OpenACC, and OpenCL. It has been anticipated these parallel programming models could be promising to achieve massive parallelism required for future Exascale computing system [17]. In any case, to use such incredible gadgets and models, a key component of the methodology is the co-structure of uses, designs and programming conditions at both equipment and programming level.

According to development to HPC Exascale computing system, China has a fast development towards HPC systems and consequently they introduced Tianhe II HPC system recently in 2014 [18]. Further they introduced the upgraded version named as Tianhe III [19]. Similarly, DEEP (Dynamical Exascale Entry Platform) by European Union in 2011 [20] started effort toward a new HPC Exascale computing system. SERT project funded by NAG took initiative to introduce first Exascale computing system in 2020 [21, 22]. In Japan, RIKEN [23] claimed to present first Exascale computing in start of 2020. Further, Indian Government also started Exascale computing development since 2018 and claimed to introduce in 2022 [24].

## III. PRELIMINARIES

MPI has many different schemes that can be used to program a cluster system. Traditionally, two prevalent methods MPI blocking (synchronous) and non-blocking (asynchronous) are being used to distribute data over a cluster system [25, 26, 27]. In legacy systems, the whole processing was performed by CPU cores using MPI blocking method. Consequently, the processing over CPU cores was very costly with respect to energy consumption and processing efficiency. Therefore, new SIMD (single instruction multiple data) based energy efficient devices (GPUs, MIC) were introduced that contains thousands of cores on it. These cores compute data in parallel and consequently, reduce processing time and power consumption.

Due to parallel computation, data processing over GPU cores is very fast which required a rapid data input. In this way, MPI non-blocking is appropriate approach to fully utilize these powerful devices and achieve maximum performance. In current study, we discussed three fundamental MPI non-blocking schemes as follows:

#### A. (S1)- MPI Non-Blocking, no Overlapping Computation

In first strategy 1 (S1) MPI non-blocking and no overlapping implemented scheme, computation does not overlap during data processing [28]. This scheme performs just like a blocking mechanism where all resources are reserved until the whole processing is completed. One disadvantage of this scheme is that many resources are reserved even though they finished their assigned tasks. Although MPI communication is capable to overlap with CUDA, but we avoided from overlapping in this implementation. During exchanging data from multiple arrays, MPI scatter and gather data for one edge while memory copying operation is proceeding for other components.

#### B. (S2)- MPI Non-Blocking, Overlapping Computation

The second implemented strategy for data distribution was (S2) MPI non-blocking but overlapping computation where CUDA copying operation was overlapped with MPI communication. In this strategy, CUDA kernel was decomposed into three portions where top and bottom edges were done from the middle. In such way, kernel was started with the edges which are going to be computed, rather than start exchanging on entire domain. Following non-blocking MPI mechanism, first portion started copying operation from device to host. Immediately after completing copy operation to host, middle portion of the domain started computation. Similarly, last part of exchanging operation started as soon middle portion complete its computation. This implementation strategy can be more significant by improving the overlapping computation of middle portion.

#### C. (S3)-MPI Non-blocking, Overlapping & GPU Computation

The final implementation was MPI (S3) non-blocking with highest amount of overlapping which is anticipated the best performing strategy for large scale cluster system [29]. Using asynchronous method, CUDA streams were enabled and started computation from middle portion that cause to for massive overlapping, MPI communication and memory operations. The important thing in this strategy is that, a very small level of changes is needed inside the CUDA kernels to perform the computations. In order to optimize the GPU threads, a flag along with grid size and number of blocks is broadcasted over the kernels to indicate a specific portion for computation.

### IV. EFFICIENT PARALLEL COMPUTING MODEL

We presented the proposed EPC model implemented in C++. Based on the predicted Exascale computing system, EPC model was categorized into three different computing environments including cluster system, compute node, and GPU computing. Each environment contained a separate layer of parallelism as presented in Fig. 2.

Programmer interacts with EPC model through the application written in C++. Before entering in parallelism zone, data is analyzed by the programmer himself statically to know that, which statement can be parallelized. Once data is analyzed and ready for parallel computation, it entered in parallel computing zones as described in following sections.

#### A. Inter-Node Computation Layer

The primary degree of parallelism of the model was accomplished between hub correspondences. In view of these parameters, developer break down and appropriate over associated framework hubs utilizing an institutionalized SIMD based Message Passing Interface (MPI) library [30]. MPI blocking (synchronous) and non-blocking (no concurrent) two pervasive components are being utilized to move and assemble information over the processors. Blocking systems is utilized when a solid synchronization is required because of reliance in information. For this situation, the assets are held utilizing some pre-characterized MPI holding up explanations until the handling is finished. In our parallel registering system, information is required just to convey over the processors that subsequently gives coarse-grain parallelism at this level, along these lines we chose "non-blocking, covering with GPU calculation" the third MPI non-blocking technique as talked about in past segment. In this procedure, when information is moved no concurrently over associated hubs, it entered in second degree of parallelism portrayed in following area.

#### B. Intra-Node Computation Layer

The proposed model provides the second level of parallelism at Intra-node computation. At this level, the distributed data through MPI processors is further communicated with CPU threads for parallel processing. At this stage, OpenMP pragmas are used that parallelize the blocks of code either fine grain or course grain computation. OpenMP threads use the system specified threads over CPU cores and complete the executions. According to new OpenMP version, we can use multiple OpenMP pragmas for multiple blocks within single block that is the reason to achieve fine grain parallelism in the code.

#### C. GPU Computation Layer Acceleration

The last level of parallelism in our proposed model is Intra-node computation. In this layer of computation, the whole processing is performed on accelerated GPU devices. In this strategy, firstly the data is transferred from CPU cores to GPU that further distributed over GPU Warps. According to the structure of GPU each warp contains 32 number of cores where the number of warps can be different from GPU structure to structure. Once the data is transferred over GPU cores, GPU kernel divide the tasks into multiple GPU warps and perform all the operations in parallel. To perform the GPU computation, we can utilize the different accelerated devices such as NVIDIA GPU, AMD GPU etc. for current study, to maintain the maximum support for C++, we selected NVIDIA GPU and implemented accordingly.

In the past, low overlapping between CPU and GPU caused the wastage of resources where GPU threads remain in idle state until the processing from other kernels is not accomplished. Usually, this inefficiency factor was found in

MPI non-blocking non-overlapping and non-blocking low-overlapping strategies that consequently waste resources utilization and decrease system efficiency.

Although MPI non-blocking was implemented in existing design as shown in Fig. 3(a) but waiting state for kernel and separate progress effected in decreasing efficiency. In each broadcasting, Isend() function/method has performed in three states including kernel initialization, kernel waiting and start sending data. During Isend() from these states, kernel stream was reserved. Once first kernel stream is complete, next one

start processing. In such way, each stream waste resource utilization during waiting state. Conversely, in our proposed design, we organized these three states for every broadcasting Isend() in such way that kernels were overlapped and initialized immediately after one. Therefore, all kernel streams are now overlapping and can start processing as soon it receives data. A minor waiting state is ignorable because data sending process can be started as soon it complete its previous stage. Fig. 3(b) shows a clear benefit of proposed design that minimize delay in processing and increase efficiency through higher overlapping.

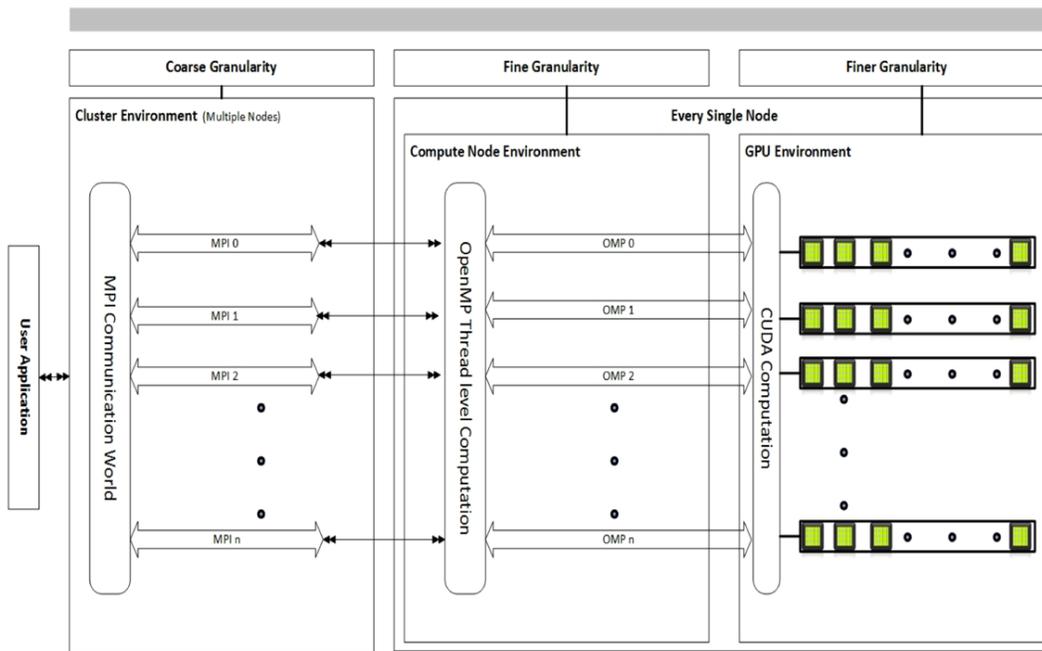


Fig. 2. EPC: A Hybrid Parallel Computational Model.

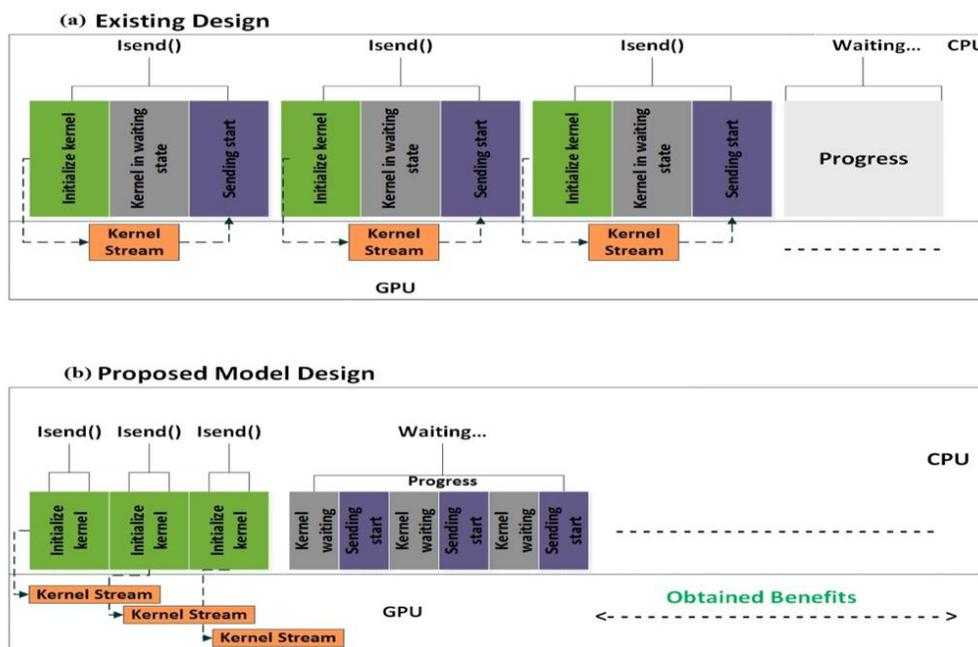


Fig. 3. Overlapping: Existing vs Proposed Design.

## V. EXPERIMENTAL SETUP

### A. Platform

To perform the experiments, we have used the Aziz supercomputer the 360th positioned in 2015 top supercomputers placed in High Performance Computing Centre (HPCC), King Abdulaziz University. The Aziz supercomputer contains Xeon CPU processors along with GPU devices [31]. Aziz comprises of complete 11904 number of cores on it including both CPU and GPU cores. Regarding the Aziz memory, 96GB hubs and 256 GB individually configured in it where each hub consists of individual processor -2.4 GHz and 12 Cores- controlled CentOS 6.4 working framework. All the nodes are connected with Infini-band medium to make the communication more efficient. With respect to overall efficiency, Aziz supercomputer is very powerful that is able to accomplish about 211 TFlops/s Linpack execution and about 228 TFlops/s overall [32].

### B. Performance Measurement

The primary factor in High performance computing systems is Performance [33]. Conventionally, the performance of a computer system is calculated in number of Flops by attaining the peak performance and the number of flops against the targeted application execution as described in equation 1. If we consider that  $F_p$  are the flops at peak floating point and  $F_m$  are the number of flops against targeted application, therefore  $F_c$  can be determined as:

$$F_c = \frac{F_p}{F_m} \quad (1)$$

Using the Aziz peak performance Aziz, we have quantified the performance by executing targeted HPC applications at different datasets described in following sections.

### C. Power Measurement

The second most important metric is the power consumption which is the primary challenge for current and emerging HPC systems. In current we have discussed the power consumption different perspectives. Conventionally the power consumption can be categorized in two ways including the power consumed at system level without running specific application and secondly the power consumption with some specific application computation [30]. Both categories have been specified the given equations 2,3 as follows.

$$P_{\text{system}}(w) = \sum_{i=1}^N P_{GPU}^i(w^i) + P_{CPU}(\sum_j^M(w^i)) + P_{mb}(w) \quad (2)$$

In above equation, the power consumed by system is the sum of power consumed by number of configured GPUs, CPUs and mainboard.

$$P_{\text{app}} = \sum_{i=1}^{N_{\text{app}}} P_{GPU}^i(w_{\text{app}}^i) + P_{CPU}(\sum_j^M w^i) + P_{mb}(w_{\text{app}}) \quad (3)$$

Similarly, the equation 3 describe the power consumed by system while running a specific application which is the sum of power consumed by number of configured GPUs, CPUs and mainboard.

## VI. EXPERIMENTAL RESULTS AND DISCUSSION

In this section we have presented all the determined results from the experiments where we implemented various numerical algorithms and discussed experimental results in this section. In first implementation, we run DMM application with multiple datasets through EPC model. A fundamental matrix multiplication method used in our implementation has been presented in below equation (6).

$$\begin{bmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{np} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mp} \end{bmatrix}$$

Sum of the given matrix can be defined as:

$$C_{ij} = \sum_{k=1}^m (A_{ik} B_{kj}) \quad (4)$$

Further to investigate the efficiency factor, we performed DMM implementation in suggested tri-level hybrid model with all MPI strategies (S1, S2 and S3) discussed in section (3).

By increasing matrix multiplication datasets, 'S3' increased the efficiency gradually and depicted the best performance compared to 'S1' and 'S2', and achieved 68% of peak performance in Tflops. Unlikely, 'S1' and 'S2' could attain the efficiency within range of 700-800 Gflops which was the initial throughput in 'S3' implementation. With large dataset computation, we observed that 'S1' declined the system efficiency which was eventually cause of over waiting during data distribution as shown in Fig. 4.

Along with performance, we quantified energy efficiency which is considered the primary metric for current and emerging HPC technologies. Likewise, the consequences in performance efficiency, 'S3' throughout increased energy efficiency at all datasets computation and accomplished 8.2 Gflops/w as shown in Fig. 5.

Further, we implemented 2-D Laplace application utilizing Jacobian iterative strategy where we run all models. By and large, the fractional differential conditions are ordered in a way like conic but here we have discussed only elliptic equation as  $U_{xx}(x,y) + U_{yy}(x,y)$  [22]. Be that as it may, the specific elliptic condition called "2-D Laplace condition" [23] utilized in current investigations is presented as follows in equation 7:

$$\frac{\partial^2 U}{\partial x^2}(x,y) + \frac{\partial^2 U}{\partial y^2}(x,y) = 0 \quad (5)$$

We implemented 2-D Laplace Jacobian iterative method in EPC proposed model using all strategies. The mesh size was increased dramatically in the range of 1000-8000. Fig. 6 and 7 demonstrate the consequences of 2-D Laplace method against both metrics (performance and energy efficiency). The similar efficiency ratio of 'S1' in matrix multiplication was found in 2-D Laplace solver method in range of 390-700 Gflops/sec. Although, efficiency increased gradually in 'S1' but we can rely on it due to poor throughput.

We also evaluated energy efficiency in 2D Laplace equation method (see Fig. 6). 'S3' provided the best energy

efficiency as compared to other strategies. We noticed that 'S2' was also prominent and achieved energy efficiency up to 8.3 Gflops/w but 'S1' wasted a lot of energy throughout the computation and achieved 7.4 Gflops/w at maximum mesh size.

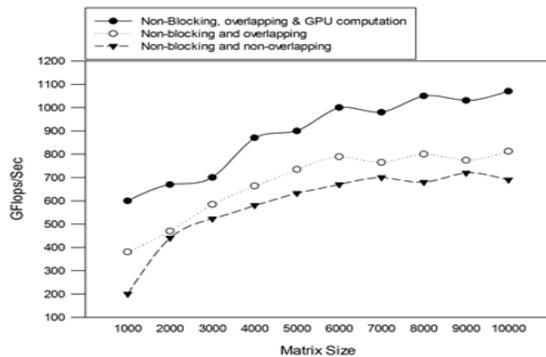


Fig. 4. Performance efficiency in all Strategies during MM Computation.

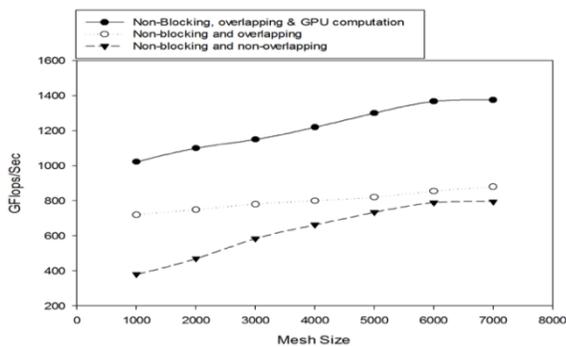


Fig. 5. Performance Efficiency in All Strategies during 2-D Laplace Solver.

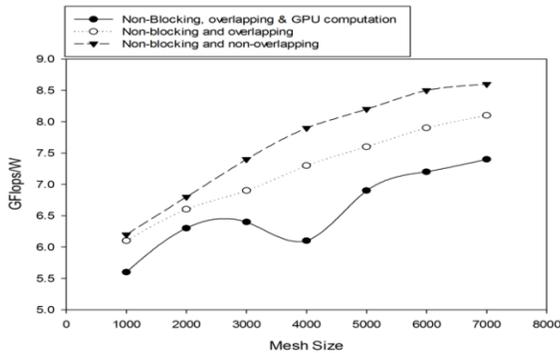


Fig. 6. Energy Efficiency 2-D Laplace.

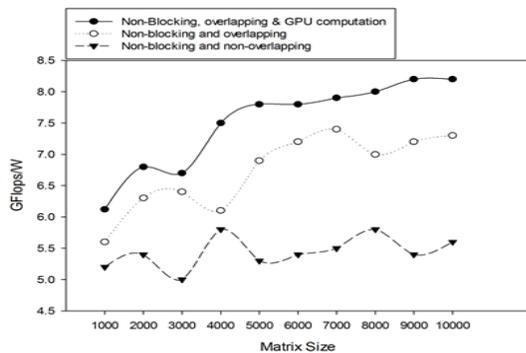


Fig. 7. Energy Efficiency in MM.

## VII. CONCLUSION

The emerging HPC models are relied upon to grow first Exaflops PC to contain a huge number of heterogeneous process hubs connected by complex systems till next half decade. This Exascale processing framework will be skilled to figure one Exaflops estimation for each subsequent which is thousands-crease increment in current Petascale framework. In current study, we have discussed the extensive constraints for Exascale systems and perspective challenges for current technologies. In this research, the proposed model is a novel secure and efficient parallel programming approach which is tri-level hybrid of MPI, OpenMP and CUDA. In MPI, we implemented different strategies (S1, S2 and S3) under non-blocking mechanism. Further to evaluate the efficiency factors, the proposed model was implemented with all these strategies in two benchmarking HPC applications including DMM and two dimensional Laplace equation. Consequently, in both applications, we found that 'S3' strategy (non-blocking, overlapping and GPU computation) performed the best in providing performance efficiency and energy efficiency comparatively to (S1 and S2). Therefore, hybrid of proposed model with 'S3' MPI strategy can be consider as promising model to achieve required performance and energy efficiency for Exascale systems. By future perspectives, this model is required to be executed a large cluster system that can meet the minimum requirement for Exascale system configurations.

## REFERENCES

- [1] Liao, G., Guo, D., Bhuyan, L. and King, S.R., 2008, November. Software techniques to improve virtualized I/O performance on multi-core systems. In Proceedings of the 4th ACM/IEEE Symposium on Architectures for Networking and Communications Systems (pp. 161-170). ACM.
- [2] Bland, B., 2012, November. Titan-early experience with the titan system at oak ridge national laboratory. In High Performance Computing, Networking, , 2012: (pp. 2189-2211). IEEE.
- [3] Liao, X., Xiao, L., Yang, C. and Lu, Y., 2014. MilkyWay-2 supercomputer: system and application. Frontiers of Computer Science, 8(3), pp.345-356.
- [4] Shalf, J., Dosanjh, S. and Morrison, J., 2010, June. Exascale computing technology challenges. In International Conference on High Performance Computing for Computational Science (pp. 1-25). Springer, Berlin, Heidelberg.
- [5] Perarnau, S., Gupta, R. and Beckman, P., 2015. Argo: An exascale operating system and runtime. In High Performance Computing, Networking, Storage and Analysis, SC15.
- [6] Bergman, K., Borkar, S., Campbell, D., Carlson, W., Dally, W., Denneau, M., Franzon, P., Harrod, W., Hill, K., Hiller, J. and Karp, S., 2008. Exascale computing study: Technology challenges in achieving exascale systems. (DARPA IPTO), Tech. Rep, 15.
- [7] Reed, D., et al., 2015. (ASCAC) Report: Exascale Computing Initiative Review. USDOE Office of Science.
- [8] Ang, J.A., Barrett, R.F., Benner, R.E., Burke, D., Chan, C., Cook, J., Donofrio, D., Hammond, S.D. Hemmert, K.S. Kelly, S.M. and Le, H., 2014, November. In Hardware-Software Co-Design for High Performance Computing (Co-HPC).
- [9] Ceruzzi, P.E., 2005. Moore's law and technological determinism: reflections on the history of technology. Technology and Culture, 46(3), pp.584-593.
- [10] Barker, Kevin J., et al. "Entering the petaflop era: the architecture and performance of Roadrunner." Proceedings of the 2008 conference on Supercomputing. IEEE Press, 2008.
- [11] Sameh, A.H., 2015. Parallel Sparse Linear System and Eigenvalue Problem Solvers: From Multicore to Petascale Computing. PURDUE UNIV LAFAYETTE IN.

- [12] Dongarra, J.J. and Walker, D.W., 2001. The quest for petascale computing. *Computing in Science & Engineering*, 3(3), pp.32-39.
- [13] Zima, H.P., 2007. From FORTRAN 77 to locality-aware high productivity languages for peta-scale computing. *Scientific Programming*, 15(1), pp.45-65.
- [14] Jin, H., Jespersen, D., Mehrotra, P., Biswas, R., Huang, L., & Chapman, B. (2011). High performance computing using MPI and OpenMP on multi-core parallel systems. 37(9), 562-575.
- [15] Roberge, V., Tarbouchi, M. and Okou, F.A., 2017. Distribution system optimization on graphics processing unit. *IEEE Transactions on Smart Grid*, 8(4), pp.1689-1699.
- [16] Satish, Nadathur, et al. "Fast sort on cpus, gpus and intel architectures." *Intel Labs* (2010): 77-80.
- [17] Bastem, B., Unat, D., Zhang, W., Almgren, A. and Shalf, J., 2017, August. Overlapping Data Transfers with Computation on GPU with Tiles. In *Parallel Processing (ICPP), 2017 46th International Conference on* (pp. 171-180). IEEE.
- [18] Collis, S.S., 2014. Computers are changing and so must our codes.. (No. SAND2014-20299PE). Sandia National Laboratories (SNL-NM), Albuquerque, NM (United States).
- [19] Lendino, J. (2016, June 29). Meet the new world's fastest supercomputer: China's TaihuLight. Retrieved from <https://www.extremetech.com/extreme/230458-meet-the-new-worlds-fastest-supercomputer-chinas-taihulight>
- [20] Kunkel, J. M., Balaji, P., & Dongarra, J. (2016). High Performance Computing: 31st International Conference, ISC High Performance 2016, Frankfurt, Germany, June 19-23, 2016, Proceedings. Basingstoke, England: Springer.
- [21] STFC, [https://www.scd.stfc.ac.uk/Pages/SCD\\_Science\\_Highlights\\_2016.pdf](https://www.scd.stfc.ac.uk/Pages/SCD_Science_Highlights_2016.pdf), 2016.
- [22] NCF, European Exascale Software Initiative.
- [23] D'Hollander, E., Dongarra, J., & Foster, I. (2013). *Transition of HPC Towards Exascale Computing*. Amsterdam, Netherlands: IOS Press.
- [24] India to Launch \$730M National Supercomputing Mission. (2014, October 7). Retrieved from <https://www.hpcwire.com/2014/10/06/india-launch-730m-national-supercomputing-mission/>
- [25] Ashraf, M. U., Fouz, F., & Alboraei Eassa, F. (2016). Empirical Analysis of HPC Using Different Programming Models. *International Journal of Modern Education and Computer Science*, 8(6), 27-34.
- [26] Hassani, A., Skjellum, A. and Brightwell, R., 2014, June. Design and evaluation of FA-MPI, a transactional resilience scheme for non-blocking MPI. In *Dependable Systems and Networks (DSN), 2014 44th Annual IEEE/IFIP International Conference on* (pp. 750-755). IEEE.
- [27] Ahmed, H., Skjellumh, A., Bangalore, P. and Pirkelbauer, P., 2017, September. Transforming blocking MPI collectives to Non-blocking and persistent operations. In *Proceedings of the 24th European MPI Users' Group Meeting* (p. 3). ACM.
- [28] Morgan, B., Holmes, D.J., Skjellum, A., Bangalore, P. and Sridharan, S., 2017, September. Planning for performance: persistent collective operations for MPI. In *Proceedings of the 24th European MPI Users' Group Meeting* (p. 4). ACM.
- [29] Hoefler, T., Squyres, J., Bosilca, G., Fagg, G., Lumsdaine, A. and Rehm, W., 2006. Non-blocking collective operations for MPI-2. *Open Systems Lab, Indiana University, Tech. Rep*, 8.
- [30] Hahnfeld, J., Cramer, T., Klemm, M., Terboven, C. and Müller, M.S., 2017, September. A Pattern for Overlapping Communication and Computation with OpenMP<sup>^</sup>\* Target Directives. In *International Workshop on OpenMP* (pp. 325-337). Springer, Cham.
- [31] Fujitsu to Provide High-Performance Computing and Services Solution to King Abdulaziz University. (n.d.). Retrieved from <http://www.fujitsu.com/global/about/resources/news/press-releases/2014/0922-01.html>.
- [32] King Abdulaziz University, [www.kau.edu.sa](http://www.kau.edu.sa)
- [33] L. Ahmad, A. Majidi, and A. Baniyadi. "IPMACC: Open source OpenACC to CUDA/OpenCL translator." (2014).
- [34] G. David, and N. S. Trudinger. *Elliptic partial differential equations of second order*. springer, 2015.
- [35] B. Richard L., and J. D. Faires. "Numerical Analysis. Brooks/Cole, Thomson Learning." Inc 206 (2001): 772.
- [36] Lilja, D. J. (2005). *Measuring Computer Performance: A Practitioner's Guide*. Cambridge, England: Cambridge University Press.

# Metaheuristic for the Capacitated Multiple Traveling Repairman Problem

Khanh-Phuong Nguyen<sup>1</sup>; Ha-Bang Ban<sup>2</sup>

School of Information and Communication Technology  
Hanoi University of Science and Technology, Hanoi, Vietnam  
Corresponding authors: Ha-Bang Ban

**Abstract**—The Capacitated Multiple Traveling Repairmen Problem (CmTRP) is an extension of the Multiple Traveling Repairmen Problem (mTRP). In the CmTRP, the number of vehicles is dispatched to serve a set of customers, while each vehicle's capacity is limited by a predefined-value as well as each customer is visited exactly once. The goal is to find a tour that minimizes the sum of waiting times. The problem is NP-hard because it is harder than the mTRP. Even finding a feasible solution is also NP-hard problem. To solve medium and large size instances, a metaheuristic algorithm is proposed. The first phase constructs a feasible solution by combining between the Nearest Neighborhood Search (NNS) and Variable Neighborhood Search (VNS), while the optimization phase develops the feasible solution by the General Variable Neighborhood Search (GVNS). The combination maintains the balance between intensification and diversification to escape local optima. The proposed algorithm is implemented on benchmark instances from the literature. The results indicate that the developed algorithm obtains good feasible solutions in a short time, even for the cases with up to 200 vertices.

**Keywords**—CmTRP; NNS; VNS; GVNS

## I. INTRODUCTION

### A. Motivation and Definition

A particular variant of the CmTRP is the Multiple Traveling Repairmen Problem (mTRP) that considers multiple vehicles or travelers to find a tour minimizing the waiting time of all customers [1], [7], [9]. Applications of the mTRP can be found in [1], [7], [9]. The mTRP is based on an assumption that there is no limit to the capacity of each vehicle. That means vehicles can carry as many goods as they want. However, real situations imply that it does not always hold because vehicles have strict regulations on capacity. In this work, the mTRP is involved the capacity constraint. In the CmTRP, the maximum capacity of each vehicle does not exceed a predefined capacity ( $Q$ ). In the CmTRP, there are  $k$  vehicles at a main depot  $s$ , and  $n$  customers. The goal is to find a tour such that each vertex is visited exactly once, while the capacity constraint is satisfied, and the total waiting time of overall customers is minimized.

In this paper, we formulate the CmTRP as followings: We consider a complete graph  $K_n$  that has the vertex set  $V = \{1, 2, \dots, n\}$  and a symmetric distance matrix  $C = \{c(i, j) \mid i, j = 1, 2, \dots, n\}$  ( $c(i, j)$  is the traveling cost between vertex  $i$  and vertex  $j$ ). Each vertex  $i$  has the demand  $d_i$ . Let  $R = (1, 2, \dots, k)$  be the number of  $k$  vehicles which begin at a main depot  $v_1$ . Let  $Q$  denote the capacity of a vehicle. Assume that a tour  $T = (R_1, \dots, R_l, \dots, R_k)$  includes a set of routes from  $k$  vehicles.  $R_l = (v_1, \dots, v_h, \dots, v_m)$  ( $1 < m \leq n$ ) is a route of

vehicle  $l$ . Let  $P(v_1, v_h)$ ,  $l(P(v_1, v_h))$  be respectively the path from vertex  $v_1$  to vertex  $v_h$  on  $R_l$ , and its length. The waiting time of  $v_h$  ( $1 < h \leq m$ ) on  $R_l$  is the cost of the path from  $v_1$  to  $v_h$ :

$$l(P(v_1, v_h)) = \sum_{i=1}^{h-1} c(v_i, v_{i+1}).$$

Let  $W(R_l)$  be the sum of waiting times of all vertices. The capacity of this route must satisfy the following constraint:

$$W(R_l) = \sum_{h=2}^m l(P(v_1, v_h)).$$
$$D(R_l) = \sum_{i=1}^m d_i \leq Q.$$

The objective function of the problem is:

$$W(T) = \sum_{l=1}^k W(R_l).$$

The CmTRP asks for a tour, which starts at a depot  $v_1$ , visits each vertex once exactly with the waiting time of all vertices being minimized.

### B. Related Works

In the general case, the CmTRP has, to the best of our knowledge, previously not been studied much, even though it is an extension of the mTRP case. However, some variants of the CmTRP are introduced in numerous works in the literature as follows:

- The mTRP is a particular case when the capacity constraint does not involve. Numerous works for mTRP can be found in [1], [7], [9]. In the metaheuristic approach, several algorithms [1], [7], [9] produce good solutions fast for instances with up to 200 vertices.
- The mTRP with distance constraint (mTRPD) is a particular case since the route length or maximum duration of each vehicle cannot exceed a predetermined limit. Metaheuristic algorithms in [3], [12] can solve the problem well for instances with up to 200 vertices.
- The mTRP with Profits (mTRPP) aims is to find a travel plan for server maximizing the total revenue. In this problem, not all customers need to be visited. Metaheuristic algorithm in [11] solves the problem well with up to 200 vertices.

- The Traveling Repairman Problem is a special case where there is only a repairman. Numerous metaheuristic algorithms [2], [4], [5], [17] for the problem have proposed in the literature.

The above algorithms are the state-of-the-art algorithms for some variants of the CmTRP case. However, they do not include the capacity constraint, and their corresponding algorithms cannot be adapted to the CmTRP. That means that we cannot use the above algorithms to solve the CmTRP.

### C. Our Methodology

Like other NP-hard problems, the CmTRP can be solved by exact or heuristic methods. Exact algorithms obtain the optimal solution but they often take exponential time in the worst case. Heuristic approaches are divided into the classical heuristic and metaheuristic approach. The classical heuristic approach finds one solution quickly, but this solution may have a large disparity in comparison with the best solution. The metaheuristic approach, on the other hand, obtains near-optimal or even global optimal solutions. Therefore, metaheuristic is usually used to reach optimal solutions for the problem with large sizes.

The CmTRP is also NP-hard because it is a generalization case of the mTRP. In many constrained optimization problems like the CmTRP, even building a feasible solution to the problem is also NP-hard problem. It indicates that obtaining a good feasible solution is a challenge. A good metaheuristic needs to ensure the balance between diversification and intensification. In [14], H. Mladenovic et al. show that the VNS generates local optima that are close to the global optimum, in a more straightforward manner than the other metaheuristics. However, the VNS only implements the intensification well. In this work, we developed a metaheuristic consisting of the constructive and optimization phase. The first phase constructs a feasible solution by combining between the Nearest Neighborhood Search (NNS) and Variable Neighborhood Search (VNS), while the optimization phase develops the feasible solution by the General Variable Neighborhood Search (GVNS). The metaheuristic uses the shaking technique to maintain diversification, while the VNS, and GVNS implement intensification. Extensive numerical experiments on benchmark instances show that the proposed algorithm reaches good feasible solutions at a reasonable amount of time, even for the instances with up to 200 vertices.

The rest of this paper is organized as follows. Section 2 introduces our algorithm. Computational evaluations are presented in Section 3, and Sections 4 and 5 discuss and concludes the work, respectively.

## II. THE PROPOSED ALGORITHM

### A. Several Variants of VNS

We describe some variants of VNS [14] such as the original VNS, VND, and GVNS, shaking technique [13], respectively.

- The Variable Neighborhood Search (VNS) algorithm is introduced by Mladenovic et al. [14]. It executes neighborhood procedures alternately, and shaking technique to escape from the local optima. At

each iteration, the best neighboring solution is chosen from neighboring solutions that are generated from a neighborhood procedure. If it is better than the current best one, the procedure is repeated. Otherwise, the search goes to the next neighborhood procedure.

- The Variable Neighborhood Descent (VND) algorithm, which is a VNS variant, is proposed by Mladenovic et al. [14]. In the VND, a change of neighborhoods is performed in a deterministic way. Assume that an initial solution is given. Local search procedures in their descent phase are used to generate neighborhoods. The final solution is the best solution in all neighborhoods. The difference between the VNS and VND is that the VNS uses Shaking.
- The General Variable Neighborhood Search (GVNS) algorithm [14] is a variant of the VNS. It includes an initial feasible solution, and a shaking procedure followed by the VND local search. The GVNS is a VNS variant where the VND is used as the improvement procedure.

### B. Neighborhoods

We use seven neighborhoods in the literature to explore the search space of the problem. Let  $N_k (k = 1, \dots, k_m)$  be a finite set of pre-selected neighborhood structures. We describe more details about seven neighborhoods:

**For Inter-route:** It is used to optimize on each route. We then describe five neighborhoods' structure in turn. Assume that,  $R$ , and  $m$  are a route and its size, respectively.

- **Remove-insert** places each vertex in the route at the end of it. Obviously, the complexity time of  $N_1(R)$  is  $O(m)$ .
- **Swap adjacent** tries to swap each pair of adjacent vertices in the route. The complexity time of  $N_2(R)$  is  $O(m)$ .
- **Swap neighborhood** attempts to swap the positions of each pair of vertices in the route. The complexity time of  $N_3(R)$  is  $O(m^2)$ .
- **3-opt neighborhood** attempts to reallocate three vertices to another position of the route. The complexity time of  $N_4(R)$  is  $O(m^3)$ .
- **4-opt neighborhood** attempts to involve deleting four edges and reconnecting the four sub-tours without changing the orientation of them. The complexity time of  $N_5(R)$  is  $O(m^4)$ .

**For intra-route:** Intra-route is used to swap vertices between two different routes or remove vertices from a route and then insert them to another. Let  $R_l, R_h, ml$ , and  $mh$  be two different routes and their sizes in  $T$ , respectively.

- **swap-2-routes** tries to exchange two vertices belonging to different routes  $R_l$  and  $R_h$ . The complexity time of  $N_6(R)$  is  $O(ml \times mh)$
- **insert-2-routes** removes one vertex  $R_l$  and inserts it at the best possible position in  $R_h$ . The complexity time of  $N_7(R)$  is  $O(ml \times mh)$ .

---

**Algorithm 1** The proposed algorithm

---

**Input:**  $T, nloop, t_{max}$  are an initial solution, the number of neighborhoods, and the number of iterations, and the maximum running time, respectively.

**Output:** the best solution  $T^*$ .

```

1: repeat
2:   {Construction phase}
3:    $T \leftarrow \text{Construction}(v_1, V)$ ;
4:   {Improvement phase}
5:    $Lvel = 1$ ;
6:    $T = \text{VND}(T)$ ;
7:   while ( $Lvel < nloop$ ) do
8:      $T' = \text{Perturbation}(T, Lvel)$ ;
9:     {implement VND}
10:     $T' = \text{VND}(T')$ ;
11:    if ( $W(T') < W(T)$ ) || ( $W(T') < W(T^*)$ ) then
12:       $T = T'$ ;
13:      if ( $W(T') < W(T^*)$ ) then
14:         $T^* = T'$ ;
15:      end if
16:    end if
17:    if ( $T$  is equal  $T'$ ) then
18:       $Lvel = 1$ ;
19:    else
20:       $Lvel ++$ ;
21:    end if
22:  end while
23: until  $time < t_{max}$ 
24: return  $T^*$ ;
```

---

**Algorithm 2** VND

---

**Input:**  $T, k_{max}$  are an initial solution, and the number of neighborhoods, respectively.

**Output:** the best solution  $T^*$ .

```

1:  $k = 1$ ;
2: repeat
3:   Find the best neighborhood  $T'$  of  $T \in N_k(T)$ ;  $\{T'$  must be feasible solution}
4:   if ( $W(T') < W(T)$ ) || ( $W(T') < W(T^*)$ ) then
5:      $T = T'$ ; {centre the search around  $T'$  and search again in the first neighborhood}
6:     if ( $W(T') < W(T^*)$ ) then
7:        $T^* = T'$ ;
8:     end if
9:      $k = 1$ ;
10:  else
11:     $k = k + 1$ ; {switch to another neighborhood}
12:  end if
13: until  $k < k_{max}$ ;
14:  $T^* = T'$ ;
15: return  $T^*$ ;
```

---

The proposed algorithm includes two phases. The construction phase finds a feasible solution, whereas the improvement phase tries to improve it. Algorithm 1 depicts the whole process.

---

**Algorithm 3** Construction

---

**Input:**  $v_1, K_n, k, \alpha$  are a depot, the graph, the number of vehicles, the length of  $NL$ , respectively.

**Output:** An initial solution  $T$ .

```

1: for ( $l = 1; l < k; l ++$ ) do
2:    $R_l = R_l \cup v_1$ ; {All routes start at  $v_1$ }
3: end for
4: while all vertices are not visited do
5:    $l = \text{random}(k)$ ; {a route randomly is chosen}
6:    $\{v_e$  is the last vertex in  $R_l\}$ 
7:   Generate  $NL$  list that includes  $\alpha$  nearest vertices to  $v_e$  in  $V$ ;
8:   Select vertex  $v \in \{v_i | v_i \in NL \text{ and } v_i \notin R_l\}$  randomly;
9:    $R_l \leftarrow \{v_i\}$ 
10: end while
11: if  $T$  is feasible then
12:   return  $T$ ;
13: else
14:    $Lvel = 1$ ;
15: end if
16: while (( $T$  is infeasible) and ( $Lvel \leq l_{max}$ )) do
17:    $T' = \text{Shaking}(T, Lvel)$ ;
18:    $T' = \text{VND}(T')$ ;
19:   if  $W(T') < W(T)$  then
20:      $T \leftarrow T'$ ;
21:   end if
22:   if  $W(T') == W(T)$  then
23:      $Lvel \leftarrow 1$ ;
24:   else
25:      $Lvel ++$ ;
26:   end if
27: end while
28: return  $T$ ;
```

---

**Algorithm 4** Perturbation( $T, Lvel$ )

---

**Input:**  $T, Lvel$  are the tour, and the parameter to control the strength of the perturbation procedure, respectively.

**Output:** a new tour  $T$ .

```

1:  $k = 1$ ;
2: while ( $k < Lvel$ ) do
3:    $T' = \text{double-bridge}(T)$ ;
4:    $T' \leftarrow \arg \min N_1(T')$ ;
5:    $\{T^*$  is the optimal solution}
6:   if ( $W(T') > (1 - \rho) \times W(T^*)$ ) then
7:      $T = T'$ 
8:   else
9:      $k ++$ ;
10:  end if
11: end while
12: return  $T$ ;
```

---

C. Construction

Algorithm 2 shows the constructive procedure. The objective function used in this procedure is the sum of all positive differences between the capacity of all vehicles and the capacity limit  $Q$ , that is,  $\min \sum_{l=1}^k \max(0, D(R_l) - Q)$ . The algorithm works until it finds a feasible solution. In the first step, a solution is created by Nearest Neighborhood Search [8]. If the solution is feasible, the construction phase stops

TABLE I. OUR RESULTS FOR THE CMTRP ON E-INSTANCES PROPOSED BY [18]

Instances	Init.Sol	Best.Sol	Aver.Sol	Improv	T
E30k3	3008.64	2419.1	2419.1	19.59	2
E30k4	2052.97	1731.43	1731.43	15.66	3
E51k5	2948.73	2769.1	2769.1	6.09	7
E76k10	3301.85	3064.68	3064.68	7.18	27
E76k14	2409.68	2261.63	2261.63	6.14	14
E76k15	2401.86	2221.59	2221.59	7.51	17
aver				10.36	11.7

TABLE II. OUR RESULTS FOR THE CMTRP ON P-INSTANCES PROPOSED BY [18]

Instances	Init.Sol	Best.Sol	Aver.Sol	Improv	T
P40k5	2095.22	1884.55	1884.55	10.05	5
P45k5	2672.9	2479.41	2479.41	7.24	8
P50k7	2262.5	1993.24	1993.24	11.90	7
P50k8	2310.09	2310.09	2310.09	0.00	4
P55k7	2325.41	2175.46	2175.46	6.45	13
P55k10	2095.43	1878.64	1878.64	10.35	6
P60k10	2511.53	2322.96	2322.96	7.51	15
P70k10	3297.38	2940.94	2940.94	10.81	17
P76k4	6509.49	6198.5	6198.5	4.78	28
P76k5	5920.72	5637.31	5637.31	4.79	29
aver				7.39	13.2

TABLE III. OUR RESULTS FOR THE CMTRP ON TAI-INSTANCES PROPOSED BY [18]

Instances	Init.Sol	Best.Sol	Aver.Sol	Improv	T
tai75a	7038	6286.77	6331.913	10.67	32
tai75b	6112.16	5061.1	5156.406	17.20	30
tai75c	5517.07	5023.56	5085.652	8.95	35
tai75d	6347.94	5856.79	5974.241	7.74	37
tai100a	11391.44	10200.9	10290.74	10.45	72
tai100b	10297.75	9755.82	9801.026	5.26	73
tai100c	6025.94	5743.6	5792.192	4.69	70
tai100d	7544.23	7138.13	7203.689	5.38	68
tai150a	16435.68	15913.87	16005.03	3.17	85
tai150b	14115.32	13627.56	13687.99	3.46	81
tai150c	14797.07	13098.63	13208.41	11.48	82
tai150d	15268.64	13530.54	13530.54	11.38	83
aver				8.32	62.3

and outputs it. On the other hand, a local search iterates until finding a feasible solution or  $l_{max}$  is reached. The solution is shaken to escape from the current local optimal. Next, the VNS is applied to obtain the best solution from neighboring solutions. If it is better than the found best solution, it is set to the current solution. Last,  $Lvel$  is increased by one if the current solution is not improved, or set to 1, otherwise.

#### D. Improvement

After the construction, the heuristic tries to improve the feasible solution created by the previous phase. In this phase, the objective function is to minimize  $W(T)$ .

Local search procedure that is developed by combining the seven neighborhoods generates various neighborhoods. The final solution should be a local minimum with respect to all neighborhoods. The order of neighborhoods is fixed.

In a preliminary experiment, the other of the neighborhoods are therefore explored in the following one, from “small” to “large” as it is common, i.e., swap-adjacent, move-up, move-down, remove-insert, swap, 2-opt, 3-opt, 4-opt, swap-2-routes, and insert-2-routes.

The Perturbation mechanism is very important to achieve success. When the mechanism has too small perturbation moves, the search can return to the previously visited solution space. Therefore, the search can get stuck into local optimal. On the other hand, large perturbation moves drive the search to undesirable space. In order to overcome these issues, we propose a new shaking technique based on the original double-bridge technique [13]. The detail is described in Algorithm 3.

The algorithm stops after  $t_{max}$  seconds or the best-solution is found ( $t_{max}$  is the parameter of the algorithm, and its value is determined from preliminary experiments).

### III. EVALUATIONS

The proposed algorithm is run on a Pentium 4 core i7 2.40 GHz processor with 8 GB of RAM. On all experiments, parameters  $\alpha, l_{max}, \rho, nloop$  are respectively set to 10, 5, 0.3, and 50. These parameters are chosen through empirical tests and, with them, the algorithm seems to produce good solutions at a reasonable amount of time in comparison with the other parameter values.

We also implement the performance of the whole implementation against the state-of-the-art algorithms. We compare both of the numerical results and computational time on the same instances.

#### A. Instances

The experiments are performed on a set of benchmark for Capacitated VRP in [12], [18]. As testing the proposed algorithm on overall instances can be computationally too expensive, some selected instances as follows: 1) to eliminate the effects of size, instances with approximately from 50 up to 200 customers are chosen; 2) in order not to bias the results by taking “easy” or “hard” instances, we randomly select them. These are: 1) Christofides et al.: This dataset includes seven instances (CMT6, CMT7, ..., CMT14); 2) Taillard et al.: Ten instances are picked randomly such as tai75a, tai75b, tai75c, tai75d, tai100a, tai100b, tai100c, tai100d, tai150a, tai150b, tai150c, and tai150d; 3) Augerat et al.: Fifteen instances of dataset  $P$  and  $E$  are selected; 4) S. Nucamendi-Guillén et al.: 150 instances from 60 to 80 vertices are used in the experiments. The optimal solutions for the instances can be extracted from [16].

#### B. Results

We define the improvement of our algorithm with respect to  $Best.Sol$  ( $Best.Sol$  is the best solution found by the proposed algorithm) in comparison with the initial solution from the construction phase as followings:

$$Improv[\%] = \frac{Best.Sol - Init.Sol}{Init.Sol} \times 100\% \quad (1)$$

In the tables,  $OPT$ ,  $Init.Sol$ ,  $Best.Sol$ ,  $Aver.Sol$  and  $T$  correspond to the optimal solution, initial solution, best solution, average

TABLE IV. THE EVOLUTION OF AVERAGE IMPROVEMENT DURING THE ITERATIONS

instances	1 iteration		10 iterations		20 iterations		30 iterations		40 iterations		50 iterations		100 iterations	
E-instance	6.85	0.2	7.65	1.48	8.45	4.32	9.78	6	10.1	8.17	10.4	11.7	10.4	17.33
P-instance	5.5	0.2	5.8	1.7	6	4.9	6.4	6.8	6.8	9.2	7.4	13.2	7.4	25.9
Tai-instances	6.7	1.1	7.1	7.9	7.4	23.0	7.8	31.9	8	43.5	8.3	62.3	8.3	187.9
aver	6.35	0.50	6.85	3.68	7.28	10.73	7.99	14.91	8.30	20.30	8.70	29.07	8.70	77.04

TABLE V. THE DIFFERENCE BETWEEN THE CmTRP AND MTRP'S OBJECTIVE FUNCTION ON SOME INSTANCES

Instances	mTRP	CmTRP	%diff
E30k3	1871.08	2419.1	29.29
E30k4	1643.30	1731.43	5.36
E51k5	2209.64	2769.1	25.32
E76k10	2310.09	3064.68	32.66
E76k14	2005.40	2261.63	12.78
E76k15	1962.47	2221.59	13.20
P40k5	1537.79	1884.55	22.55
P45k5	1912.31	2479.41	29.66
P50k7	1547.89	1993.24	28.77
aver			22.18

solution, and the average time in seconds of ten executions obtained by our algorithm, respectively. In this work, we choose several state-of-the-art metaheuristic algorithms for some variants of the CmTRP [1], [6], [10], [12], [15], [16] as a baseline in our research.

Tables I to III show the average gap of the improvement phase in comparison with the construction is 8.70%. The average gap value is not too large. This indicates that the construction phase gives good feasible solutions fast. The proposed algorithm consumes much time for the instances with up to 200 vertices. Therefore, the first way to decrease the running time is only to run the construction phase. In this case, the proposed algorithm suffers from a slightly loss of 8.70% solution quality on average. Although we cannot compare the results directly to other algorithms in the literature (note that in the algorithms for some variants of the mTRP, the capacity constraint is removed), we succeed in producing feasible solutions for instances with 200 customers. It is an important contribution when finding feasible solutions is also NP-hard.

Table IV shows the evolution of average gap during the iterations. The average gaps are 6.35%, 6.85%, 7.28%, 7.99%, 8.30%, 8.70%, and 8.70% in comparison with the initial solution, obtained by one, five, ten, twenty, thirty, forty, fifty, and one-hundred iterations, respectively. No improvement obtains from fifty to one-hundred iterations. Therefore, additional iterations give a minor improvement while it consumes much time. Hence, the second way to reduce the running time is to use no more than fifty iterations, and the improvement reaches about 8.70%. The fastest option is to run the construction phase and then improve it by using a single iteration, which obtains an average gap of 6.35% and average time of 0.5 seconds.

Table V shows the difference between the objective function of two problems on the same instances. The average

TABLE VI. OUR RESULTS FOR THE MTRP-INSTANCES WITH 60 VERTICES PROPOSED BY LUO

Instances	OPT	Init.Sol	Best.Sol	Aver.Sol	T
pr1002 60 0	530946.01	660211.35	530946.01	530946.01	2.98
pr1002 60 1	356469.79	455893.41	356469.79	356469.79	2.84
pr1002 60 2	344118.14	467498.53	344118.14	344118.14	2.98
pr1002 60 3	429604.2	579392.35	429604.2	429604.2	2.92
pr1002 60 4	435655.25	540342.11	435655.25	435655.25	2.82
pr1002 60 5	668129.73	779776.11	668129.73	668129.73	2.86
pr1002 60 6	406678.77	495022.53	406678.77	406678.77	2.9
pr1002 60 7	311254.73	414296.52	311254.73	311254.73	2.98
pr1002 60 8	469816.84	591638.26	469816.84	469816.84	3
pr1002 60 9	277336.06	377249.41	277336.06	277336.06	2.84
brd14051 60 0	213420.42	267899.2375	213420.42	213420.42	3
brd14051 60 1	218315.68	312468.7714	218315.68	218315.68	2.98
brd14051 60 2	151666.85	207799.2353	151666.85	151666.85	2.9
brd14051 60 3	172199.83	232433.3597	172199.83	172199.83	2.96
brd14051 60 4	133952.5	167792.608	133952.5	133952.5	2.84
brd14051 60 5	203145.14	290606.4286	203145.14	203145.14	2.88
brd14051 60 6	136233.51	171636.975	136233.51	136233.51	2.98
brd14051 60 7	171879.58	248180.3	171879.58	171879.58	2.96
brd14051 60 8	191580.79	241067.3882	191580.79	191580.79	2.98
brd14051 60 9	128178.58	174326.1925	128178.58	128178.58	2.94
fnl4461 60 0	156260.54	194032.1583	156260.54	156260.54	2.8
fnl4461 60 1	103190.13	131569.4961	103190.13	103190.13	2.96
fnl4461 60 2	109739.93	149525.6149	109739.93	109739.93	2.98
fnl4461 60 3	100792.2	136198.0575	100792.2	100792.2	2.94
fnl4461 60 4	149638.18	185947.0777	149638.18	149638.18	2.96
fnl4461 60 5	158679.44	185251.4379	158679.44	158679.44	2.96
fnl4461 60 6	122266.92	149102.6283	122266.92	122266.92	2.88
fnl4461 60 7	107469.11	142108.8532	107469.11	107469.11	2.94
fnl4461 60 8	100531.72	127280.3749	100531.72	100531.72	2.84
fnl4461 60 9	135829.76	183343.8156	135829.76	135829.76	2.94
d15112 60 0	684939.42	851498.8482	684939.42	684939.42	2.8
d15112 60 1	644759.99	819500.5493	644759.99	644759.99	2.86
d15112 60 2	425069.33	583381.5404	425069.33	425069.33	2.82
d15112 60 3	528177.95	662371.45	528177.95	528177.95	2.82
d15112 60 4	586915.82	736112.95	586915.82	586915.82	2.96
d15112 60 5	422195.61	494729.4263	422195.61	422195.61	2.94
d15112 60 6	518793.6	633637.8578	518793.6	518793.6	2.86
d15112 60 7	616918.44	776732.675	616918.44	616918.44	2.98
d15112 60 8	397619.37	500495.3875	397619.37	397619.37	2.8
d15112 60 9	673840.81	910298.6184	673840.81	673840.81	2.9
nrrw1379 60 0	64359.77	80086.56654	64359.77	64359.77	2.88
nrrw1379 60 1	83410.67	104646.3375	83410.67	83410.67	2.96
nrrw1379 60 2	52858.87	70986.81333	52858.87	52858.87	2.96
nrrw1379 60 3	62341.36	84434.20476	62341.36	62341.36	2.84
nrrw1379 60 4	56012.13	69680.90881	56012.13	56012.13	2.9
nrrw1379 60 5	58083.8	72973.325	58083.8	58083.8	2.9
nrrw1379 60 6	52224.66	65749.025	52224.66	52224.66	2.92
nrrw1379 60 7	58402.97	73290.6375	58402.97	58402.97	2.94
nrrw1379 60 8	52145.08	66101.85821	52145.08	52145.08	2.96
nrrw1379 60 9	49026.52	66572.84307	49026.52	49026.52	2.86

difference of 22.18% indicates the capacity constraint also

TABLE VII. OUR RESULTS FOR THE MTRP-INSTANCES WITH 70 VERTICES PROPOSED BY LUO

Instances	OPT	Init.Sol	Best.Sol	Aver.Sol	T
pr1002 70 0	429557.7	429557.7	429557.7	530946.01	3.12
pr1002 70 1	430048.06	430048.06	430048.06	356469.79	3.08
pr1002 70 2	377233.86	377233.86	377233.86	344118.14	2.68
pr1002 70 3	429562.01	429562.01	429562.01	429604.2	2.64
pr1002 70 4	435659.17	435659.17	435659.17	435655.25	2.96
pr1002 70 5	429584.16	429584.16	429584.16	668129.73	3.34
pr1002 70 6	344534.44	344534.44	344534.44	406668.77	2.82
pr1002 70 7	393558.46	393558.46	393558.46	311254.73	3.02
pr1002 70 8	397072.39	397072.39	397072.39	469816.84	2.74
pr1002 70 0	429557.7	429557.7	429557.7	277336.06	3.12
brd14051 70 0	191843.35	191843.35	191843.35	213420.42	2.76
brd14051 70 1	169340.01	169340.01	169340.01	218315.68	2.98
brd14051 70 2	216195.95	216195.95	216195.95	151666.85	3.12
brd14051 70 3	229328.9	229328.9	229328.9	172199.83	3.28
brd14051 70 4	302498.42	302498.42	302498.42	133952.5	3.34
brd14051 70 5	179470.31	179470.31	179470.31	203145.14	3
brd14051 70 6	231693.74	231693.74	231693.74	136233.51	2.66
brd14051 70 7	284960.31	284960.31	284960.31	171879.58	2.66
brd14051 70 8	167533.17	167533.17	167533.17	191580.79	2.76
brd14051 70 9	253499.74	253499.74	253499.74	128178.58	3.26
fnl4461 70 0	154805.67	154805.67	154805.67	156260.54	2.76
fnl4461 70 1	104585.82	104585.82	104585.82	103190.13	3.24
fnl4461 70 2	161892.44	161892.44	161892.44	109739.93	2.74
fnl4461 70 3	99122.23	99122.23	99122.23	100792.2	3.32
fnl4461 70 4	157106.13	157106.13	157106.13	149638.18	2.84
fnl4461 70 5	112094.64	112094.64	112094.64	158679.44	2.72
fnl4461 70 6	121521	121521	121521	122266.92	2.74
fnl4461 70 7	175859.51	175859.51	175859.51	107469.11	3.06
fnl4461 70 8	122141.15	122141.15	122141.15	100531.72	2.94
fnl4461 70 0	154805.67	154805.67	154805.67	135829.76	2.76
d15112 70 0	517426.18	517426.18	517426.18	684939.42	3.24
d15112 70 1	715678.26	715678.26	715678.26	644759.99	3.02
d15112 70 2	688605.9	688605.9	688605.9	425069.33	3
d15112 70 3	625623.9	625623.9	625623.9	528177.95	3.32
d15112 70 4	532088.98	532088.98	532088.98	586915.82	2.78
d15112 70 5	500455.25	500455.25	500455.25	422195.61	3.18
d15112 70 6	497229.6	497229.6	497229.6	518793.6	3.18
d15112 70 7	599776.85	599776.85	599776.85	616918.44	2.86
d15112 70 8	576957.51	576957.51	576957.51	397619.37	3.02
d15112 70 9	775176.3	775176.3	775176.3	673840.81	2.6
nrw1379 70 0	66839.83	66839.83	66839.83	64359.77	2.58
nrw1379 70 1	65103.43	65103.43	65103.43	83410.67	2.98
nrw1379 70 2	63480.7	63480.7	63480.7	52858.87	3.2
nrw1379 70 3	59273.92	59273.92	59273.92	62341.36	3.32
nrw1379 70 4	70594.56	70594.56	70594.56	56012.13	2.66
nrw1379 70 5	73884.17	73884.17	73884.17	58083.8	3.02
nrw1379 70 6	64306.14	64306.14	64306.14	52224.66	2.94
nrw1379 70 7	90554.87	90554.87	90554.87	58402.97	2.54
nrw1379 70 8	91738.43	91738.43	91738.43	52145.08	2.82
nrw1379 70 9	68024.3	68024.3	68024.3	49026.52	2.68

TABLE VIII. OUR RESULTS FOR MTRP-INSTANCES WITH 80 VERTICES PROPOSED BY LUO

Instances	OPT	Init.Sol	Best.Sol	Aver.Sol	T
pr1002 80 0	491764.64	656239.68	491764.64	491764.64	9.48
pr1002 80 1	442164.21	613287.46	442164.21	442164.21	9.18
pr1002 80 2	505524.17	609954.56	505524.17	505524.17	9.32
pr1002 80 3	436752.96	614611.29	436752.96	436752.96	9.1
pr1002 80 4	453609.46	587470.83	453609.46	453609.46	9.36
pr1002 80 5	599492.4	771733.95	599492.4	599492.4	9.16
pr1002 80 6	619003.36	805206.35	619003.36	619003.36	9.4
pr1002 80 7	508186.51	658640.85	508186.51	508186.51	9.42
pr1002 80 8	409733.88	518052.51	409733.88	409733.88	9.44
pr1002 80 9	557220.48	670387.49	557220.48	557220.48	9.28
brd14051 80 0	336983.07	403178.34	336983.07	336983.07	9.04
brd14051 80 1	277861.02	348787.38	277861.02	277861.02	9.14
brd14051 80 2	265370.92	321922.47	265370.92	265370.92	9.56
brd14051 80 3	189815.69	240361.74	189815.69	189815.69	9.1
brd14051 80 4	206068.45	275228.43	206068.45	206068.45	9.5
brd14051 80 5	303621.75	348578.27	303621.75	303621.75	9.32
brd14051 80 6	213405.23	266958.37	213405.23	213405.23	9.6
brd14051 80 7	263737.93	308039.16	263737.93	263737.93	9.04
brd14051 80 8	232967.83	298574.84	232967.83	232967.83	9.26
brd14051 80 9	317790.55	368183.51	317790.55	317790.55	9.06
fnl4461 80 0	153124.51	194685.8	153124.51	153124.51	9.58
fnl4461 80 1	174975.64	224516.74	174975.64	174975.64	9
fnl4461 80 2	162755.5	197782.39	162755.5	162755.5	9.46
fnl4461 80 3	160819.04	192927.87	160819.04	160819.04	9.5
fnl4461 80 4	151790.69	187440.09	151790.69	151790.69	9.52
fnl4461 80 5	131045.47	172293.83	131045.47	131045.47	9.04
fnl4461 80 6	125405.93	166418.99	125405.93	125405.93	9.24
fnl4461 80 7	125228.91	164627.71	125228.91	125228.91	9.16
fnl4461 80 8	185280.87	228208.31	185280.87	185280.87	9.48
fnl4461 80 9	130304.95	165022.1	130304.95	130304.95	9.26
d15112 80 0	551900.43	753989.49	551900.43	551900.43	9.56
d15112 80 1	815029.39	979921.76	815029.39	815029.39	9.1
d15112 80 2	828114.32	1080571.45	828114.32	828114.32	9.16
d15112 80 3	689450.94	964458.54	689450.94	689450.94	9.1
d15112 80 4	560385.47	737417.48	560385.47	560385.47	9.08
d15112 80 5	821959.4	1030515.79	821959.4	821959.4	9.52
d15112 80 6	715206.03	882086.8	715206.03	715206.03	9.36
d15112 80 7	958278.86	1155190.13	958278.86	958278.86	9.32
d15112 80 8	990277.77	1174384.17	990277.77	990277.77	9.1
d15112 80 9	672457.47	931587.71	672457.47	672457.47	9.5
nrw1379 80 0	64831.76	96656.39	64831.76	64831.76	9.36
nrw1379 80 1	64967.83	88394.72	64967.83	64967.83	9.22
nrw1379 80 2	73858.13	96499.97	73858.13	73858.13	9.3
nrw1379 80 3	100592.83	131733.34	100592.83	100592.83	9.24
nrw1379 80 4	98228.29	126451.61	98228.29	98228.29	9.04
nrw1379 80 5	75984.21	99492.47	75984.21	75984.21	9.14
nrw1379 80 6	79165.6	105024.23	79165.6	79165.6	9.08
nrw1379 80 7	73194.55	105328.52	73194.55	73194.55	9.1
nrw1379 80 8	83492.62	115793.31	83492.62	83492.62	9.14
nrw1379 80 9	67034.31	92380.75	67034.31	67034.31	9.24

affects the quality of solutions. However, the best solutions for the mTRP are not feasible solutions for the CmTRP. Therefore, the methods designed for the mTRP instances may not be adapted easily to solve the CmTRP.

From Tables VI to VIII show that the efficiency of the proposed algorithm is very good for the mTRP-instances since it can reach the optimal solutions for the instances with up to 80 vertices at a reasonable amount of time. In Table IX, in comparison with Ban et al.'s [1], Ezzine et al.'s [6], and Nucamendi-Guillén's [16] algorithm, our solutions are better than those of Ban et al.'s, and Ezzine et al. in all cases while

they are comparable with Nucamendi-Guillén's solutions in the most of instances. In many cases, our algorithm obtains the optimal solution for the instance with 76 vertices. Moreover, for the CCVRP in Table X, our algorithm reaches the known best solutions for the instances with 100 vertices (note that: the best solutions are extracted from [10], [15]).

The running time of the proposed algorithm grows quite moderate compared to the Nucamendi-Guillén's algorithm [16] while it is comparable with those of Ban et al.'s [1], and Ezzine et al.'s algorithm [6].

TABLE IX. COMPARISONS WITH THE PREVIOUS ALGORITHMS FOR MTRP-INSTANCES PROPOSED IN [18]

Instances	Ban et al.	Ezzine et al.	Nucamendi-Guillén et al.	Our results	
				Best.Sol	T
E30k3	2108.26	-	-	2097.3	0.25
E30k4	2623.65	-	-	2595.11	0.22
E51k5	2623.65	3320	2209.64*	2209.64	0.41
E76k10	2786.07	4094	2310.09*	2419.89	0.78
E76k14	2201.13	3762	2005.4*	2005.4	0.71
E76k15	2400.17	-	-	2377.5	0.81
E101k8	-	6383	-	4051.47	2.52
E101k14	-	5048	-	3288.53	2.50
P40k5	1793.14	-	1537.79*	1580.21	0.27
P45k5	2336.43	-	1912.31*	1912.31	0.32
P50k7	1878.81	-	1547.89*	1590.41	0.58

Note that: symbol '\*' is the optimal value

TABLE X. COMPARISONS WITH THE PREVIOUS ALGORITHMS FOR CCVRP-INSTANCES PROPOSED BY [18]

Instances	BKS	Best.Sol	T
CMT1	2230.35	2230.35	1.70
CMT2	2391.63	2429.18	6.19
CMT3	4045.42	4073.12	17.93
CMT4	4987.52	4987.52	85.94
CMT5	5809.59	5838.32	295.67
CMT11	7314.55	7314.55	31.09
CMT12	3558.92	3559.43	15.84

#### IV. DISCUSSIONS

Due to NP-hard problem, metaheuristic approach is suitable approach to solve the problem with large sizes in a short time. Currently, several algorithms for the close variants of CmTRP have been proposed. However, these algorithms cannot apply to the CmTRP because the capacity constraint does not include in them. The results in Table V indicate that the algorithms for the mTRP produce infeasible solutions for the CmTRP. Therefore, developing efficient algorithm for the CmTRP is our contribution. The VNS, and GVNS [14] are a general schemes that are used widely to solve many optimization problems. However, to apply them for a specific problem, they require many efforts. Specifically, how to use and combine neighborhoods to explore and exploit good solution space, how many neighborhoods are used to balance between solution quality and running time, and how to balance between diversity and intensification. Therefore, in the literature, many algorithms are developed on the VNS scheme for a specific problem, but the solution quality of them is different. Applying the VNS scheme to solve the problem effectively is our contribution in this work.

For a metaheuristic approach, a classical and successful recipe is to combine 1) efficient "intensification" procedures via local searches with 2) "diversification" methods. Diversification means to generate diverse solutions to explore the search space on a global scale, while intensification means to focus on the search in a local region by exploiting the information that a current good solution is found in this region. Our algorithm uses the VNS, GVNS to implement intensification while shaking maintains diversification. Though the initial solution is generated far from the global minima, the explored

solution region is extended. Therefore, the chance for finding good solutions is high. The CmTRP is harder than the mTRP because it is the general case of the mTRP. For constrained NP-hard problem like the CmTRP, finding feasible solution is also NP-hard. The new contribution in this work is applied the VNS in both of two phases. The first phase generates feasible solution while the post one improve solution quality.

With respect to the instances, it indicates that the proposed algorithm can be used as follows:

- The first way is to only run the construction phase with a rather loss of 8.7% solution quality on average. It is the fastest option.
- The second way is to run the construction phase and then enhance it by using one iteration. As the result, our algorithm obtains an average improvement of 6.35%, and an average time of 0.5 seconds. It balances between solution quality and running time.
- The last is to run the construction phase and improve it no more than fifty iterations, and the average improvement reaches about 8.7%, and an average time of 29.07 seconds. It is the best option in terms of solution quality.

Moreover, in comparison with some close variants of the mTRP, the proposed algorithm obtains the optimal solutions for the instances with up to 80 vertices in a short time. Moreover, it also reaches the better solutions than the algorithms in [1], [6] while it is comparable with the algorithms in [10], [12], [15], [16]. It shows that our algorithm is still effective for various problems.

#### V. CONCLUSIONS

In this paper, we propose a metaheuristic algorithm which applies the VNS in both of phases. The first phase creates a feasible solution while the post one improve it. The optimal solutions can be reached for the problem with up to 80 vertices in several seconds. The solution's quality is comparable with the previous algorithms for the other cases. Moreover, we give three options to use the proposed algorithm effectively. However, the running time needs to enhance to meet practical situations. It will be our aim in future research.

#### ACKNOWLEDGMENT

This research is funded by Hanoi University of Science and Technology (HUST) under grant number T2018-PC-208.

#### REFERENCES

- [1] H.B. Ban, "An effective GRASP+VND metaheuristic for the k-Minimum Latency Problem", Proc. RIVE, 2016, pp. 31-36.
- [2] H.B. Ban, "General Variable Neighborhood Search for the Quote-Travelling Repairman Problem", J. IJACSA, No. 4, 2020 (To appear).
- [3] H.B. Ban, "A GRASP+VND Algorithm For The Multiple Traveling Repairman Problem With Distance Constraints", J. JCC, Vol.33, No.3, 2017, pp. 272-288.
- [4] H.B. Ban, and D.N. Nguyen, "Improved genetic algorithm for Minimum Latency Problem", Proc. SOICT, 2010, pp. 9-15.
- [5] H.B. Ban, K. Nguyen, M.C. Ngo, and D.N. Nguyen, "An efficient exact algorithm for Minimum Latency Problem", J. PI, No.10, 2013, pp. 1-8.

- [6] I. O. Ezzine, and Sonda Elloumi, "Polynomial Formulation and Heuristic Based Approach For The k-Travelling Repairman Problem", *Int. J. Mathematics In Operational Research*, Vol. 4, No. 5, 2012, pp. 503-514.
- [7] F. Jittat, C. Harrelson, and S. Rao, "The k-Traveling Repairman Problem", *Proc. ACM-SIAM*, 2003, pp.655-664.
- [8] D. S. Johnson, and L. A. Mcgeoch, "The Traveling Salesman Problem: A Case Study In Local Optimization In Local Search In Combinatorial Optimization", E. Aarts and J. K. Lenstra, Eds., pp. 215-310.
- [9] R. Jothi, and B. Raghavachari, "Minimum Latency Tours and The k-Traveling Repairmen Problem", *Proc. LATIN*, 2004, pp. 423-433.
- [10] L. Ke, and Z. Feng, "A Two-Phase Metaheuristic For The Cumulative Capacitated Vehicle Routing Problem", *J. Computers and Operations Research*, Vol. 40, No. 2, pp. 633-638, 2013.
- [11] Y. Lua, U. Benlic, Q. Wua, and B. Peng, "Memetic algorithm for the multiple traveling repairman problem with profits", *J. EAAI*, Vol. 80, 2019, pp. 35-47.
- [12] Z. Luo, H. Qin, and A. Lim, "Branch-and-Price-and-Cut for The Multiple Traveling Repairman Problem With Distance Constraints", *J. Operations Research*, Vol. 234, No. 1, 2013, pp. 49-60.
- [13] O. Martin, S. W. Otto, and E. W. Felten, "Large-Step Markov Chains For The Traveling Salesman Problem", *J. Complex Systems*, Vol. 5, No. 3, 1991, pp. 299-326.
- [14] N. Mladenovic, and P. Hansen, "Variable Neighborhood Search", *J. Operations Research*, Vol.24, No. 11 24, 1997, pp.1097-1100.
- [15] SU. Ngueveu, C. Prins, and R. Wolfler Calvo, An Effective Memetic Algorithm for the Cumulative Capacitated Vehicle Routing Problem, *J. Computers and Operations Research*, Vol 37, No. 11, pp. 1877-1885, 2010.
- [16] S. Nucamendi-Guillén, I. Martínez-Salazar, F. Angel-Bello, and J. M. Moreno-Vega, "A Mixed Integer Formulation and An Efficient Metaheuristic Procedure For The K-Travelling Repairmen Problem", *J. JORS*, Vol. 67, No. 8, 2016, pp. 1121-1134.
- [17] A. Salehipour, K. Sorensen, P. Goos, and O.Brassy, "Efficient GRASP+VND and GRASP+VNS Metaheuristics for the Traveling Repairman Problem", *J. Operations Research*, Vol. 9, No. 2, 2011, pp.189-209.
- [18] NEO (2013), <http://neo.lcc.uma.es/vrp/vrp-instances/capacitated-vrp-instances>.

# An Efficient Methodology for Water Supply Pipeline Risk Index Prediction for Avoiding Accidental Losses

Muhammad Shuaib Qureshi<sup>1</sup>, Ayman Aljarbouh<sup>2</sup>,  
Muhammad Fayaz<sup>3</sup>  
Department of Computer Science, University of Central  
Asia, Naryn 722918, Kyrgyz Republic

Muhammad Bilal Qureshi<sup>4</sup>  
Department of Computer Science  
Shaheed Zulfikar Ali Bhutto Institute of Science and  
Technology, Islamabad, 46000, Pakistan

Wali Khan Mashwani<sup>5</sup>  
Institute of Numerical Sciences  
Kohat University of Science and Technology  
Kohat, Pakistan

Junaid Khan<sup>6</sup>  
Department of Computer Science  
International Islamic University, Islamabad  
Pakistan

**Abstract**—The accidents happening to buildings and other human facilitation sectors due to poor water supply pipelining system is a random phenomenon, but an efficient estimation system can help to escape from such accidents. Such a system can be useful in assisting the caretakers to take the initiative measures to avoid the occurrence of the accidents or at least reduce the associated risk. In this paper, we target this issue by proposing a water supply pipelines risk estimation methodology using feed forward backpropagation neural network (FFBPNN). For validation and performance evaluation, real data of water supply pipelines collected in Seoul, Republic of South Korea from 1987 to 2010 is used. A comprehensive analysis is performed in order to get reasonable results with both original and pre-processed input data. Pre-processing consists of two steps: data normalization and statistical moments computation. Statistical moments are mean, variance, kurtosis and skewness. Significant improvement in prediction accuracy is observed with data pre-processing in terms of selected performance metrics, such as mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean squared error (RMSE).

**Keywords**—Neural networks; normalization; risk index; mean square error; statistical moments

## I. INTRODUCTION

Underground facilities are very important to be monitored, because these facilities are unforeseen hazards to buildings, bridges, railway lines, etc. In terms of reducing the hazards, underground water supply pipelines are one the most important facility which contributes many underground risks and difficulties. Water is the superlative part of life without which no one can survive in the world. Water facilities are fully dependent on pipe, without pipeline water supply is impossible. In order to provide and supply water to homes, buildings and commercial areas, a bundle of pipes is installed underground. In recent age, the cities become more and more congested and these water supply pipelines are danger warning to constructed buildings as well as more risks to be taken to supply water in these congested areas [1,2].

The affection of water supply pipelines is affected by different types of parameters such as leakage, age, depth and

height, quality of the pipe, water temperature, soil electrical resistivity, soil temperature, soil moisture, etc. One of the most important features for estimation of water supply pipeline risk is aging. The fitness of pipelines may be degraded with the passage of time, and also depends on the quality of materials. The degradation of the water supply pipeline can affect and destruct the underground structure of the buildings; hence it is very important to fix and restore damaged water supply pipelines on time and reduce the chances of occurrence of accidents and risks [3]. In this paper, we examine the depth, leakage, length and age which are a very important factors in water supply pipeline failure or damage.

Water supply pipeline leakage is a very remarkable factor that can cause for unexpected underground menaces such as urban sinkhole and abrupt road-side subsidence due to water pipes leaks. This water supply pipeline leakage slowly and steadily destructs the underground structure of congested buildings, sub-ways, bridges, railways, etc. and because of this alarming situation, underground water facilities permanently remain a major and serious threat. Failure and depth of water supply pipeline also ultimately damages the underground structural damage or failure eventually. When a water pipeline is spread closely to the surface then any human activity or moving of vehicles can damage the surface over the pipe and can even break the pipe when heavy load is exerted on the pipe, if a pipe under the ground is buried deeper, affection does not cause on the surface because of deepness. Suddenly rupture in the water supply pipeline can bring serious damage to the near advantage people. Comparably another parameter that is length, plays an essential role in the protection of the water supply pipeline and ultimately at underground risk [4]. To analyze water supply risks, researchers have proposed different new techniques.

The objective of this paper is to compute the accurate estimated risk failure of water supply pipelines by using FFBPNN on data with statistical moments, original data and normalized data. We selected the FFBPNN for water supply pipeline risk estimation because it is the most important model

for estimation and prediction [5]. The data with statistical moments is also very important to increase the performance of ANN [5]. Normalization also increases the performance of an algorithm. In this study the pipeline failure risk is considered in term that how it impacts the underground structure.

Rest of the paper is organized as: Section 2 presents related work, Section 3 proposed model for water supply pipelines, Section 4 explains the experimental results and discussion. The conclusion of the paper is given in Section 5.

## II. RELATED WORK

Numerous algorithms have been developed based for water supply risk index assessment in literature.

Hussam et al. in [6] suggested a hierarchical based fuzzy model for water supply pipeline hazard evaluation. 16 risk factors were considered in that system. They inferred that the age of the pipe extends a solid effect on failure. They further added that pipe segments length, diameter and material are also very important elements of water supply pipelines. Yan and Vairavamoorthy in [7] proposed a decision-making technique to assess pipeline conditions. The fuzzy numbers are the output of this model that reflects the state of each pipeline. Kleiner et al. recommended a technique for buried pipelines to model the decline process; the model is based on a fuzzy rule-based non-homogeneous Markova process [8]. They also proposed a fuzzy logic method for pipeline risk evaluation. They have three main chunks, namely failure possibility, failure consequences, and a mixture of both [9].

Kleta et al., [10] used to review the system of lining surface images by video recording in which cameras are moved for damage assessment and integrity observation. The recording and camera monitoring systems are only limited to the visible parts of the surface. Different tools are designed to inspect large diameters and water mains. The most accurate tool available to detect the pockets trapped gas, leak and structural faults complicated networks of huge diameter water mains is the Sahara Pipeline Inspection System. Meng et al., [11] recommended a quantitative risk assessment (QRA) model which is a novel approach used for evaluation of non-homogeneous road tunnels risks, because the QRA models are inappropriate to apply to assess risk in road tunnels. The tunnel segmentation principle is used in this model in which the dissimilar urban road tunnel is segmented into numerous similar segments. The separate risk for road tunnel segments along with the combined risk indication for the entire road tunnel is elucidated. Duzgun et al., [12] suggested the decision analysis method for evaluation and managing the risk of underground coal mines and falling of coal mine roof. Possible consequences and cost of consequences, the probability was used for the risk assessment. Ustinovichius et al., [13] discussed various risk assessment methods. Assessment of risk can help decision-makers for ranking existing risk to take proper reaction suitably. Fault trees, monte carlo simulation, failure mode and effective analysis, event trees, game theory, fuzzy set, grey systems and multi-criteria verbal analysis are available numerous risk assessment methods. Multilayer perceptron (MLP) is an artificial neural network (ANN) with more than one hidden layer and a bias layer. For different types of modeling of ANN, different types

of architectures have been used for many years in different research areas including mathematics, engineering, medicine, neurology, meteorology, economics, hydrology, psychology and different other areas [14-17]. The ANN has many variants like multilayer perceptron (MLP), self-organization map (SOM), support vector machine (SVM), recurrent neural network (RNN) and feed forwarded neural network (FFNN). In this work, feed forward back-propagation neural network (FFBPNN) is used which is a very famous ANN model for prediction and estimation [18].

Fayaz et al. [19] proposed a model called blended hierarchical fuzzy logic for water supply risk index assessment. The purposed of the proposed model was to reduce the number of rules in the developed model. Another model named as the cohesive hierarchical fuzzy inference system was developed in [20] to assess water supply risk index. The aim of this model was similar to the previous model and both models have the potential to decrease the number of rules as well as to improve accuracy.

## III. PROPOSED METHODOLOGY OF WATER SUPPLY PIPELINES RISK INDEX ESTIMATION

Fig. 1 depicts the proposed method, comprised of four different kind of layers; data acquisition layer, pre-processing layer, estimation layer based on the neural network and performance evaluation layer. Each layer has its own functionality. Data layer contains the data related to water supply pipeline risks. In the pre-processing layer, statistical moments and normalization are used to pre-process the acquired data. In the estimation layer, FFBPNN is used for water supply pipeline risk index estimation. The performance of the FFBPNN is measured in the performance evaluation layer. The three-performance measurement used for evaluation of neural network such as mean absolute error (MAE), root means absolute error (RMSE) and mean absolute percentage error (MAPE), respectively.

### A. Data Layer

The datasets used in this work are real datasets that have been acquired from Electronics and Telecommunications Research Institute (ETRI) working on the underground projects. This institute completed a lot of underground project globally. For our research, we collected water supply pipelines data from 1989-2010 at different places in Seoul, the Republic of South Korea. It is observed in the literature when some tweaks are added to the original data, the performance of machine learning algorithms improves, therefore in this study, we normalize the data and calculate statistical moments of the data to get better results.

### B. Pre-Processing Layer

First, we take the dataset in pre-processing layer. The datasets comprise of the leakage, age, depth and height parameters of water supply pipelines. Using this data, first we calculate the statistical moments and then concatenate with the original data. The first four parameters, namely variance, mean, kurtosis and skewness [17] can be calculated using below Equations (1-4).

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

$$\sigma = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad (2)$$

$$S = \frac{1}{n} \sum_{i=1}^n \left[ \frac{(x_j - \mu)}{\sigma} \right]^3 \quad (3)$$

$$K = \frac{1}{n} \sum_{i=1}^n \left[ \frac{(x_j - \mu)}{\sigma} \right]^4 \quad (4)$$

where  $\sigma, \mu, K, S$  represent variance, mean, kurtosis and skewness, respectively.  $x_i$  denotes values of leak, depth, length and age of pipes values,  $i = 1, 2, \dots, 4$ .

For trial and test purposes, the normalized data can be computed by using Equation (5).

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (5)$$

Where the output normalized value is denoted by  $x_{new}$ , the current value is indicated by  $x$ , the minimum value in the set is represented by  $x_{min}$ , and the maximum value is denoted by  $x_{max}$  [5].

### C. Water Supply Risk Index Estimation Layer

The ANN method is characterized as the regression method, which signifies the state of the art nonlinearity between the dependent and independent variables [5]. In the recent decade, researchers have deployed NNs for analyzing different kinds of estimation problems in a variety of situations. The model we used in the proposed work is the FFNN model with back-error propagation as depicted in Fig. 2 and 3, respectively for original and normalized data, and for data with statistical moments. The ANN model, bind with the error propagation algorithm (FFBPNN) is a very popular ANN model for prediction and estimation [18]. ANN has normally three layers model such as input layer, hidden layer, and output layer as depicted in Fig. 2 and 4. Researcher always use more than one hidden layer, and a bias node can also be added to the hidden layer to reduce error in the model.

If we compute the hidden layer, we can use the below Equation (6).

$$v_j = \left( 1 + \exp(-1 \times \sum_{i=1}^1 x_i w_{ij}) \right)^{-1} \quad (6)$$

Where  $j$  nodes in the hidden layer can be denoted by  $v_j$ , node  $i$  in the input layer is denoted by  $x_j$ ,  $w_{ij}$  denotes the weight between the nodes,  $y$  represents the output layer node and can be computed by (7).

$$y = \left( 1 + \exp(-1 \times \sum_{j=1}^j v_j w_{ij}) \right)^{-1} \quad (7)$$

The output layer node is denoted by  $y$ , (we have only taken two output nodes in this research work, multiple numbers of output nodes can be taken according to the requirements). Error can be computed between computed data and observed data by using equation (8):

$$\text{Error} = 0.5(d - y)^2 \quad (8)$$

The observed data propagation from the output layer is represented by  $d$ , and the hidden layer can be calculated by using equations 9 and 10, respectively.

$$\delta_y = (d - y)(1 - y) \quad (9)$$

$$\delta_y = v_j(d - v_j)(1 - y)\delta_y w_{j1}, j = 1, \dots, J \quad (10)$$

The input and hidden layers and the adjustment of weight  $w$  between hidden layers and output layers can be computed by the following formulae (11, 12), respectively.

$$\Delta w_{ij} = \alpha \delta_y v_j, i = 1, \dots, I; j = 1, \dots, J \quad (11)$$

$$\Delta w_{ij}^n = \alpha \delta_y v_j, j = 1, \dots, J \quad (12)$$

The learning rate is represented by, and also momentum can be computed as (13-14);

$$\Delta w_{ij}^n = \alpha \delta_y v_j + \beta \Delta w_{j1}^{n-1}, j = 1, \dots, J \quad (13)$$

$$\Delta w_{ij}^n = \alpha \delta_y v_j + \beta \Delta w_{j1}^{n-1}, i = 1, \dots, I; j = 1, \dots, J \quad (14)$$

The iteration of error backpropagation is indicated by  $n$ , and momentum constant is represented by  $\beta$ . This momentum method accelerates the weights to avoid any fluctuations in the training process of error surface in the flat region.

For the validation of each of the models developed for a different number of inputs and hidden neurons, the percentage split method is applied in which the total data are separated, and experimentation is done by 70/30 random training-test splits, 70% for training the data and 30% for testing to validate the samples. This ratio is the standard ratio for splitting training and testing data [15].

### D. Performance Evaluation Layer

Different parameters are available to calculate the performance of the model. Below three performance measurement equations have been used to measure the performance that is mean absolute error (MAE), root mean square error (RMSE) and the mean absolute percentage error (MAPE) [15]. For the assessment of regression accuracy, these performance measurement matrices are normally used in the literature. Numerical equations of MAE, RMSE, and MAPE are calculated by using below Equations (15, 16, 17), alternatively.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=0}^n (A - E)^2} \quad (15)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^n |A_i - E_i| \quad (16)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^n \frac{|A_i - E_i|}{A_i} \times 100 \quad (17)$$

Where the total number of observations is represented by  $N$ , actual values are denoted by  $A$ , and estimated values is represented by  $P$ . As illustrated in Table I, we assessed the performance of the FFBPNN by using MAE, RMSE, and MAPE on different types of data, the values get from MAE, RMSE and MAPE show that the FFBPNN outperforms on data with statistical moments and normalized data.

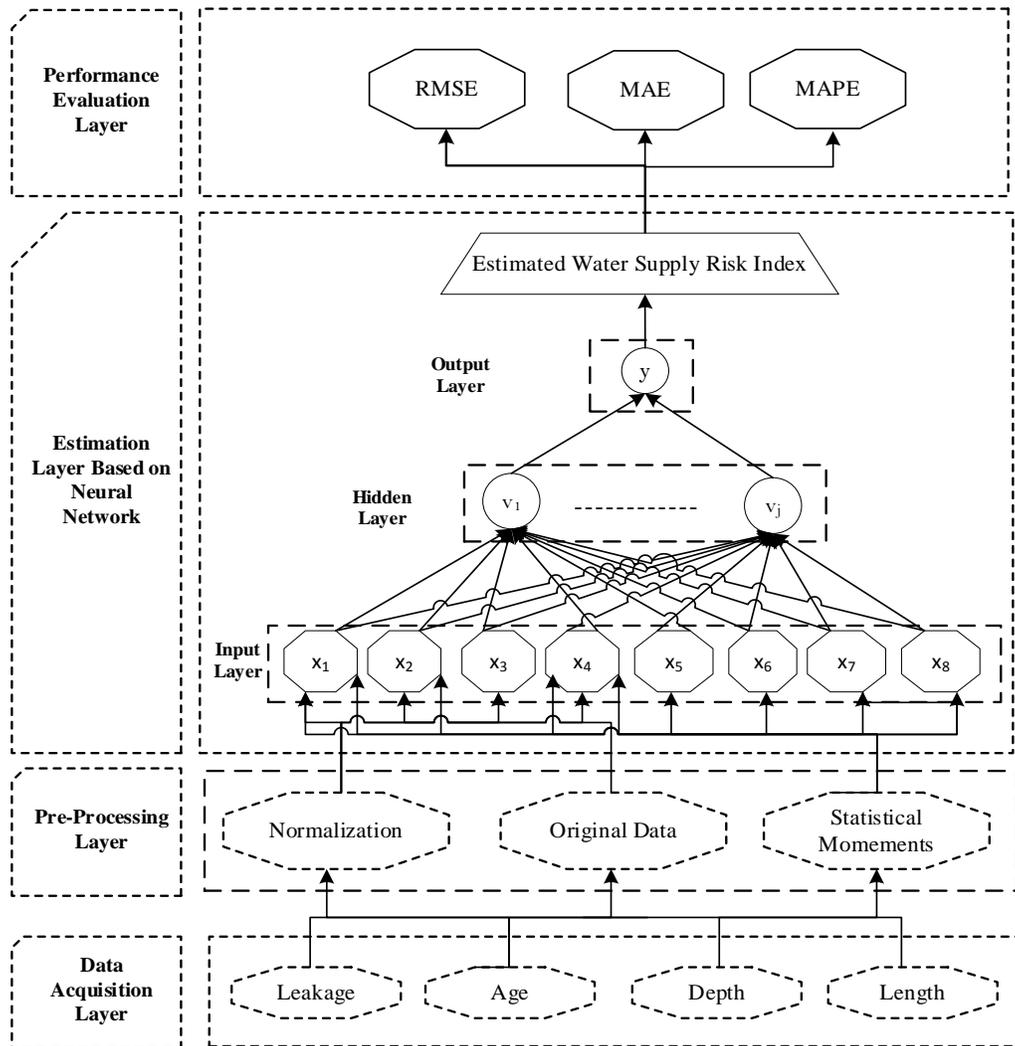


Fig. 1. Proposed Water Supply Pipeline Risk Index Estimation Methodology.

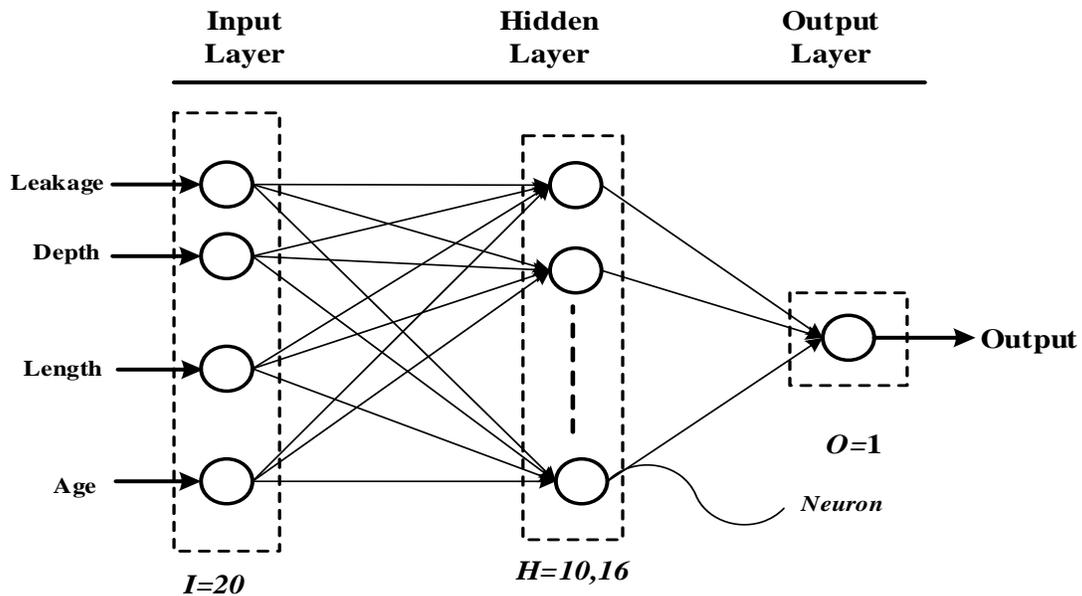


Fig. 2. Structure of Model M1 for Four Inputs.

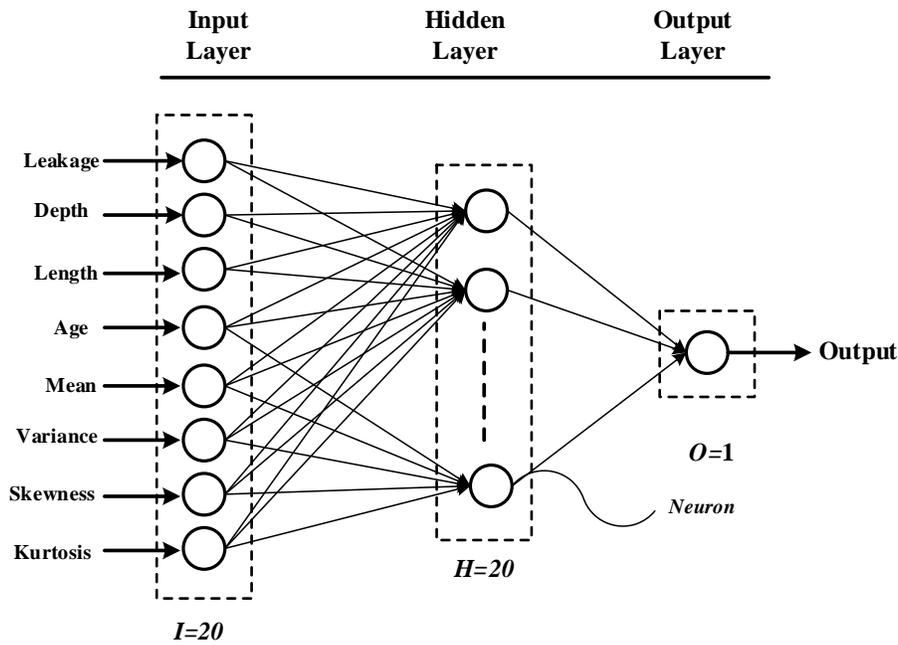


Fig. 3. Structure of Model M2 for Eight Inputs.

#### IV. EXPERIMENTAL RESULTS AND DISCUSSION

##### A. Experimental Setup

The proposed scheme experimentation is performed using windows 10 operating system with an Intel Core i5 processor using MATLAB R2019b version 9.7.0.1216025. In this study different types of experiments have been performed on the data in order to calculate the best estimated risk index for water supply pipelines. Typically input parameters play the most important role in the performance of any kind of machine learning algorithm technique. Therefore, in this research, first the leakage, depth, length and age values of water supply pipelines with the error correction neural network have been given as inputs to the feed forward backpropagation. The model we have tried in this research composed of input layer with different combination of neurons in the hidden layer and output layer. The number of neurons which is best suited to the proposed method is the combination of ten (10) neurons in hidden layer with four (4) neurons in input layer, and single neuron in output layer have been selected as shown in Fig. 4.

Secondly, Different sets of experiments are accomplished with normalized data. Four neurons in the input layer, sixteen (16) neurons in the hidden layer and single neuron in the output layer have been applied as shown in Fig. 5. In the same case, we also tried different combinations of maximum and minimum number of neurons in the hidden layer. We found that this combination (10 neurons in hidden layer) is best fit in combination with four inputs and one output layer.

Third, the experiment is performed by combining original data with statistical moments as shown in Fig. 6. Eight (8)

neurons in the input layer and Twenty (20) neurons are configured in the hidden layer with one neuron in the output layer. Age, depth, height, leakage, variance, mean, kurtosis and skewness are inputs to the neural networks. In order to find the better combination of the number of neurons in the hidden layer with the input and output layer, we tested different number of neurons in the hidden layer and we concluded that this combination is more accurate combination as shown in Fig. 6, thus selected.

##### B. Results and Discussion

The graphical representations of the estimated results are presented in the following section. The actual risk, estimated risk and the errors observed in estimation for water supply pipelines using originally collected data are shown in Fig. 7 and 8. The estimated risk, actual risk and the errors observed in estimated risk and actual risk using FFBPNN on normalized data are shown in Fig. 9 and 10. The actual risk, estimated risk and the errors observed in estimation using FFBPNN for collecting water supply pipeline data with statistical moments are shown in Fig. 11 and 12.

Performance using three measurement; mean absolute percentage error (MAPE), mean absolute error (MAE) and root mean square error (RMSE) is calculated for FFBPNN on the normalized data (ND), original data (SD) and data with statistical moments (SMD) is shown in Table I and Fig. 13. The outcomes show that FFBPNN outperforms on both normalized data (ND) and data with statistical moments (SMD) as compared to original data that is provided to FFBPNN.

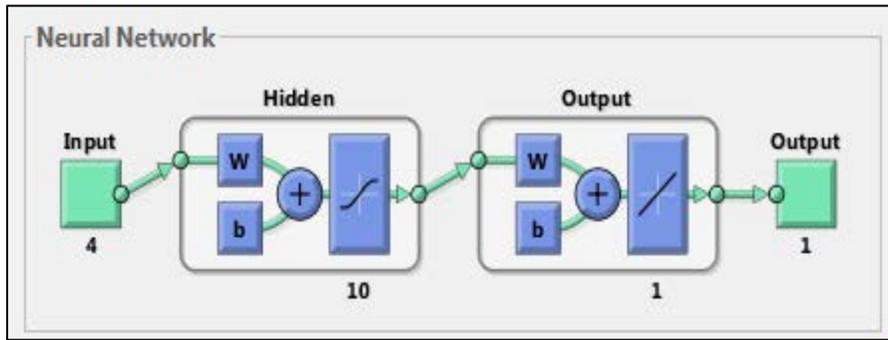


Fig. 4. Deployment of ANN Configuration on Original Data.

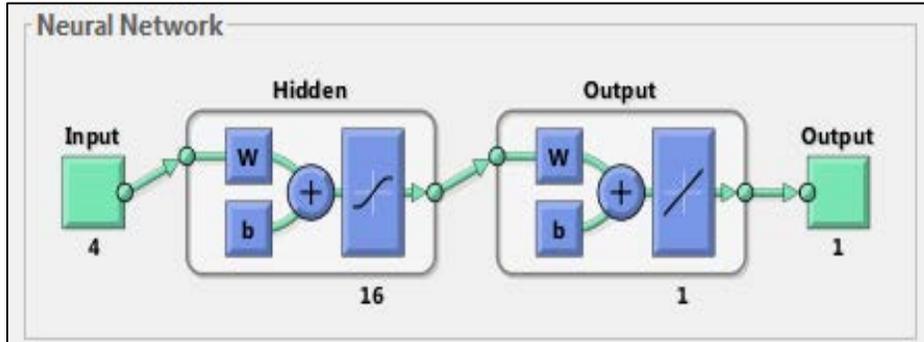


Fig. 5. Deployed ANN Configuration on Normalized Data.

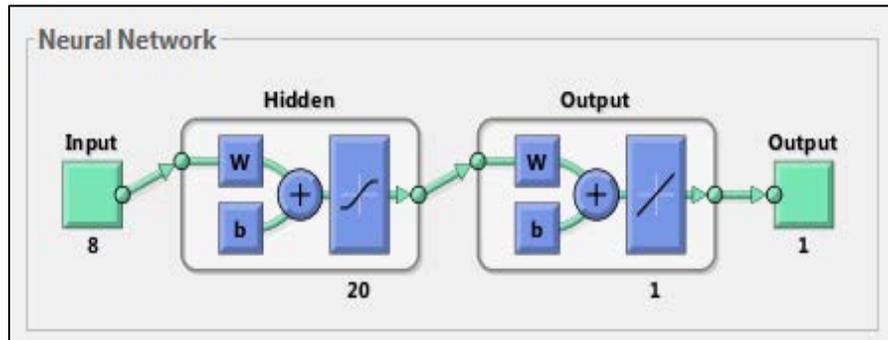


Fig. 6. ANN Configuration Applied on Data with Statistical Moments.

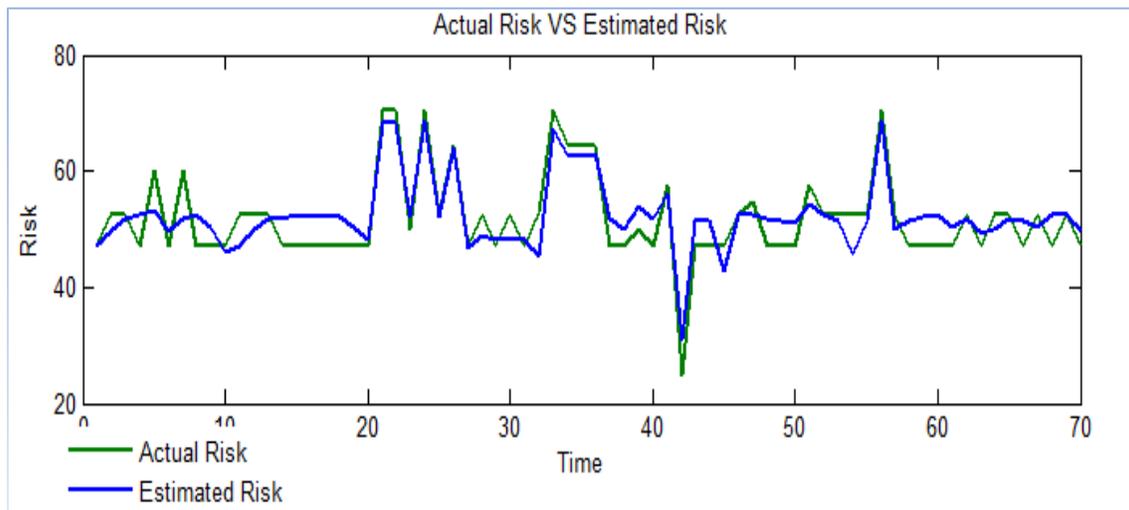


Fig. 7. Actual Risk and Estimated Risk Values of Water Supply Pipelines using FFBPNN on Original Data.

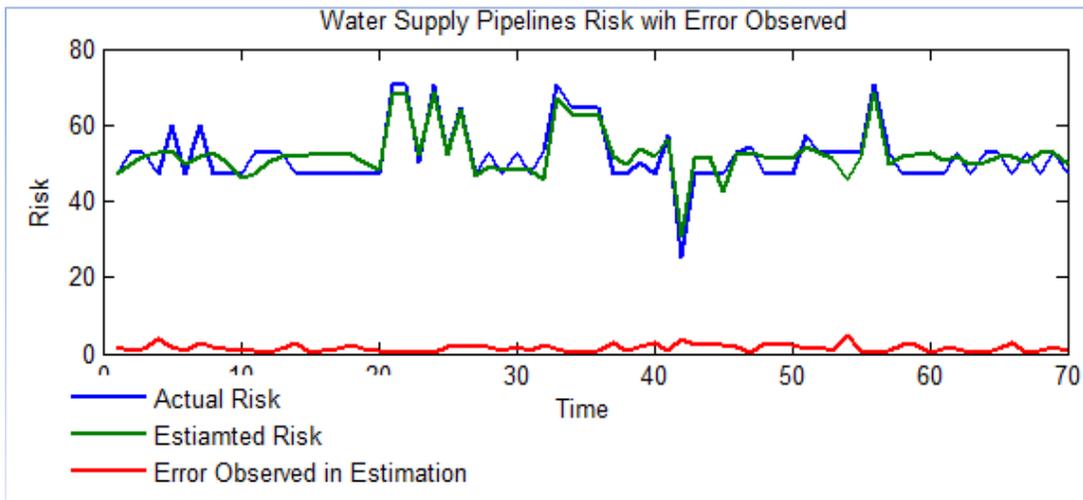


Fig. 8. Actual and Estimated Risk Values of Water Supply Pipelines using FFBPNN on Original Data with Errors Observed.

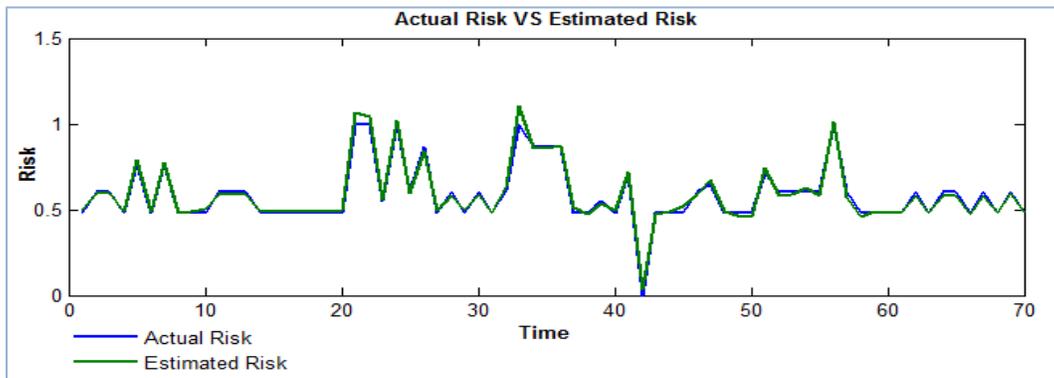


Fig. 9. Actual Risk and Estimated Water Supply Pipelines Risk Values using Neural Network (NN) on Normalized Data.

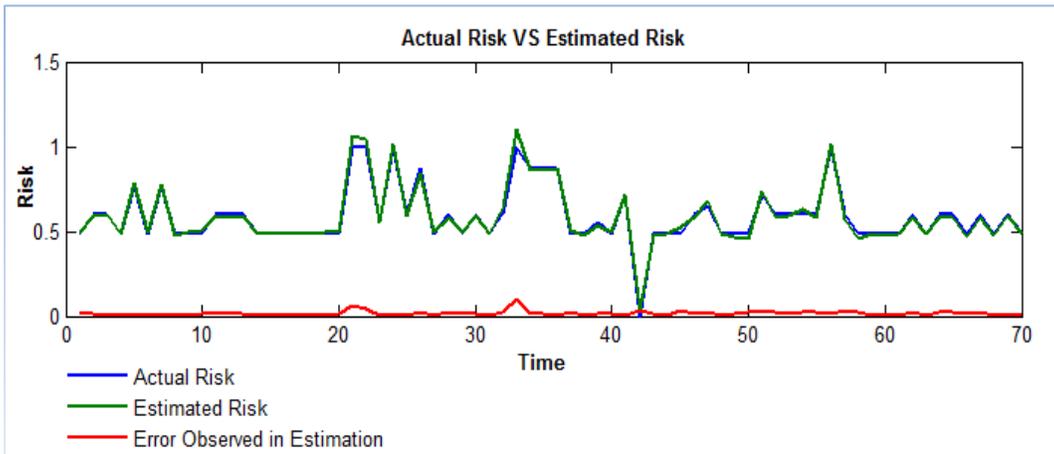


Fig. 10. Estimated Water Supply Pipeline Risk using Neural Network (NN) on Normalized Data with Errors Observed.

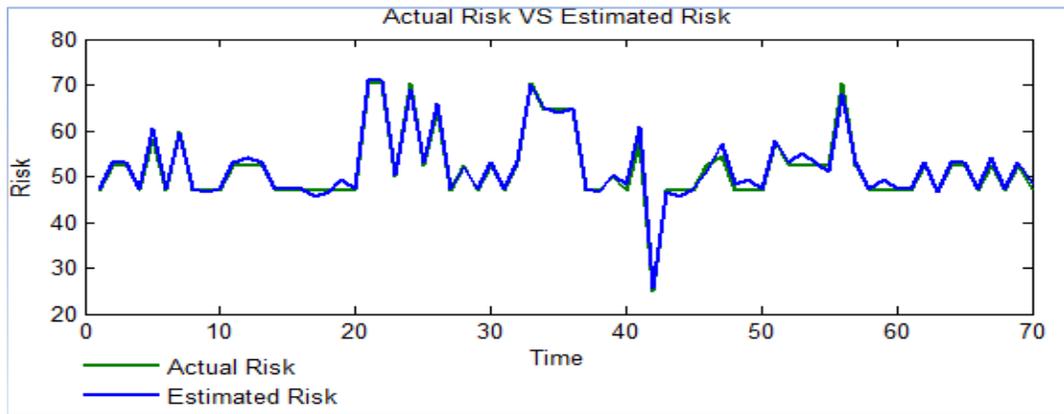


Fig. 11. Actual and Estimated Risk Values of Water Supply Pipeline using Neural Network (NN) on Data with Statistical Moments.

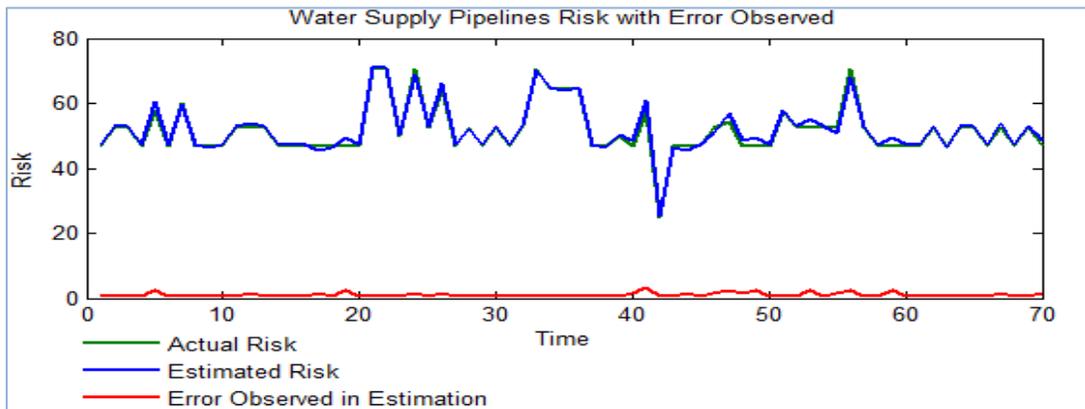


Fig. 12. Actual, Estimated Risk Values of Water Supply Pipelines using Neural Network (NN) on Data with Statistical Moments with Errors Observed.

TABLE I. PERFORMANCE OF FFBPNN ON ORIGINAL AND PREPROCESSED DATA

	Original Data (SD)	Data with Statistical Moments (SMD)	Normalized Data (ND)
MAPE	2.763	2.3965	2.4948
MAE	1.846	1.4619	1.5219
RMSE	1.3907	1.2106	1.3016

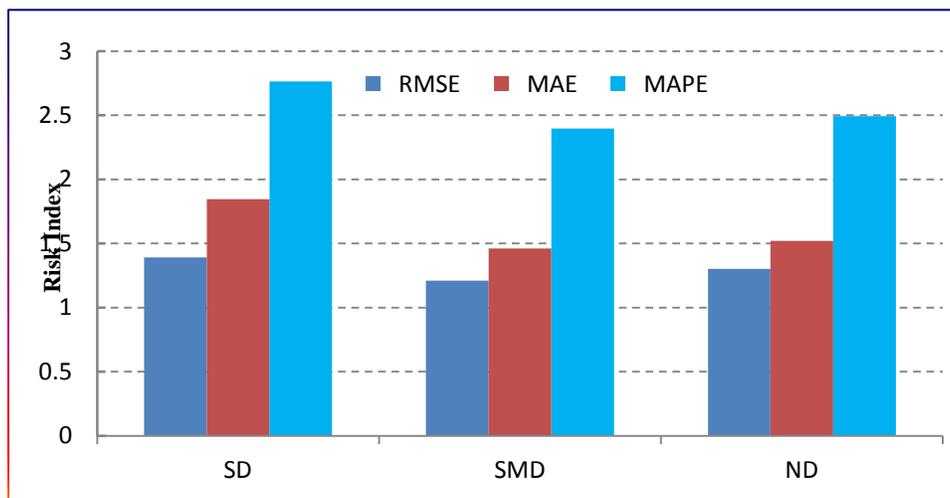


Fig. 13. Graphical Comparison of Estimation Performance of FFBPNN on different Types of Data.

## V. CONCLUSION

In this paper, a multi-layer perceptron is applied to predict water supply pipeline risk. The multi-layer perceptron was randomly trained tested using historical data. The collected data are from 1989-2010 about water supply pipelines fitted in Seoul, Republic of Korea. Experimentation is done by 70/30 random training-test splits, where 70% for training and 30% for testing to validate the samples. For the performance and accuracy evaluation of the models, the root means square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) were used. The FFBPNN was applied to original collected data and we noticed that the outcomes attained for the original data were not prominent. Then we added a tweak by calculating the statistical moments of the original data and merged with the original data. After this process, we applied FFBPNN on this new data. The performance measure matrices indicate that the results provided by FFBPNN using the new data with statistical moments are comparatively better. Further, we normalized the original data and applied FFBPNN using this normalized data with statistical moments where noticeable outperformance is achieved. Overall the performance of FFBPNN on statistical moment data is slightly prominent as compared to normalized data and far better than original data. In future, we may apply some more tweaks on the data and may test more machine learning algorithms.

## REFERENCES

- [1] Kleiner, Y.; Rajani, B. In Considering time-dependent factors in the statistical prediction of water main breaks, AWWA Infrastructure Conference Proceedings, 2000; pp 12-15.
- [2] Fayaz, M.; Pham, Q.B.; Linh, N.T.T.; Nhi, P.T.T.; Khoi, D.N.; Qureshi, M.S.; Shah, A.S.; Khalid, S. A Water Supply Pipeline Risk Analysis Methodology Based on DIY and Hierarchical Fuzzy Inference. *Symmetry* 2020, 12, 44.
- [3] Tang, X.; Liu, Y.; Zheng, L.; Ma, C.; Wang, H. In Leak detection of water pipeline using wavelet transform method, Environmental Science and Information Application Technology, 2009. ESIAT 2009. International Conference on, 2009; IEEE: pp 217-220.
- [4] Caleyó, F.; Velázquez, J.; Valor, A.; Hallen, J. Probability distribution of pitting corrosion depth and rate in underground pipelines: A monte carlo study. *Corrosion Science* 2009, 51, 1925-1934.
- [5] Fayaz, M.; Shah, H.; Aseere, A.M.; Mashwani, W.K.; Shah, A.S. A Framework for Prediction of Household Energy Consumption Using Feed Forward Back Propagation Neural Network. *Technologies* 2019, 7, 30.
- [6] Fares, H.; Zayed, T. Hierarchical fuzzy expert system for risk of failure of water mains. *Journal of Pipeline Systems Engineering and Practice* 1, 53-62.
- [7] Yan, J.; Vairavamoorthy, K. Fuzzy approach for pipe condition assessment. In *New pipeline technologies, security, and safety*, 2003; pp 466-476.
- [8] Kleiner, Y.; Sadiq, R.; Rajani, B. Modeling failure risk in buried pipes using fuzzy markov deterioration process. In *Pipeline engineering and construction: What's on the horizon?* 2004; pp 1-12.
- [9] Kleiner, Y.; Rajani, B.; Sadiq, R. Failure risk management of buried infrastructure using fuzzy-based techniques. *Journal of Water Supply: Research and Technology-AQUA* 2006, 55, 81-94.
- [10] Kleta, H.; Heyduk, A. Image processing and analysis as a diagnostic tool for an underground infrastructure technical condition monitoring. *Underground Infrastructure of Urban Areas* 3, 39.
- [11] Meng, Q.; Qu, X.; Wang, X.; Yuanita, V.; Wong, S.C. Quantitative risk assessment modeling for nonhomogeneous urban road tunnels. *Risk Analysis* 31, 382-403.
- [12] Duzgun, H. Analysis of roof fall hazards and risk assessment for zonguldak coal basin underground mines. *International Journal of Coal Geology* 2005, 64, 104-115.
- [13] Ustinovichius, L.; Barvidas, A.; Vishnevskaja, A.; Ashikhmin, I.V. Multicriteria verbal analysis for the decision of construction problems. *Technological and Economic Development of Economy* 2009, 15, 326-340.
- [14] Kumar, M.; Raghuvanshi, N.; Singh, R.; Wallender, W.; Pruitt, W. Estimating evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering* 2002, 128, 224-233.
- [15] Wahid, F.; Kim, D.H. Short-term energy consumption prediction in korean residential buildings using optimized multi-layer perceptron. *Kuwait Journal of Science* 44.
- [16] Chow, T.; Zhang, G.; Lin, Z.; Song, C. Global optimization of absorption chiller system by genetic algorithm and neural network. *Energy and buildings* 2002, 34, 103-109.
- [17] Firat, F.; Arslan, A.K.; Colak, C.; Harputluoglu, H. Estimation of risk factors associated with colorectal cancer: An application of knowledge discovery in databases. *Kuwait Journal of Science* 43.
- [18] Gibbs, M.S.; Morgan, N.; Maier, H.R.; Dandy, G.C.; Nixon, J.B.; Holmes, M. Investigation into the relationship between chlorine decay and water distribution parameters using data driven methods. *Mathematical and Computer Modelling* 2006, 44, 485-498.
- [19] Fayaz, M., Ahmad, S., Ullah, I., & Kim, D. A blended risk index modeling and visualization based on hierarchical fuzzy logic for water supply pipelines assessment and management. *Processes*, 2018, 6, 61.
- [20] Fayaz, M., Ahmad, S., Hang, L., & Kim, D. Water supply pipeline risk index assessment based on cohesive hierarchical fuzzy inference system. *Processes*, 2019, 7, 182.

# Modeling of Coronavirus Behavior to Predict it's Spread

Shakir Khan<sup>1</sup>, Amani Alfaifi<sup>2\*</sup>

College of Computer and Information Sciences  
Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia

**Abstract**—With the increasing presence and feast of infectious diseases and their fatalities in densest areas, many academics and societies have become fascinated in discovering new behaviors to predict these diseases' feast behaviors. This media will help them to plan and contain the disease better in trivial provinces and thus decrease the beating of human lives. Some cases of an indeterminate cause of pneumonia occurred in Wuhan, Hubei, China, in December 2019, with clinical presentations closely resembling viral pneumonia. In-depth analyzes of the sequencing from lower respiratory tract samples discovered a novel coronavirus, called 2019 novel coronavirus (2019-nCoV). Current events showed us how easily a coronavirus could take root and spread—such viruses transmitted easily between persons. To cure with these infections, we applied time series forecasting model in this paper to predict possible coronavirus events. The forecasting model applied is SIR. The results of the implemented models compared with the actual data.

**Keywords**—COVID-19; coronavirus; SIR model; data mining; R Software; forecasting model

## I. INTRODUCTION

On 30 January 2020, the WHO announces the outbreak of COVID-19 as a public health emergency of international concern (PHEIC) by the WHO [14]. This continuing outbreak has since 3 December 2019 spread to over 50 other countries [13]. There are over 1 million cases of confirmed COVID-19 worldwide and deaths over 50 000 as of 20 February [15]. Together with MERSnCoV and SARS-nCoV, it is the 7th member of the coronavirus group which can spread to humans [1,2,16].

Human coronaviruses, which include hCoV-229E, OC43, NL63, and HKU1, cause light respiratory disease. Fatal coronavirus infections that have occurred over the past two decades are extreme coronavirus acute respiratory syndrome (SARS-CoV) and coronavirus respiratory syndrome in the Middle East [3]. Coronavirus disease (COVID-19) is a recently discovered coronavirus-caused infectious disease. The majority of people diagnosed with COVID-19 have mild to moderate respiratory diseases [4]. This way has drawn considerable interest not only in China but globally.

A brief definition for the term “data mining” is to extract the useful information and patterns from large data sources” [11]. Across different areas, data mining used to help to increase the quality and efficacy of pattern detection and analysis of such events using existing statistical evidence [32]. Data mining technology is the process by which we perform all sorts of analyzes on vast volumes of data [33, 34]. This

paper goals to use perceptions made possible from data mining methods to relief in predicting coronavirus feast. Large volumes of data can be cumbersome and repetitive to compile and analyze. However, the underlying patterns in the data identified, such that the predicted occurrence of coronavirus is known beforehand. Forward-thinking techniques joining from the grounds of computer science, mathematics, and data science are necessary for discovering these underlying patterns [12]. Because data mining is a vast assembly of procedures and has extensive solicitation, and agreed data mining concept differs slightly based on the source. The data mining can also be named like "Big Data" or "Data Science" as the alternative name of data mining [19].

Various imminent characteristics of our lives rely on historical data arithmetic analysis. For example, prediction of illness, changes in stock market activities, weather prediction, etc. can be forecasted only if we can find a pattern in historical data due to time and it can be any ways for example daily, weekly, monthly, or annually. This form of forecast is commonly called Forecasting of the Time Series. Observations were sequentially taken in time, usually called time series [5]. Mathematical models have been applied to study a variety of communicable disease outbreaks [8], [9], [10].

The formal description of predictive modeling given by [18] is “the method of evolving a mathematical tool or prototypical that produces a perfect prediction”. Predictive demonstrating could be engaged to compute information accessible and to create better conversant result based on different information finding. Modeling techniques offer computational capabilities in circumstances with vast quantities of evidence to build prototypes with the prognostic implication that help in hands-on conclusions. Because of overlapping algorithms to discover unseen facts in data, prognostic modeling closely correlated with data mining [19].

Data spring-cleaning, structure, statistical analysis, prognostic modeling, and statistics picturing performed for this project with R software system and R-studio GUI have used widely. R is a programming language open-source platform with different statistical functionalities. The R environment is interactive computing, graphical display, statistics, visualization, and data manipulation suite. Robert Gentleman and Ross Ihaka of the Department of Statistics at the University of Auckland initially wrote R. Currently, R is a product of user assistances around the world. Apart from users who build innovative features, there are thousands of built

\*Corresponding Author

packages open on the Robust R Archive Network (CRAN) with countless features and capabilities [20].

For mathematical epidemiology, the SIR model is perhaps the most commonly used [23, 24 and 25]. It widely used for a variety of purposes. Often as the core of further multifarious epidemiological prototypes for the transmission of communicable diseases [24, 26, 27], and other times to research the propagation of phenomena in other regions, such as rumor spread [28], computer viruses [29], Dissemination of information in Web fora [30], or the behavior of investors in stock markets [31].

For this paper, we used statistical models of time series to forecast possible cases of coronavirus. Because coronavirus is one of the contagious viruses, it is essential to estimate the number of cases so that the government can take the appropriate steps and measures to avoid its spread to treat or avoid viruses.

## II. RELATED WORKS

S. Eubank et al. [6] analyzed the algorithmic and structural properties of Portland, Oregon, from extensive social communication networks. A bipartite graph composed of individuals and places created. The individuals are nodes, while edges reflect places. This binary configuration does not provide users with node to node interaction information. Using a CL-model they use the method of a random graph. The CL-model correctly mimics the dataset's critical features. Alternatively, algorithms of rapid approximation developed to measure basic structural properties. Such studies investigated the impact of political decisions on the management of large-scale urban diseases. It study demonstrates the effect of different algorithms that deal with the pattern of spreading the disease but do not visualize its effect.

An epidemiological model developed by K. Wang et al. to explain the flu when its transmission through a network of human contacts. The aim of this study is to build an ABM method combining GIS and civic ecosystems to mimic the spread of influenza. Using JAVA and GIS software, the model was developed using the Repast Symphony framework. A system developed to simulate the spread and control of influenza in a specific area. the model defined influenza using a mathematical relationship among the probability of transmission, the distance between two persons, the latent duration, The time between infections and death, the rate of cure. For example, users could modify the results of the simulation by changing the time-value to the hospital. [7].

The research, as mentioned earlier, and models offer useful information which allows medical leaders to take better outbreak protocol decision making. Nonetheless, each model mentioned above lacks a particular feature, whether it is computational capacity, necessary variables that are essential in calculating an accurate pattern of virus spread, susceptibility to disease expansion, or the incapability to measure over one form of spreading disease.

SEIR points respectively to the Susceptible, Exposed, Infectious, and Removed or Recovered. These results are designed on the basis SIR but gives a variable to the container

Exposed. Susceptible relates to persons that may acquire the infection and be carriers if infect, the exposed are persons already infected but asymptomatic, the infectious are persons that display signs of infection. They may spread the disease, eliminate, or recover the virus are previously infected persons who are no longer infectious and who are now immune to the virus [17].

## III. METHODOLOGY

### A. Dataset

The dataset used is the "COVID 19 data.csv" file, which the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) periodically updates. The dataset contains data about COVID-19, which is the Province/State and Country/Region also the date and confirmed, deaths, and recovered cases number the data was collected from the starting of the virus spread as shown (screenshot) in Fig. 1.

### B. Data Visualization

From Fig. 2 to 7, data were visualized as most of the countries where COVID19 is spread, the rates of Infect, recovery, and death caused by the epidemic, as well as the rates of new cases in the most affected countries.

By using the R program, we visualize major outbreaks for the top 10 countries in Fig. 2.

SNo	ObservationDate	Province.State	Country.Region	Last.Update	Confirmed	Deaths	Recovered
1	2020-01-22	Anhui	Mainland China	1/22/2020 17:00	1	0	0
2	2020-01-22	Beijing	Mainland China	1/22/2020 17:00	14	0	0
3	2020-01-22	Chongqing	Mainland China	1/22/2020 17:00	6	0	0
4	2020-01-22	Fujian	Mainland China	1/22/2020 17:00	1	0	0
5	2020-01-22	Gansu	Mainland China	1/22/2020 17:00	0	0	0
6	2020-01-22	Guangdong	Mainland China	1/22/2020 17:00	26	0	0
7	2020-01-22	Guangxi	Mainland China	1/22/2020 17:00	2	0	0
8	2020-01-22	Guizhou	Mainland China	1/22/2020 17:00	1	0	0
9	2020-01-22	Hainan	Mainland China	1/22/2020 17:00	4	0	0
10	2020-01-22	Hebei	Mainland China	1/22/2020 17:00	1	0	0
11	2020-01-22	Heilongjiang	Mainland China	1/22/2020 17:00	0	0	0
12	2020-01-22	Henan	Mainland China	1/22/2020 17:00	5	0	0
13	2020-01-22	Hong Kong	Hong Kong	1/22/2020 17:00	0	0	0
14	2020-01-22	Hubei	Mainland China	1/22/2020 17:00	444	17	28
15	2020-01-22	Hunan	Mainland China	1/22/2020 17:00	4	0	0
16	2020-01-22	Inner Mongolia	Mainland China	1/22/2020 17:00	0	0	0
17	2020-01-22	Jiangsu	Mainland China	1/22/2020 17:00	1	0	0
18	2020-01-22	Jiangxi	Mainland China	1/22/2020 17:00	2	0	0
19	2020-01-22	Jilin	Mainland China	1/22/2020 17:00	0	0	0
20	2020-01-22	Liaoning	Mainland China	1/22/2020 17:00	2	0	0
21	2020-01-22	Macao	Macao	1/22/2020 17:00	1	0	0
22	2020-01-22	Ningxia	Mainland China	1/22/2020 17:00	1	0	0
23	2020-01-22	Qinghai	Mainland China	1/22/2020 17:00	0	0	0

Fig. 1. Dataset.

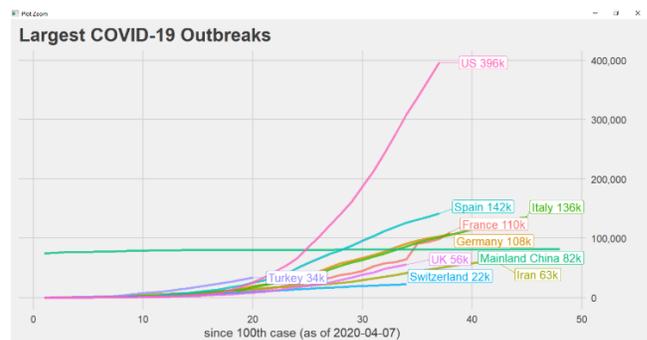


Fig. 2. Major Outbreaks.

And daily new cases for the top 20 countries in Fig. 3:

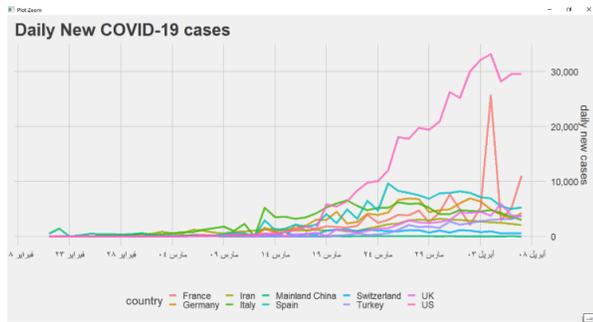


Fig. 3. Daily New Cases.

Daily new COVID-19 cases in the 12 most-affected countries in Fig. 4:

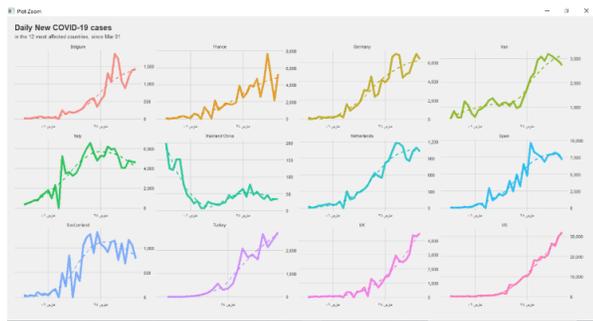


Fig. 4. Daily New COVID-19 Cases in the 12 Most-Affected Countries.

Daily new deaths in the 12 most-affected countries in Fig. 5:

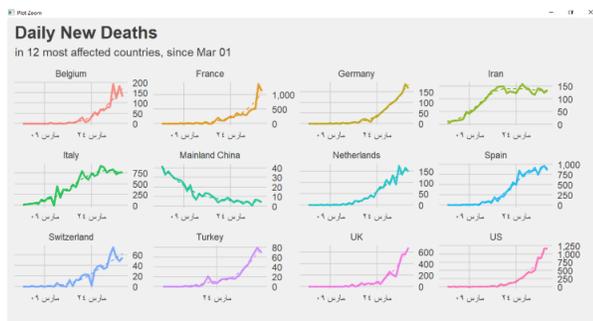


Fig. 5. Daily New Deaths in 12 Most Affected Countries.

Finally, daily recovered in 12 most affected countries in Fig. 6:

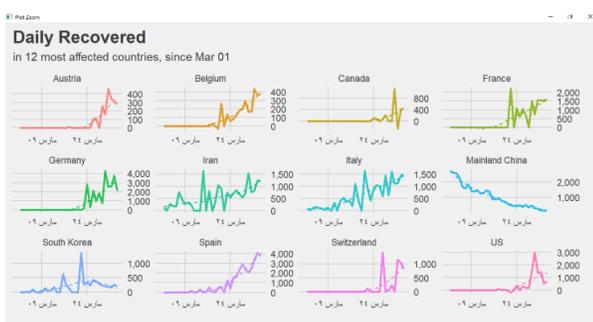


Fig. 6. Daily Recovered in 12 Most Affected Countries.

### C. SIR Model

Here a briefly explain the features of the essential Susceptible-Infected- Recovered (SIR) system which used define the recent COVID-19 outbreak. The original SIR model, in which Kermack and McKendrick modified a Malthusian growth model, is a model known to simulate epidemic growth using ODE.

The SIR model describes three-stage rules for the infection. The first rule is the unsafe condition (S) in which an agent is likely to become contaminated at any given point in time. The second is infection (I) when many neighbors are also in this state. The agent will switch to that state. An agent moves to the third state that is recovered (R) after a given period [21]. The SIR diagram in Fig. 7 shows how individuals move.

According to [22], S, I, and R are susceptible, infectious, and removed individuals, and where parameters  $\beta$  and  $\pi$  are the rate of infection and the rate of recovery. The equations for every time t are defined as follows:

$$dS(t) / dt = -\beta S(t) I(t),$$

$$dI(t) / dt = \beta S(t) I(t) - \gamma I(t),$$

$$dR(t) / dt = \gamma I(t)$$

We first created a time series of that data, divided into infected and recovered data. We can visualize the generated time-series in Fig. 7.

It is interesting to note that there two sub-waves of infection, which may be the time the coronavirus left China and spread worldwide. The two different growths in the number of reported cases (indicating spread) can be more clearly seen in the graph in Fig. 8 after calculating the rate of infection and recovery.

The first wave considered to run from the beginning of the dataset until 14 February, and the second wave is from 15 February onwards until four days before the current date. This cut-off lets us later compare the output of the model with the actual data. At the 95% confidence interval, both the two-sided t-test and the KS-test conducted with the null hypothesis that the two infection growth rates are the same as the output in Table I.

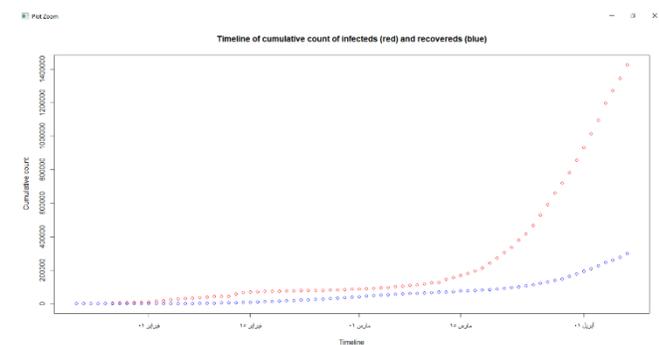


Fig. 7. Timeline of Cumulative Count of Infected and Recovered.

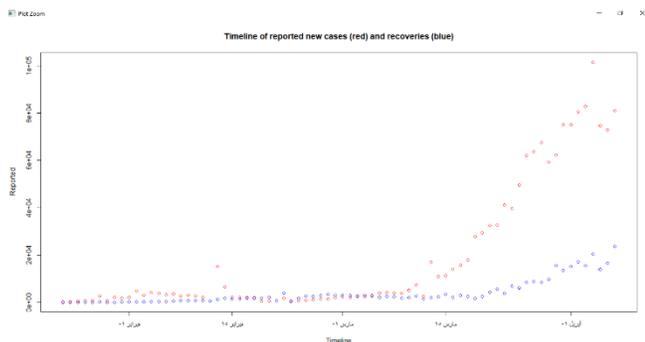


Fig. 8. Timeline of Reportes New Cases and Recoveries.

TABLE I. T-TEST, KS-TEST

t-test	<p>Welch Two Sample t-test</p> <p>data: log(wave.1.results\$rate) and log(wave.2.results\$rate)</p> <p>t = -4.3879, df = 61.827, p-value = 4.543e-05</p> <p>alternative hypothesis: true difference in means is not equal to 0</p> <p>95 percent confidence interval:</p> <p>-2.1750871 -0.8135133</p> <p>sample estimates:</p> <p>mean of x mean of y</p> <p>7.469233 8.963533</p>
K S-test	<p>Two-sample Kolmogorov-Smirnov test</p> <p>data: log(wave.1.results\$rate) and log(wave.2.results\$rate)</p> <p>D = 0.45652, p-value = 0.001595</p> <p>alternative hypothesis: two-sided</p>

So, the inference is that the second wave is more aggressive on average than the first (rejecting  $H_0$  at 5 percent t-test significance) since the KS-Test's p-value is above 0.05. In contrast, the t-test is only marginally lower, while the overall shape of the distribution is similar (not enough evidence to reject  $H_0$  at 5% concerning the KS-test).

#### IV. APPLYING THE SIR MODEL AND DISCUSSION

$N$  reflects the total considered population. One condition the SIR equations have to satisfy for all  $t$  is:

$$N = S+I+R$$

We will observe linear growths in  $I$  in order to evaluate  $b$  and  $c$ , hence the motivation to divide the data into two waves, as discussed earlier. As we have found that both waves behave similarly, we carry out the modeling using the second wave only. The output is shown in Table II.

The p-value is small for the coefficient, which shows that we have sufficient evidence to reject the null hypothesis and conclude that there is a linear correlation between growth and time. The R-squared values are also very similar to 1 which means that the above regression also provides a good fit for the variance.

TABLE II. REGRESSION SUMMARY

The regression summary of infected	<p>Residuals:</p> <p>Min 1Q Median 3Q Max</p> <p>-29196 1494 4987 8254 10342</p> <p>Coefficients:</p> <p>Estimate Std. Error t value Pr(&gt; t )</p> <p>x -0.18938 0.00549 -34.49 &lt;2e-16 ***</p> <p>---</p> <p>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</p> <p>Residual standard error: 7588 on 49 degrees of freedom</p> <p>Multiple R-squared: 0.9604, Adjusted R-squared: 0.9596</p> <p>F-statistic: 1190 on 1 and 49 DF, p-value: &lt; 2.2e-16</p>
The regression summary of recovery	<p>Residuals:</p> <p>Min 1Q Median 3Q Max</p> <p>-5552.2 94.7 793.4 1667.3 3634.0</p> <p>Coefficients:</p> <p>Estimate Std. Error t value Pr(&gt; t )</p> <p>x 0.019180 0.000649 29.55 &lt;2e-16 ***</p> <p>---</p> <p>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</p> <p>Residual standard error: 1628 on 49 degrees of freedom</p> <p>Multiple R-squared: 0.9469, Adjusted R-squared: 0.9458</p> <p>F-statistic: 873.3 on 1 and 49 DF, p-value: &lt; 2.2e-16</p>

So, the  $c$  and  $b$  are determined, which means we have the necessary coefficients to build the model. That is where the last equation will come in. The model is building on the already present data, predicting 150 days in the future. Fig. 9 shows the output of the SIR model along with the data.

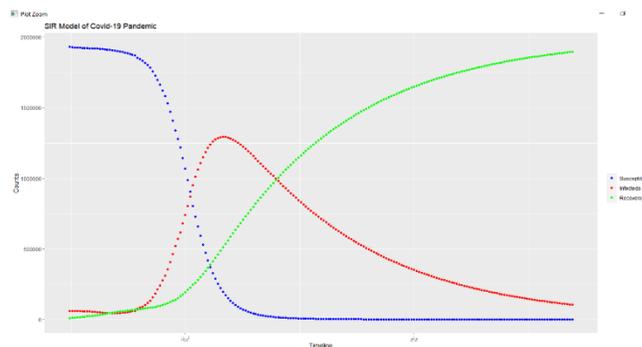


Fig. 9. SIR Modle of COVID-19.

The result below shows the model predictions of five days in the past and five days in the future:

Dates	Susceptible	Infected	Recovered
50	2020-04-04	802595.0	951253 246152.0
51	2020-04-05	727885.0	1012103 260012.0
52	2020-04-06	658128.7	1062447 279424.2
53	2020-04-07	591920.1	1108278 299802.0
54	2020-04-08	529803.4	1149138 321058.9
55	2020-04-09	472155.6	1184745 343099.5
56	2020-04-10	419188.5	1214989 365823.0
57	2020-04-11	370962.9	1239911 389126.6
58	2020-04-12	327410.0	1259682 412908.1
59	2020-04-13	288357.5	1274574 437069.0
60	2020-04-14	253556.4	1284928 461515.4

In Table III, a comparison between actual data and the data from the SIR model application, which proved the efficiency of the model as the numbers are very close to each other, while some numbers constitute a perfect match. It is important to note that the recovery rate tends to be reliable in the short term, at least. In contrast, the infected rate causes underestimation when this model previously tested. With more data available due to disease progression, the predictions will get better. The results are generally reassuring that the epidemic, according to the model, will soon be limited.

TABLE III. COMPARING PREDICTED INFECTIONS AND RECOVERY NUMBERS WITH THE ACTUAL NUMBERS

	Dates	Predicted_Infecteds	Predicted_Recovereds	Actual_Infecteds	Actual_Recovereds
1	2020-04-05	1012103	260012.0	1012103	260012
2	2020-04-06	1062447	279424.2	1068586	276515
3	2020-04-07	1108278	299802.0	1126042	300054

### V. CONCLUSION

Coronavirus has become a concern of many recently, and the focus on it became intense due to the rapid spread and lethality of people. In this paper, the SIR model for predicting 150 days in the future was applied, and the numbers compared with the real numbers, which proved the efficiency of the model. As the numbers are very close to each other, while some numbers constitute a perfect match. The results are generally reassuring that the epidemic, according to the model, will soon be limited.

### ACKNOWLEDGMENT

The researcher is master student of big data course and wants to thank instructor Dr. Shakir Khan for his valuable support in this research.

### REFERENCES

- [1] WHO. Novel coronavirus – China. Jan 12, 2020. <http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/> (Jan 12, 2020).
- [2] WHO. Novel coronavirus – Thailand (ex-China). Jan 14, 2020. <http://www.who.int/csr/don/14-january-2020-novel-coronavirusthailand/en/> (Jan 14, 2020).
- [3] Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China [published correction appears in *Lancet*. 2020 Jan 30;]. *Lancet*. 2020;395(10223):497-506. doi:10.1016/S0140-6736(20)30183-5
- [4] WHO. Novel coronavirus – [https://www.who.int/health-topics/coronavirus#tab=tab\\_1](https://www.who.int/health-topics/coronavirus#tab=tab_1) (2020).
- [5] S. Nashreen, and N. Sharma, "Statistical Models for Predicting Swine Flu Incidences in India." 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC). IEEE, 2018.
- [6] S. Eubank, V.S. Kumar, M. V. Marathe, A. Srinivasan, N. Wang, "Structural and algorithmic aspects of massive social networks." Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms.718-727, 2004
- [7] J.Wang, J. Xiong, K. Yang, S. Peng, Q. Xu, "Use of GIS and Agent-Based Modeling to Simulate the Spread of Influenza," 2010 18th International Conference on Geoinformatics. IEEE, pp. 1-6, 2010
- [8] R. M. Anderson, R. M, "Infectious Diseases of Humans", Oxford Science Publications, May 1992.
- [9] N. G. Becker, L. R. Egerton, "A Transmission Model of HIV", *Mathematical Biosciences*, vol. 119, pp. 205-224, 1994.
- [10] J. D. Murray, "Mathematical Biology Second Corrected Edition", Berlin: Springer- Verlag, 1993.
- [11] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, "From Data Mining to Knowledge Discovery in Databases", *AI Mag.*, vol. 17, no. 3, pp. 37, 1996.
- [12] M. Campion et al., "Predicting West Nile Virus (WNV) occurrences in North Dakota using data mining techniques," 2016 Future Technologies Conference (FTC), San Francisco, CA, pp. 310-317, doi: 10.1109/FTC.2016.7821628, 2016.
- [13] World Health Organization (WHO), "Coronavirus disease 2019 (COVID-19) Situation Report - 35," WHO, 2020.
- [14] World Health Organization (WHO), "Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)," WHO, 2020.
- [15] World Health Organization (WHO), " Coronavirus disease 2019 (COVID-19): Situation Report - 75 (4 April 2020)", WHO, 2020.
- [16] D. Hui et al, "The continuing 2019-nCoV epidemic threat of novel coronavirus to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China," *International Journal of Infectious Diseases*, vol. 91, pp. 264-266, 2020.
- [17] A. Rachah and D. F. M. Torres, "Analysis, simulation and optimal control of a SEIR model for Ebola virus with demographic effects," *Commun. Fac. Sac. Univ. Ank. Series A1*, vol. 67, no. 1, pp. 179-197, 2018.
- [18] M. Kuhn, K. Johnson, *Applied Predictive Modeling*, New York City: Springer New York Heidelberg Dordrecht London, 2013.
- [19] Campion, Mitch, et al. "Predicting West Nile Virus (WNV) occurrences in North Dakota using data mining techniques." 2016 Future Technologies Conference (FTC). IEEE, (pp. 310-317), 2016.
- [20] R. Gentleman, R. Ihaka, "R: The R Project for Statistical Computing", 2016, <https://www.r-project.org/>.
- [21] N. Guizani, and A. Ghafoor. . Modeling and evaluation of disease spread behaviors. In 2014 International Wireless Communications and Mobile Computing Conference (IWCMC) (pp. 996-1003). IEEE. 2014.
- [22] Y. Maki, & H. Hirose, "Infectious disease spread analysis using stochastic differential equations for SIR model". In 2013 4th International Conference on Intelligent Systems, Modelling and Simulation (pp. 152-156). IEEE. 2013.
- [23] B. Norman, "The mathematical theory of infectious diseases and its applications. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE. 1975.

- [24] R. M. Anderson, and M. M. Robert, "Infectious diseases of humans: dynamics and control". Oxford university press, 1992.
- [25] K. Matt, and P. Rohani, "Modeling infectious diseases in humans and animals". Princeton University Press, 2011.
- [26] K. Sattar, T. Ahmad, Abdulghani HM, Khan S, John J, Meo S. Social networking in medical schools: medical student's viewpoint. *Biomed Res* 2016; 27: 1378-84
- [27] H. Hans, et al. "Modeling infectious disease dynamics in the complex landscape of global health." *Science* 347.6227: aaa4339, 2015.
- [28] D.J. Daley, D.G. Kendall, "Epidemics and rumours. *Nature*", 204(4963), 1118-1118. 1964.
- [29] B.K Mishra, D Saini. "Mathematical models on computer viruses". *Applied Mathematics and Computation*, 187(2), 929-936. 2007
- [30] J. Woo, H. Chen. Epidemic model for information diffusion in web forums: experiments in marketing exchange and political dialog. *Springer Plus*, 5(1), 66. 2016
- [31] S. Shive, "An epidemic model of investor behavior," *Journal of Financial and Quantitative Analysis*, 45(1), 169-198. 2010.
- [32] S. Khan. "How Data Mining Can Help Curb City Crime", *International Journal of Control Theory and Applications (IJCTA)* 9 (23), 483-488, 2016
- [33] M.F. AlAjmi, & S. Khan. "Data Mining-Based, Service-Oriented Architecture (SOA) in e-Learning". In *ICERI2012 Proceedings* pp. 3023-3023, IATED. 2012
- [34] S. Khan, M. F. AlAjmi. "Impact of Medical Technology on Expansion in Healthcare Expenses", *International Journal of Advanced Computer Science & Applications* 4 (4), pp. 150-152. 2013.

# Automatic Segmentation of Hindi Speech into Syllable-Like Units

Ruchika Kumari<sup>1</sup>

Department of ECE

Maharaja Surajmal Institute of Technology

GGSIU,

New Delhi, India

Research Scholar

Indira Gandhi Delhi Technical University for Women

New Delhi, India

Amita Dev<sup>2</sup>

Vice-Chancellor

Indira Gandhi Delhi Technical University for Women

New Delhi, India

Ashwani Kumar<sup>3</sup>

Department of ECE

Indira Gandhi Delhi Technical University for Women

New Delhi, India

**Abstract**—To develop the high-quality Text-to-Speech (TTS) system, appropriate segmentation of continuous speech into the syllabic units placed an important role. The research work has been implemented for automatic syllable based speech segmentation technique for continuous speech for the Hindi language. The experiments were conducted by using the energy convex hull approach for clean, continuous speech for Hindi. In this method, the Savitzky-Golay filter was applied on the short term energy (STE) signal to increase the signal to noise ratio (SNR), followed by applying the median filter to preserve the boundaries, hence smoothing the energy curve. Also, the Hamming sliding-window was applied twice on speech signal to get the more accurate depth of convex hull valleys. Further, the algorithm was tested on 50 unique utterances chosen from the travel domain. The accuracy of the proposed algorithm has been calculated and obtains that 76.07% syllables have time-error less than 30 ms with manual segmentation reference. The performance of the proposed algorithm is also analyzed and gives better-segmented accuracy as compared to the existing group delay segmentation technique for fricatives or nasal sounds. The syllable base segmented database is suitable for the speech technology system for Hindi in the travel domain.

**Keywords**—Database; short term energy; convex hull; speech segmentation; syllable

## I. INTRODUCTION

Speech is considered as quasi-periodic signal since the characteristic of the signal changes over time. Segmentation is the process of splitting the speech signal into several parts. Speech can be segmented into various units, such as words, syllables, and phones. TTS is the ability of a machine to convert the given text in a language to spoken speech.

The accurate segmentation and label play a vital role in developing the TTS. The speech synthesis system makes use of various speech and language technology. It is being used to enhance human-machine interactions such as in mobile communication, screen reader, remote access to online information. The various application of speech synthesis includes talking aids, health care, banks, travel and tourism, visual and speech impairment, etc. Building a TTS for any

language requires a corpus, which is a labor-intensive and time-consuming task. The research aim is to develop and analyze continuous speech segmentation as syllable like units for the Hindi language. Hindi is one of the official languages of India. It is a primary communication language for a large number of Indian populations and in other parts of the world. Most of the research has been done in other languages, such as European, English, Mandarin, Arabic, etc. However, less work has been done in the Hindi language due to a lack of standard database and pronunciation rule. As Hindi is syllable-centric in nature, the syllable is considered as an appropriate segment to a label. Several advance works have been reported to the phoneme level segmentation technique but still lacking on syllable base level.

The objective of the paper is to propose a time-domain automatic segmentation technique based on STE and convex hull approach for the Hindi language. Moreover, applied Savitzky-Golay filter [13] and median filter to get smoother energy curve and also apply Hamming sliding-window twice on STE to get a smoother curve and more profound valleys to make it easy to set the threshold boundary. The performance of resultant syllable units is calculated in terms of time duration, which is compared with the existing group delay and manual segmentation techniques.

The remaining paper is organized as follows: Section II describes the literature review. Section III describes the methods and procedures. Section IV explains the information about acoustic-phonetic features in Hindi. Section V describes the energy convex hull algorithm approach. Section VI gives experimentation based on the proposed algorithm. In Section VII, the result and time error analysis are discussed. Section VIII gives a subjective evaluation. Section IX describes the conclusion of the paper.

## II. LITERATURE REVIEW

The accurate segmentation of speech is an essential factor in creating a high quality of TTS. Zhao and O'Shaughnessy [1] implemented algorithms of the convex hull in speech segmentation. Similarly, Ling and colleagues [2] used speech

segmentation to cleft palate speech of the Mandarin language using a convex hull. They initially extracted syllables from the speech utterances and classified as "quasi-unvoiced" or "quasi-voiced" and estimated the segmentation accuracy, which came out to be high. K. Prasad et al. [3] and Hema A Murthy [4] have performed an algorithm based on short-term energy and group delay processing of the magnitude spectrum for determining segmented syllable boundaries for the Indian languages and TIMIT database. Panda and Nayak [5] carried out successful automated speech segmentation of Hindi, Bengali, and Odia languages using vowel offset point identification technique along with Zero Crossing Rate (ZCR) segmentation method with the manual segmentation approach. Similarly, Stan et al. [6] used an ALISA tool to segment sentence-level alignment of speech with imperfect transcripts. This method helped in the creation of a new speech corpora. This method found that utilizing the speech segmentation tools and transcribing speech data is reduced. Hamza Frihia and Halima Bahi [7] reported the Hidden Markov Model (HMM) and support vector machine (SVM) model to generate the phoneme-based speech segmentation for the Arabic language for application of speech recognition. Sandrine Brognaux and Thomas Drugman [8] presented the HMM algorithm speech segmentation on the phone level for English, French, or under-score Language. Jon Ander Gomez and Marcos Calvo [9] shown the segmentation technique with a combination of HMM and DTW (Dynamic Time Wrapping) to achieved phone boundaries on the Albayzin and TIMIT database. Asaf Rendel et al. [10] shown that the HMM-GMM modeling technique is applied to the TIMIT corpus to get phoneme speech segmentation, and SVM is used to refine the obtained phone boundaries. The accuracy of the above modeling technique is 96%. Fréjus A. A. Laleye [11] published the algorithm based on STE & Zero crossing rate (ZCR) and perform the machining phase using the set of Fuzzy rules to get the syllable and phone boundaries on Fongbe language spoken in Benin, Tago, and Nigeria. Balyan et al. [12] built a medium-sized database for passenger rail information systems for the Hindi language in the phoneme level using HMM. The database consists of 630 utterances with 12674 words to facilitate the researcher in TTS and automatic speech recognition (ASR). Arum Bobby et al. [21] presented the speech segmentation for Indian language consider as a phone level by using deep neural network (DNN) and convolutional neural network (CNN) framework. Md. Mijanur Rahman and Md Al-Amin Bhuiyan have created the database on time and frequency domain approach on word level and achieve a segmentation accuracy rate of 96.25 for Bangla Language [22]. Yahia Hasan Jazyah [23] has reported the segmentation of audio data such as human speech in both English and Arabic languages by using Dynamic Windows and Thresholds. The algorithm achieved a segmentation accuracy rate up to 91.6% in average for English and 89.0% for the Arabic language.

### III. METHODS AND PROCEDURES

The following steps are carried out to design a Speech corpus.

- Selection of text sentences from news domains

- Recording of the selected text
- Syllabification of the speech signal

#### A. Selection of Sentences

The selection of the 150 sentences has been manually selected from various sources relevant to Metro travel information announcements in Delhi Rail for building the speech synthesis system. Adequate care has been taken to include all types of the required information so that the recording has enough occurrence of each type of Hindi sound [14].

#### B. Recording of Speech Corpus

The steps followed for recording the speech wav files were as follows:

- Professional male speaker voice has been recorded to maintain constant pitch and prevent stress phenomenon in noise and echo-free studio.
- The speaker has clear pronunciation and no articulatory defect.
- The sampling frequency was set to 16 kHz store in 16-bit PCM with Mono mode type.
- The speaker is required to read each text sentence, and the recorded sample was saved as wav files.

### IV. ACOUSTIC- PHONETIC FEATURES IN HINDI

The acoustic-phonetic of Hindi differs from the European languages. Hindi is mostly phonetic in nature, i.e., there is one to one correspondence between written symbols and the spoken sentences. Hindi phonemes can be divided into vowels and Consonants. The Hindi alphabet consists of 10 pure vowels (/ə/, /a/, /i/, /ɪ/, /u/, /ʊ/, /æ/, /e/, /o/, /ɔ:/) including two diphthongs namely; /æ/ and /ɔ:/. All these vowels have their nasalized form also. Creaky and whispered vowels are rarely used [15]. The Hindi consonants consist of 4 semivowels, 4 fricatives, and 25 stop consonants (including 5 nasals). The stop consonants are ordered systematically in the Hindi language, and this order may suggest ideas for developing a recognition/synthesis system [17, 18]. Classification of Hindi consonants and vowels are presented in Table I.

TABLE I. DESCRIPTION OF HINDI PHONEME

Shorts Vowels				Long vowels						
अ	इ	उ	ए	ओ	आ	ई	ऊ	ऐ	औ	
Unvoiced					Voiced					
Unaspirated				Aspirated		Unaspirated		Aspirated		Nasals
क		ख		ग		घ		ण		
च		छ		ज		झ		ञ		
ट		ठ		ड		ढ		ण		
त		थ		द		ध		न		
प		फ		ब		भ		म		
Semi Vowel					Fricatives					
य	र	ल	व	श	ष	स	ह			

### V. SYLLABLE BASE SEGMENTATION ALGORITHM

The syllables are identified from the speech database. The fundamental of the database is multiple forms of the unit phoneme, syllable, and words. In the Hindi language, the syllable types are CV, CVC, VC, V, CCV, and CCVC [14, 16]. The database distribution of syllables is mentioned below in Table II.

The | syllable likes boundary identification is performed by using an energy convex hull approach. The steps are as follows:

- Let's  $x(t)$  is the represented continuous speech signal, and  $x[n]$  be digitized speech signal.
- Determine the Short-term energy (STE) by applying the overlapped Hamming window (N= 400).

$$Q(n) = \sum_{m=-\infty}^{\infty} [X(m)]^2 w(n - m)$$

$$E(n) = 10 * \log Q(n)$$

$$W(n) = .54 - .46 \cos\left(\frac{2\pi n}{N-1}\right); 0 \leq n < N$$

- Apply the Savitzky-Golay smoothing filter and Median filter to reduce the noise and preserve boundaries.

- Estimate the initial syllable decision threshold for initial syllable detection.
- Apply the Hamming window for refining the boundaries of syllable- like units.

$$D(n) = 10 * \log[Q(n) + 1]$$

$$p(n) = \sum_{m=-\infty}^{\infty} D(m)w(n - m)$$

- Reset the threshold on  $p(n)$  to obtain correct syllable boundary

The block diagram in Fig. 1 shows the steps involved to obtain of syllable – like segmented speech.

TABLE II. DISTRIBUTION OF VARIOUS SYLLABLE IN HINDI

Syllables	Relative Frequency (%age)
CV	69.69
CVC	22.00
VC	2.78
V	3.60
CVCC	1.18
CCVC	0.89
CCV	0.48

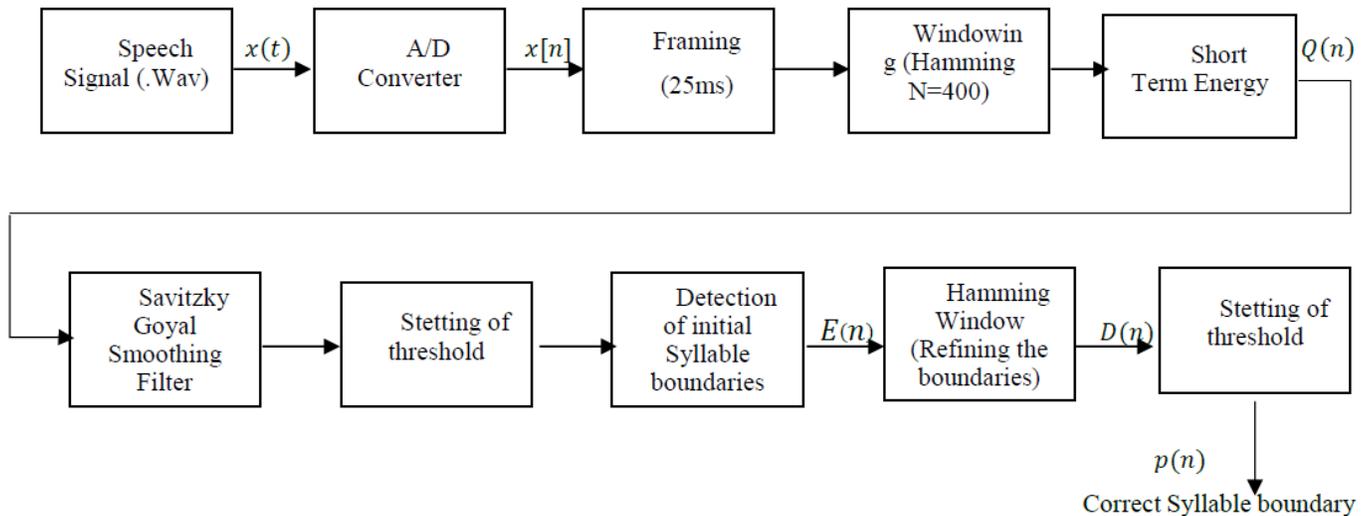


Fig. 1. Block Diagram Showing Steps Taken in Finding Syllable Boundaries.

## VI. EXPERIMENTATION

The experiment is done on the word and sentence level of medium size database consisting of 150 sentences of the duration of approx. 45 mins spoken by a single male speaker and obtained 1175 syllables units.

The 50 sentences of a syllable are processed manually by using PRAAT [19] speech analysis to check the performance of the proposed techniques. Fig. 2 shows the manually segmented output of the input wav file "Yahhan line do ke liye badle". This input wav file consists of 9 syllable units.

### A. Initial Boundary Detection

On STE  $Q(n)$ , the Savitzky–Golay [12] filter is applied for signal smoothing, and the SNR ratio is improved. Further, the median filter is used to preserve the boundaries and the smoothing energy curve.

To detect the initial boundary, a threshold is required to be estimated in the short term energy curve. To get the threshold in training set in the average STE of utterance was calculated.

However, the threshold can't be set to this value. For example, in Fig. 3, the utterance contained five possible syllable boundaries points A to E when the energy threshold was set to the average STE curve of a speech signal. The threshold value is -17 dB. If the threshold were kept higher than -17 dB, more valley points might be obtained, which are incorrect. If the threshold were kept lower, then the valley points E and C would be removed. The threshold value was reset from -17 dB to -32 dB to obtain the correct boundaries based on the above observation. After experimentation with a Hindi training set, it was seen that the threshold value between -28 dB to -38 dB gives more accurate segmentation boundaries.

### B. Convex Hull Boundary Detection Analysis

In this approach, a sliding Hamming window is applied on  $Q(n)$  shown in Fig. 4 to obtain  $P(n)$ . It is seen, the STE curve is smoother deeper valley is obtained, which makes it easy to set convex-hull threshold value.

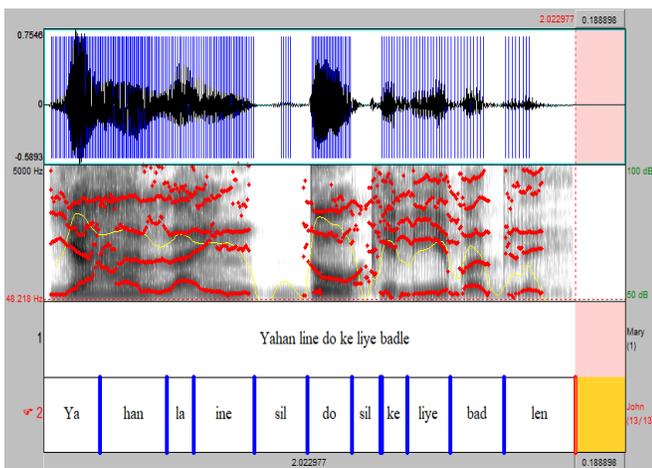


Fig. 2. Manual Segmentation of Continuous Speech at the Syllable Level.

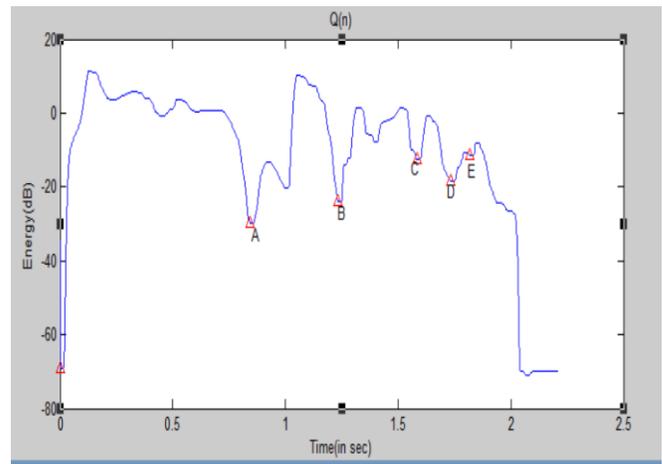


Fig. 3. STE Curve Syllable Points in an utterance.

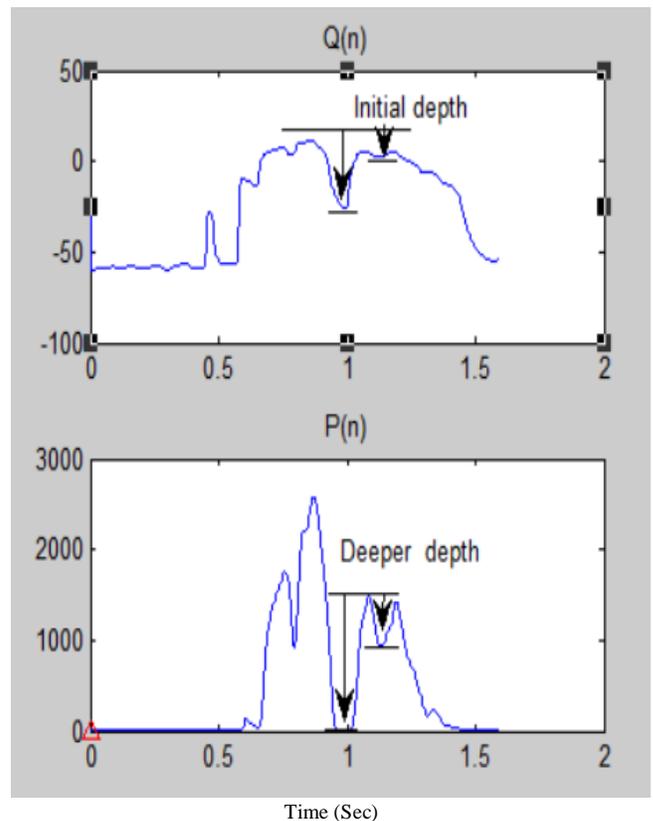


Fig. 4. Comparison of the Valley of the Energy Curve and Convex Hull Curve.

Fig. 5 shows the output of the segmentation algorithm for the input speech utterance "यहाँ लाइन दो के लिए बदले" ("yahan line do ke liye badlen"). It is seen that the input speech signal is segmented into three initial syllable units "यहाँ लाइन", "दो" and "के लिए बदले" ("yahan line", "do" and "ke liye badlen"). On the application of the convex hull approach, the speech is re-segmented into nine syllables units. "य", "हाँ", "ला", "इन", "दो", "के", "लिए", "बद" and "ले" ("ya", "han", "la", "ine", "do", "ke", "liye", "bad" and "len"). The same process has been applied for 50 utterances and obtained 402 syllables like units.

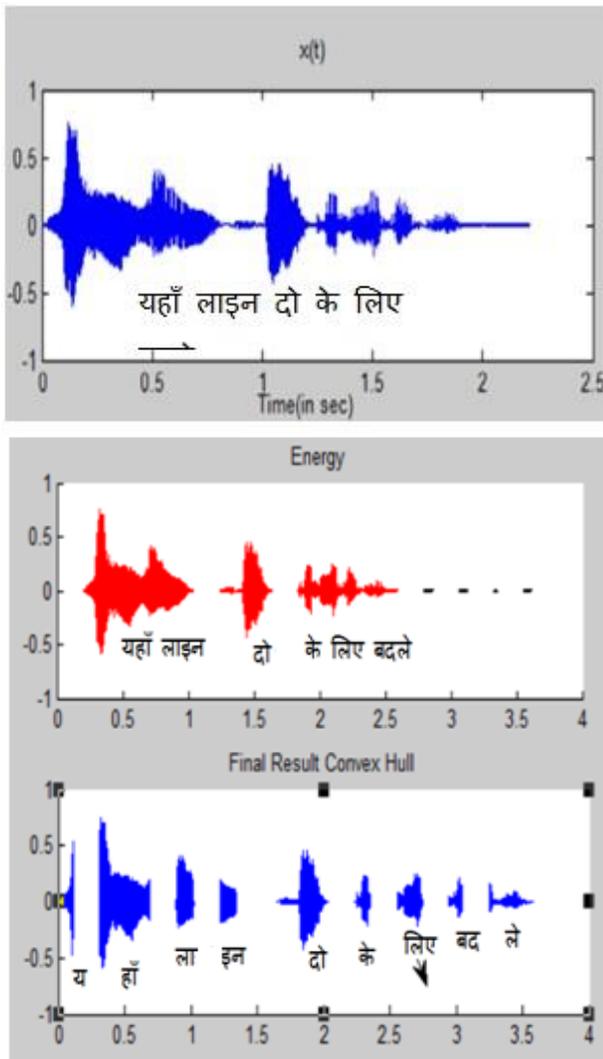


Fig. 5. The Waveform of Input Speech  $x(t)$  and Segmented Output Syllable units of STE and Convex Hull.

Below examples are shown in Table III to obtain as syllable boundaries units for a few input wav files.

TABLE III. EXAMPLES ILLUSTRATING SYLLABLES SEGMENTS

Input Wav file	Obtained Syllable output
सफदरजंग	सफ् दर जंग्
सेवा में नहीं	से वा मे न ही
कृपया दरवाजो से दूर हट कर खड़े हो	कृप् या दर वा जो से दूर हट् कर ख डे हो
लाजपत नगर	लाज् पत् न गर्

## VII. RESULT

The performance of the segmentation algorithm is analyzed on a set of 50 test samples. Time error analysis is calculated to test the accuracy of the segmented syllable-likes unit for each syllable. The research also includes silence occurrence in the sentence as discuss:

$$\text{Time\_Error} = |\text{duration of manual segment boundary} - \text{duration of the automatic segment boundary}|$$

Table IV shows the result of the segmented output and the calculated error rate of the proposed algorithm & existing group delay technique [20]. The error rate obtained in the energy convex hull algorithm performs better as it has a lower value.

Experiments performed in Fig. 6 demonstrate in the graph that the energy convex hull segmentation technique achieves better results that are closer to the outcome achieved by manual segmentation. But, the group delay based method shows a high degree of variation in syllable durations compared to the energy convex hull approach.

The same process has applied a set of words and sentences to find overall performance segmented syllable like units of continuous speech by using proposed and group delay segmentation techniques.

The performance results are shown in Table V and found that the group delay-based algorithm approach shows an accuracy rate of 63.05%. The proposed algorithm energy convex hull approach achieves an accuracy rate of 76.12% of segmented speech in less than 30 ms.

In the proposed algorithm, the final segmentation result is obtained after applying the double sliding widow along with the reset of the threshold value. After analysis, it is observed that if the threshold is set between 2200-2800 for Hindi speech, it gives an accurate syllable boundary. During the experiment, it was found that the duration of time error was higher for fricative and nasal sound, but it provided better results as compared to group-delay segmentation. The threshold value for fricative sound {e.g., shakur basti (शाकुर बस्ती), safdarjung (सफदरजंग), udghoshnaa (उदघोषना), Station (स्टेशन), Shalimar (शालीमार), etc.} is set at approx. 2600 to 2700 as these sounds are high energy signals. For nasal sound (e.g., mangolpuri (मंगोलपुरी), nagar (नगर), anand (आनंद), nirmal (निर्मल), etc.) the threshold is set at approx. 2300 to 2400.

TABLE IV. DURATION OF SEGMENTED OUTPUT BY USING MANUAL SEGMENTATION (PRAAT TOOL), GROUP DELAY ALGORITHM AND ENERGY CONVEX HULL ALGORITHM

Obtained syllable units	Duration of manual segmentat ion (sec)	Duration of group delay algorithm (sec)	Error rate (msec)	Duration of energy convex hull algorithm (sec)	Error rate (msec)
य (ya)	0.19	0.12	44.00	0.15	21.05
हाँ (han)	0.18	0.20	11.54	0.18	2.25
ला (la)	0.21	0.243	16.48	0.22	4.35
इन (ine)	0.33	0.35	5.06	0.34	1.20
दो (do)	0.34	0.30	10.19	0.32	3.57
के (ke)	0.11	0.12	6.25	0.16	45.45
लिए (liye)	0.18	0.20	10.53	0.19	5.56
बद (bad)	0.21	0.15	39.52	0.17	21.23
ले (len)	0.16	0.15	4.35	0.16	1.26

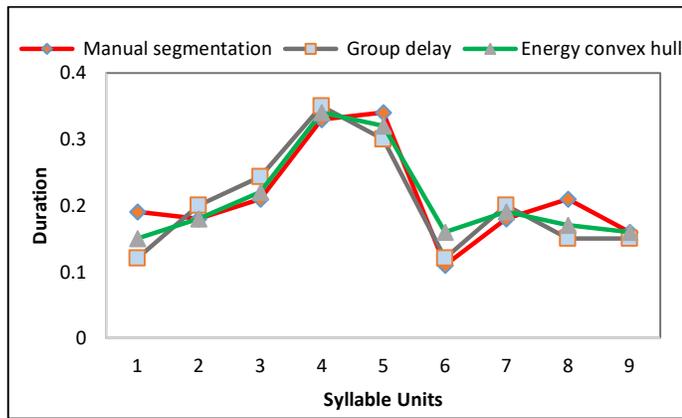


Fig. 6. Duration of Syllable units Obtained by manual Analysis and Segmentation Algorithm.

TABLE V. TIME ERROR ANALYSIS OF OVERALL SEGMENTATION CONTINUOUS SPEECH

Algorithm	Time-error (msec)	≤ 30	31-40	41-50	> 50	Total no. of segments
Proposed algorithm	Number of segments	306	18	16	62	402
	Performance (in %age)	76.12	4.47	3.98	15.42	
Group Delay	Number of segments	275	51	28	82	
	Performance (in %age)	63.07	11.69	6.40	18.08	

VIII. SUBJECTIVE EVALUATION

Accuracy is an essential factor in measuring the performance of segmented speech. In this work, five subjects were considering for perception evaluation of segmented speech. Subjects were asked to access the accuracy on a 5 points scale (1-Unsatisfactory, 2-Poor, 3-Fair, 4-Good, and 5-Excellent) for each of the segmented sentences. The test is carried out for the segmented sentences generated by group delay and energy convex hull approach. The mean opinion score (MOS) is calculated for the accuracy of segmented speech. Table VI shows that the segmented accuracy rate is improved in the convex hull approach.

TABLE VI. MEAN OPINION SCORE FOR THE QUALITY OF SEGMENTED CONTINUOUS SPEECH

Algorithm	No of Test samples	Accuracy rate
Energy Convex hull	50	4.18
Group Delay	50	4.02

IX. CONCLUSION

In this paper, the energy convex hull algorithm is proposed for segmenting the speech signal into syllable-like units for improving the segmentation performance. The algorithm is applied to speech corpus, and segmented syllabic units are obtained. The algorithm calculated the time duration of each syllable unit and obtained a time error rate about manual segmentation syllable units to validate the accuracy of the

proposed algorithm. After a comprehensive analysis, it is found that the segmented boundary errors are ≤ 30 ms for 76.07% of the total syllables. The performance of the algorithm gives an accurate result as compared to the existing group delay segmentation technique. Hence the proposed algorithm is highly useful to create syllable like speech units as it takes a few milliseconds to obtain syllabic units over manually labelling process of speech segmentation, which is a very time-consuming and strenuous task.

This algorithm may also be extended over large databases for building the high quality of TTS by the researcher for the limited and unlimited domain. Further, the research may be extended to reduce errors by applying various optimization techniques - machine learning (DNN, CNN, or hybrid models) and fuzzy-based algorithms.

REFERENCES

- [1] X. Zhao, and D. O’Shaughnessy, “A New hybrid approach for automatic speech signal segmentation using silence signal detection, energy convex hull, and spectral variation,” IEEE International Conference, pp. 000145-000148, 2008.
- [2] J. Li, and F. Shen, “Automatic segmentation of Chinese Mandarin speech into syllable-like,” Asian Language Processing (IALP) 2015 International Conference, pp. 57-60, 2015.
- [3] K. Prasad, T. Nagarajan, and H. A. Murthy, “Automatic segmentation of continuous speech using minimum phase group delay function,” Speech Communication Vol. 42, pp. 1883-6, 2004.
- [4] H. A. Murthy, and B. Yegnanarayana, “Group delay functions and its applications in speech technology,” Indian Academy of Sciences, pp. 745-782, 2011.
- [5] S. P. Panda, and A. K. Nayak, “Automatic speech segmentation in syllable centric speech recognition,” International Journal of speech technology, pp. 9-18, 2015.
- [6] A. Stan, Y. Mamiya, J. Yamagishi, P. Bell, O. Watts, R. A. J. Clark, and S. King, “ALISA: An automatic lightly supervised speech segmentation and alignment tool,” Computer Speech & Language, 35, pp. 116 – 133, 2016.
- [7] H. Frihia, and H. Bahi, “HMM/SVM segmentation and labelling of Arabic speech for speech recognition application,” International Journal of Speech Technology 20(3), pp-563-573, 2017.
- [8] S. Brognaux, and T. Drugman, “HMM-based speech segmentation: Improvements of fully automatic approaches,” IEEE/ ACM Transactions on Audio, Speech, and Language Processing, 24(1), pp. 5-15, 2016.
- [9] J. A. G’omez, and M. Calvo, “Improvements on Automatic Speech Segmentation at the Phonetic Level,” Springer-Verlag Berlin Heidelberg, pp. 556-564, 2011.
- [10] A. Rendel, A. Sorin, R. Hoory and, A. Breen, “Towards Automatic Phonetic Segmentation for TTS,” International Conference on Acoustics, Speech and Signal Processing, pp. 4533-4536, 2012.
- [11] F. A. A. Laleye, E. C. Ezin, and C. Motamed, “Fuzzy-based algorithm for Fonjbe continuous speech segmentation,” Pattern Analysis and Application 20, pp. 855-864, 2017.
- [12] A. Balyan, S. S. Agrawal, and A. Dev, “Automatic phonetic segmentation of Hindi speech using hidden Markov model,” AI & Soc Springer-Verlag London Limited, pp. 543-549, 2012.
- [13] A. Savitky and M. J. E. Goyal, “Soothing and differentiation of data by simplified least square procedure,” Anal Chem., vol. 36, no.9, pp. 1627-1639, 1964.
- [14] A. Balyan, S. S. Agrawal, A. Dev, “Building Syllable dominated Speech corpora for Metro Rail Information system,” International Conference of O-COCOSDA-2008, pp. 135-140, Kyoto, Japan, Nov 25-27, 2008.
- [15] K. Arora, Sunita, K. Verma, and S. S. Agrawal, “Automatic extraction of phonetically rich sentences from large Text corpus of Indian Languages,” In INTERSPEECH, pp. 2885-2888, 2004.

- [16] S. S. Agrawal, "Emotions in Hindi speech-analysis, perception and recognition," International Conference on Speech Database and Assessments (Oriental COCODA), pp.7-13, 2011.
- [17] B. Aartil, and S. K. Kopparapu, "Spoken Indian language identification: a review of features and databases," Indian Academy of Sciences, pp. 1-14, 2018.
- [18] S. Bhatt, A. Dev, and A. Jain, "Confusion analysis in phoneme based speech recognition in Hindi," Journal of Ambient Intelligence and Humanized Computing, DOI:10.1007/s12652-020-01703-x, 2020.
- [19] P. Boersma, and D. Weenik, Praat: A System for Doing Phonetics by Computer. <http://www.praat.org/> , 2001.
- [20] A. Balyan, A. Dev, R. Kumari, and S. S. Agrawal, "Labelling of Hindi Speech," IETE Journal of Research, vol 62, issue 2, pp. 146-153, 2015.
- [21] A. Baby, J. J. Prakash, and H. A. Murthy, " A Hybrid approach to Neural Networks Based Speech Segmentation," International Symposium Frontiers of Research on Speech and Music, 15-16, National Institute of Technology (NIT) Rourkela, 2017.
- [22] Md. M. Rahman, and Md. A. Bhuiyan, "Continuous Bangla Speech Segmentation using Short-term Speech Features Extraction Approaches," International Journal of Advanced Computer Science and Applications, Vol. 3, No. 11, pp. 131-138, 2012.
- [23] Y. H. Jazyah, "Speech segmentation using dynamic window and thresholds or Arabic and English Language," Journal of Computer Science, Volume 14, Issue 4, pp. 485-490, 2018.

# Identification and Assessment of the Specific Absorption Rate (SAR) Generated by the most used Telephone in Peru in 2017

Natalia Indira Vargas-Cuentas<sup>1</sup>

Image Processing Research Laboratory (INTI-Lab)  
Universidad de Ciencias y Humanidades (UCH)  
Lima, Peru

Avid Roman-Gonzalez<sup>3</sup>

Image Processing Research Laboratory (INTI-Lab)  
Universidad de Ciencias y Humanidades (UCH)  
Lima, Peru

Mark Clemente-Arenas<sup>2</sup>

CSE-HF  
Universidad Nacional Tecnológica de Lima Sur (UNTELS)  
Lima, Peru

Roxana Moran-Morales<sup>4</sup>

Instituto Nacional de Investigación y Capacitación de  
Telecomunicaciones INICTEL-UNI  
Universidad Nacional de Ingeniería (UNI), Lima, Peru

**Abstract**—According to the World Health Organization (WHO) it is estimated that between 5 and 10% of the population is electro sensitive, the excessive or prolonged exposure to electromagnetic waves can damage health. Currently the electromagnetic radiation generated by wireless mobile telephony involves our daily lives, since it is reported that there are more than 5 billion cell phone users. Each country establishes its own relative national standards on exposure to electromagnetic fields, which Peru lacks. Nevertheless, they are based in standards that has not been revised since 1996. This contribution seeks to identify the Specific Absorption Rate (SAR) generated by the most used mobile phone in Peru in 2017 using the ComoSAR measuring system in the GSM (Global System for Mobile communications) band at a frequency of 900 MHz. The results obtained will evaluate the behavior of electromagnetic waves by affecting the emulated tissues of body density and dielectric constant. The maximum SAR values recorded in the measurement were 0.05 W/Kg for a 1 g cube and 0.02 W/Kg for a 10 g cube. On the other hand, the average values obtained were 0.046 W/Kg for a 1 g cube and 0.019 W/Kg for a 10 g cube. The SAR values measured in the conditions of the experiment are below that what is indicated by the US standard and the European standard of SAR values.

**Keywords**—*Electromagnetic radiation; SAR; ComoSAR; GSM*

## I. INTRODUCTION

Excessive or prolonged exposure to electromagnetic waves generated by radio frequency (RF) can damage human health [1], currently globally according to the Association of Internet and Cellular Telecommunications there were more than 400 million cell phone subscribers in the United States in 2017. Globally, there are more than 5 billion cell phone users.

Although it is not appreciated, excessive or prolonged exposure to electromagnetic waves generated by radio frequency (RF) can damage health and its effects are cumulative [2].

The effects induced by electromagnetic radiation can be thermal and non-thermal, which are the ones that generate the greatest risk when producing biological changes [3]. According to the World Health Organization (WHO) [4] it is estimated that between 5 and 10% of the population is electro sensitive and among the most frequent symptoms appear headaches, insomnia, irritability, depression or increased risk of cancer, among others.

In order to set the exposure limits, scientific research must identify the threshold at which the first effects on health are manifested, but no tests can be performed on humans, therefore other alternatives for experimentation were found by the Institute of Electrical and Electronics Engineers (IEEE) [5] and the International Electrotechnical Commission (IEC) [6, 7]. It is important to mention that there is no single level above which exposure becomes dangerous to health; on the contrary, the potential risk to health increases gradually as the level of exposure of people increases [8].

The electromagnetic radiation generated by wireless mobile telephony involves our daily lives, much more today, because in the last decade it has increased exponentially in urban centers and in each of the homes.

Each country establishes its own relative national standards on exposure to electromagnetic fields. However, most of these national standards are based on the recommendations of the International Commission for Non-Ionizing Radiation Protection (ICNIRP). This non-governmental organization, formally recognized by WHO, evaluates the results of scientific studies conducted worldwide.

Table I summarizes the recommended exposure limits corresponding to the types of technologies that have caused concern in society. The last update of these guidelines was made in April 1998.

TABLE I. RECOMMENDED EXPOSURE LIMITS

Frequency	Frequency of the European electricity grid	Frequency of mobile phone base stations		Frequency of microwave ovens	
	50 Hz	50 Hz	900 MHz	1,8 GHz	2,45 GHz
	Electric field (V/m)	Magnetic field ( $\mu$ T)	Power density (W/m <sup>2</sup> )	Power density (W/m <sup>2</sup> )	Power density (W/m <sup>2</sup> )
Exposure limits for the population	5000	100	4.5	9	10
Occupational Exposure Limits	10000	500	22.5	45	.

Given this, the Council of Europe establishes maximum exposure levels for wireless telephony in indoor areas of 0.6 volts per meter (0.1 microwatts per square centimeter or  $\mu$ W / cm<sup>2</sup>).

The Federal Communications Commission (FCC) is the entity that decides how much radiation it is allowed to emit to cell phones in the United States. Currently, the limit defined by the FCC is 1.6 W / Kg per 1 g of tissue. However, this standard has not been revised since 1996 [9].

Specific Absorption Rate (SAR) values indicate the highest level of exposure to electromagnetic radiation measured. In Europe, the SAR limit for mobile devices is set at 2 W/Kg per 10 g of tissue. This standard conforms to the ICNIRP 1998 guidelines, as well as the IEC standards and is determined by CENELEC [10].

In Peru, the established SAR value is the recommended by ICNIRP of 2 W/Kg averaged over a tissue mass of 10 g (0.02 W absorbed in any mass of 10 g of tissue in the head) [11].

This project seeks to identify the Specific Absorption Rate generated by the most used mobile phone in Peru using SAR measuring equipment, for the study of the distribution and levels of the electromagnetic field perceived by the human body when exposed to the electromagnetic waves generated by radio frequency.

The main importance of this research is that the results obtained will evaluate the behavior of electromagnetic waves by affecting the emulated tissues of body density and determine the impact that these mobile devices have on humans.

Based on the results obtained, consumers can be provided with information on the possible risks of prolonged use of mobile equipment and formulate safety advice to protect sensitive subjects such as children and the elderly.

Finally, the current radiation standards for cell phones can be reviewed and it will be important to try to propose and set standards to limit cell phone radiation in Peru.

## II. IDENTIFICATION OF THE MOST USED PHONE IN PERU IN 2017

According to the Lima Chamber of Commerce (CCL) in 2016 and 2017, the third most imported product to the country was mobile phones [12], as can be seen in Table II.

It can be seen the total imports by companies in Table III, we can see in the top positions companies such as Samsung, LG, Huawei, among others [13].

In 2016, for example, 12.3 million smartphones were imported, of which five telephone brands were the most consumed in the Peruvian market [14].

According to the CCL, there is leadership in total imports of the following brands: Samsung (22.2%), Huawei (21.3%), Apple (13.6%), LG (10.6%) and Motorola (7.8%) [15].

According to Device Atlas in 2017, the three most used phones were Samsung and Motorola [9]. In the first place is the Samsung Galaxy J7, second to the Samsung Galaxy J5 and third is the Motorola Moto G [16].

As can be seen in Table IV, in five South American countries, the most used devices are Samsung and Motorola, among which the most preferred models are the Samsung Galaxy J5, Galaxy J7, Galaxy J1 Ace and the Motorola Moto G.

According to Table III, the most used cell phone in Peru in 2017 was the Samsung Galaxy J7, in this sense in this study that equipment will be used to be able to measure and identify the specific percentage of absorption generated by this mobile device.

TABLE II. MAIN IMPORTED PRODUCTS IN PERU - JANUARY TO DECEMBER 2016 / 2017

RK 17	RK 16	Item	Main Imported Products	2016	2017
1	1	2709000000	Crude of oil or bituminous ore	1,633	2,471
2	2	2710192111	Diesel 2, with a sulfur content less than or equal to 50 ppm	1,293	1,540
3	3	8517120000	Mobile phones (cell phones) and other wireless networks	1,127	1,097
4	4	8703239020	Other vehicles assembled with a spark ignition reciprocating piston or reciprocating piston engine, for the carriage of persons, with a cylinder capacity greater than 1,500 cc but less than or equal to 3,000 cc.	793	743
5	7	1005901100	Yellow hard corn, except for sowing	581	626

TABLE III. TOTAL IMPORTS BY COMPANIES - JANUARY TO DECEMBER 2016 / 2017

RK 17	RK 16	Importing companies	2016	2017	VAR.% 17/16	Part.% 2017
1	2	Refinería La Pampilla S.A.A.	1,860	2,705	45.4%	6.8%
2	1	Petróleos Del Peru Petroperu S.A.	2,305	2,607	13.1%	6.5%
3	4	Toyota Del Peru S.A.	608	592	-2.7%	1.5%
4	7	Motorindustria S.A.	469	527	12.2%	1.3%
5	6	Samsung Electronics Peru S.A.C.	474	468	-1.3%	1.2%
6	8	Alicorp S.A.A.	371	454	22.5%	1.1%
7	12	Pure Biofuels Del Peru S.A.C.	305	434	42.1%	1.1%
8	10	Lg Electronics Peru S.A.	356	377	5.7%	0.9%
9	17	Corporación Aceros Arequipa S.A.	272	333	22.8%	0.8%
10	19	Huawei Del Peru Sac	238	321	35.1%	0.8%
11	15	Diveimport S.A.	284	320	12.8%	0.8%
12	9	America Movil Peru S.A.C	362	319	-11.7%	0.8%
13	14	Automotores Gildemeister-Peru S.A.	299	311	4.1%	0.8%
14	13	Southern Peru Copper	299	279	-6.7%	0.7%
15	16	Adm Andina Peru S.R.L.	283	269	-4.9%	0.7%
16	38	Molinos & Cia S.A.	121	260	115.9%	0.7%
17	21	Contilatin Del Peru S.A.	232	256	10.1%	0.6%
18	18	Gloria S.A.	271	232	-14.3%	0.6%
19	30	Cargill Americas Peru S.R.L.	184	232	25.7%	0.6%
20	27	Volvo Peru S.A.	198	229	16.0%	0.6%
Other importing companies			27,658	28,384	2.6%	71.1%
TOTAL			37,449	39,909	6.6%	100.0%

TABLE IV. MOST USED CELL PHONES IN SOUTH AMERICA 2017

Country	#1	#2	#3
Argentina	Motorola Moto G	Samsung Galaxy J7	Samsung Galaxy J1 Ace
Brazil	Motorola Moto G	Samsung Galaxy J5	Motorola Moto G 2nd Gen
Chile	Samsung Galaxy J5	Motorola Moto G	Samsung Galaxy J7
Ecuador	Samsung Galaxy J5	Samsung Galaxy J7	Samsung Galaxy J1 Ace
Peru	Samsung Galaxy J7	Samsung Galaxy J5	Motorola Moto G

### III. SPECIFIC ABSORPTION RATE (SAR) MEASUREMENT

The specifications of the mobile equipment to be measured must be found. Perform 3D scanning of the distribution of electromagnetic fields in the human body in the 900 MHz frequency band using the ComoSAR Specific Absorption Rate measuring equipment. Use the 900 MHz Head liquid in the device phantom, perform the measurement with the mobile phone connected to a call, position the phone on the right side and on the left side of the phantom in the cheek and tilt positions.

Obtain the results of the 2D scan area by finding the position of the hot spots and find the results of the 3D zoom scan by finding the results of the 10 g cube and the 1 g cube.

#### A. Characteristics of the Mobile Device

The Samsung Galaxy J7 in Fig. 1, is a mid-range Smartphone of the Galaxy J line, among its many features [17] we highlight an eight-core 1.5 GHz processor, Android 6.0

operating system, the battery capacity is 3000 mAh and it is of the Li-ion type (lithium ions).

According to the technical data sheet, the Samsung Galaxy J7 has a Qualcomm Snapdragon 615 SoC (System on a Chip). This SoC has an RF system of 3rd generation power-efficient LTE, with 7-mode bands, and lower power and PCB reduction. [18].

The operating system of the Samsung Galaxy J7 is the Android 6.0.1 Marshmallow, the device has proximity, light and accelerometer sensors.

It also has an internal storage of 16GB, external microSD storage of up to 128GB and 2GB of RAM [17].

As can be seen in Fig. 2, the dimensions of this mobile device are 152.2 x 79.1 x 7.9 mm, its weight is 168 g and its volume is 95.11 cm<sup>3</sup>. On the other hand it has a resolution of 1920 x 1080 and a density of 401 ppi [18].



Fig. 1. Samsung Galaxy J7.

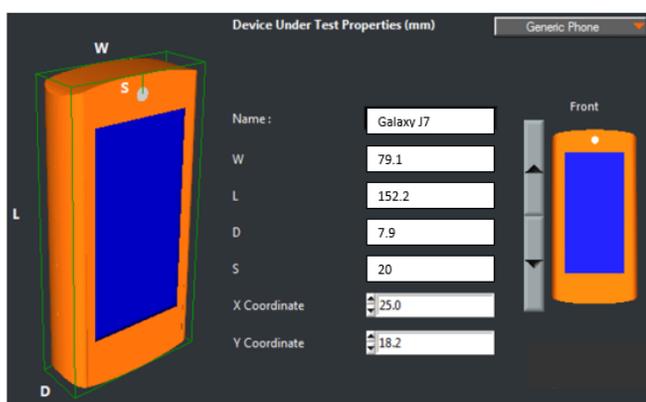


Fig. 2. Samsung Galaxy J7 Dimensions.

The mobile network technologies that the phone supports are the following: GSM (Global System for Mobile communications), TD-SCDMA (Time Division Synchronous Code Division Multiple Access), UMTS (Universal Mobile Telecommunications System) y LTE (Long Term Evolution).

According to [17] the Specific Absorption Rate (SAR) of this mobile device when the device is held in the head next to the ear in conversation position is 0.18 W / Kg per 10 g of tissue, on the other hand the SAR value of this mobile equipment when the device is kept close to the body at the level of the hip is 0.265 W / Kg for 10 g of tissue. On the other hand, as specified by [18] the SAR values for every 10 g of tissue is 0.503 W / Kg for the head and 0.373 W / Kg for the body, in addition the SAR values for every 1 g of tissue is 1.14 W / Kg for the head and 1.32 W / Kg for the body.

The Federal Office for Radiation Protection (BfS) that is a German entity that has conducted measurements of the radiation values of the mobile phones following standardized procedures and has published the information.

Measurements made by BfS [19] of the Galaxy J7 mobile phone, indicate that the SAR values per 10 g of tissue is 0.57 W / Kg for the head and 1.33 W / Kg for the body at a measuring distance of 0.5 cm.

It is important to mention that devices not exceeding 0.6 W/Kg during operation fulfil an important criterion for obtain the "Blue Angel" eco label. As can be observed the Samsung Galaxy J7 only complies with one part of the requirement, since the SAR value in the body is very high.

### B. ComoSAR Equipment Description

To measure SAR from the Samsung Galaxy J7, the ComoSAR 30-6000 MHz system with double phantom inside of anechoic chamber is used. The equipment was obtained through the scientific equipment grant with code RNR N° 285-INNOVATEPERU-EC-2017 of INICTEL-UNI and UCH, funded by Ministry of Production, through the Innóvate-Perú Program.

Mainly the Microwave Vision Group (MVG) ComoSAR measurement system and certification system has three main components: The SAR bench, the instrumentation and the remote computer, as can be seen in the following block diagram in Fig. 3.

The instrumentation component has the necessary equipment to perform system calibration tests, noise evaluation, liquid properties measurement and finally the realization of SAR measurements of some device.

On the other hand, the remote computer component mainly has a computer that will have the OpenSAR software, which is the element that allows interaction between the Bench SAR, devices such as the probe, the multimeter, the communication tester and the end user.

The ComoSAR system has a semi anechoic chamber, which is a shielded room that provides an isolation above 100dB. In other words, it reduces and blocks any external electromagnetic field. This is necessary to perform standardized SAR measurements.

The SAR bench is located inside the semi anechoic chamber. It is composed of different subsystems, such as the KUKA arm robot, the E-field probe, two SAM (Specific Anthropomorphic Mannequin) phantom tables, liquids with human like dielectric properties, a video positioning system in order to guide the probe and DUT (Device Under Test) holders for positioning the mobile phone.

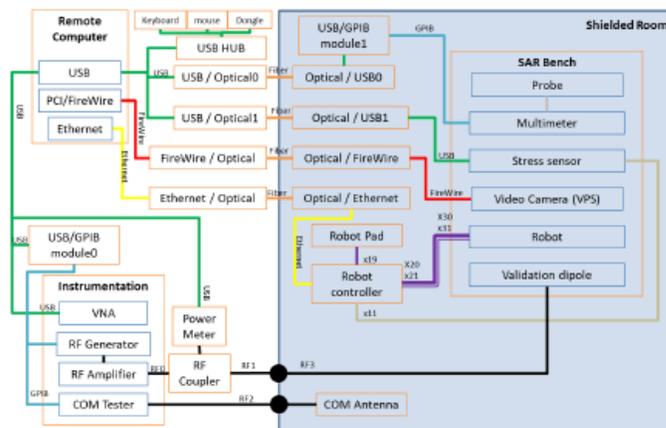


Fig. 3. Block Diagram of the ComoSAR Measurement System.

### C. Mobile Device Measurement with the ComoSAR

For measurement purposes, the first step is to select and to place the handset in a correct position. The positions displayed in Fig. 4, must be taken into account.

The first position is the called cheek position in which the headset is in contact with the ear and the front of the phone is fully in contact with the cheek of the phantom. On the other hand the second position is called tilt position, where the mobile device moves away from the mouth and forms an angle of 15° with the head of the phantom.

As can be seen in Fig. 5, it is necessary to use the DUT holder and match it with the vertical marks that the phantom has to make sure that the mobile device is in the correct position to perform the corresponding measurements.

Once we have the mobile device well located, we can proceed to perform SAR measurements, for which we will have the configuration observed in Fig. 6.

The ComoSAR system will allow to determine the distribution of SAR within the phantom that represents the human body, as can be seen in the figure to make the corresponding measurements it will be necessary to use the main computer of the system, the OpenSAR software, the six-axis robot, the data acquisition system, the E-field probe, the tissue simulator liquid, the mobile device (DUT) and the DUT holder.

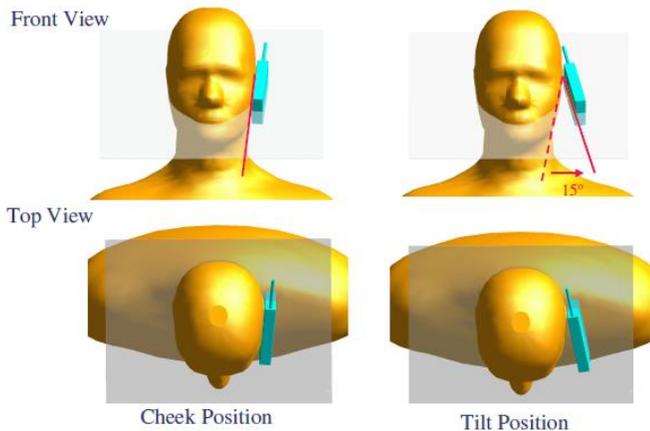


Fig. 4. Cheek and Tilt Positions [20].

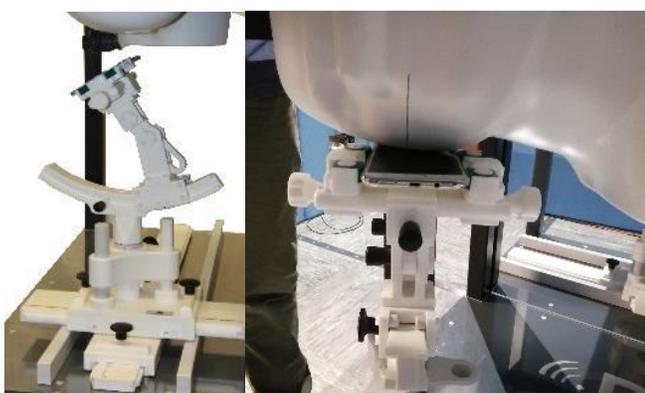


Fig. 5. Correct Position of the Device under Test (DUT) Holders.

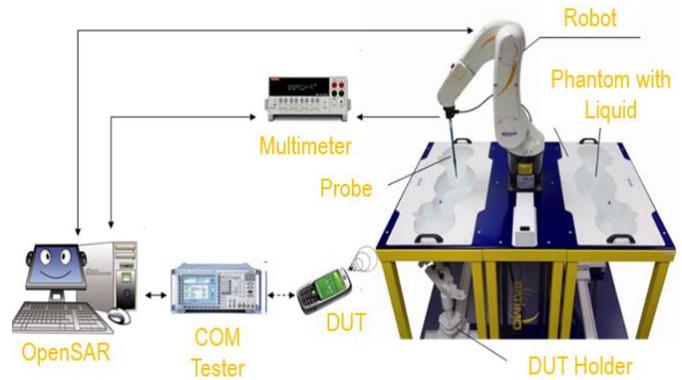


Fig. 6. DUT SAR Evaluation [21].

Specific Absorption Rate (SAR) measurements will be made on the Samsung Galaxy J7 device, in a 900 MHz GSM network, with a head simulating liquid that has a relative permittivity of 41.5 in addition to a conductivity of 0.97 (S / m). It is important that the phone battery is well charged during each measurement.

For the measurement procedure a call is established with the mobile device, then the desired measurement must be configured in the OpenSAR software, as can be seen in Fig. 7.

In the OpenSAR software, a connection with the robot, the multimeter and the communication tester must be established. The type of probe and phantom to be used must be selected, on the other hand the measurement parameters must be indicated, such as the type of device and its dimensions, the position of the device, the section of the phantom to be evaluated, the type of liquid, the frequency band of study, the standard and the channel.

The OpenSAR software computes the results to give a SAR value in a 1g or 10g mass, these results will be observed in the next section.

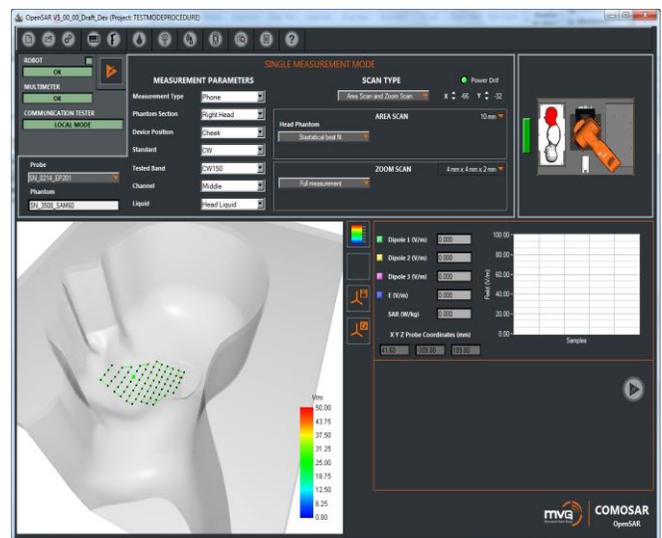


Fig. 7. Software OpenSAR Configuration.

#### IV. RESULTS

To obtain the SAR measurements, a scan area was first performed on the projection given by the dimensions of the mobile device. From this first scan the maximum points or hot spots were identified, then the SAR is evaluated in a cube above these maximum local, as can be observed in Fig. 8.

Table V shows the experimental conditions configured to perform the corresponding SAR measurements on the selected device.

For a GSM band at a frequency of 900 MHz, for a medium channel, in a section of the phantom that represented the right profile of the head, with a cheek type position the following SAR values have been obtained for the Samsung mobile device Galaxy J7, as can be observed in Table VI.

The maximum SAR values recorded in the measurement were 0.05 W / Kg for a 1 g cube and 0.02 W / Kg for a 10 g cube. On the other hand, the average values obtained were 0.046 W / Kg for a 1 g cube and 0.019 W / Kg for a 10 g cube.

Table VII shows the comparison of the results obtained from SAR with the values allowed by the European and American standards.

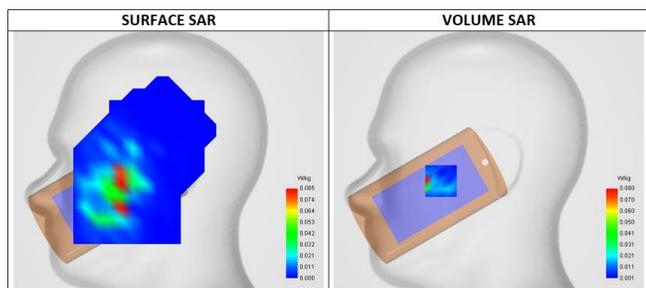


Fig. 8. Surface Scan and Volume Scan Results.

TABLE V. EXPERIMENTAL CONDITIONS

Phantom	Right head
Device position	Cheek
Band	GSM900
Channels	Middle
Signal	TDMA (Crest factor: 8.0)

TABLE VI. MEASURED SPECIFIC ABSORPTION RATE (SAR)

Specific Absorption Rate (SAR)	Measured value (W/Kg)
SAR 10 g	0.019266
SAR 1 g	0.046403

TABLE VII. COMPARISON OF THE MEASURED SAR VALUES WITH THE STANDARDS

	USA	Europe/Asia	Peru	Measured values (Galaxy J7)
SAR 10 g (W/Kg)	-	2	2	0.019266
SAR 1 g (W/Kg)	1.6	-	-	0.046403

As can be seen in the previous table, the SAR values measured in the conditions of the experiment are well below what is indicated by both the US standard and the European standard of SAR values for 10 g and 1 g of tissue, respectively.

#### V. CONCLUSIONS

Based on the results obtained, safety advice can be formulated for the use of cell phones, especially to protect sensitive subjects such as children and the elderly. On the other hand, consumers can be provided with information on the possible risks of prolonged use of mobile equipment.

Radiation standards for cell phones can also be examined in order to protect the health of consumers and to provide useful information to the consumer.

Finally, it will be important to try to propose and establish technical standards to cell phone importation in order to verify if wireless handsets are working below the limits on exposure to non-ionizing radiation in Peru.

It is important to continue with this type of research because as mobile communications are evolving really fast, and new frequency spectrum is being used, standardized testing needs to be updated accordingly. The future work of this research will aim to establish the experiment with a base station simulator, in order to establish a call with the maximum output power for several of the new frequency bands that are being used for emerging technologies. In addition to performing SAR measurements on other new mobile devices of interest. On the other hand, in order to be able to continue with future work in the same research track and to acquire complementary equipment and carry out additional studies, it is possible to apply for different research funds.

#### ACKNOWLEDGMENT

This research work was supported by INNOVATE Perú, ‘Contrato RNR N° 285-INNOVATEPERU-EC-2017’.

#### REFERENCES

- [1] Takei, R., Nagaoka, T., Saito, K., Watanabe, S., & Takahashi, M. (2017). SAR variation due to exposure from a smartphone held at various positions near the torso. *IEEE Transactions on Electromagnetic Compatibility*, 59(2), 747-753.
- [2] Guideline, ICNIRP. (1998). Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz). *Health phys*, 74(4), 494-522.
- [3] Hardell, L. (2017). World Health Organization, radiofrequency radiation and health-a hard nut to crack. *International journal of oncology*, 51(2), 405-413.
- [4] World Health Organization (WHO), *Electromagnetic fields (EMF)*, December 2017.
- [5] IEEE Standard 1528-2013 IEEE Recommended Practice for Determining the Peak Spatial-Average Specific Absorption Rate (SAR) in the Human Head from Wireless Communications Devices: Measurement Techniques.
- [6] IEC 62209-1 Human exposure to radio frequency fields from hand-held and body-mounted wireless communication devices—Human models, instrumentation, and procedures—Part 1: Procedure to determine the specific absorption rate (SAR) for hand-held devices used in close proximity to the ear (frequency range of 300 MHz to 3 GHz).
- [7] IEC 62209-2 Human exposure to radio frequency fields from hand-held and body-mounted wireless communication devices—Human models, instrumentation, and procedures—Part 2: Procedure to determine the specific absorption rate (SAR) in the head and body for 30 MHz to 6

- GHz handheld and body-mounted devices used in close proximity to the body.
- [8] Parthasarathy, S. R., & Tukimin, R. (2018, January). Residential exposure from extremely low frequency electromagnetic field (ELF EMF) radiation. In IOP Conference Series: Materials Science and Engineering (Vol. 298, No. 1, p. 012007). IOP Publishing.
- [9] IEC International Electrotechnical Commission. (2005). Human Exposure to Radio Frequency Fields from Hand-Held and Body-Mounted Wireless Communication Devices–Human Models, Instrumentation, and Procedures to Determine the Specific Absorption Rate (SAR) for Hand-Held Devices Used in Close Proximity to the Ear (Frequency Range of 300MHz to 3 GHz). International Standard, 62, 209.
- [10] Peyman, A., & Gabriel, C. (2007). Development and characterisation of tissue equivalent materials for frequency range 30-300MHz. Electronics Letters, 43(5), 19-20.
- [11] Grandolfo, M. (2009). Worldwide standards on exposure to electromagnetic fields: an overview. The Environmentalist, 29(2), 109-117.
- [12] Cámara de Comercio de Lima, “Boletín mensual de importaciones en Perú”. Boletín informativo Cámara de Comercio de Lima, vol. 48, pp. 10–14, 2017, [Online] Available at: <https://camaralima.org.pe/RepositorioAPS/0/0/par/BOLETINIMPO2017-12/DICIEMBRE%20IMPORTACION11.pdf> (Accessed: 15 May 2019).
- [13] Diario Gestión, “¿Cuáles son las marcas de telefonía móvil más consumidas por los peruanos?”, August 2017, [Online] Available at: <https://camaralima.org.pe/RepositorioAPS/0/0/par/BOLETINIMPO2017-12/DICIEMBRE%20IMPORTACION11.pdf> (Accessed: 13 May 2019).
- [14] Melly F., “Estas serían las marcas de móviles más consumidas por los peruanos”, PeruSmart, August 2017, [Online] Available at: <http://www.perusmart.com/5-marcas-moviles-peruanos/> (Accessed: 14 May 2019).
- [15] Diario El Comercio, “Smartphones: estos son los tres teléfonos más utilizados en el Perú”, El Comercio, September 2017, [Online] Available at: <https://elcomercio.pe/tecnologia/moviles/smartphones-son-tres-telefonos-utilizados-peru-noticia-459367> (Accessed: 15 May 2019).
- [16] DeviceAtlas, “The Mobile Web Intelligence Report Q2 2017”, August 2017
- [17] Device Specifications comparisons, news and reviews, Samsung Galaxy J7 [Online] Available at: <https://www.devicespecifications.com/es/model/6b4f34a8>
- [18] Movil Zona, Samsung Galaxy J7 [Online] Available at: <https://www.movilzona.es/samsung/galaxy-j7/>
- [19] Bundesamt für Strahlenschutz, SAR search for Samsung Galaxy J7, 2019 [Online] Available at: [http://www.bfs.de/SiteGlobals/Forms/Suche/Bfs/EN/SARsuche\\_Formular.html](http://www.bfs.de/SiteGlobals/Forms/Suche/Bfs/EN/SARsuche_Formular.html)
- [20] Microwave Vision Group (MVG), User Manual OpenSAR V5 Software, Brest, France, 2013.
- [21] Butet R., Microwave Vision Group (MVG), Presentation ComoSAR, Instituto Nacional de Investigación y Capacitación de Telecomunicaciones (INICTEL-UNI), Lima, Peru, Dec. 2018.

# Multi Focus Image Fusion using Image Enhancement Techniques with Wavelet Transformation

Sarwar Shah Khan<sup>1</sup>, Dr. Muzammil Khan\*<sup>2</sup>, Dr. Yasser Alharbi<sup>3</sup>

College of Information Science and Technology, Beijing University of Chemical Technology, China<sup>1</sup>

Department of Computer & Software Technology, University of Swat, Pakistan<sup>1,2</sup>

College of Computer Science & Engineering, University of Hail, Saudi Arabia<sup>3</sup>

**Abstract**—Multi-focus image fusion produces a unification of multiple images having different areas in focus, which contain necessary and detailed information in the individual image. The paper is proposing a novel idea in a pre-processing step in the image fusion environment in which sharpening techniques applied before fusion in the pre-processing step. This article is proposing multi-focus hybrid techniques for fusion, based on image enhancement, which helps to identify the key features and minor details and then fusion performed on the enhanced images. In image enhancement, we introduced a new hybrid sharpening method that combines Laplacian Filter (LF) with a Discrete Fourier Transform (DFT) and also performs sharpening using the Unsharp sharpen approach. Then fusion is performed using Stationary Wavelet Transformation (SWT) technique to fused the enhanced images and obtaining more detail of the resultant image. The proposed approach is applied to two image sets, i.e., the “planes” and “clocks” image sets. The quality of the output image evaluated using both qualitative and quantitative approaches. Four will know quantitative metrics used to assess the performance of the novel technique. The experimental results of the novel methods showed efficient, improved outcomes and better for multi-focused image fusion. The SWT (LF+DFT) and SWT (Unsharp Mask) are 2.6 %, 1.8%, and 0.62%, 0.61% better than the best baseline measure, i.e., SWT, considering RMSE (Root Mean Square Error) for both image sets.

**Keywords**—Multi focus image fusion; image enhancement; unsharp masking; Laplacian Filter (LF); Stationary Wavelet Transforms (SWT); frequency domain technique

## I. INTRODUCTION

The perfect image should contains the complete elements of the view that are totally transparent and required all the necessary information for the particular application. Due to the intrinsic limitations of the capturing system, an image may cannot comprise all the essential information and the objects description in the scene. For example, the restraint of the limited depth of the focus of optical lenses, that is Complementary Metal-Oxide Semiconductor (CMOS) / Charge-Coupled Device (CCD), digital cameras prompt to prepare the refined image from various focused (multi-focused) of the same scene using the image fusion method, which combines all the focus information from the source images to produced well-informative image [16].

In image fusion, produced the resultant image which captured the complete necessity information in the source images. The fused image is more accurate and informative than any of the individual input image in image fusion. The primary goal of the image fusion is to construct an image from various

images that are more appropriate for the specific application or scenario and more understandable, which also reduce the size of image [1]. The image fusion approaches are involved in many important applications such as object detection, image analysis, monitoring, robotics, remote sensing, hyperspectral image fusion [8], [14], military and medical [16].

The multi focus image fusion is important research field from last couple decades, and the researchers are continuously developing methods that can generate improved results for combining images into the fused image. Basically the image fusion is depends upon two domains, frequency domain and spatial domain, and it's also known as spectral domain and time domain respectively. In spatial domain includes Minimum/Maximum Selection [22], Averaging [15], Principal Component Analysis (PCA) [20] and Intensity Hue Saturation (IHS) [11] methods. All these methods generate poor results because of spectral distortions in resultant image, and generate image with low contrast, which contain less information comparatively [23]. On the other side, the methods such as Discrete Cosine Transform (DCT) [6], Stationary Wavelet Transform (SWT), and Discrete Wavelet Transform (DWT) [7] and the most common frequency domain methods used in multi-focused image fusion. In image fusion, DWT is advantageous method in wavelet transformation [1] but with the following drawbacks:

- It keeps the vertical and horizontal characteristics only
- It suffers through ringing artefacts and reduces the quality of the fused image
- Lack of shifting invariance
- Lack of shifting dimensionality
- Not good for edge places due to missing the edges during in fusion

The method of discrete wavelet transform is not time-invariant transformation method, which means that “with periodic signal extension, the DWT of a translated version of a signal X is not, in general, the translated version of the DWT of X.” The typical DWT method lost to restore the translation invariance, which is marginally covered-up with SWT method by averaging slightly different DWTs, also called  $\varepsilon$ -decimated DWT [24].

From last many year the scientists performed the fusion with simple multi-focused images such the objects that are only located in the special depth of focus are clear, and the others are blurred. In this Paper, we are introducing the

new concept as a pre-processed step before fusion. The pre-processed step is based on image enhancement, which helps to identify the key features and minor details. The Laplacian filter (LF) and Discrete Fourier transform (DFT) is developed as the new hybrid method for sharpening the images (pre-processed) before fusion. In the novel hybrid method, initially, the images are enhanced with LF + DFT sharpen method and then combined the enhanced image with the SWT fusion method. Similarly, the Unsharp sharpening method is also introduced as a new approach for pre-processing. The Unsharp method enhanced the images and then fused by the SWT method. The pre-processing step is firstly introduced in the image fusion environment. The new approaches produced encouraging results using both qualitatively and quantitatively evaluation approaches and compared the results with traditional techniques using two datasets.

The paper is organized as follows: Section 2 describes the novel approach, LF+DFT and Unsharp sharpening methods, and SWT fusion method in detail. Section 3 describes the motivations of the proposed sharpening technique, Section 4 describes the performance metrics, Section 5 providing the experimental results and its comparison with exiting techniques, The conclusion is drawn in Section 6.

## II. PROPOSED APPROACH

In this work, we are introducing a new multi-focus hybrid approach for fusion based on image enhancement, which helps to identify the key features, minor details and then fusion is performed on the enhanced images. The novel framework is presented in Fig. 1, and both the sharpening method i.e., LF + DFT and Unsharp masking with SWT method, are described as follows:

### A. Laplacian Filter (LF)

LF is a spatial filtering method often applied to the images and used to identify the meaningful discontinuities in image, i.e., grey level or colour images, by detecting edges. The edges are formed among two different parts, i.e., having different intensities by calculating the Laplacian using second derivatives and convoluting it with the image [13], [3]. The calculation of the Laplacian equation as follows;

$$\Delta^2 I = \left( \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} \right) \otimes I(\mathbf{x}, \mathbf{y}) \quad (1)$$

The zero-crossings of the second derivative in Fig. 2, corresponding to the edges of the objects [18].

### B. Discrete Fourier Transform (DFT)

The DFT [20] is the equivalent of the continuous Fourier Transform for signals known only at N instants classified by sample times T (i.e., a finite sequence of data). The Fourier Transform of the original signal,  $f(t)$ , as follows;

$$F(i\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \quad (2)$$

The inverse discrete fourier transform is used as;

$$F[n] = \sum_{k=0}^{N-1} f[k] e^{\frac{-i2\pi}{N}nk} \quad (3)$$

### C. Unsharp Mask

An “unsharp mask” is a simple image operator, contrary to what its name might lead you to believe. The name is derived from the fact that it sharpens edges through a process that subtracts an unsharp mask of an image from the reference image, and then detects the presence of edges [4]. Sharpening can demonstrate the texture and details of the images. This is probably the common type of sharpening and can be executed with nearly any image. In a sharpened image, the resolution of the image doesn’t change. In the unsharp mask method, the sharpen image  $a(x, y)$  will be produced from the input image  $b(x, y)$  as

$$a(x, y) = b(x, y) + \lambda c(x, y) \quad (4)$$

Whereas is the correction signals are calculated as the output of a high pass filter and is a positive scaling element that control the level of contrast and an enhanced image achieved at the output [19].

### D. Stationary Wavelet Transform (SWT)

The SWT is a wavelet transform developed to get the better of the lack of translation invariance of the DWT method. The stationary wavelet transform is the whole shift-invariant transformation and overcome the down sampling step of the decimated technique and alternative of up sampling the filter by putting zeros among the filters coefficients [17]. The design is simple and provides better time-frequency localization. Appropriate high pass and low pass filters are applied to the data at each level, and it generate two sequences at the next level. In the decimated algorithm, the filters are used for the rows at the first and second for the columns [5], [12]. The benefit of SWT are: No sub-sampling of input, Translation invariant, providing better time-frequency localization, providing the freedom to carry out a design [10]. The detail of stationary wavelet transform is in Reference [17]. The SWT filter bank structure is shown in Fig. 3.

## III. MOTIVATION OF USING SHARPENING TECHNIQUE

In sharpening technique, the apparent sharpness of an image is increased, which is the merger pair of factors, that is, resolution and acutance. Resolution is straightforward and not subjective which means the size of the images in terms of the number of pixels. With all other factors remain equal, the higher the resolution of the image - the more pixels it has - the sharper it can be. Acutance, which is a measure of the contrast at an edge, is subjective and a little complicated comparatively. There’s no unit for acutance - you either think an edge has contrast or think it doesn’t. Edges that have more contrast appear to have more defined edge to the human visual system. Sharpness comes down to how defined the details in an image are especially the small details.

## IV. PERFORMANCE MEASURES

To properly evaluate the performance of the novel hybrid approaches, To considered four known and common performance measures, i.e., Mean Absolute Error (MAE), Percentage Fit Error (PFE), Root Mean Square Error (RMSE) and Entropy (E) as briefly discussed below;

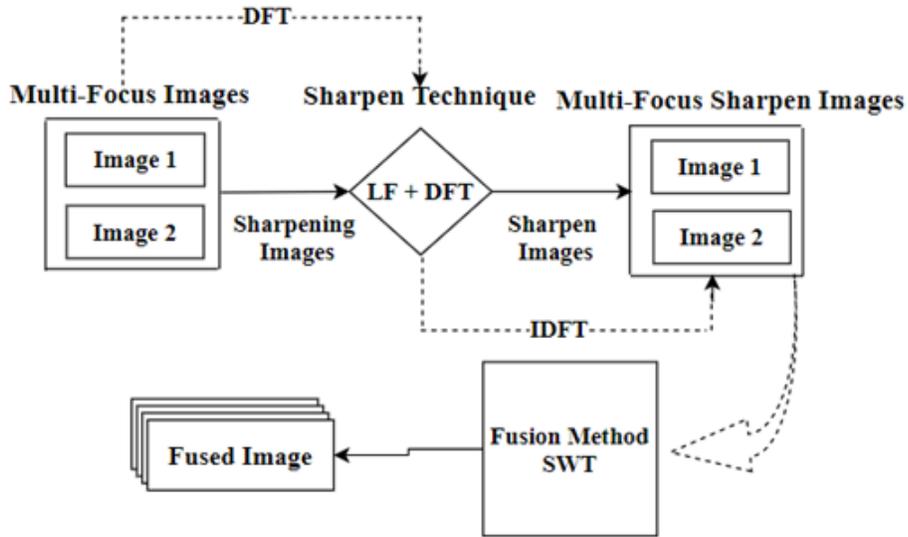


Fig. 1. Framework of the Novel Approach

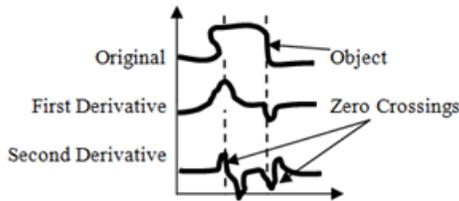


Fig. 2. Edge Detection Using Laplacian Filter

**Mean Absolute Error (MAE)** It gives the MAE of the corresponding pixels in the true image and resultant image, as defined in eq.(5). Lower MAE value indicates higher image quality [2]. It is zero when the reference image and resultant image are equal.

$$MAE = \frac{1}{XY} \sum_{i=1}^X \sum_{j=1}^Y |l_x(i, j) - l_f(i, j)| + \frac{1}{XY} \sum_{i=1}^X \sum_{j=1}^Y |l_y(i, j) - l_f(i, j)| \quad (5)$$

**Percentage Fit Error (PFE)** It is computed as the norm of the difference between the corresponding pixels of the true image and resultant image to the norm of the true image [9]. The smallest values are showing good results. PFE as defined in eq. (6)

$$PFE = \left[ \frac{norm(l_x, l_f)}{norm(l_x)} + \frac{norm(l_y, l_f)}{norm(l_y)} \right] \quad (6)$$

where the norm operator is calculate the highest singular value.

**Root mean square error (RMSE)** is generally applied to compare the difference among the true image and resultant image by instantly calculating the variations in pixel values [21]. The resultant image is close to the true image when the RMSE value is near zero or zero. RMSE is indicating the

spectral quality of the resultant image.

$$RMSE = \sqrt{\frac{1}{XY} \sum_{i=1}^X \sum_{j=1}^Y (I_r(i, j) - I_f(i, j))^2} \quad (7)$$

**Entropy (E)** is an significant metric applied to measure the information content of the resultant image [21]. Entropy as define in eq.(8)

$$E = - \sum_{j=1}^{L-1} P_j \log P_j \quad (8)$$

Where 'L' is the number of grey levels of the fused image. "P<sub>i</sub>" is given by the ratio of the number of pixels.

## V. EXPERIMENTS

In this section, we are discussing the experiments which are performed by the proposed hybrid approach on two image sets, such as "Clocks" and "Planes". These image sets are used as testing multi-focus images for the experimental evaluation of the proposed techniques. The size of the image set (test images) is 512 × 512. The performance of both the proposed approaches such as SWT + Unsharp and SWT + (DFT + LF) methods are compared with the will performed traditional and advanced techniques, which include the average method (AM), minimum method (MM), DWT and SWT methods. The algorithms are implemented using MATLAB 2016b application software tool, and the simulations are performed using a computer of Intel (R) Core(TM) i7-6700K CPU at 4.00 GHz machine with 8GB of RAM to carry out the experiments. The resultant images are evaluated in two ways, i.e., quantitatively and qualitatively. The qualitative analysis is a significant evaluation metric in multi-focus image fusion, which is used to visually observed the changes or improvement in the fused images after applying a techniques. Similarly, quantitative analysis techniques are used to evaluate the effectiveness of a technique statistically. Here, we are using four well know performance matrices for evaluation, such as

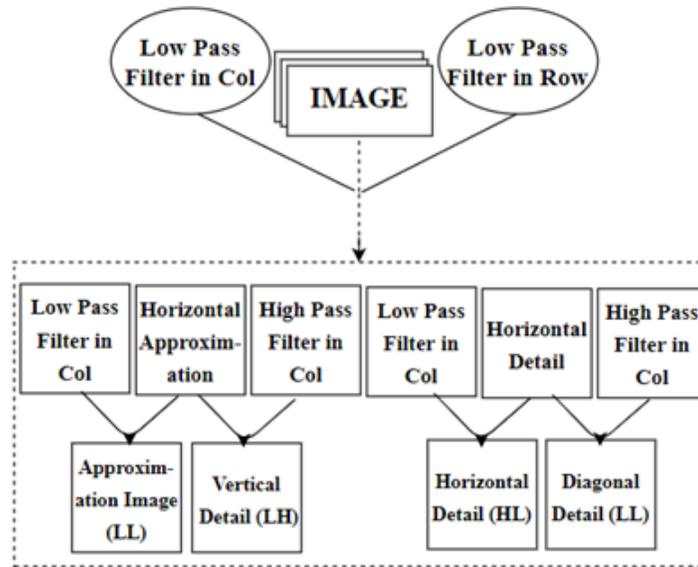


Fig. 3. SWT Filter Bank Structure

Mean Absolute Error (MAE), Percentage Fit Error (PFE), Root Mean Square Error (RMSE) and Entropy (E). The quality of the fused images of the new methods is compared against the baseline techniques using two image sets. In this article, the new concept is introduced as a pre-processing step before fusion and implemented on the Plane image set and Clock image set. The pre-processed step is involved as sharpening the images, and two image sharpening techniques are used as a pre-process like LF+DFT and Unsharp mask.

In Fig. 4, image (a) and (b) are source images of clocks dataset, which are enhanced in pre-process step, i.e., the details of edges are sharpened by Unsharp masking and LF+DFT sharpening techniques showing as (c), (d) and (e), (f) reflectively.

The source and enhanced images are fused by traditional and advanced methods in Fig. 5, image (a)-(f) by average technique, minimum technique, DWT technique, SWT technique, SWT+ Unsharp technique, and SWT(LF+DFT) technique, respectively. Both the proposed techniques are comparatively sharpened and more informative images (showing the detail information) than the existing techniques.

The four common performance matrices are used, and the results are demonstrated in Table I. To easily observed, the best results of the proposed technique against the known techniques are bold. The smallest values indicate good performance for three performance metrics, i.e., RMSE, PFE, and MAE, which can be observed for both the proposed techniques. While the largest value for entropy performance metric and demonstrated impressive results by the SWT with LF+DFT technique.

In Fig. 6, image (a) and (b) are source images of plan dataset, which are enhanced in pre-process step, i.e., the details of edges are sharpened by Unsharp masking and LF+DFT sharpening methods showing as (c),(d) and (e),(f) reflectively. The source and enhanced images are fused by traditional and advanced methods in Fig. 7, image (a)-(f) by average technique, minimum technique, DWT technique, SWT technique,

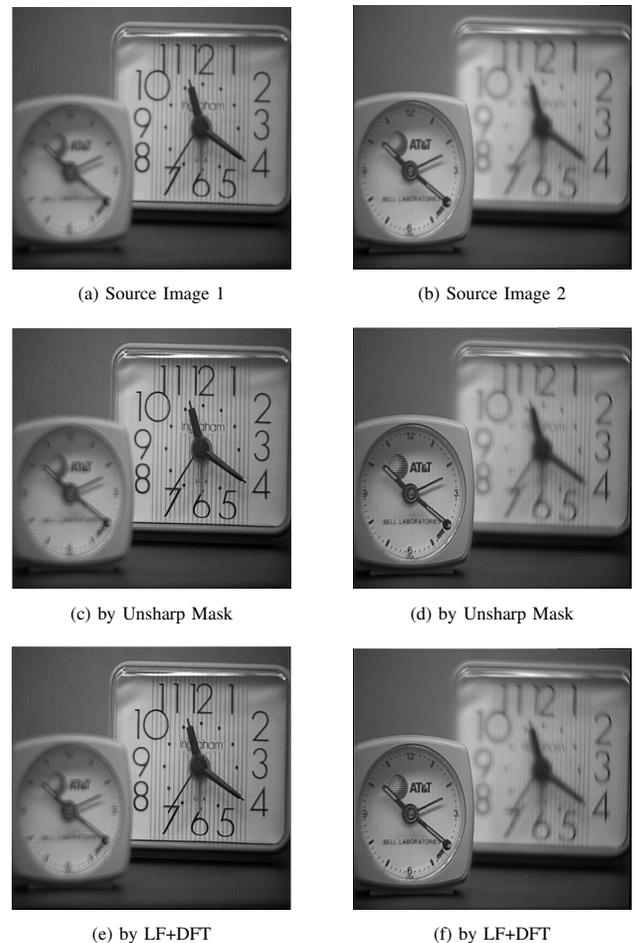


Fig. 4. (a) and (b) are two source images of “Clocks image set”, (c) and (d) are sharp images by unsharp method and (e) and (f) are sharp images by LF+DFT Sharpen images

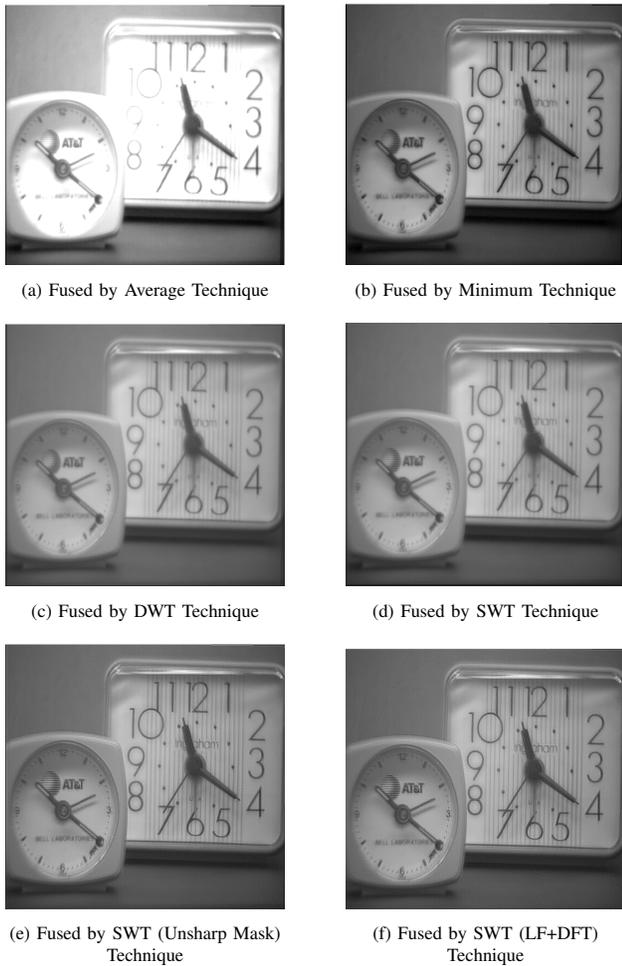


Fig. 5. Fused image of six different techniques on clocks image set (a) Average method (AM) (b) Minimum method (c) DWT (d) SWT (e) SWT + Unsharp method (Proposed) (e) SWT + LF + DFT (Proposed)

TABLE I. RMSE, PEF, MAE, AND ENTROPY PERFORMANCE METRICS COMPARISON OF VARIOUS IMAGE FUSION TECHNIQUES WITH PROPOSED TECHNIQUES ON THE CLOCKS IMAGE SET

Techniques	RMSE	PFE	MAE	Entropy
Average Method	28.4166	23.8202	9.8278	1.9823
Minimum Method	11.5217	10.5229	4.4813	4.8810
DWT	7.7077	7.0396	0.4880	7.8322
SWT	7.5158	6.8643	0.4835	8.3824
SWT+(Unsharp)	6.9049	3.9811	0.4101	8.7321
SWT+(LF+DWT)	<b>5.6761</b>	<b>3.4278</b>	<b>0.4010</b>	<b>9.0121</b>

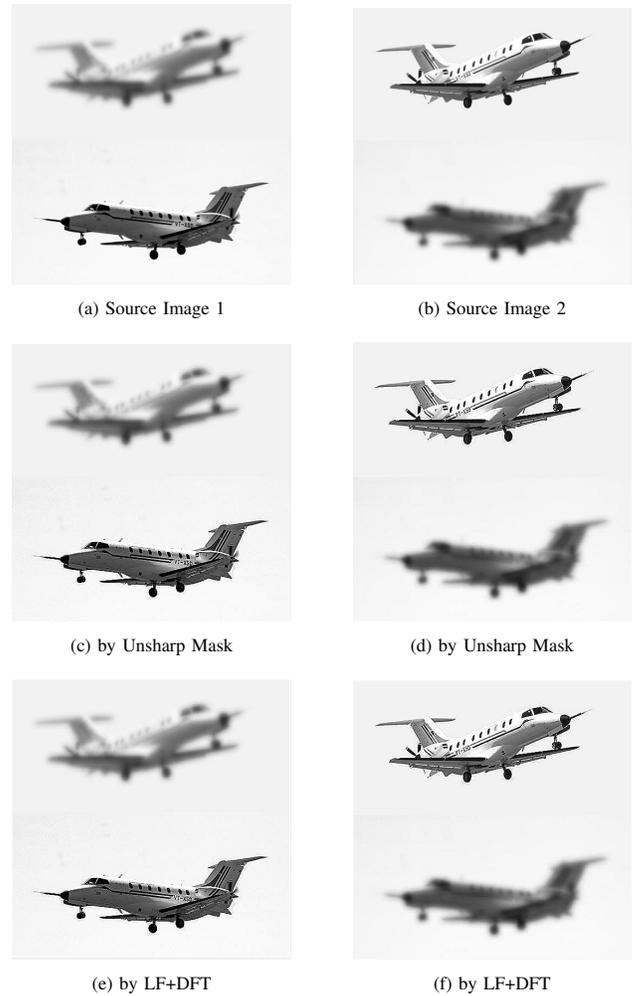


Fig. 6. (a) and (b) are two Source images of “Planes image set”, (c) and (d) are sharp images by unsharp method and (e) and (f) are sharp images by LF+DFT Sharpen images.

SWT+ Unsharp technique, and SWT(LF+DFT) technique, respectively. Both the proposed techniques are comparatively sharper and more informative images (showing the detail information) than the existing methods.

The four known performance matrices are used, and the results are shown in Table II. To easily observed, the best results of the novel technique against the known techniques are bold. The smallest values indicate good performance for three performance metrics, i.e., RMSE, PFE, and MAE, which can be observed for both the proposed techniques, i.e., LF+DFT and SWT with LF+DFT demonstrated good results for entropy performance metric.

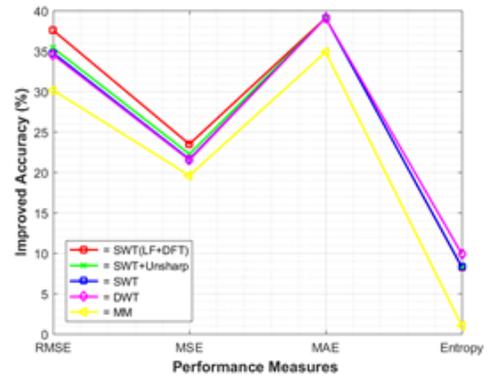
To present the improvement of the proposed techniques, we calculate the improvement of the techniques in terms of accuracy percentage. The percentage is calculated from one of the weak performance matrices in the baselines, i.e. average technique against all comparative techniques as shown in Table III. The proposed technique SWT (Unsharp Mask) outclass all the baseline techniques and improved 35.38% from Average technique as the SWT, DWT, and MM is 34.76, 34.52, and



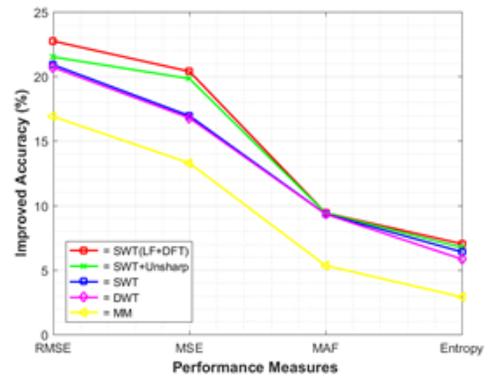
Fig. 7. Fused image of six different techniques on Planes image set (a) Average method (AM) (b) Minimum method (MM) (c) DWT (d) SWT (e) SWT + Unsharp method (Proposed) (e) SWT + LF + DFT (Proposed)

TABLE II. RMSE, PEF, MAE, AND ENTROPY PERFORMANCE METRICS COMPARISON OF VARIOUS IMAGE FUSION TECHNIQUES WITH PROPOSED TECHNIQUES ON THE PLANE IMAGE SET

Techniques	RMSE	PFE	MAE	Entropy
Average Method	46.0270	26.5667	39.1296	0.0027
Minimum Method	15.8744	6.9720	4.2543	0.1032
DWT	11.5027	5.0520	0.0195	<b>0.9920</b>
SWT	11.2614	4.9460	0.0195	0.8329
SWT+(Unsharp)	10.6395	4.2973	0.0198	0.8317
SWT+(LF+DWT)	<b>8.4261</b>	<b>3.0921</b>	<b>0.0182</b>	0.8243



(a) Planes Image Set



(b) Clocks Image Set

Fig. 8. Accuracy improved on Planes and Clocks image sets

30.15, respectively, for Planes image set. Similarly, 21.51% from Average technique as the SWT, DWT, and MM is 20.9, 20.7, and 16.85, respectively, for Clock image set. While the proposed technique SWT (LF+DFT) outperform all the comparative baseline techniques and one of the proposed technique SWT (Unsharp), the comparison can also be observed in the given Fig. 8.

TABLE III. RMSE, PEF, MAE, AND ENTROPY PERFORMANCE METRICS COMPARISON OF VARIOUS IMAGE FUSION TECHNIQUES WITH PROPOSED TECHNIQUES ON THE PLANE IMAGE SET

Dataset	Techniques	RMSE	PFE	MAE	Entropy
Planes	Minimum Method	30.15	19.59	34.87	1.0
	DWT	34.52	21.51	39.10	9.89
	SWT	34.76	21.62	39.10	8.30
	SWT+(Unsharp)	35.38	22.26	39.10	8.29
	SWT+(LF+DWT)	37.56	23.47	39.10	8.21
Clocks	Minimum Method	16.89	13.29	5.34	2.89
	DWT	20.70	16.78	9.33	5.84
	SWT	20.90	16.95	9.34	6.40
	SWT+(Unsharp)	21.51	19.83	9.41	6.74
	SWT+(LF+DWT)	22.74	20.39	9.42	7.02

## VI. CONCLUSION AND FUTURE WORK

Image fusion techniques are essential to get a more informative image from multi-focused images. To fused to fuse multi-focused images and get a more informative resultant image, we proposed hybrid approaches. In which the source images are sharpened in the pre-processing step and then applied two new techniques, i.e., SWT (Unsharp Mask) or SWT (LF+DFT). The results of the novel techniques are compared against four know baseline techniques, i.e., RMSE, PFE, MAF, and Entropy, to assess the proposed techniques. The proposed techniques show good results comparatively by applying both qualitative matric and quantitative matrices to two image sets. The accuracy is keenly analyzed using the RMSE performance matric from Table III. The SWT(LF+DFT), and SWT (Unsharp Mask) shows improved results and outperformed all the comparative techniques, i.e., SWT (Unsharp Mask), SWT, DFT, MM, and Average) by 2.18%, 2.6%, 2.84%, 7.21%, and 37.56% for Plane image set, and 1.23%, 1.84%, 2.04%, 5.85%, and 22.74% for Clock image set.

Currently, we are working to assess the effectiveness of the proposed techniques for other greyscale image sets, and color image sets. In the future, the proposed methods will be extended and improved by other advanced fusion methods such as DWT or DCT. The different performance metrics will validate the new approaches because each metric has its own situational properties. A third evaluation technique will be introduced beside both qualitative and quantitative measures in the future.

## REFERENCES

- [1] M Amin-Naji and A Aghagolzadeh. Multi-focus image fusion in dct domain using variance and energy of laplacian and correlation coefficient for visual sensor networks. *Journal of AI and Data Mining*, 6(2):233–250, 2018.
- [2] Radek Benes, Pavel Dvorak, Marcos Faundez-Zanuy, Virginia Espinosa-Duró, and Jiri Mekyska. Multi-focus thermal image fusion. *Pattern Recognition Letters*, 34(5):536–544, 2013.
- [3] Satish Bhairannawar, Apeksha Patil, Akshay Janmane, and Madhuri Huilgol. Color image enhancement using laplacian filter and contrast limited adaptive histogram equalization. In *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, pages 1–5. IEEE, 2017.
- [4] Muhammad Shahid Farid, Arif Mahmood, and Somaya Ali Al-Maadeed. Multi-focus image fusion using content adaptive blurring. *Information fusion*, 45:96–112, 2019.
- [5] Reham Gharbia, Aboul Ella Hassanien, Ali Hassan El-Baz, Mohamed Elhoseny, and M Gunasekaran. Multi-spectral and panchromatic image fusion approach using stationary wavelet transform and swarm flower pollination optimization for remote sensing applications. *Future Generation Computer Systems*, 88:501–511, 2018.
- [6] Mohammad Bagher Akbari Haghighat, Ali Aghagolzadeh, and Hadi Seyedarabi. Real-time fusion of multi-focus images for visual sensor networks. In *2010 6th Iranian Conference on Machine Vision and Image Processing*, pages 1–6. IEEE, 2010.
- [7] Maruturi Haribabu and Ch Hima Bindu. Visibility based multi modal medical image fusion with dwt. In *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, pages 1561–1566. IEEE, 2017.
- [8] Maryam Imani and Hassan Ghassemian. An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges. *Information fusion*, 59:59–83, 2020.
- [9] P Jagalingam and Arkal Vittal Hegde. A review of quality metrics for fused image. *Aquatic Procedia*, 4(Icwrcoe):133–142, 2015.
- [10] S Janani, R Marisuganya, and R Nivedha. Mri image segmentation using stationary wavelet transform and fcm algorithm. *International Journal of Computer Applications*, pages 0975–8887, 2013.
- [11] Umer Javed, Muhammad Mohsin Riaz, Abdul Ghafoor, Syed Sohaib Ali, and Tanveer Ahmed Cheema. Mri and pet image fusion using fuzzy logic and image local features. *The Scientific World Journal*, 2014, 2014.
- [12] Sarwar Shah Khan, Muzammil khan, and Qiong Ran. Multi-focus color image fusion using laplacian filter and discrete fourier transformation with qualitative error image metrics. In *Proceedings of the 2nd International Conference on Control and Computer Vision*, pages 41–45, 2019.
- [13] Sarwar Shah Khan, Muzammil khan, and Qiong Ran. Pan-sharpening framework based on laplacian sharpening with brovey. In *IEEE International Conference on Signal, Information and Data Processing 2019*, 2019.
- [14] Sarwar Shah Khan, Qiong Ran, Muzammil Khan, and Mengmeng Zhang. Hyperspectral image classification using nearest regularized subspace with manhattan distance. *Journal of Applied Remote Sensing*, 14(3):032604, 2019.
- [15] Liang Kou, Liguozhang, Kejia Zhang, Jianguo Sun, Qilong Han, and Zilong Jin. A multi-focus image fusion method via region mosaicking on laplacian pyramids. *PLoS one*, 13(5), 2018.
- [16] Xiaosong Li, Fuqiang Zhou, and Juan Li. Multi-focus image fusion based on the filtering techniques and block consistency verification. In *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pages 453–457. IEEE, 2018.
- [17] Tian Lianfang, J Ahmed, Du Qiliang, Bhawani Shankar, and Saifullah Adnan. Multi focus image fusion using combined median and average filter based hybrid stationary wavelet transform and principal component analysis. *International Journal of Advanced Computer Science and Applications*, 9(6):34–41, 2018.
- [18] Ioannis Pitas. *Digital image processing algorithms and applications*. John Wiley & Sons, 2000.
- [19] Andrea Polesel, Giovanni Ramponi, and V John Mathews. Image enhancement via adaptive unsharp masking. *IEEE transactions on image processing*, 9(3):505–510, 2000.
- [20] Senthil Kumar Sadhasivam, Mahesh Bharath Keerthivasan, and S Mutan. Implementation of max principle with pca in image fusion for surveillance and navigation application. *ELCVIA: electronic letters on computer vision and image analysis*, 10(1):1–10, 2011.
- [21] Abdul Basit Siddiqui, M Arfan Jaffar, Ayyaz Hussain, and Anwar M Mirza. Block-based feature-level multi-focus image fusion. In *2010 5th International Conference on Future Information Technology*, pages 1–7. IEEE, 2010.
- [22] Ias Sri Wahyuni. *Multi-focus image fusion using local variability*. PhD thesis, 2018.
- [23] Wencheng Wang and Faliang Chang. A multi-focus image fusion method based on laplacian pyramid. *JCP*, 6(12):2559–2566, 2011.
- [24] Yudong Zhang, Zhengchao Dong, Lenan Wu, Shuihua Wang, and Zhenyu Zhou. Feature extraction of brain mri by stationary wavelet transform. In *2010 International Conference on Biomedical Engineering and Computer Science*, pages 1–4. IEEE, 2010.

# Multiclass Pattern Recognition of Facial Images using Correlation Filters

Nisha Chandran S<sup>1</sup>, Charu Negi<sup>2</sup>, Poonam Verma<sup>3</sup>

School of Computing  
Graphic Era Hill University  
Dehradun, India

**Abstract**—Pattern Recognition comes naturally to humans and there are many pattern recognition tasks which humans can perform admirably well. However, human pattern recognition cannot compete with machine speed when the number of classes to be recognized becomes tremendously large. In this paper, we analyze the effectiveness of correlation filters for pattern classification problems. We have used Distance Classifier Correlation Filter (DCCF) for pattern classification of facial images. Two essential qualities of a correlation filter are distortion tolerance and discrimination ability. DCCF transposes the feature space in such a way that the images belonging to the same class gets closer and the images from different class moves far apart; thereby increasing the distortion tolerance and the discrimination ability. The results obtained demonstrate the effectiveness of the approach for face recognition applications.

**Keywords**—Pattern recognition; correlation filter; multiclass recognition

## I. INTRODUCTION

There are many daily pattern recognition tasks that come naturally to humans. For example, we can recognise a close friend of ours even after a gap of many years though his features have changed a lot. We can understand a familiar voice even if it is slightly distorted. However, human pattern recognition suffers from three main drawbacks: poor speed, difficulty in scaling, and inability to handle some recognition tasks. Not surprisingly, humans cannot match machine pattern recognition tasks where good pattern recognition algorithms exist. Also, human pattern recognition has limitations when the number of classes to be recognized becomes large. Although humans have evolved to perform well on some recognition tasks such as face or voice recognition, except for a few trained experts, most humans cannot tell whose fingerprint they are looking at. Thus, there are many interesting pattern recognition tasks for which we need machines.

The main goal of pattern recognition is to assign an observation maybe a signal, or an image or a high dimensional object into one of the multiple classes. An important class of pattern recognition applications is the use of biometric signatures like face image, fingerprint image, iris image etc. for person identification [1-.3] The use of two-dimensional (2-D) correlation to detect, locate, and classify targets in observed scenes has been a topic of research for a long time [4-10].

In this paper we analyze the possibility of using correlation filters for solving multiclass classification problems. DCCF design uses a global transformation (correlation filter) to transform the feature space to decrease the intra class distance and to increase the inter class distance. Results of the experiments conducted on benchmark dataset demonstrate the robustness of the proposed method. The rest of this paper is organized as follows. We begin with a discussion on some related work on correlation filters in Section 2. In Section 3, we discuss the salient features of DCCF and outlines the strategies adopted to apply DCCF for a multiclass facial recognition problem. Section 4 provides the experimental results and finally Section 5 concludes the paper.

## II. RELATED WORK

Correlation filters have been widely used for several pattern recognition tasks and visual tracking of objects [11]. The advantage of using correlation filters for object tracking tasks is that it can track objects that are rotated or occluded or are with several photometric and geometric challenges. Pattern recognition of complex objects which are partially occluded are also done efficiently using multiple correlation filters [12] [13]. Composite correlation filters [14-17] gives superior performance to Matched Filters as they are designed from multiple reference images. If the reference image set are well represented to incorporate all the possible distortions that are likely to be encountered by an object, then the resulting filter will be distortion tolerant. While designing correlation filters three questions need to be considered: (1) How good is the ability of the filter to suppress clutter and noise. (2) How easy is it to detect a correlation peak. (3) How tolerant is the filter to the distortion of the object.

One of the earliest composite correlation filters proposed was the Synthetic Discriminant Function (SDF) [18]. SDF uses a linear combination of reference images to create a composite image. When the designed filter is correlated with a test image, a peak will be present in the correlation plane if the test image corresponds to the TRUE class to be recognized. For all other inputs belonging to FALSE class, there will be no peak in the correlation plane. For digital pattern recognition applications, an SDF synthesized in the computer is correlated with the test image digitally, whereas for optical pattern recognition applications, the SDF designed digitally is converted to a hologram using multiple-exposure holographic techniques.

Initially, design of SDF did not consider any noise and hence the filters were not noise tolerant. Minimum Variance SDF (MVSDF) [19] was one of the earliest attempts to introduce noise analysis in SDF filter design by maximizing the noise tolerance of the SDF. The original SDF design considered only the cross-correlation values at the origin. This could not ensure that the output had its peak at the origin. Since shift in the reference images resulted in shift in the peak location, there was an ambiguity in peak location when the reference image shifts were unknown. This problem was addressed in the Minimum Average Correlation Energy (MACE) filters [20]. MACE filters could produce sharp correlation peaks at the origin and were more likely to produce a correlation peak at the same location as the shifted input.

However, it was realized that by exclusively focussing on the correlation-peak values, one neglects the information in the other regions of the correlation plane. In Minimum-squared-error synthetic discriminant function the averaged squared error between the resulting correlation outputs and the desired one is minimized to obtain a desired correlation plane. Maximum-average-correlation-height filter essentially uses this idea to achieve a correlation shape that yield the smallest squared error. Distance Classifier Correlation Filters (DCCF) [21] essentially incorporates the two ideas. In DCCF rather than considering just the peak the entire correlation plane was considered. Applications of initial DCCF had limitations as the approach was limited to just two classes at a time.

### III. MULTICLASS PATTERN RECOGNITION OF FACIAL IMAGES

A facial recognition system is a computer application for automatically identifying or verifying a person from a digital image or a video frame obtained from a video source. One of the ways to do this is by comparing selected facial features from the image and a facial database. DCCFs were first proposed by Mahalanobis *et al.* DCCF uses a correlation filter  $h$ , designed from a set of training images, to classify a test image into one of a set of predefined classes. As illustrated in Fig. 1, DCCF design uses a global transformation (correlation filter)  $h$  which transforms the space in such a way that same class gets closer and images belonging to different classes move apart. This improves distortion tolerances as well as discrimination ability of the correlation filter.

#### A. Formulation of DCCF Filter

Let  $X_{ik}$  represent the 2D Fourier transform of the  $i^{\text{th}}$  training image of class  $k$  ordered as a vector and  $X_{ik}$  a diagonal matrix with  $X_{ik}$  as its diagonal elements. Let  $m_k$  represents the mean vector of class  $k$  and  $M_k$  the diagonal matrix with  $m_k$  as its diagonal elements. If the transformation  $h$  has to make the inter-class distance large, then the distance between the mean correlation peak values between the different classes is made as large as possible. This is formulated as the measure  $h^+ M h$  where  $M$  is given as in equation (1)[16].

$$M = \frac{1}{c} \sum_{k=1}^c (m - m_k) (m - m_k)^+ \quad (1)$$

In the above equation  $m_k$  represents the mean vector of class 'k' and  $m$  is the mean of all the classes given as in equation (2).

$$m = \frac{1}{c} \sum_{k=1}^c m_k \quad (2)$$

Simultaneously, the transformation  $h$  makes each class compact. The compactness of a class after applying  $h$  is measured by the metric average similarity measure [9] which is a metric measure of the similarity between the training images and the mean value of the class given by  $h^+ S h$  where  $S$  is intra-class scatter matrix given as in equation (3).

$$S = \frac{1}{c} \sum_{k=1}^c \frac{1}{N} \sum_{i=1}^N h^+ |X_{ik} - M_k| |X_{ik} - M_k|^* h \quad (3)$$

It follows that the correlation filter  $h$  should be so designed that it maximizes the metric.

$$J(h) = \frac{h^+ M h}{h^+ S h} \quad (4)$$

The  $J(h)$  is maximized when  $h$  is the dominant eigenvector of  $S^{-1}M$ .

Once the correlation filter  $h$  is designed using the training image set, a test image is classified by correlating with  $h$ . The distance metric is calculated as the difference between two correlation peaks. The first correlation peak corresponds to the correlation of the filter with the test image. As shown in Fig. 1, this corresponds to the transformation of the test image. The second correlation peak corresponds to the correlation of the filter with the mean vector of class 'k' which corresponds to the transformation of the class 'k'. The test image is assigned to the class which gives the minimum distance.

#### B. Classification using the DCCF Filter: Method 1

Let  $z$  be the test image. The correlation peak that corresponds to the correlation between the test image and the correlation filter is given by  $H^* z$ . The correlation peak that corresponds to the correlation between the mean vector of class 'k' and the correlation filter is given by  $H^* m_k$ . The distance metric  $d_k$  that gives the distance between the transformed test image and transformed mean vector of class 'k' is given by.

$$d_k = |H^* z - H^* m_k|^2 \quad (5)$$

The given test image is assigned to the class that gives the minimum value of  $d_k$

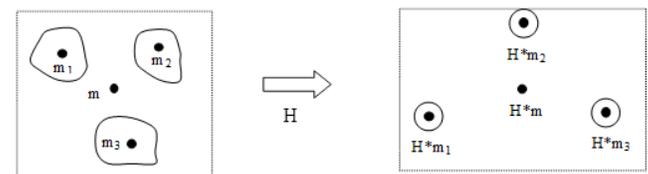


Fig. 1.  $H$  Transforms the Feature Space in Such a Way that the Interclass Distance Increases Whereas the Intraclass Distance Decreases. Here  $m_1, m_2, m_3$  Represent the mean Value of the Fourier Transform of the Images within each Class and  $m$  that of All the Images Combined.  $H^+ m, H^+ m_1, H^+ m_2,$  and  $H^+ m_3$  Represent the Correlation Peak Values of the Respective Classes.

C. Classification using the DCCF Filter: Method 2

The above equation for  $d_k$  can also be written as follows

$$d_k = a + b_k - z^+ h_k - h_k^+ z \quad 1 \leq k \leq C \quad (6)$$

In the above equation,  $a$  is the transformed input image and is given as  $a = |H^*Z|^2$  and  $b_k = |H^*m_k|^2$  is the energy of the transformed  $k^{\text{th}}$  class mean and  $h_k = HH^*m_k$  is considered as the effective filter for class  $k$ . This gives us an alternate strategy for classification of a test image using 2D correlation techniques. The third and the fourth term in the expression (6) is implemented as a 2D correlation of the test image 'z' and the effective filter  $h_k$  for the class 'k'. The term  $d_k$  is minimized when the third and fourth term is maximized which corresponds to the peak correlation value of  $z$  with  $h_k$ . Shown in Fig. 2 is the schematic for implementing this method using 2D correlation. The test image is correlated with effective filter of each class and the peak value which is the value at the origin of the correlation plane is used to compute the distance  $d_k$  using expression (6). As discussed in the previous section, coherent optical processing systems can be used to implement these techniques.

D. Facial Recognition using DCCF

In this section we discuss the classification of 'c' different objects, in this case facial images, using the DCCF. Facial images of one person with different facial expressions form the members of one class. The training images are used to design the filter, which is used to classify a test image. The algorithms for the classification of the facial images using DCCF are given in Table I.

TABLE I. ALGORITHMS FOR MULTICLASS PATTERN CLASSIFICATION USING DCCF

Algorithm to calculate DCCF filter	
Step 1:	Formulate the vectors $X_{jk}$ , $m_k$ and $m$
Step 2:	Calculate M using Eqn (1)
Step 3:	Calculate Average Similarity Measure S using Eqn (3)
Step 4:	Calculate h as the dominant Eigenvector of $S^{-1}M$
Algorithm to calculate the dominant eigenvector of $S^{-1}M$	
Step 1:	Formulate the matrix E as $E =  e_1 e_2 e_3 \dots e_C $
Step 2:	Formulate V as $E^+E$
Step 3:	Find the eigenvectors and Eigen values of V.
Step 4:	Formulate the matrix P whose columns consist of the normalized eigenvectors of V.
Step 5:	Formulate the square matrix $\Delta$ whose diagonal consist of non-negative eigenvectors of V
Step 6:	Compute $\Phi = E P \Delta^{-1/2}$
Step 7:	Compute $\Delta \Phi^+ S^{-1} \Phi$
Step 8:	Calculate the dominant eigenvector $a_{\max}$ of $\Delta \Phi^+ S^{-1} \Phi$
Step 9:	Compute the dominant eigenvector of $S^{-1}M$ as $S^{-1} \Phi a_{\max}$
Algorithm for classification of the test image (Method 1)	
Step 1:	Formulate the square matrix H whose diagonal consist of the vector h
Step 2:	Calculate the distance metric of the test image vector z to the class k mean vector $m_k$ as $d_k =  H^+z - H^+m_k ^2$ for $k=1$ to C
Step 3:	Classify the vector z to that class that gives the minimum value of distance metric.
Algorithm for classification of the test image (Method 2)	
Step 1:	Calculate $h_k$ and $b_k$ for each $k=1$ to C as given by expression (6)
Step 2:	Perform the 2D correlation of test image 'z' with the effective filter $h_k$ for each class.
Step 3:	Compute the distance metric $d_k$ for each class 'k' using expression (6).
Step 4:	Classify the vector z to that class that gives the minimum value of distance metric.

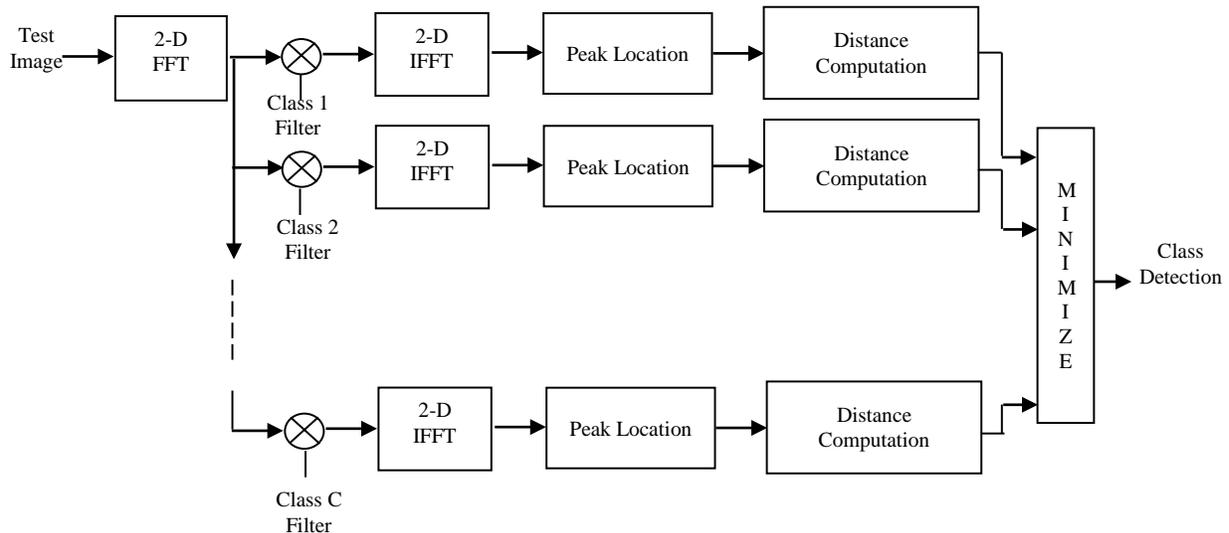


Fig. 2. Schematic Diagram to Illustrate the Design and the use of DCCF's.

#### IV. EXPERIMENTAL EVALUATION

Experiments are carried out on Essex face database [22] which consists of 3040 facial images of 152 persons. A total of 20 images are available for training from each class. The subjects are set at fixed distance from the camera and were asked to speak, whilst a sequence of 20 images was taken. The speech was used to introduce facial expression variation. The resolution of each image was 180 x 200 pixels. The algorithm discussed in Section 2 was used to design a DCCF that could classify any image in the database to one of the 152 classes. The 20 images in each class were divided into two sets. One set of images were used to train the filter referred to as training images. The other set of images were used to test the filter referred to as test images. Representative images from the database is shown in Fig. 3.

The robustness of the designed filter would depend on the size of the training image set and how well the training image set represents the possible distortions that occur. Obviously, the larger the training image set, the robust the filter will be. But usually there are a limited number of images available. Hence, if most of the images are used to train the filter, we are left with few images to test the filter. If the number of images used to train the filter is less, then the classification error increases. Hence the training and test image set must be chosen judiciously.

Table II gives the results obtained from the study varying the size of the training images set and test image dataset. In each case the total number of images used (i.e.training+testing) is kept a constant. The filter is generated using the training images. The other images are used for testing. The distance of the transformed test image from the transformed mean image of each class is calculated as described in the previous chapter. The test image is classified to the class which gives the minimum value of distance metric.

The total classification error for all classes as a percentage of total facial images are plotted against total number of training images as a percentage of total facial images and is shown in Fig 4. It is observed that when more than 40% images are used for training the classification error is significantly low. Fig. 5 shows the distribution of errors when 95% of the images are used for training. There are only 5 errors out of 152 images used for testing (one image from each of 152 classes) and these 5 errors fall in 5 different classes. In other words, the errors are distributed evenly. Fig. 6 shows the distribution of errors when 90% images of the image dataset is used for training. The 16 errors are distributed over 10 classes with the maximum error per class being 2. Fig. 7 shows the distribution of errors when 85% of the images are used for training. The 25 errors are distributed over 13 classes with the maximum error per class being 3. Fig. 8 shows the distribution of errors 38 errors over 14 classes when 80% of the images are used for training with maximum error per class being 4 and minimum error per class being 1. Fig. 9-15 shows the distribution of errors when the training images decreases from 75% to 45%. The classification errors correspondingly increase from 50 to 175. For the case when training images are 75% of the image dataset, the errors fall in

15 classes with maximum error per class being 5 and minimum error per class being 1. When the training images are 45%, the errors fall into 40 classes with maximum error per class being 11 and minimum error per class being 1. It is observed that for all the cases the error is uniformly distributed across all classes. Fig. 16-23 shows the distribution of errors when the training images decreases from 40% to 5%. The total errors increase from 219 to 2868. These errors fall in 46 classes and 152 classes, respectively. In the worst case when the training images are just 5%, the errors are evenly distributed in all the 152 classes with the average error per class being 19. It may be seen that as the number of training images decreases, the error increases, as expected, and these errors are evenly distributed across multiple classes.



Fig. 3. Representative Images from the Face Database.

TABLE II. CLASSIFICATION ERROR OBTAINED BY VARYING THE NUMBER OF TRAINING IMAGES

Sr. No.	Training Images per class	Total Training images	Test Images Per class	Total Test images	% Training image	Class Error	% Error
1	19	2888	1	152	95	5	3.3
2	18	2736	2	304	90	16	5.3
3	17	2584	3	456	85	25	5.5
4	16	2432	4	608	80	38	6.3
5	15	2280	5	760	75	50	6.6
6	14	2128	6	912	70	67	7.3
7	13	1976	7	1064	65	79	7.4
8	12	1824	8	1216	60	102	8.4
9	11	1672	9	1368	55	129	9.4
10	10	1520	10	1520	50	158	10.4
11	9	1368	11	1672	45	175	10.5
12	8	1216	12	1824	40	219	12
13	7	1064	13	1976	35	305	15.4
14	6	912	14	2128	30	400	18.8
15	5	760	15	2280	25	539	23.6
16	4	608	16	2432	20	751	30.9
17	3	456	17	2584	15	1096	42.4
18	2	304	18	2736	10	1892	69.1
19	1	152	19	2888	5	2868	99.3

The experimental results presented shows that DCCF successfully classifies facial images of 152 persons from the Essex database. The facial images belong to persons from quite different ethnic backgrounds. There are 20 images of each person with different facial expressions. It is seen that the DCCF filter shows good discrimination ability to classify facial expressions of 152 persons while maintaining good distortion tolerance towards a variety of facial expressions present in each class. It is also seen that the classification error is below 10% when at least 40% images from the available dataset are used to train the filter.

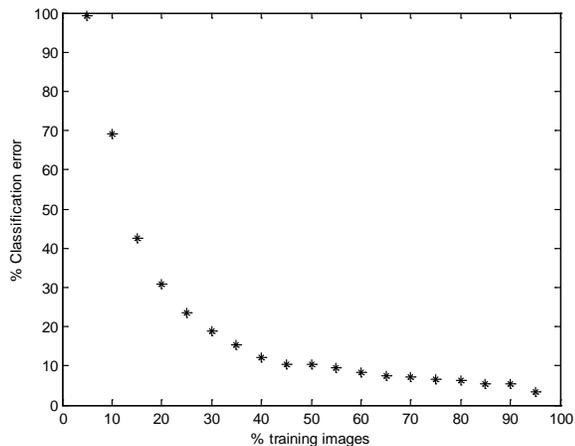


Fig. 4. Graph Showing the Decrease in Classification Error as the Number of Training Images Increases.

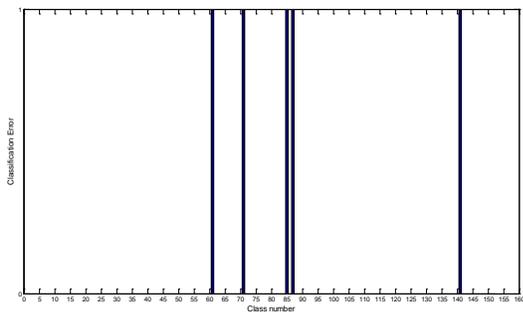


Fig. 5. 95% Training Images.

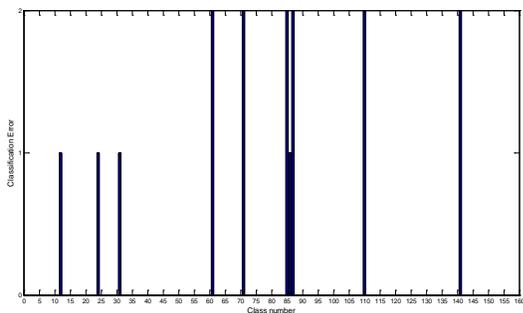


Fig. 6. 90% Training Images.

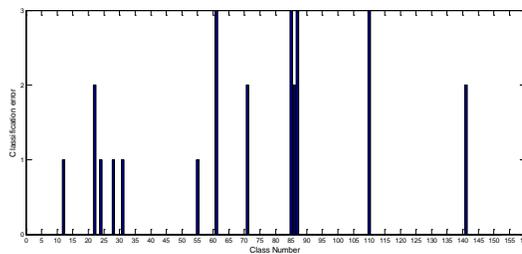


Fig. 7. 85% Training Images.

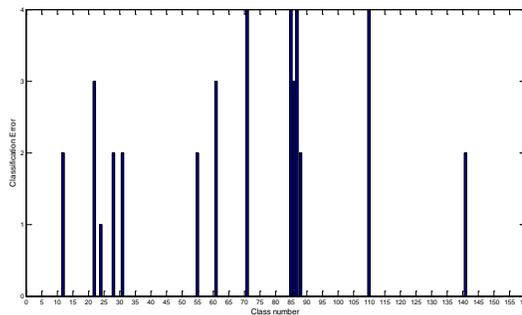


Fig. 8. 80% Training Images.

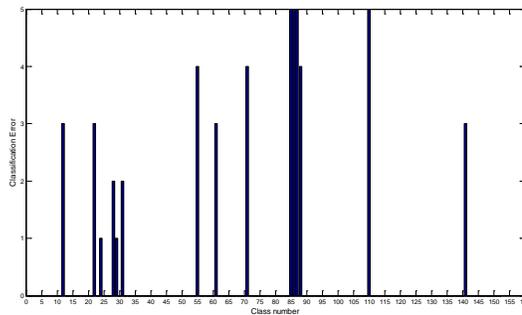


Fig. 9. 75% Training Images.

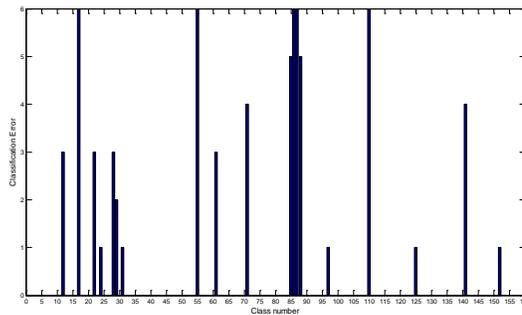


Fig. 10. 70% Training Images.

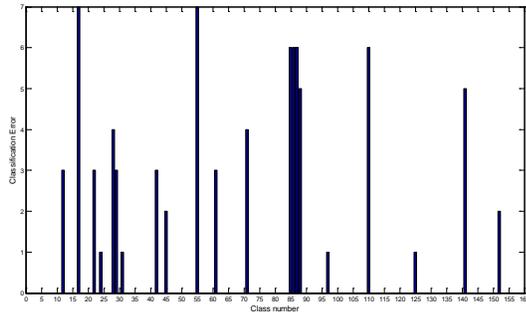


Fig. 11. 65% Training Images.

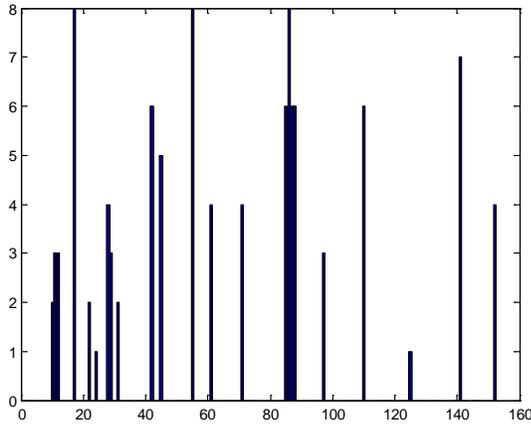


Fig. 12. 60% Training Images.

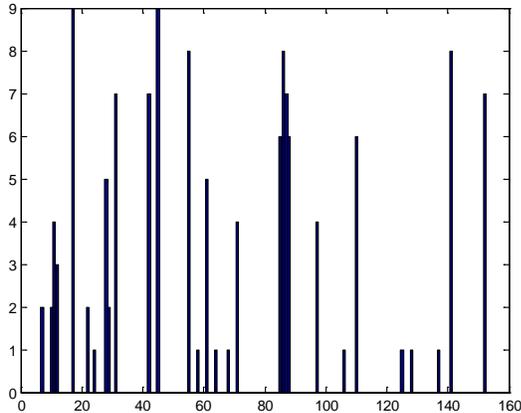


Fig. 13. 55% Training Images.

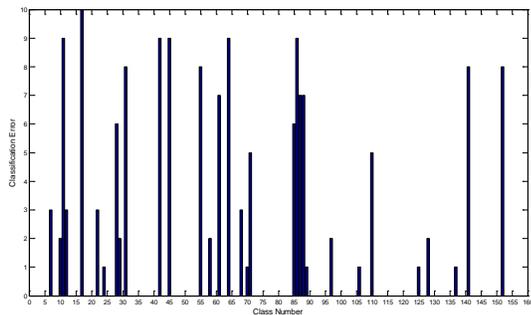


Fig. 14. 50% Training Images.

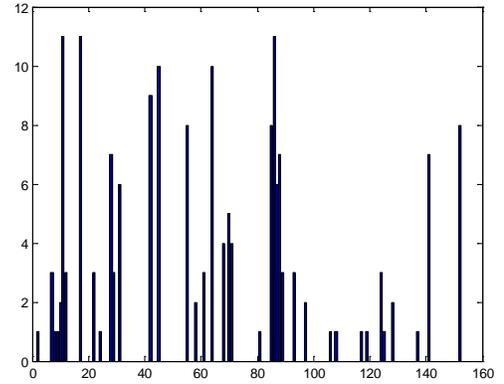


Fig. 15. 45% Training Images.

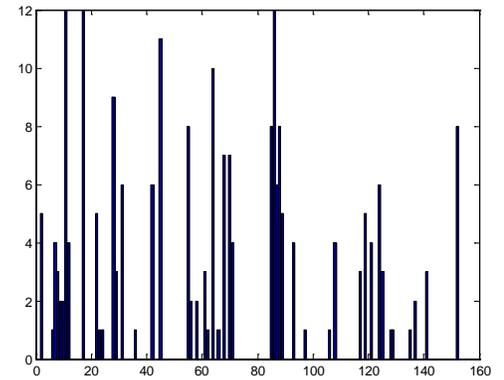


Fig. 16. 40% Training Images.

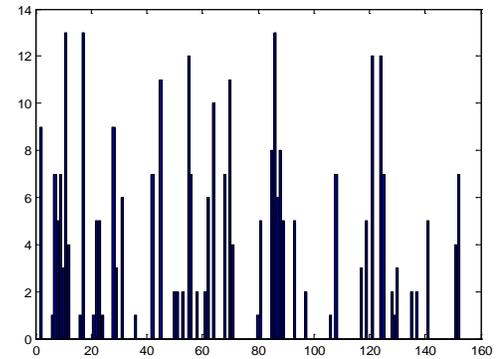


Fig. 17. 35% Training Images.

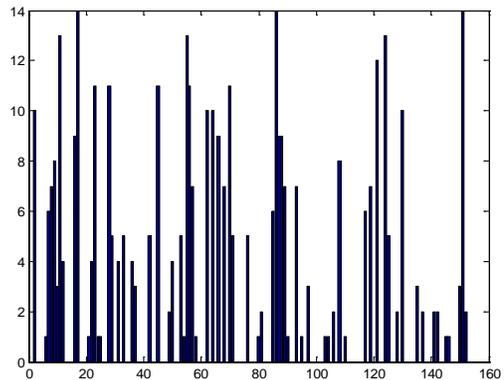


Fig. 18. 30% Training Images.

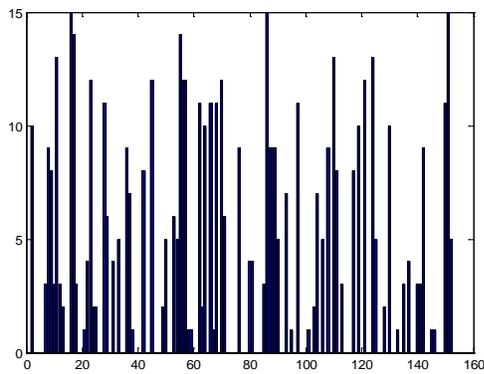


Fig. 19. 25% Training Images.

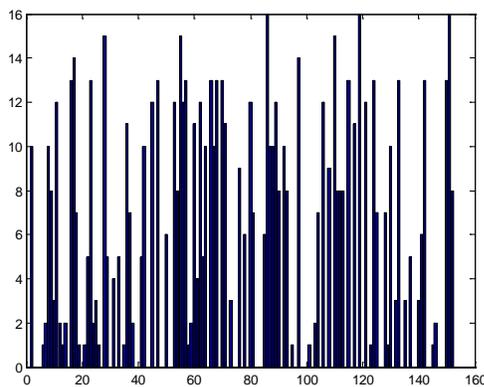


Fig. 20. 20% Training Images.

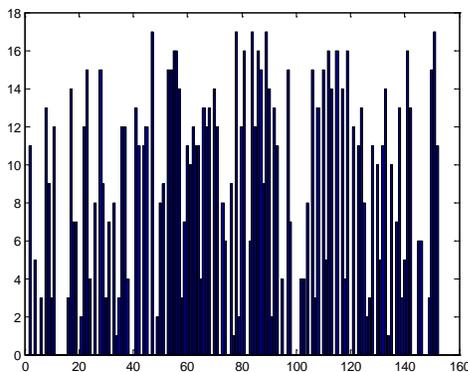


Fig. 21. 15% Training Images.

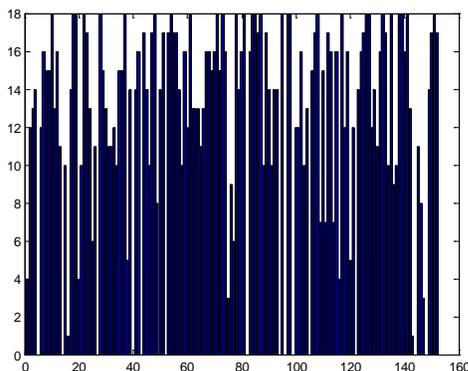


Fig. 22. 10% Training Images.

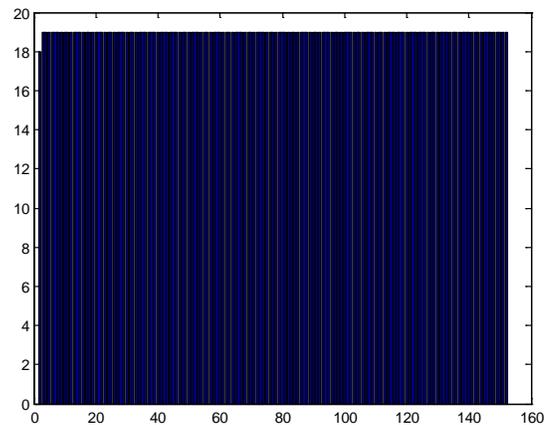


Fig. 23. 5% Training Images.

## V. CONCLUSION

In this paper correlation based pattern recognition is adopted for classification of facial images. A good correlation filter must have distortion tolerance and discrimination ability in equal measure. Distance Classifier Correlation Filter is designed, with a global transformation  $H$  which maximises the separation between different classes and which minimizes the spread of each class is found. Experimental results show DCCFs work perfectly well for face recognition applications. However, the training set used to develop the filter should be large and well represented. Percentage of training images must be at least 40% of the available dataset for errors to be reasonably low. We find that as the images used to train the filter decreases the increase in classification error is distributed evenly across multiple classes. Our future work would involve the study of DCCF based classification technique for various other biometrics like fingerprint, iris, etc.

## REFERENCES

- [1] R.O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, 2<sup>nd</sup> edn. New York, John Wiley, 2001.
- [2] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2<sup>nd</sup> edn., New York, Academic Press, 1990.
- [3] G. F. Schils and D. W. Sweeney, "Optical processor for recognition of three-dimensional targets viewed from any direction," *J. Opt. Soc. Am. A* 5, 1309-1321 (1988).
- [4] M. Turk and A. Pentland, "Eigen faces for recognition," *Journal of Cognitive Neuroscience*, 3, 1991, 71-86.
- [5] B.V.K. Vijaya Kumar, "Tutorial survey of composite filter designs for Optical correlators," *Applied Optics*, 31, 1992, 4773-4801.
- [6] B. V. K. Vijayakumar, A. Mahalanobis, and R. D. Juday, *Correlation Pattern Recognition*, Cambridge University Press, Cambridge, 2005.
- [7] B. Braunecker, R. W. Hauck, and A. W. Lohmann, "Optical character recognition based on no redundant correlation measurements," *Appl. Opt.* 18, 1979, 2746-2753.
- [8] C. Hester and D. Casasent, "Multivariate technique for multiclass Pattern recognition," *Applied Optics*, 19, 1980, 1758-1761.
- [9] R. R. Kallman, "Direct construction of phase-only filters," *Appl. Opt.* 26, 1987, 5200-5201.
- [10] S.M. Kay, *Fundamentals of Signal Processing: Estimation Theory Upper* Saddle River, NJ, Prentice Hall, 1993.
- [11] D.S. Bolme, J.R. Beveridge, B.A. Draper, and Y.M. Lui, "Visual object tracking using adaptive correlation filters", in *proc. IEEE computer society conference on computer vision and pattern recognition*, 13-18 June 2010 San Francisco, CA, USA, pp. 2544-2550.

- [12] T. Liu, G. Wang and Q. Yang, "Real-time part-based visual tracking via adaptive correlation filters", in proc. IEEE Conference on Computer Vision and Pattern Recognition, 7-12 June 2015, Boston. pp. 4902-4912.
- [13] H. Kiani Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking", in Proc. IEEE international conference on computer vision, 22 -29 Oct 2017 ,Venice, Italy, pp. 1135-1143.
- [14] H.L. Van Trees, Detection, Estimation and Modulation Theory: Part I New York, John Wiley, 1968.
- [15] C.R. Rao, "Information and accuracy attainable in the estimation of Statistical Parameters," Bulletin of the Calcutta mathematical Society, 37, 1945, 81-91.
- [16] C.R. Rao, Linear Statistical Inference and its Applications, New York, John Wiley, 1973.
- [17] H. J. Caulfield and W. T. Maloney, "Improved discrimination in optical character recognition", Appl. Opt., 8, 1969, 2354-2356.
- [18] Z. Bahri and B. V. K. ViJaya Kumar, "Generalized synthetic discriminant functions," J. Opt. Soc. Am. A 5, 1988, 562-571.
- [19] B.V.K. Vijaya Kumar, "Minimum variance synthetic discriminant Functions," Journal of the Optical Society of America A., 3, 1986, 1579-84.
- [20] A. Mahalanobis, B.V.K. Vijaya Kumar, D. Casasent," Minimum Average Correlation Energy filters," Applied Optics, 26, 1987, 3633-3640.
- [21] A. Mahalanobis, B.V.K.Vijaya Kumar, and S.R.F. Sims, "Distance Classifier correlation filters for multi-class automatic target recognition," Applied Optics, 35, 1996, 3127- 3133.
- [22] Face Database Available [ Online] <http://cswww.essex.ac.uk/mv/allfaces/faces94.html>.

# Analytical Comparison between the Information Gain and Gini Index using Historical Geographical Data

Dr. Majid Zaman<sup>1</sup>  
Directorate of IT & SS  
University of Kashmir  
Srinagar, India

Sameer Kaul<sup>2</sup>  
Department of Computer Science,  
University of Kashmir  
Srinagar, India

Dr. Muheet Ahmed<sup>3</sup>  
Department of Computer Science,  
University of Kashmir  
Srinagar, India

**Abstract**—The historical geographical data of Kashmir province is spread across two disparate files having attributes of Maximum Temperature, Minimum Temperature, Humidity measured at 12 A.M., Humidity measured at 3 P.M., rainfall besides auxiliary parameters like date, year etc. The parameters Maximum Temperature, Minimum Temperature, Humidity measured at 12 A.M., Humidity measured at 3 P.M. are continuous in nature and here, in this study, we applied Information Gain and Gini Index on these attributes to convert continuous data into discrete values, their after we compare and evaluate the generated results. Of the four attributes, two have same results for Information Gain and Gini Index; one attribute has overlapping results while as only one attribute has conflicting results for Information Gain and Gini Index. Subsequently, continuous valued attributes are converted into discrete values using Gini index. Irrelevant attributes are not considered and auxiliary attributes are labeled accordingly. Consequently, the data set is ready for the application of machine learning (decision tree) algorithms.

**Keywords**—Geographical data mining; information gain; Gini index; machine learning; decision tree

## I. INTRODUCTION

### A. Splitting Rules

Decision tree is built by recursively splitting data partitions into smaller partitions according to splitting rules or criteria. Attribute selection measure or splitting rules is a heuristic for choice of criteria that best splits class labeled training dataset into separate classes. Attribute selection measure should be such that split should produce pure partitions i.e. all the records in given partition belong to same class.

The attribute selection measure gives a score/value for each attribute, best describing given class labeled training dataset, the attribute having best score/value is chosen as splitting attribute for given partition. In this paper we have used Information Gain for the attribute selection measure.

### B. Information Gain and Gini Index

ID3 uses information gain as its attribute selection measure. For a given node that holds tuples of partition D, the attribute with highest information gain (score/value) is chosen as splitting attribute for the given node [1][6]. The chosen attribute requires least information for classifying records in the resultant partitions besides discloses least impurity in these partitions, thus resulting in minimum number of tests required to classify a given record and generation of (simple) decision

tree, accordingly information required for classification of a record in D is given by (1).

$$\text{Info}(D) = -\sum_{i=1}^m p_i \log_2(p_i) \quad [5] \quad (1)$$

and Information still required to arrive at an exact classification is measured by (2).

$$\text{Info}_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} * \text{Info}(D) \quad [5] \quad (2)$$

Information Gain is the difference between the original information requirement and the new requirement, that is

$$\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D) \quad [5] \quad (3)$$

Thus, Gain(A) is the gain if A is chosen for branching, accordingly Gain is calculated for all the attributes of the training set and attribute with the highest information gain is chosen as splitting attribute for the given node[2][3][7]. Thus calculation of information gain enables us to choose the attribute that would do the best classification, further most the amount of information still required for classifying records is minimal.

The Gini Index is used by CART. The Gini index measures the impurity in D[10][11]. The Gini index considers binary split for each attribute; accordingly weighted sum of impurity of each resulting partition is calculated, thus binary split on A partitions D into D1 & D2 i.e. [5].

$$\text{Gini}(D) = 1 - \sum_{i=1}^m p_i^2 \quad (4)$$

$$\text{and Gini}_A(D) = \frac{|D_1|}{|D|} \text{Gini}(D_1) + \frac{|D_2|}{|D|} \text{Gini}(D_2) \quad (5)$$

The reduction in impurity that would be incurred by a binary split on a discrete on attribute A is

$$\text{Gini}(A) = \text{Gini}(D) - \text{Gini}_A(D) \quad (6)$$

The process is repeated for every attribute and the attribute that has minimum Gini index is chosen as splitting attribute [2][3][8].

### C. Continuous Valued Attributes

For an attribute “A” that has continuous values e.g. temperature, humidity etc. the best split point is to be determined for “A”. All the possible unique values of A are sorted in ascending order, the midpoint between two adjacent values is considered [5].

$$\frac{ai+ai+1}{2} \quad (7)$$

for the given unique u values of attribute A, u-1 values will be generated, for each generated value infoA(D) is calculated with number of partitions two [4][9][12]. The mid-point with minimum value is chosen as the split point of A where

D1 is set of records satisfying

D2 is set of records satisfying

The other possible solution is to calculate Gini index for every mid-point (Gini index is calculated instead of infoA(D)) and minimum Gini index for a give attribute is taken as split point of the attribute.

## II. RELATED WORK

Gini index and Information gain have been used extensively used over the years, however most relevant work done in the recent past on the comparison of Gini index and Information gain is presented below.

In their research paper entitled “Theoretical comparison between the Gini Index and Information Gain criteria” Laura Elena Raileanu and Kilian Stoffel proposed a formal methodology to compare multiple split criteria and also presented a formal description of how to select between split criteria for a given data set, they concluded that Information Gain and Gini Index disagree only in 2% of all cases [13].

Mohammed A. Muharram and George D. Smith compared the performance of classifiers in their paper “Evolutionary Feature Construction Using Information Gain and Gini Index” to ascertain if C5 or CART was in any way benefiting from the inclusion of an attribute evolved using Information gain or Gini index respectively, they found no evidence that any algorithm has an advantage over the other classifiers and according to them all classifiers benefit from the inclusion of an evolved attribute [14].

Theoretical and empirical comparison of different split measures for induction of decision tree in Random forest and its effect on the accuracy of Random forest was done by Vrushali Y. Kulkarni, Manisha Petare and P. K. Sinha in their work entitled Analyzing Random Forest Classifier with Different Split Measures. The empirical results put forth by them, show that there is not much / significant variation in accuracy obtained except Chi Square, further Information gain and Gain ratio give comparable results for almost all datasets and Gini index slightly lags in the results with most of the datasets [15].

## III. DATA

The data used in this paper is split across two CSV files, which has been collected from NDC Pune (India Meteorological department), agency of Ministry of earth sciences, Government of India. It is the principal agency responsible for meteorological observations, weather forecasting and seismology. IMD is one of the six regional specialized meteorological centers of the world meteorological organization.

The weather parameters in both data files are taken for the 3 regions of Kashmir division i.e. Gulmarg (North Kashmir), Srinagar (Central Kashmir) and Qazigund (South Kashmir). Gulmarg is geographically located at 34.05°N 74.38°E and has an average elevation of 2,650 m (8,690 ft.), Srinagar (Central) is located at 34.5°N 74.47°E and has an average elevation of 1,585 m (5,200 ft.), and Qazigund (South) is located at 33.59°N 75.16°E. It has an average elevation of 1,670 m (5,480 ft.).

The first data file (Fig. 1), shown below consists of 12190 instances of relative humidity (in %) measured every day at time 12 AM and 3 PM from year 2012 to 2017, for all the three stations.

The second data file (Fig. 2), shown below consists of 6117 instances of Maximum temperature (°C), Minimum temperature (°C) and Rainfall (in mm) measured every day from year 2012 to 2017, for all the three stations.

The two data files are integrated into single holistic dataset, discrepancies are resolved, data for each attribute is cleaned, transformed and loaded for formation of single dataset, shown below (Fig 3). The integrated data has Maximum temperature (tmax), Minimum temperature (tmin) and Rainfall (rfall), humidity measured 12 AM (humid12) and 3 PM (humid3) for every day (with exception) from year 2012 to 2017, for all the three stations.

station_id	year	mnt	hr	dt	rhumid
42026	2012	1	3	1	100
42026	2012	1	3	2	100
42026	2012	1	3	3	96
42026	2012	1	3	4	100
42026	2012	1	3	5	100
42026	2012	1	3	6	100
42026	2012	1	3	7	100
42026	2012	1	3	8	100
42026	2012	1	3	9	100
42026	2012	1	3	10	86
42026	2012	1	3	11	87
42026	2012	1	3	12	100
42026	2012	1	3	13	100
42026	2012	1	3	14	100
42026	2012	1	3	15	100
42026	2012	1	3	16	100
42026	2012	1	3	17	100

Fig. 1. Instances of Relative Humidity at 12 am and 3pm.

station_id	year	mnt	dt	tmax	tmin	rfall
42026	2012	1	1	5.5	-8	0
42026	2012	1	2	5.4	-7.6	0
42026	2012	1	3	4.2	-8	0
42026	2012	1	4	4	-7.2	0
42026	2012	1	5	-1	-9.1	1.1
42026	2012	1	6	-2	-8	17.9
42026	2012	1	7	-1	-10.5	6.8
42026	2012	1	8	1	-16.5	12.6
42026	2012	1	9	-2.8	-14.5	0
42026	2012	1	10	-2.5	-16.2	0
42026	2012	1	11	-7.8	-14.8	0
42026	2012	1	12	-8.2	-16.4	0
42026	2012	1	13	-7.5	-16.5	0
42026	2012	1	14	-7.5	-15.2	0
42026	2012	1	15	-1.5	-9.6	16
42026	2012	1	16	-3	-6.7	21

Fig. 2. Instances of Maximum Temperature, Minimum Temperature and Rainfall.

station_id	year	mnth	dt	tmax	tmin	rfall	humid3	humid12
42026	2012	1	1	5.5	-8	0	100	100
42026	2012	1	2	5.4	-7.6	0	100	100
42026	2012	1	3	4.2	-8	0	96	90
42026	2012	1	4	4	-7.2	0	100	100
42026	2012	1	5	-1	-9.1	1.1	100	100
42026	2012	1	6	-2	-8	17.9	100	100
42026	2012	1	7	-1	-10.5	6.8	100	100
42026	2012	1	8	1	-16.5	12.6	100	100
42026	2012	1	9	-2.8	-14.5	0	100	83
42026	2012	1	10	-2.5	-16.2	0	86	94
42026	2012	1	11	-7.8	-14.8	0	87	100
42026	2012	1	12	-8.2	-16.4	0	100	100
42026	2012	1	13	-7.5	-16.5	0	100	100
42026	2012	1	14	-7.5	-15.2	0	100	100
42026	2012	1	15	-1.5	-9.6	16	100	100

Fig. 3. Cleaned and Integrated Dataset.

A. Data Attributes

Of the nine attributes five are geographical parameters, they are Maximum Temperature, Minimum Temperature, Rainfall, Humidity at 12 & Humidity at 3 termed as tmax, tmin, rfall, humid12 & humid3 respectively, while as four parameters are auxiliary/dependent parameters they are station id, year, month and date termed as station\_id, year, mnth & dt. In order to implement decision tree for the prediction of rainfall we have to evaluate each attribute of the resultant data independently.

1) *Rainfall*: As per the resultant dataset the rainfall in Kashmir province varies from no rainfall to above 100 mm of rainfall in one day. The broader inspection of rain data of five years recorded in 5951 entries is that there is no rainfall in 4026 instances and rainfall in 1925 instances, thus the inference is that we can divide rain data in to two classes that is presence and absence of rain, accordingly dataset is to be modified with new column “rfall” which will be marked as “Y” in case of rainfall (1925 entries) and “N” in case of no rainfall (4026 entries). The Decision Tree is trained to predict presence or absence of rain on a given day.

2) *Maximum temperature*: Maximum Temperature (tmax) is continuous valued rather than discrete valued, in this case we must determine the “best” split-point for Maximum Temperature (tmax), where the split-point is a threshold on Maximum Temperature (tmax), this can be determined by employing either of the two techniques, Information Gain used by ID3 or Gini Index used by CART, in this paper we use both the techniques to determine the split-point, we will compare the results from the two techniques (Information Gain & Gini Index) and decide accordingly. In order to calculate Information Gain or Gini index, we need to determine unique values of Maximum Temperature (tmax) and then these unique values are to be sorted in ascending order. In the dataset of 5951 records there are 380 unique values of Maximum Temperature (tmax) recoded, varying from -8.2°C to 35.4°C. Their after mid-point between each pair of adjacent values is considered as possible split-point., the snap shot of first 10, middle 10 and last 10 sorted records with mid points are shown in Fig. 4.

rno	tmax	spltptnt
1	-8.2	0
2	-7.8	-8
3	-7.6	-7.7
4	-7.5	-7.55
5	-6.7	-7.1
6	-5.5	-6.1
7	-5	-5.25
8	-4.5	-4.75
9	-4.4	-4.45
10	-4	-4.2
141	10.6	10.55
142	10.7	10.65
143	10.8	10.75
144	10.9	10.85
145	11	10.95
146	11.1	11.05
147	11.2	11.15
148	11.3	11.25
149	11.4	11.35
150	11.5	11.45
371	33.6	33.55
372	33.7	33.65
373	33.8	33.75
374	33.9	33.85
375	34	33.95
376	34.1	34.05
377	34.2	34.15
378	34.4	34.3
379	34.6	34.5
380	35.4	35

Fig. 4. Unique Values of Maximum Temperatures and their Split Points.

Therefore given 380 values of Maximum Temperature (tmax), 379 possible splits will be evaluated, accordingly there shall be no mid-point generated for first recoded temperature - 8.2°C because there is no prior temperature value. For example, the mid-point between the values of 33.8 and 33.9 of Maximum Temperature (rno 373 & 374) is 33.85, which is listed in the table for rno 374 against the value of 33.9.

$$\frac{33.8 + 33.9}{2} = 33.85$$

For each possible split-point for Maximum Temperature, we will evaluate  $Info_{tmax}(D)$  and  $Gini_{tmax}(D)$  but first we have to determine the prerequisites, for possible split value of 33.85 we have to determine the following:

- 1) fyes: No. of days there was rain for  $tmax \leq 33.85$
- 2) fno: No. of days there was no rain for  $tmax \leq 33.85$
- 3) syes: No. of days there was rain for  $tmax > 33.85$
- 4) sno: No. of days there was no rain for  $tmax > 33.85$

These values have to be generated for all possible split-points, the snap shot of first 10, middle 10 and the last 10 records with necessary values are shown below (Fig. 5).

Again first row shall not be considered because it has no mid-point, for every other possible point we have generated necessary values.

For each possible split-point for Maximum Temperature, we will calculate  $Info_{spltptnt}(D)$  and  $Gini_{spltptnt}(D)$  using following equations

$$Info(D) = -\sum_{i=1}^m p_i \log_2(p_i) \quad [5] \quad (8)$$

$$Info_{spltptnt}(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} * Info(D) \quad (9)$$

$$and \quad Gini(D) = 1 - \sum_{i=1}^m p_i^2 \quad (10)$$

$$Gini_{spltptnt}(D) = \frac{|D_1|}{|D|} Gini(D1) + \frac{|D_2|}{|D|} Gini(D2) \quad (11)$$

rno	tmax	spltpnt	fyes	fno	syas	sno
1	-8.2	0	0	0	0	0
2	-7.8	-8	0	1	1925	4025
3	-7.6	-7.7	0	2	1925	4024
4	-7.5	-7.55	0	3	1925	4023
5	-6.7	-7.1	0	5	1925	4021
6	-5.5	-6.1	0	6	1925	4020
7	-5	-5.25	1	7	1924	4019
8	-4.5	-4.75	2	7	1923	4019
9	-4.4	-4.45	3	7	1922	4019
10	-4	-4.2	3	8	1922	4018
141	10.6	10.55	647	858	1278	3168
142	10.7	10.65	653	872	1272	3154
143	10.8	10.75	655	876	1270	3150
144	10.9	10.85	665	887	1260	3139
145	11	10.95	666	896	1259	3130
146	11.1	11.05	674	922	1251	3104
147	11.2	11.15	674	933	1251	3093
148	11.3	11.25	678	951	1247	3075
149	11.4	11.35	678	965	1247	3061
150	11.5	11.45	681	983	1244	3043
371	33.6	33.55	1922	4005	3	21
372	33.7	33.65	1923	4008	2	18
373	33.8	33.75	1923	4009	2	17
374	33.9	33.85	1923	4013	2	13
375	34	33.95	1923	4017	2	9
376	34.1	34.05	1924	4019	1	7
377	34.2	34.15	1925	4021	0	5
378	34.4	34.3	1925	4022	0	4
379	34.6	34.5	1925	4024	0	2
380	35.4	35	1925	4025	0	1

Fig. 5. Possible Splitpoints for Maximum Temperature.

For example, we will generate Info(D) for a possible split-point of 10.85 listed in above table for rno 144 for tmax of 10.9.

$$Info_{spltpnt}(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} * Info(D) \tag{12}$$

$$Info_{10.9}(D) = \frac{1552}{5951} * \left( -\frac{665}{1552} * LOG2\left(\frac{665}{1552}\right) - \frac{887}{1552} * LOG2\left(\frac{887}{1552}\right) \right) + \frac{4399}{5951} * \left( -\frac{1260}{4399} * LOG2\left(\frac{1260}{4399}\right) - \frac{3139}{4399} * LOG2\left(\frac{3139}{4399}\right) \right)$$

$$Info_{10.9}(D) = 0.895652633$$

And we will generate Gini(D) for a possible split-point of 10.85 listed in above table for rno 144 for tmax of 10.9.

$$Gini_{spltpnt}(D) = \frac{|D_1|}{|D|} Gini(D1) + \frac{|D_2|}{|D|} Gini(D2) \tag{5}$$

$$Gini_{10.9}(D) = \frac{1552}{5951} * \left( 1 - \left(\frac{665}{1552}\right)^2 - \left(\frac{887}{1552}\right)^2 \right) + \frac{4399}{5951} * \left( 1 - \left(\frac{1260}{4399}\right)^2 - \left(\frac{3139}{4399}\right)^2 \right)$$

$$Gini_{10.9}(D) = 0.429897835$$

Likewise we generate Info(D) and Gini(D) for each possible split-point for Maximum Temperature, the snap shot of first 10, middle 10 and last 10 records with necessary values are shown in Fig. 6.

In this way we generate Info(D) and Gini(D) for every possible split-point, with exception to rno 1 because it has no

split point, further of 379 possible split-points 9 possible split-points do not generate info(D), show below (Fig. 7).

This is because one of the values of fyes, fno, syas, sno is zero. We have generated Information Gain and Gini Index for every split point; we now compare the two results.

Case 1: Information Gain

The point with minimum expected information requirement for Maximum Temperature (tmax) is to be selected as the split point for Maximum Temperature (tmax), the five best cases with minimum Information Gain are shown below (Fig. 8).

The above table is regenerated with Gini Index for the above split-points (Fig 9).

rno	tmax	spltpnt	fyes	fno	syas	sno	info	gini
1	-8.2	0	0	0	0	0	0	0
2	-7.8	-8	0	1	1925	4025	0	0.437643
3	-7.6	-7.7	0	2	1925	4024	0	0.437608
4	-7.5	-7.55	0	3	1925	4023	0	0.437572
5	-6.7	-7.1	0	5	1925	4021	0	0.437502
6	-5.5	-6.1	0	6	1925	4020	0	0.437467
7	-5	-5.25	1	7	1924	4019	0.907914	0.437572
8	-4.5	-4.75	2	7	1923	4019	0.908035	0.437647
9	-4.4	-4.45	3	7	1922	4019	0.908117	0.437676
10	-4	-4.2	3	8	1922	4018	0.908104	0.437668
141	10.6	10.55	647	858	1278	3168	0.895845	0.43001
142	10.7	10.65	653	872	1272	3154	0.896015	0.430121
143	10.8	10.75	655	876	1270	3150	0.896035	0.430134
144	10.9	10.85	665	887	1260	3139	0.895653	0.429898
145	11	10.95	666	896	1259	3130	0.896035	0.430141
146	11.1	11.05	674	922	1251	3104	0.896625	0.430519
147	11.2	11.15	674	933	1251	3093	0.897178	0.430868
148	11.3	11.25	678	951	1247	3075	0.897694	0.431196
149	11.4	11.35	678	965	1247	3061	0.898353	0.431611
150	11.5	11.45	681	983	1244	3043	0.898915	0.431966
371	33.6	33.55	1922	4005	3	21	0.9075	0.437359
372	33.7	33.65	1923	4008	2	18	0.907442	0.437341
373	33.8	33.75	1923	4009	2	17	0.907511	0.437373
374	33.9	33.85	1923	4013	2	13	0.907769	0.437495
375	34	33.95	1923	4017	2	9	0.907985	0.437604
376	34.1	34.05	1924	4019	1	7	0.907914	0.437572
377	34.2	34.15	1925	4021	0	5	0	0.437502
378	34.4	34.3	1925	4022	0	4	0	0.437537
379	34.6	34.5	1925	4024	0	2	0	0.437608
380	35.4	35	1925	4025	0	1	0	0.437643

Fig. 6. Information Gain and Gini for each Possible Split-Point for Maximum Temperature.

rno	tmax	spltpnt	fyes	fno	syas	sno	info	gini
2	-7.8	-8	0	1	1925	4025	0	0.437643
3	-7.6	-7.7	0	2	1925	4024	0	0.437608
4	-7.5	-7.55	0	3	1925	4023	0	0.437572
5	-6.7	-7.1	0	5	1925	4021	0	0.437502
6	-5.5	-6.1	0	6	1925	4020	0	0.437467
377	34.2	34.15	1925	4021	0	5	0	0.437502
378	34.4	34.3	1925	4022	0	4	0	0.437537
379	34.6	34.5	1925	4024	0	2	0	0.437608
380	35.4	35	1925	4025	0	1	0	0.437643

Fig. 7. Split-Points where Information Gain is not Generated for Maximum Temperature.

rno	spltpnt	info
286	25.05	0.891764
285	24.95	0.892011
284	24.85	0.892055
288	25.25	0.892174
283	24.75	0.892284

Fig. 8. Five Best Cases with Minimum Information Gain for Maximum Temperature.

rno	spltpnt	info	Gini
286	25.05	0.891764	0.428311
285	24.95	0.892011	0.428436
284	24.85	0.892055	0.428453
288	25.25	0.892174	0.428548
283	24.75	0.892284	0.428572

Fig. 9. Gini Index for each Respected Split-Point.

and in accordance to the rule of Information Gain we have to choose 25.05 as split-point for Maximum Temperature (tmax) since it has the lowest Information Gain, split-point 25.05 with all the attributes is shown below: (Fig. 10).

Case 2: Gini Index

The point giving the minimum Gini index for a given attribute Maximum Temperature (tmax) is to be taken as a split-point for the Maximum Temperature (tmax), the five best cases with minimum Gini Index are shown below: (Fig. 11).

The above table is regenerated with Information Gain for the above split-points (Fig. 12).

And in accordance to the rule we have to choose 8.05 as split-point for Maximum Temperature (tmax) since it has the lowest Gini Index, split-point 8.05 with all the attributes is shown below (Fig. 13).

rno	tmax	spltpnt	fyes	fno	syes	sno	info	gini
286	25.1	25.05	1624	2853	301	1173	0.891764	0.428311

Fig. 10. Split-Point with Lowest Information Gain for Maximum Temperature.

rno	spltpnt	info
116	8.05	0.893182
120	8.45	0.893214
113	7.75	0.893338
112	7.65	0.893355
121	8.55	0.893285

Fig. 11. Five Best Cases with Minimum Gini Index for Maximum Temperature.

rno	spltpnt	gini	info
116	8.05	0.428188	0.893182
120	8.45	0.42823	0.893214
113	7.75	0.428271	0.893338
112	7.65	0.42828	0.893355
121	8.55	0.428284	0.893285

Fig. 12. Information Gain for Each Respected Split-Point.

rno	tmax	spltpnt	fyes	fno	syes	sno	info	gini
116	8.1	8.05	494	551	1431	3475	0.893182	0.428188

Fig. 13. Split-Point with Lowest Gini Index for Maximum Temperature.

The results of Information Gain and Gini Index do not corroborate, and hence we have to choose one of the values, either as per Information Gain (25.05) or as per Gini Index (8.05).

3) *Minimum Temperature*: Minimum Temperature (tmin) is again continuous valued rather than discrete valued, in this case we must determine the “best” split-point for Minimum Temperature (tmin), where the split-point is a threshold on Minimum Temperature (tmin), again we use both the techniques to determine the split-point, we will compare the results from the two techniques (Information Gain & Gini Index) and decide accordingly.

We determine unique values of Minimum Temperature (tmin) and then these unique values are sorted in ascending order. In the dataset of 5951 records there are 354 unique values of Minimum Temperature (tmin) recorded, varying from -16.5°C to 23.8°C. Their after mid-point between each pair of adjacent values is generated as possible split-point., the snap shot of first 10, middle 10 and last 10 sorted records with mid points are shown below (Fig. 14).

Therefore given 354 values of Minimum Temperature (tmin), 353 possible splits will be generated and evaluated, there is no mid-point generated for the first minimum recorded temperature -16.5°C.

rno	tmin	spltpnt
1	-16.5	0
2	-16.4	-16.45
3	-16.2	-16.3
4	-15.2	-15.7
5	-14.8	-15
6	-14.6	-14.7
7	-14.5	-14.55
8	-14.4	-14.45
9	-14.2	-14.3
10	-14	-14.1
178	5.1	5.05
179	5.2	5.15
180	5.3	5.25
181	5.4	5.35
182	5.5	5.45
183	5.6	5.55
184	5.7	5.65
185	5.8	5.75
186	5.9	5.85
187	6	5.95
345	21.8	21.75
346	21.9	21.85
347	22	21.95
348	22.2	22.1
349	22.5	22.35
350	22.6	22.55
351	22.8	22.7
352	22.9	22.85
353	23.1	23
354	23.8	23.45

Fig. 14. Unique Values of Minimum Temperatures and their Split Points.

For each possible split-point for Minimum Temperature (tmin), we calculate values of fyes, fno, syes, and sno. These values have to be generated for all possible split-points, the snap shot of first 10, middle 10 and last 10 records with necessary values are shown below (Fig. 15).

Again first row shall not be considered because it has no mid-point, for every other possible point we have generated necessary values.

For each possible split-point for Minimum Temperature, we will calculate  $Info_{spltpnt}(D)$  and  $Gini_{spltpnt}(D)$  using following equations. (13)(14)(15)(16).

$$Info(D) = -\sum_{i=1}^m pi \log_2(pi) \tag{13}$$

$$Info_{spltpnt}(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} * Info(D) \tag{14}$$

$$\text{and } Gini(D) = 1 - \sum_{i=1}^m pi^2 \tag{15}$$

$$Gini_{spltpnt}(D) = \frac{|D_1|}{|D|} Gini(D1) + \frac{|D_2|}{|D|} Gini(D2) \tag{16}$$

The snap shot of first 10, middle 10 and last 10 records with necessary values are shown in Fig. 16.

We generate Info(D) and Gini(D) for every possible split-point with exception to rno 1 because it has no split point, further of 353 possible split-points 12 possible split-points do not generate info(D), this is because one of the values of fyes, fno, syes, sno is zero, as shown in Fig. 17.

We have generated Information Gain and Gini Index for every split point; we now compare the two results.

rno	tmin	spltpnt	fyes	fno	syes	sno
1	-16.5	0	0	0	0	0
2	-16.4	-16.45	1	1	1924	4025
3	-16.2	-16.3	1	2	1924	4024
4	-15.2	-15.7	1	4	1924	4022
5	-14.8	-15	1	5	1924	4021
6	-14.6	-14.7	1	6	1924	4020
7	-14.5	-14.55	3	6	1922	4020
8	-14.4	-14.45	3	7	1922	4019
9	-14.2	-14.3	3	8	1922	4018
10	-14	-14.1	3	9	1922	4017
178	5.1	5.05	837	1978	1088	2048
179	5.2	5.15	839	1981	1086	2045
180	5.3	5.25	848	2004	1077	2022
181	5.4	5.35	853	2011	1072	2015
182	5.5	5.45	866	2027	1059	1999
183	5.6	5.55	871	2038	1054	1988
184	5.7	5.65	881	2057	1044	1969
185	5.8	5.75	888	2073	1037	1953
186	5.9	5.85	897	2093	1028	1933
187	6	5.95	900	2098	1025	1928
345	21.8	21.75	1925	4002	0	24
346	21.9	21.85	1925	4007	0	19
347	22	21.95	1925	4013	0	13
348	22.2	22.1	1925	4014	0	12
349	22.5	22.35	1925	4018	0	8
350	22.6	22.55	1925	4020	0	6
351	22.8	22.7	1925	4021	0	5
352	22.9	22.85	1925	4022	0	4
353	23.1	23	1925	4024	0	2
354	23.8	23.45	1925	4025	0	1

Fig. 15. Possible Splitpoints for Minimum Temperature.

rno	tmin	spltpnt	fyes	fno	syes	sno	info	gini
1	-16.5	0	0	0	0	0	0	0
2	-16.4	-16.45	1	1	1924	4025	0.908088	0.437657
3	-16.2	-16.3	1	2	1924	4024	0.90812	0.437678
4	-15.2	-15.7	1	4	1924	4022	0.908074	0.437652
5	-14.8	-15	1	5	1924	4021	0.908028	0.437628
6	-14.6	-14.7	1	6	1924	4020	0.907974	0.437601
7	-14.5	-14.55	3	6	1922	4020	0.90812	0.437678
8	-14.4	-14.45	3	7	1922	4019	0.908117	0.437676
9	-14.2	-14.3	3	8	1922	4018	0.908104	0.437668
10	-14	-14.1	3	9	1922	4017	0.908083	0.437656
178	5.1	5.05	837	1978	1088	2048	0.906094	0.436451
179	5.2	5.15	839	1981	1086	2045	0.906116	0.436464
180	5.3	5.25	848	2004	1077	2022	0.906044	0.43642
181	5.4	5.35	853	2011	1072	2015	0.906107	0.436458
182	5.5	5.45	866	2027	1059	1999	0.906302	0.436576
183	5.6	5.55	871	2038	1054	1988	0.906294	0.436571
184	5.7	5.65	881	2057	1044	1969	0.906327	0.436591
185	5.8	5.75	888	2073	1037	1953	0.906305	0.436577
186	5.9	5.85	897	2093	1028	1933	0.906285	0.436565
187	6	5.95	900	2098	1025	1928	0.906307	0.436578
345	21.8	21.75	1925	4002	0	24	0	0.43683
346	21.9	21.85	1925	4007	0	19	0	0.437008
347	22	21.95	1925	4013	0	13	0	0.43722
348	22.2	22.1	1925	4014	0	12	0	0.437255
349	22.5	22.35	1925	4018	0	8	0	0.437396
350	22.6	22.55	1925	4020	0	6	0	0.437467
351	22.8	22.7	1925	4021	0	5	0	0.437502
352	22.9	22.85	1925	4022	0	4	0	0.437537
353	23.1	23	1925	4024	0	2	0	0.437608
354	23.8	23.45	1925	4025	0	1	0	0.437643

Fig. 16. Information Gain and Gini for each Possible Split-Point for Minimum Temperature.

rno	tmin	spltpnt	fyes	fno	syes	sno	info	gini
343	21.6	21.55	1925	3994	0	32	0	0.436546
344	21.7	21.65	1925	3997	0	29	0	0.436653
345	21.8	21.75	1925	4002	0	24	0	0.43683
346	21.9	21.85	1925	4007	0	19	0	0.437008
347	22	21.95	1925	4013	0	13	0	0.43722
348	22.2	22.1	1925	4014	0	12	0	0.437255
349	22.5	22.35	1925	4018	0	8	0	0.437396
350	22.6	22.55	1925	4020	0	6	0	0.437467
351	22.8	22.7	1925	4021	0	5	0	0.437502
352	22.9	22.85	1925	4022	0	4	0	0.437537
353	23.1	23	1925	4024	0	2	0	0.437608
354	23.8	23.45	1925	4025	0	1	0	0.437643

Fig. 17. Split-Points where Information Gain is not Generated for Minimum Temperature.

Case 1: Information Gain

The point with minimum expected information requirement for Minimum Temperature (tmin) is to be selected as the split point for Minimum Temperature (tmin), the five best cases with minimum Information Gain are shown below: (Fig 18).

The above table is regenerated with Gini Index for the split-points (Fig. 19).

rno	spltpnt	info
124	-0.35	0.900033
125	-0.25	0.900106
123	-0.45	0.900434
122	-0.55	0.900468
120	-0.75	0.900554

Fig. 18. Five Best cases with Minimum Information Gain for Minimum Temperature.

rno	spltpnt	info	gini
124	-0.35	0.900033	0.432954
125	-0.25	0.900106	0.432992
123	-0.45	0.900434	0.43319
122	-0.55	0.900468	0.433212
120	-0.75	0.900554	0.433268

Fig. 19. Gini Index for each Respected Split-Point.

And in accordance to the rule of Information Gain we have to choose -0.35 as split-point for Minimum Temperature (tmin) since it has the lowest Information Gain, split-point -0.35 with all the attributes is shown below: (Fig 20).

Case 2: Gini Index

The point giving the minimum Gini index for a given attribute Minimum Temperature (tmin) is to be taken as a split-point for the Minimum Temperature (tmin), the five best cases with minimum Gini Index are shown below: (Fig. 21).

The table is regenerated with Information Gain for the above split-points: (Fig. 22).

And in accordance to the rule of Gini Index we have to choose -0.35 as split-point for Minimum Temperature (tmin) since it has the lowest Gini Index, split-point -0.35 with all the attributes is shown below: (Fig. 23).

The results of Information Gain and Gini Index are exactly the same, hence split-point -0.35 will be chosen in either case, and there is no conflict at all.

rno	tmin	spltpnt	fyes	fno	syas	sno	info	gini
124	-0.3	-0.35	349	1115	1576	2911	0.900033	0.432954

Fig. 20. Split-Point with Lowest Information Gain for Minimum Temperature.

rno	spltpnt	gini
124	-0.35	0.432954
125	-0.25	0.432992
123	-0.45	0.43319
122	-0.55	0.433212
120	-0.75	0.433268

Fig. 21. Five best cases with Minimum Gini Index for Minimum Temperature.

rno	spltpnt	gini	info
124	-0.35	0.432954	0.900033
125	-0.25	0.432992	0.900106
123	-0.45	0.43319	0.900434
122	-0.55	0.433212	0.900468
120	-0.75	0.433268	0.900554

Fig. 22. Information Gain for each Respected Split-Point.

rno	tmin	spltpnt	fyes	fno	syas	sno	info	gini
124	-0.3	-0.35	349	1115	1576	2911	0.900033	0.432954

Fig. 23. Split-point with lowest Gini Index for Minimum Temperature.

4) Humidity Measured at 12:00 A.M: Like Maximum Temperature (tmax) & Minimum Temperature (tmin) Humidity Measured at 12:00 A.M (humid12) is continuous valued rather than discrete valued, and in accordance with the methodology used for the determination of best split-point for maximum and minimum temperature, we use same procedure for determination of best split-point for humidity12 as well. In the dataset of 5951 records there are 82 unique values of Humidity Measured at 12:00 A.M (humid12) recoded, varying from 18 to 100. The snap shot of first 10, middle 10 and last 10-sorted records with mid points are shown below (Fig. 24), 81 possible split-points will be evaluated.

The snap shot of first 10, middle 10 and last 10 records with necessary values of fyes, fno, syes & sno are shown below (Fig. 25).

rno	humid12	spltpnt
1	18	0
2	19	18.5
3	20	19.5
4	21	20.5
5	22	21.5
6	23	22.5
7	24	23.5
8	25	24.5
9	26	25.5
10	27	26.5
42	59	58.5
43	60	59.5
44	61	60.5
45	62	61.5
46	63	62.5
47	64	63.5
48	65	64.5
49	66	65.5
50	67	66.5
51	68	67.5
73	90	89.5
74	91	90.5
75	92	91.5
76	93	92.5
77	94	93.5
78	95	94.5
79	96	95.5
80	97	96.5
81	98	97.5
82	100	99

Fig. 24. Unique Values of Humidity at 12 am and their Split Points.

rno	humid12	spltpnt	fyes	fno	syas	sno
1	18	0	0	0	0	0
2	19	18.5	0	1	1925	4025
3	20	19.5	0	5	1925	4021
4	21	20.5	0	7	1925	4019
5	22	21.5	0	14	1925	4012
6	23	22.5	0	23	1925	4003
7	24	23.5	0	37	1925	3989
8	25	24.5	0	47	1925	3979
9	26	25.5	1	57	1924	3969
10	27	26.5	1	75	1924	3951
42	59	58.5	452	2341	1473	1685
43	60	59.5	483	2437	1442	1589
44	61	60.5	516	2543	1409	1483
45	62	61.5	551	2609	1374	1417
46	63	62.5	583	2699	1342	1327
47	64	63.5	600	2755	1325	1271
48	65	64.5	630	2817	1295	1209
49	66	65.5	653	2880	1272	1146
50	67	66.5	685	2944	1240	1082
51	68	67.5	711	3003	1214	1023
73	90	89.5	1406	3802	519	224
74	91	90.5	1437	3825	488	201
75	92	91.5	1493	3852	432	174
76	93	92.5	1529	3877	396	149
77	94	93.5	1578	3894	347	132
78	95	94.5	1622	3907	303	119
79	96	95.5	1643	3917	282	109
80	97	96.5	1682	3944	243	82
81	98	97.5	1731	3955	194	71
82	100	99	1741	3957	184	69

Fig. 25. Possible Splitpoints for Humidity at 12 am.

For each possible split-point for Minimum Temperature, we will calculate  $Info_{splitpnt}(D)$  and  $Gini_{splitpnt}(D)$  using following equations.(17)(18)(19)(20).

$$Info(D) = - \sum_{i=1}^m pi \log_2(pi) \tag{17}$$

$$Info_{splitpnt}(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} * Info(D) \tag{18}$$

$$\text{and } Gini(D) = 1 - \sum_{i=1}^m pi^2 \tag{19}$$

$$Gini_{splitpnt}(D) = \frac{|D_1|}{|D|} Gini(D1) + \frac{|D_2|}{|D|} Gini(D2) \tag{20}$$

The snap shot of first 10, middle 10 and last 10 records with Information Gain & Gini Index values are shown below (Fig. 26).

We generate Info(D) and Gini(D) for every possible split-point with exception to rno 1 because it has no split point, further of 81 possible split-points 8 possible split-points do not generate info(D), this is because one of the values of fyes, fno, syes, sno is zero, as shown below (Fig. 27).

We have generated Information Gain and Gini Index for every split point; we now compare the two results.

Case 1: Information Gain

The point with minimum expected information requirement for Humidity Measured at 12:00 A.M (humid12) is to be selected as the split point; the five best cases with minimum Information Gain are shown in Fig. 28.

rno	humid12	spltpnt	fyes	fno	syes	sno	info	gini
1	18	0	0	0	0	0	0	0
2	19	18.5	0	1	1925	4025	0	0.437643
3	20	19.5	0	5	1925	4021	0	0.437502
4	21	20.5	0	7	1925	4019	0	0.437431
5	22	21.5	0	14	1925	4012	0	0.437184
6	23	22.5	0	23	1925	4003	0	0.436866
7	24	23.5	0	37	1925	3989	0	0.436369
8	25	24.5	0	47	1925	3979	0	0.436012
9	26	25.5	1	57	1924	3969	0.903642	0.435832
10	27	26.5	1	75	1924	3951	0.90198	0.435186
42	59	58.5	452	2341	1473	1685	0.828693	0.391461
43	60	59.5	483	2437	1442	1589	0.825976	0.389539
44	61	60.5	516	2543	1409	1483	0.822266	0.386989
45	62	61.5	551	2609	1374	1417	0.823413	0.387333
46	63	62.5	583	2699	1342	1327	0.820683	0.38537
47	64	63.5	600	2755	1325	1271	0.818063	0.383605
48	65	64.5	630	2817	1295	1209	0.817818	0.383169
49	66	65.5	653	2880	1272	1146	0.815476	0.381504
50	67	66.5	685	2944	1240	1082	0.815059	0.380949
51	68	67.5	711	3003	1214	1023	0.813581	0.379789
73	90	89.5	1406	3802	519	224	0.846623	0.397544
74	91	90.5	1437	3825	488	201	0.848767	0.398902
75	92	91.5	1493	3852	432	174	0.855596	0.403295
76	93	92.5	1529	3877	396	149	0.858094	0.404911
77	94	93.5	1578	3894	347	132	0.865226	0.409533
78	95	94.5	1622	3907	303	119	0.87197	0.413917
79	96	95.5	1643	3917	282	109	0.874283	0.415426
80	97	96.5	1682	3944	243	82	0.876471	0.416886
81	98	97.5	1731	3955	194	71	0.884496	0.422116
82	100	99	1741	3957	184	69	0.886153	0.423198

Fig. 26. Information Gain and Gini for each Possible Split-Point for Humidity at 12 am.

rno	humid12	spltpnt	fyes	fno	syes	sno	info	gini
2	19	18.5	0	1	1925	4025	0	0.437643
3	20	19.5	0	5	1925	4021	0	0.437502
4	21	20.5	0	7	1925	4019	0	0.437431
5	22	21.5	0	14	1925	4012	0	0.437184
6	23	22.5	0	23	1925	4003	0	0.436866
7	24	23.5	0	37	1925	3989	0	0.436369
8	25	24.5	0	47	1925	3979	0	0.436012

Fig. 27. Split-Points where Information Gain is not Generated for Humidity at 12 am.

rno	spltpnt	info
53	69.5	0.809583
54	70.5	0.810877
52	68.5	0.810984
55	71.5	0.812556
57	73.5	0.812772

Fig. 28. Five Best cases with Minimum Information Gain for Humidity at 12 am.

The above table is regenerated with Gini Index for the above split-points (Fig 29).

And in accordance to the rule of Information Gain we have to choose 69.5 as split-point for Humidity Measured at 12:00 A.M (humid12) since it has the lowest Information Gain, split-point 69.5 with all the attributes is shown below: (Fig. 30).

Case 2: Gini Index

The point giving the minimum Gini index for a given attribute Humidity Measured at 12:00 A.M (humid12) is to be taken as a split-point; the five best cases with minimum Gini Index are shown below: (Fig 31).

The above table is regenerated with Information Gain for the above split-points (Fig. 32).

And in accordance to the rule of Gini Index we have to choose 69.5 as split-point for Humidity Measured at 12:00 point 69.5 with all the attributes is shown below (Fig. 33).

rno	spltpnt	info	gini
53	69.5	0.809583	0.37666
54	70.5	0.810877	0.377227
52	68.5	0.810984	0.37783
55	71.5	0.812556	0.378023
57	73.5	0.812772	0.37766

Fig. 29. Gini Index for each Respected Split-Point.

rno	humid12	spltpnt	fyes	fno	syes	sno	info	gini
53	70	69.5	779	3151	1146	875	0.809583	0.37666

Fig. 30. Split-Point with Lowest Information Gain for Humidity at 12 am.

rno	spltpnt	gini
53	69.5	0.37666
54	70.5	0.377227
57	73.5	0.37766
52	68.5	0.37783
55	71.5	0.378023

Fig. 31. Five Best Cases with Minimum Gini Index for Humidity at 12 am.

rno	spltpnt	gini	info
53	69.5	0.37666	0.809583
54	70.5	0.377227	0.810877
57	73.5	0.37766	0.812772
52	68.5	0.37783	0.810984
55	71.5	0.378023	0.812556

Fig. 32. Information Gain for each Respected Split-Point.

rno	humid12	spltptnt	fyes	fno	syes	sno	info	gini
53	70	69.5	779	3151	1146	875	0.809583	0.37666

Fig. 33. Split-Point with Lowest Gini Index for Humidity at 12 am.

The results of Information Gain and Gini Index are exactly the same, hence split-point 69.5 will be chosen in either case, and there is no conflict at all.

5) *Humidity Measured at 03:00 P.M.*: Like the earlier three cases Humidity Measured at 03:00 P.M (humid3) is also continuous valued rather than discrete valued, and accordingly best split- point for humidity3 is generated and evaluated as well.

In the dataset of 5951 records there are 80 unique values of Humidity Measured at 03:00 P.M (humid3) recoded, varying from 16 to 100. The snap shot of first 10, middle 10 and last 10-sorted records with mid points are shown below (Fig 34), 79 possible split-points will be evaluated.

The snap shot of first 10, middle 10 and last 10 records with necessary values of fyes, fno, syes & sno are shown in Fig 35.

rno	humid3	spltptnt
1	16	0
2	17	16.5
3	20	18.5
4	22	21
5	24	23
6	25	24.5
7	26	25.5
8	27	26.5
9	28	27.5
10	30	29
41	61	60.5
42	62	61.5
43	63	62.5
44	64	63.5
45	65	64.5
46	66	65.5
47	67	66.5
48	68	67.5
49	69	68.5
50	70	69.5
71	91	90.5
72	92	91.5
73	93	92.5
74	94	93.5
75	95	94.5
76	96	95.5
77	97	96.5
78	98	97.5
79	99	98.5
80	100	99.5

Fig. 34. Unique Values of Humidity at 3 pm and their Split Points.

rno	humid3	spltptnt	fyes	fno	syes	sno
1	16	0	0	0	0	0
2	17	16.5	0	2	1925	4024
3	20	18.5	0	3	1925	4023
4	22	21	0	6	1925	4020
5	24	23	0	7	1925	4019
6	25	24.5	0	9	1925	4017
7	26	25.5	0	10	1925	4016
8	27	26.5	0	11	1925	4015
9	28	27.5	0	14	1925	4012
10	30	29	0	15	1925	4011
41	61	60.5	54	752	1871	3274
42	62	61.5	64	813	1861	3213
43	63	62.5	76	890	1849	3136
44	64	63.5	93	953	1832	3073
45	65	64.5	113	1036	1812	2990
46	66	65.5	126	1128	1799	2898
47	67	66.5	147	1222	1778	2804
48	68	67.5	168	1314	1757	2712
49	69	68.5	177	1420	1748	2606
50	70	69.5	211	1523	1714	2503
71	91	90.5	1101	3522	824	504
72	92	91.5	1193	3593	732	433
73	93	92.5	1272	3690	653	336
74	94	93.5	1371	3767	554	259
75	95	94.5	1468	3817	457	209
76	96	95.5	1522	3853	403	173
77	97	96.5	1603	3929	322	97
78	98	97.5	1710	3946	215	80
79	99	98.5	1730	3949	195	77
80	100	99.5	1731	3949	194	77

Fig. 35. Possible Splitpoints for Humidity at 3 pm.

For each possible split-point for Minimum Temperature, we will calculate  $Info_{spltptnt}(D)$  and  $Gini_{spltptnt}(D)$  using following equations (21)(22)(23)(24).

$$Info(D) = -\sum_{i=1}^m p_i \log_2(p_i) \tag{21}$$

$$Info_{spltptnt}(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} * Info(D) \tag{22}$$

$$\text{and } Gini(D) = 1 - \sum_{i=1}^m p_i^2 \tag{23}$$

$$Gini_{spltptnt}(D) = \frac{|D_1|}{|D|} Gini(D1) + \frac{|D_2|}{|D|} Gini(D2) \tag{24}$$

The snap shot of first 10, middle 10 and last 10 records with Information Gain & Gini Index values are shown below: (Fig. 36).

We generate Info(D) and Gini(D) for every possible split-point with exception to rno 1 because it has no split point, further of 79 possible split-points 15 possible split-points do not generate info(D), this is because one of the values of fyes, fno, syes, sno is zero, as shown below: (Fig. 37).

We have generated Information Gain and Gini Index for every split point; we now compare the two results.

Case 1: Information Gain

The point with minimum expected information requirement for Humidity Measured at 03:00 P.M (humid3) is to be selected as the split point for Humidity Measured at 03:00 P.M (humid3) the five best cases with minimum Information Gain are shown below: (Fig 38).

rno	humid3	spltpnt	fyes	fno	syes	sno	info	gini
1	16	0	0	0	0	0	0	0
2	17	16.5	0	2	1925	4024	0	0.437608
3	20	18.5	0	3	1925	4023	0	0.437572
4	22	21	0	6	1925	4020	0	0.437467
5	24	23	0	7	1925	4019	0	0.437431
6	25	24.5	0	9	1925	4017	0	0.437361
7	26	25.5	0	10	1925	4016	0	0.437326
8	27	26.5	0	11	1925	4015	0	0.43729
9	28	27.5	0	14	1925	4012	0	0.437184
10	30	29	0	15	1925	4011	0	0.437149
41	61	60.5	54	752	1871	3274	0.865622	0.417068
42	62	61.5	64	813	1861	3213	0.863979	0.415986
43	63	62.5	76	890	1849	3136	0.861462	0.414452
44	64	63.5	93	953	1832	3073	0.861832	0.414211
45	65	64.5	113	1036	1812	2990	0.86187	0.413424
46	66	65.5	126	1128	1799	2898	0.856965	0.411126
47	67	66.5	147	1222	1778	2804	0.855041	0.409773
48	68	67.5	168	1314	1757	2712	0.853037	0.408397
49	69	68.5	177	1420	1748	2606	0.845842	0.404508
50	70	69.5	211	1523	1714	2503	0.84627	0.404191
71	91	90.5	1101	3522	824	504	0.828948	0.386998
72	92	91.5	1193	3593	732	433	0.837883	0.392434
73	93	92.5	1272	3690	653	336	0.838367	0.392463
74	94	93.5	1371	3767	554	259	0.845895	0.397129
75	95	94.5	1468	3817	457	209	0.857442	0.40452
76	96	95.5	1522	3853	403	173	0.861853	0.407349
77	97	96.5	1603	3929	322	97	0.86224	0.407678
78	98	97.5	1710	3946	215	80	0.882094	0.420539
79	99	98.5	1730	3949	195	77	0.885633	0.42285
80	100	99.5	1731	3949	194	77	0.88584	0.422986

Fig. 36. Information Gain and Gini for each Possible Split-Point for Humidity at 3 pm.

rno	humid3	spltpnt	fyes	fno	syes	sno	info	gini
2	17	16.5	0	2	1925	4024	0	0.437608
3	20	18.5	0	3	1925	4023	0	0.437572
4	22	21	0	6	1925	4020	0	0.437467
5	24	23	0	7	1925	4019	0	0.437431
6	25	24.5	0	9	1925	4017	0	0.437361
7	26	25.5	0	10	1925	4016	0	0.437326
8	27	26.5	0	11	1925	4015	0	0.43729
9	28	27.5	0	14	1925	4012	0	0.437184
10	30	29	0	15	1925	4011	0	0.437149
11	31	30.5	0	20	1925	4006	0	0.436972
12	32	31.5	0	22	1925	4004	0	0.436901
13	33	32.5	0	25	1925	4001	0	0.436795
14	34	33.5	0	28	1925	3998	0	0.436689
15	35	34.5	0	40	1925	3986	0	0.436262
16	36	35.5	0	53	1925	3973	0	0.435797

Fig. 37. Split-Points where Information Gain is not Generated for Humidity at 3 pm.

rno	spltpnt	info
63	82.5	0.817457
70	89.5	0.818059
67	86.5	0.819001
64	83.5	0.819493
68	87.5	0.819696

Fig. 38. Five Best cases with Minimum Information Gain for Humidity at 3 pm.

The above table is regenerated with Gini Index for the above split-points (Fig. 39).

And in accordance to the rule of Information Gain we have to choose 82.5 as split-point for Humidity Measured at 03:00 P.M (humid3) since it has the lowest Information Gain, split-point 82.5 with all the attributes is shown in Fig. 40.

rno	spltpnt	info	gini
63	82.5	0.817457	0.38305
70	89.5	0.818059	0.380388
67	86.5	0.819001	0.382286
64	83.5	0.819493	0.38386
68	87.5	0.819696	0.382274

Fig. 39. Gini Index for each Respected Split-Point.

rno	humid3	spltpnt	fyes	fno	syes	sno	info	gini
63	83	82.5	619	2798	1306	1228	0.817457	0.38305

Fig. 40. Split-Point with Lowest Information Gain for Humidity at 3 pm.

### Case 2: Gini Index

The point giving the minimum Gini index for a given attribute Humidity Measured at 03:00 P.M (humid3) is to be taken as a split-point for the Humidity Measured at 03:00 P.M (humid3) the five best cases with minimum Gini Index are shown below: (Fig. 41).

The above table is regenerated with Information Gain for the above split-points (Fig. 42).

And in accordance to the rule of Gini Index we have to choose 89.5 as split-point for Humidity Measured 03:00 P.M (humid3) since it has the lowest Gini Index, split-point 89.5 with all the attributes is shown below (Fig. 43).

As per Information Gain choice of split-point is 82.5, while as per the choice of Gini Index the split-point is 89.5. In order to make decision on the choice of split-point we compare the two generated list, as shown below (Fig. 44).

rno	spltpnt	gini
70	89.5	0.380388
68	87.5	0.382274
67	86.5	0.382286
69	88.5	0.382833
63	82.5	0.38305

Fig. 41. Five best cases with Minimum Gini Index for Humidity at 3 pm

rno	spltpnt	gini	info
70	89.5	0.380388	0.818059
68	87.5	0.382274	0.819696
67	86.5	0.382286	0.819001
69	88.5	0.382833	0.821271
63	82.5	0.38305	0.817457

Fig. 42. Information Gain for each Respected Split-Point.

rno	humid3	spltpnt	fyes	fno	syes	sno	info	gini
70	90	89.5	1013	3463	912	563	0.818059	0.380388

Fig. 43. Split-Point with Lowest Gini Index for Humidity at 3 pm.

Information Gain			VS	Gini Index		
rno	spltpt	info	rno	spltpt	gini	
63	82.5	0.817457	70	89.5	0.380388	
70	89.5	0.818059	68	87.5	0.382274	
67	86.5	0.819001	67	86.5	0.382286	
64	83.5	0.819493	69	88.5	0.382833	
68	87.5	0.819696	63	82.5	0.38305	

Fig. 44. Comparison between Information Gain and Gini Index.

From the comparison shown above, there is a visible overlap between the two results, we choose 89.5 as split-point for Humidity Measured at 03:00 P.M (humid3), because it is first choice as per Gini Index and it is second choice of Information Gain.

**B. Evaluation -- Information Gain vs. Gini Index**

Four attributes are continuous valued rather than discrete valued, we employed Information Gain used by ID3 and Gini Index used by CART to determine best possible split-point, the results are shown below (Table I).

TABLE I. BEST POSSIBLE SPLITS USING ID3 AND CART

Attribute	Information Gain	Gini Index	Class One	Class Two
TMAX	25.05	8.05	8.05<= is H1	>8.05 is H2
TMIN	-0.35	-0.35	-0.35<= is L1	>-0.35 is L2
HUMID12	69.5	69.5	69.5<= is T1	>69.5 is T2
HUMID3	82.5	89.5	89.5<= is U1	>89.5 is U2

Of the four attributes, Tmin and Humid12 have same results for Information Gain and Gini Index. Humid3 has overlapping results for Information Gain and Gini Index, as already discussed we choose 89.5 as split-point for Humid3. It is the attribute Tmax where the results of Information Gain and Gini Index do not corroborate, and hence we have to choose one of the values, either as per Information Gain (25.05) or as per Gini Index (8.05). We chose Gini Index over Information Gain primarily because the split-point of three attributes (Tmin, Humid12, Humid3) is as per Gini Index while as split point of two attributes (Tmin & Humid12) is as per Information Gain, thus we choose to go with the majority i.e. Gini Index over Information Gain accordingly split-point of Tmax is 8.05.

**C. Rest of Data Attributes**

Off the rest of the data attributes, Station\_id, Year, Month and date, we decide not to consider recording station (Station\_id) as part of decision tree for prediction of rainfall, since all the stations belong to the same province. Further, a year is 365 days or 12 month or 4 seasons, thus we split the months into season as shown below: (Table II).

TABLE II. SPLITTING MONTHS IN RESPECTED SEASONS

Months	Season
12, 1, 2	Winter
3, 4, 5	Spring
6, 7, 8	Summer
9, 10, 11	Autumn

Thus we use seasons instead of months, and decide not to use year and date as part of decision table, this will also maximize information dissemination.

1) *Resultant Dataset*: Consequent upon conversion of continuous valued attributes into discrete valued and conversion of months into seasons besides not considering some irrelevant attributes, the snapshot of the resultant dataset is shown below: (Fig. 45).

season	ctmax	ctmin	chumid12	chumid3	crfall
spring	H1	L1	T2	U1	N
spring	H1	L1	T2	U2	Y
spring	H2	L1	T1	U1	N
spring	H2	L2	T1	U1	N
spring	H2	L2	T2	U1	Y
spring	H2	L2	T2	U2	Y
spring	H2	L2	T2	U1	Y
spring	H2	L2	T2	U2	Y
spring	H1	L2	T2	U2	Y
spring	H2	L2	T2	U1	Y
spring	H2	L2	T2	U2	Y
spring	H1	L2	T1	U1	Y
spring	H2	L1	T2	U1	Y
spring	H2	L2	T2	U2	Y
spring	H1	L1	T2	U2	Y
spring	H1	L1	T2	U1	Y
spring	H1	L1	T2	U2	Y
spring	H2	L1	T1	U1	N
spring	H2	L2	T1	U1	N

Fig. 45. Labelled Resultant Dataset.

Where

Ctmax = H1 if tmax <= 8.05

Ctmax = H2 if tmax > 8.05

Ctmin = L1 if tmin <= -0.35

Ctmin = L2 if tmin > -0.35

Chumid12 = T1 if humid12 <= 69.5

Chumid12 = T2 if humid12 > 69.5

Chumid3 = U1 if humid3 <= 89.5

Chumid3 = U2 if humid3 > 89.5

Further months have been converted into seasons as per the table shown above and crfall is Y if rfall >0 and crfall is N if rfall =0.

**IV. CONCLUSION AND FUTURE WORK**

In this paper two techniques are employed i.e. Information Gain and Gini index to convert continuous data into discrete valued data. This is preliminary and prerequisite step in order to apply machine learning algorithm Decision tree on the geographical data set. Besides having prepared historical geographical data for the application of Decision tree algorithm we have also compared the results from two varying techniques applied on the same dataset.

Whilst this study was primarily aimed at the comparison of Information Gain and Gini index, a fuller work is underway in which two separate dataset shall be generated on the basis of Information Gain and Gini index thereafter decision tree

algorithms shall be employed on these two generated data sets this will enable us to compare the performance of Information Gain and Gini index at the individual level of implementation.

#### REFERENCES

- [1] Han, J., Kamber, M.: Data Mining Concepts and Techniques. China Machine Press, Beijing (2007).
- [2] Zhang Quancheng, You Kun, Ma, Gang. Application of ID3 Algorithm in Exercise Prescription[C]. The International Conference on Electric and Electronics, Nanchang, China, June 22, 2011. 99(3): 669-675.
- [3] LI Shoubang. Application Study on Mining of University Students' Physical Fitness Test Data Based on Classification Rules: Taking the Juniors of Xi'an Shiyou University as an Example [J]. Journal of Xi'an Shiyou University (Natural Science Edition), 2018, 33(5) : 120- 126.
- [4] J.R. Quinlan. Improved use of continuous attributes in c4.5. Journal of Artificial Intelligence Approach, (4):77-90, 1996. 211, 212, 216.
- [5] Data Mining: Concepts and Techniques, 3rd Edition Jiawei Han, Micheline Kamber, Jian Pei.
- [6] Ashraf, Mudasir, Majid Zaman, and Muheet Ahmed. "Performance analysis and different subject combinations: An empirical and analytical discourse of educational data mining." 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2018.
- [7] Ashraf, Mudasir, Majid Zaman, and Muheet Ahmed. "Using Ensemble StackingC Method and Base Classifiers to Ameliorate Prediction Accuracy of Pedagogical Data." Procedia computer science 132 (2018): 1021-1040.
- [8] Mirza, Shuja, Sonu Mittal, and Majid Zaman. "A Review of Data Mining Literature." International Journal of Computer Science and Information Security (IJCSIS) 14.11 (2016).
- [9] Shuja, Mirza, Sonu Mittal, and Majid Zaman. "Diabetes Mellitus and Data Mining Techniques: A survey.". International Journal of Computer Sciences and Engineering 7 (2019): 858.
- [10] Mirza, Shuja, Sonu Mittal, and Majid Zaman. "Decision Support Predictive model for prognosis of diabetes using SMOTE and Decision tree." International Journal of Applied Engineering Research 13.11 (2018): 9277-9282.
- [11] Shuja M., Mittal S., & Zaman M. (2018). Decision Support System for Prognosis of Diabetes using Non-Clinical Parameters and Data Mining Techniques, International Journal of Database Theory and Applications, 11(3), 39-48.
- [12] Shuja, Mirza, Sonu Mittal, and Majid Zaman. "Effective Prediction of Type II Diabetes Mellitus Using Data Mining Classifiers and SMOTE." Advances in Computing and Intelligent Systems. Springer, Singapore, 2020. 195-211.
- [13] Raileanu, Laura Elena, and Kilian Stoffel. "Theoretical comparison between the Gini index and Information gain criteria." Annals of Mathematics and Artificial Intelligence 41.1 (2004): 77-93.
- [14] Muharram M.A., Smith G.D. (2004) Evolutionary Feature Construction Using Information Gain and Gini Index. In: Keijzer M., O'Reilly UM., Lucas S., Costa E., Soule T. (eds) Genetic Programming. EuroGP 2004. Lecture Notes in Computer Science, vol 3003. Springer, Berlin, Heidelberg.
- [15] Kulkarni, Vrushali Y., Manisha Petare, and P. K. Sinha. "Analyzing random forest classifier with different split measures." Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012. Springer, New Delhi, 2014.

# Disparity of Stereo Images by Self-Adaptive Algorithm

Md. Abdul Mannan Mondal<sup>1</sup>, Mohammad Haider Ali<sup>2</sup>

Dept. of Computer Science and Engineering  
University of Dhaka  
Dhaka, Bangladesh

**Abstract**—This paper introduces a new searching method named “Self Adaptive Algorithm (SAA)” for computing stereo correspondence or disparity of stereo image. The key idea of this method relies on the previous search result which increases searching speed by reducing the search zone and by avoiding false matching. According to the proposed method, stereo matching search range can be selected dynamically until finding the best match. The searching range  $-d_{max}$  to  $+d_{max}$  is divided into two searching regions. First one is  $-d_{max}$  to 0 and second one is 0 to  $+d_{max}$ . To determine the correspondence of a pixel of the reference image (left image), the window costs of the right image are computed either for  $-d_{max}$  to 0 region or for 0 to  $+d_{max}$  region depending only on the matching pixel position. The region where the window costs will be computed- will be automatically selected by the proposed algorithm based on previous matching record. Thus the searching range is reduced to 50% within every iteration. The algorithm is able to infer the upcoming candidate’s pixel position depending on the intensity value of reference pixel. So the proposed approach improves window costs calculation by avoiding false matching in the right image and reduces the search range as well. The proposed method has been compared with the state-of-the-art methods which were evaluated on Middlebury standard stereo data set and our SAA outperforms the latest methods both in terms of speed and gain enhancement with no degradation of accuracy.

**Keywords**—Stereo correspondence; stereo matching; window cost; adaptive search; disparity; sum of absolute differences

## I. INTRODUCTION

Stereo correspondence is required in applications such as autonomous vehicle and robot navigation, virtual reality, stereo image coding in 3D-TV and pedestrian detection. The difference between the coordinates of left image and right image of the corresponding pixels is known as stereo correspondence or disparity, which is inversely proportional to the distance of the object from the camera. Stereo correspondence is a common tool in computer or robot vision, essential for determining three-dimensional depth information of an object using a pair of left and right images from a stereo camera system. For a pixel in the left image, its correspondence has to be searched in the right image based on epipolar line and maximum disparity. Stereo correspondence is conventionally determined by matching windows of pixels using Sum of Square Differences (SSD), Sum of Absolute Differences (SAD), or Normalized Correlation Techniques (COR).

Window-based stereo correspondence estimation technique is widely used due to its efficiency and ease of implementation. However, there is a well-known problem in the selection of an appropriate size and shape of window [1- 2]. If the window is small and does not cover enough intensity variation, it gives erroneous result due to the low signal to noise ratio. If the window is large, it includes a bigger region where the disparity varies or discontinuity of disparity happens and the result becomes erroneous due to different projective distortions in the left and right images. To overcome this problem, many researchers proposed adaptive window techniques using windows of different shapes and sizes [3-7]. In the existing adaptive window technique involves comparing the window costs for different window sizes and shapes, so the computation time is relatively higher than that of fixed window technique. In case of gray scale stereo images, the use of color stereo images brings a substantial gain in accuracy with the expense of computation time [8]. The Virtual Masking System has been employed in [9]. The Virtual Masking System is based on the stereo matching constraint states the corresponding pixels should be similar in color or intensity [10]. The Virtual Masking System technique improves the computational efficiency by excluding unlikely correspondences. Two-stage Approximation Algorithm, Arbitrary Window Pixel, Diagonal Mask Searching were also reported in [11-12] and survived the minimum values among the three searches to overcome the window-based problems.

The main achievement of the proposed algorithm is to improve the huge incremental computational schemes by dynamically readjusting the search range and eliminating redundant computational calculations. The proposed adaptive search method (SAA) is able to avoid false matching in the right image pixels that causes to increase the visual quality of the estimated stereo correspondence. SAA performs according to needful in the two new searching zones of the right image. Thus reduces the computational time that ensures the better speed of experimental dense disparity images as shown in Fig. 15 and Fig. 16. In fact, the proposed SAA approach optimizes the computational time and speed. The main contribution of this paper is fully self-adaption based adaptive function with the following advantages.

- Experimental results show that the 3D reconstruction of output disparity map is very similar to ground truth dense disparity.

- We observe 535 *fps* for input images with 384×288 (Tsukuba head pair) pixel resolution and 377 *fps* for input images of Venus stereo pair with 434×383 resolution.
- The proposed method is applied on the Middlebury standard stereo images and the results show that it reduces 70.53% computational time for Venus stereo pair and 99.93% computational time for Tsukuba stereo pair with no degradation of accuracy.
- Numerical evaluations confirm that the bad pixel is only 6.2% with error threshold 1 for Tsukuba head.

## II. RELATED WORK

Many researchers do research on a dense two-frame stereo in many ways. They try to optimize the dense disparity in locally or globally on a stereo pair. So the dense matching algorithms are divided into local and global ones [36]. The best classifications have presented by Scharstein and Szeliski [13] and many new methods have been proposed here. Local methods are also known as area based stereo matching that can perform better speed compare to global methods. According to this, disparity is being calculated at a point in a fixed window. Global methods are also known as intensity or energy based stereo matching that can perform better accuracy compare to local methods. In this method, the global cost function is reduced to minimum as possible. This cost function synthesizes image data and smoothness terms. Neural adaptive stereo matching [15] are done by trained neural networks based on window size and shape. The effects of shape of the stereo image pair presented in [16]. Yoon et al. [17] used SAD method and a left-right consistency check that performed like real time system. Yoon's method can process 7 *fps* for 320×240 pixels images with 32 disparity levels. The experiment has been implemented by using an Intel Pentium 4 at 2.66GHz Processor. The uses of Cellular Automata (CA) are presented in [18]. Real-time extraction of disparity maps was demonstrated by this method. It can process the input image with a speed of 40 *fps*. This method also based on SAD matching cost. A window-based method uses different support-weights in [19]. The support-weights of the pixels in a given support window are adjusted based on geometric proximity and color similarity to reduce the image ambiguity in [20]. The running time for the Tsukuba image pair with a 35×35 pixels support window is about 0.016 *fps* on an AMD 2700+ processor. The research work presented in [20] based on unified framework that supports the fusion of any partial knowledge such as matching features and surfaces about disparities. Physical processes and systems are designated by a cell array which expressed the new state of a cell depending on the states of its neighbors [21].

Bayesian estimation theory with a prior Markov Random Fields model for the assigned disparities is described in [22]. According to this method, the continuity, coherence, occlusion constraints and the adjacency principle are taken into considerations. Image color segmentation is reported in [23]. By this method, disparity map is estimated using an adapting window based technique. The segments are combined in larger layers iteratively. Bidirectional matching is used in [24] to

discard the false matching and the authors employed the technique to calculate the disparities of stereo images without known epipolar geometry. In [25] pruning technique is used to estimate the disparities.

The recent works of related problem to the matching costs are stated in [26] and [27]. The authors used bilateral filter to determine the cost aggregation and in order to reduce the computational cost they also limit the label space. The work in [28] can be considered as a cost aggregation method by guided image filter. The average runtime [28] of the four standard Middlebury data sets (including Tsukuba, Venus, Teddy and Cones data sets) is 960 milliseconds reported in [31]. So the run time of single image pair like Tsukuba or Venus is about 240 milliseconds. Disparity space image (DSI) structure and gradient information has been combined as a new technique is first time introduced by Nadia Baha and Slimane Larabi [29]. They used DSI technique with adaptive window – support. Another approach is introduced by themselves as DSI and refinement. The experimental results take time 0.2 *second* and 0.39 *second* respectively for processing Tsukuba head image pair. A new geodesic  $o(1)$  filter is employed in [30] for the reliable disparity propagation. Such type of filter is very effective for the cost matching. As it is state-of-the-art method and the speed of this method has been justified on the Middlebury standard data set, so we can compare this paper to our proposed SAA method. Xun Sun et al. [30] perform the experiment on PC equipped with a 3.0 GHz Intel i5 CPU, 8 GB of memory and a Geforce GTX 580 graphics card. The processing time on Middlebury standard data set is only 9 milliseconds. A cost aggression has been adaptively estimated on a tree structure derived from the stereo image pair to preserve depth edges [31]. This latest idea is launched by Q. Yang [31] in which shortest distances measure the similarity between two pixels on the tree. The average runtime of the four standard Middlebury data sets (including Tsukuba, Venus, Teddy and Cones data sets) is about 90 milliseconds using the tree filtering method. But He et al. [31] mentioned in same section that the runtime is about 7 milliseconds on average on the Middlebury data sets. For identical comparison to our proposed method we consider his second result that takes 7 milliseconds on average on the Middlebury data sets. Q. Yang tested his experiment on a MacBook Air laptop with a 1.8 GHz Intel Core i7 CPU and 4 GB memory. Another recent method achieves state-of-the-art result on Middlebury stereo data sets that performs stereo matching as a two steps energy-minimization algorithm [32]. The running time of this method is 3 seconds only for Tsukuba data set and 20 seconds for Teddy data set on a computer containing an Intel Core i5-4300U 1.9-GHz CPU and a 6-GB RAM. Semi-global matching and cost is refined by cross-based aggression [33] has been introduced by J. Zbontar and Y. LeCun. Y. LeCun et al. [33] also uses left-right consistency check to eliminate the errors. The experiment performs on KITTI stereo data set. Depth of information is estimated by learning a similar measure on image patches [34]. At very recent, Fusing Adaptive Support Weights [38] has been launched by Wenhuan Wu and others. Local and global support windows used for each pixel in [38]. Self-guided cost aggression [39] has been determined by deep learning method that depends on two sub-networks. A pyramid stereo matching network [40] also consists of two modules

based on pyramid and 3D CNN that have been tested on KITTI 2012 and 2015 data set. Adaptive Weighted Bilateral Filter [41] is used as main filter at cost aggregation step for edge preserve factor.

With the above reviews we found that some researchers employed adaptive window-based techniques to calculate the matching costs. But in our proposed method we employ *self-adaptive function* to calculate the matching costs dynamically. This is one of the main distinguishable points between the proposed SAA method and existing state-of-the-art methods. The mentioned recent methods are very similar to our proposed method but differing mostly in optimal searching technique. Besides these analysis, the work in [30] requires preprocess and the works in [29], [31], [33] and [35] needed post processing steps like refinement, filtering and histogram equalization. The proposed SAA method also runs without preprocessing and post processing. The experimental disparity maps are directly eligible to compare with ground truth dense disparity. So considering the adaptive similarity, identical stereo data set (Middlebury Standard data set) and hardware platform we can consider the papers of [14], [24], [25], [29], [30], [31], [32], [38], [39] and [40] to compare between the state-of-the-art method and our proposed SAA method.

### III. COMPARISON WITH EXISTING MATCHING ALGORITHMS

The proposed matching mechanism is compared with the early established four stereo matching methods: (1) A Fast Area based Algorithm (2) Bidirectional matching or left-right checking (3) Hierarchical disparity estimation and (4) Window-based fast algorithm. First one approach is kept trace previously matched points, while the second one calculates every possible combination of matches from left to right and right to left. So it is also called left-right consistency checking. The hierarchical disparity method calculates disparities either for rectified stereo images or uncalibrated pairs of stereo images without known epipolar geometry [24]. Additionally this method also uses bidirectional matching to remove false matches. Window-based fast algorithm uses the same idea of first method but it additionally uses the threshold technique. The coloring area defines the probable matching points of right image. According to bidirectional matching, first scan takes place in the direct phase i.e. from left to right and find the all scores associated with the reference image. Suppose the matching pixels are marked by white circle along the horizontal color path in Fig. 1. Secondly in reverse phase it is searched to find the best match again and a match is accepted if it is the only one the matching pixels of first search i.e. left to right. It is noted that during the reverse phase bidirectional searching, checks all the potential matches are along the middle line of horizontal path but only those are allowed marked by black circle with white center and those fall in the white circle area of the middle horizontal line [14] for  $R(x,y)$ . The comparisons of the proposed method and recent works [26-40] are explained briefly in the last portion of section-II.

On the contrary in our proposed adaptive matching method, the best match occurs by minimum computational cost with dynamically readjusted the searching starting and ending pixels. According to the main concepts matching pixels are

those which are the nearest neighbors of  $R(x,y)$ . The searching points are dynamically readjusted based on the said main concepts of underlying algorithm. Accordingly, 1<sup>st</sup> search occurs on all the lying pixel ranges from  $R\{x +(- d_{max}), y\}$  to  $R\{x +(+d_{max}), y\}$ .

First search for first reference pixel is to be accomplished through the whole area of searching regions. Second, third and consecutive searches are to be adapted. By the completion of first search the proposed algorithm keeps track the position of matching pixel also. So the second matching may occur surrounding the first matching pixel, as we use the concepts that neighbor pixels have the same photometric properties. So depending on the position of matching pixel, the successive search areas are reselected. It will be either  $- d_{max}$  to  $0$  or  $0$  to  $+ d_{max}$  region. This adaptive matching procedure will be continued for every candidate's pixel within the search range. The main advantages of proposed method (SAA) compared to the above mentioned existing established methods relies on its lowest computational costs. The *left-right check* has proven to be particularly effective in detecting and discarding the erroneous matches but it requires two matching phases (direct and reverse). This implies doubling the computational complexity of the matching process. The fast area based algorithm is based on a matching core that it does not require a reverse matching phase but some details such as for example the lamp's wire (Tsukuba pair), the lamp's switch and the two roads that sustain the lamp, have been vanished. Moreover, the disparity map shows the *border-localization* problem, i.e. the objects' borders are not perfectly localized with respect to their original position. This algorithm requires only direct matching phase. The hierarchical method executed on DirectX 8.1 class 3D hardware (ATI Radeon 9000 Mobility). The disparity map is verified by bidirectional procedure. Window-based fast algorithm [25] uses the different threshold values like 10,20,30,40 and 50. From the experiment we explore that as soon as the threshold value increased the searching range also increased causes to high computational costs. Conversely, the proposed SAA method shows lowest computational cost because it involves only minimum matching spaces. Suppose there are  $n$  candidate pixels appear on the searching range. According to the above analysis, the fast area based algorithm requires  $n^2$  numbers of matching iterations and the bidirectional or hierarchical search requires  $2n^2$  matching iterations. The proposed method requires only  $\{(n+1) + (n/2) \times (n-1)\}$  matching iterations.

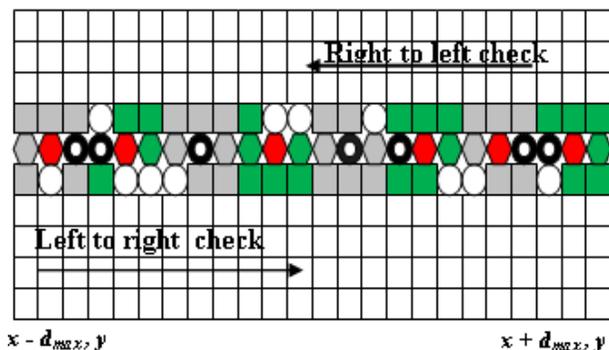


Fig. 1. Computational Path of Bidirectional and Unidirectional Matching from the Computational Point of View.

IV. PROPOSED SELF-ADAPTIVE SEARCH METHOD

A new stereo imaging search technique has been introduced in this section. In traditional window based searching algorithm, a particular pixel  $L(x,y)$  is selected by search method along the corresponding epipolar line in the right image within a search range from  $-d_{max}$  to  $+d_{max}$ . Assuming the left image as the reference image, for most pixels in the left image, there is a corresponding pixel in the right image within a search range from  $-d_{max}$  to  $+d_{max}$ .

Accordingly, the first search of first reference pixel of  $L(x-d_{max})$  is searched to the right image from  $R(x+(-d_{max}), R[x+(-d_{max}+1)], R[x+(-d_{max}+2)] \dots R[x+0]] \dots R[x+(d_{max}-1)]$  to  $R(x+d_{max})$ . During the matching process the algorithm finds the best candidate pixel by evaluating its window costs function within the interval  $[R\{x+(-d_{max}), y\} \dots R\{x+(+d_{max}), y\}]$ . The method is visually explained in Fig. 2 and Fig. 3 by mortars its coordinate's pixel. Suppose the window cost function  $f(wc) = \{wc_1, wc_2, wc_3 \dots wc_n\}$ . Let  $wc_2 < wc_1$  so the best match occurs for the cost function  $f(wc_2)$  and the function associated with the corresponding pixel of right image say  $R(x_2, y)$  to indicate that this match from left to right has been established. Assume another pixel  $R(x_4, y)$  is associated with the cost function  $f(wc_4)$ . If  $f(wc_4)$  has better score than the previous  $f(wc_2)$  .i.e.,  $f(wc_4) < f(wc_2)$ , this algorithm will reject the score of  $wc_2$  and will accept  $wc_4$ . Therefore, the function  $f(wc_4)$  associated with the corresponding pixel of right image say  $R(x_4, y)$  indicates the new matching establishment. Thus the coordinate distance from  $R(x,y)$  to  $R(x_4,y)$  is the final disparity of reference pixel of  $L(x-d_{max})$ . The process is then repeated for the successive pixels of reference image along with the scan line from left to right of the whole image.

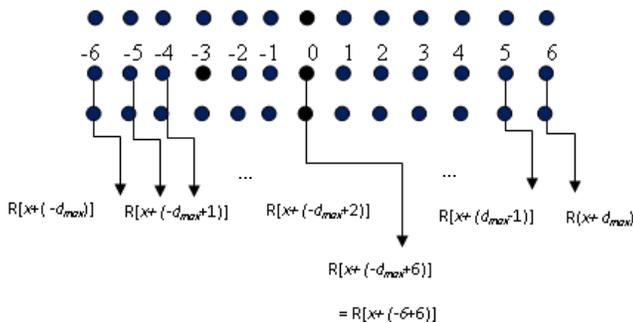


Fig. 2. The Total Search Regions of Right Image for the Particular Pixel of  $L(x,y)$ .

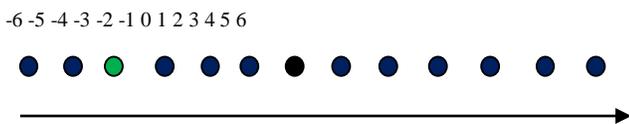


Fig. 3. First Search Interval  $\{R(x+(-d_{max})) \dots R(x+(+d_{max}))\}$ .

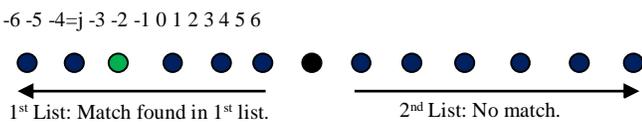


Fig. 4. Search Range Separated by 1<sup>st</sup> List and 2<sup>nd</sup> List with their Candidate's Pixels.

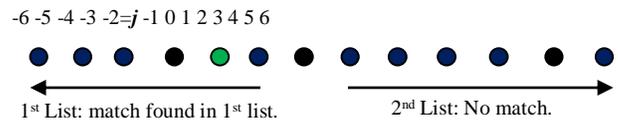


Fig. 5. Second Search Occurs in 1<sup>st</sup> List too, and the Matching Pixel is Indicated by Green Color also.

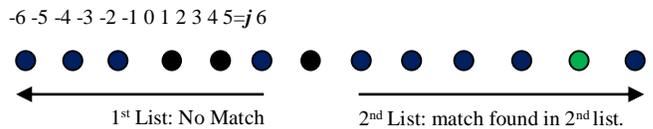


Fig. 6. 4<sup>th</sup> Search Occurs in 2<sup>nd</sup>List, Match Indicated by Green Color.

First search matching occurs into in 1<sup>st</sup> list and the matching pixel indicated by the green color at position  $j=-2$ .

According to SAA method, if the matching pixel point  $R(x_4,y)$  of first search is “  $j$  ” indicates the  $x$ -position of matching point in the 1<sup>st</sup> list. This procedure is illustrated in Fig. 4. In the proposed matching approach we use prior knowledge-based adaptive search technique for 2<sup>nd</sup> reference pixel matching. As per sequence of proposed method, 2<sup>nd</sup> reference pixel obviously resides surrounding to the  $j^{\text{th}}$  position i.e., within the 1<sup>st</sup> list, as the neighbor pixel's photometric properties are approximately same as  $L(x-d_{max})$ . So the second search for 2<sup>nd</sup> reference pixel performs in the 1<sup>st</sup> list too. This search matching procedure is depicted in Fig. 5. In case of critical situation occurs in second search; suppose the 2<sup>nd</sup> reference pixel does not match with the candidate's pixels of 1<sup>st</sup> list. Then the SAA search sequence goes to new searching zone i.e. in the 2<sup>nd</sup> list with readjusting the search interval. Accordingly, the SAA algorithm calculates the window costs of 2<sup>nd</sup> reference pixel in 2<sup>nd</sup> list. This procedure is outlined in Fig. 6. In this case, the candidates' pixels of 1<sup>st</sup> list are not taken into consideration for matching process. Next search for 3<sup>rd</sup> reference pixel occurs again in 2<sup>nd</sup> list too (which is not shown here). In case of no match in 2<sup>nd</sup> list for 3<sup>rd</sup> reference pixel, the program search sequence goes to 1<sup>st</sup> list by resetting the new starting and ending points of 1<sup>st</sup> list. The searching procedure of 3<sup>rd</sup> reference pixel will be the same as illustrated in Fig. 5. Next search for 4<sup>th</sup> reference pixel occurs in 1<sup>st</sup> list also as per base criteria of the proposed approach. So the matching procedure occurs either in 1<sup>st</sup> list or in 2<sup>nd</sup> list. In essence, if no match occurs in 1<sup>st</sup> list, the program sequence goes to 2<sup>nd</sup> list and if no match occurs in 2<sup>nd</sup> list, the program sequence goes to 1<sup>st</sup> list and vice-versa. The above stated SAA approach is repeated for the successive pixels of reference image along with the scan line from left to right on the whole image. The algorithm divides automatically the search interval  $[R\{x+(-d_{max}), y\} \dots R\{x+(+d_{max}), y\}]$  into two regions; 1<sup>st</sup> list and 2<sup>nd</sup> list. 1<sup>st</sup> list ranges from  $-d_{max}$  to 0 while 2<sup>nd</sup> list dominates from 0 to  $+d_{max}$ . The capability of proposed SAA algorithm is to adapt itself the search range automatically. This process reduces the iterations and tracks the exact position of next matching pixel. The proposed approach relies on  $j^{\text{th}}$  position of  $x$  axis. According to the proposed approach the algorithm explores the new searching starting and ending points of first and second list. This procedure is iterated for successive pixels in interval  $[R\{x+(-d_{max}), y\} \dots R\{x+(+d_{max})$

,  $y$ ] along with the scan line for the whole image. In this paper  $(2n+1) \times (2n+1)$  mask size is used to estimate the window cost over the every scan line; where  $n=1,2,3...k$ . The right image is scanned by this mask from left to right and from top to bottom during the matching process.

Suppose, if  $d_{max}=6$  and  $j=0$ , the algorithm resets the starting point at  $d_{max1}=-d_{max}$  and ending point  $d_{max2}=+d_{max}$ . In this case matching process occurs on the full scan line. Actually it is treated as 1<sup>st</sup> search as shown in Fig. 7 in which window costs calculation are shown here. There are 13 candidates' pixels in right image along with the scan line. For each pixel, window cost is calculated according to SSD method. So there are thirteen window costs that have been extracted at 1<sup>st</sup> search, although five window costs are shown here for simplicity. The proposed algorithm arranges the window costs function in ascending order. Suppose the order for 1<sup>st</sup> search is like as  $f(wc) = \{wc_2 < wc_1 < wc_3 < \dots < wc_{13}\}$ . So the proposed method rejects all other window cost except  $wc_2$ . Best match occurs for 2<sup>nd</sup> window due to its minimum window cost. Thus the reference pixel  $L(x-d_{max})$  matches with the candidate's pixel (center pixel) of the second window of right image.

As the first matching pixels position is  $j=-2$  (since  $j < 0$ ) illustrated in Fig. 4 and Fig. 8, the algorithm resets the search interval  $d_{max1}=-d_{max}$  to  $d_{max2}=0$ . In this region only six pixels are participants, rest of the pixels of 2<sup>nd</sup> list are ignored due to  $j < 0$ . This idea is exactly illustrated in Fig. 8.

So there are only six window costs has been extracted from 2<sup>nd</sup> search, though there three window costs are shown in Fig. 8 for simplicity. The proposed algorithm arranges the window costs function in ascending order. Suppose the order for 2<sup>nd</sup> search is like as  $f(wc) = \{wc_1 < wc_4 < wc_3 < wc_2 < wc_1 < wc_6\}$ . Hence the proposed method rejects all other window costs except  $wc_1$ . Best match occurs for 1<sup>st</sup> window cost in first list due to its minimum window cost. Thus the reference pixel  $L(x-d_{max}+1)$  matches with the candidates pixel (center pixel) of the first window of right image. Since the reference pixel  $L(x-d_{max}+1)$  matching position is  $j=-5(j < 0)$  shown in Fig. 8, so third search (not shown here) will be continued into 1<sup>st</sup> list also. If it falls into the critical situation in 4<sup>th</sup> search, i.e. pixel  $L(x-d_{max}+1)$  does not match with 1<sup>st</sup> list, the algorithm readjusts the new searching zone's interval  $d_{max1}=0$  to  $d_{max2}=+d_{max}$  and the searching sequence goes to 2<sup>nd</sup> list. The proposed approach calculates the window costs function following the same procedures applied in 2<sup>nd</sup> search. The calculations of window costs for 4<sup>th</sup> search are outlined in Fig. 9 in which the starting and ending points are quite different from 2<sup>nd</sup> and 3<sup>rd</sup> search and three window costs are shown for simplicity. Actually there are six window costs has been estimated. So the proposed SAA method achieves the prior knowledge of matching pixel position in  $x$  axis and accordingly it redirects the program sequence in the 1<sup>st</sup> list or the 2<sup>nd</sup> list. As the matching procedure occurs either in 1<sup>st</sup> list that contains half of candidates pixels or in 2<sup>nd</sup> list that also contains half of candidate's pixels, so it reduces 50% or above the iterations process. In next section, the iterative reductions in percentage are figured out.

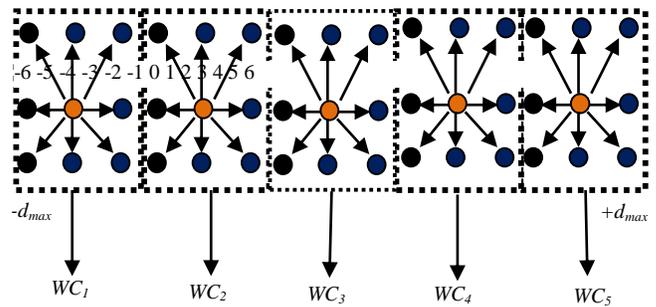


Fig. 7. Window Cost Estimation Process Over the Scan Line (1<sup>st</sup> Search).

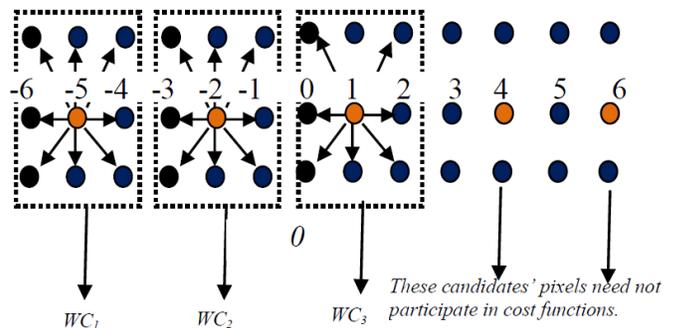


Fig. 8. Window Cost Calculation Process for 1<sup>st</sup> List Only (2<sup>nd</sup> Search).

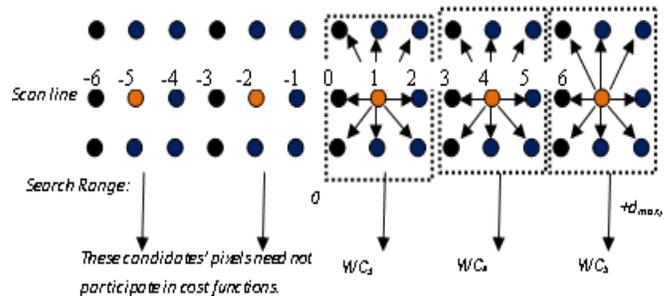


Fig. 9. Window Cost Calculation Process for 2<sup>nd</sup> List Only (4<sup>th</sup> Search).

#### A. Disparity estimation Algorithm of SAA method for a pixel $(x,y)$ .

1. Start.
2. Repeat step 2 to 10 for each pixel  $(x,y)$
3. If  $flag=0$ ,
4. Set:  $d_{max1}=-d_{max}, d_{max2}=d_{max}$ ;
5. else if  $(flag < 0)$
6. Set:  $d_{max1}=-d_{max}, d_{max2}=0$ ;
7. else
8. Set:  $d_{max1}=0, d_{max2}=d_{max}$ ;
- [end of if ]
9. for  $d' = d_{max1}$  to  $+d_{max2}$  do  
Calculate  $W_c((x,y,d'))$
10. Find best  $W_c(x,y,d) \in W_c(x,y,d')$
11. [End of the step 2]
12. Disparity of  $(x,y) = d$ ;
13. if  $(j < 10)$
14. Set:  $flag = -1$ , goto step 6;
- else
15. Set:  $flag=1$ , goto step 8;
16. Stop.

The above procedure shows the calculation of window cost for one pixel only. The next matching position range  $\lambda'$  is being selected by setting the flag pointer either -1 or +1 with the following statements:

```
if(j<10)
flag = -1;
else
flag = 1;
```

Flag pointer controls the reduced matching position range  $\lambda'$  and thus both the computational costs and number of iterations have to reduce always.

V. OPTIMIZATION OF SELF-ADAPTIVE SEARCH

The main highlight of SAA approach is optimization technique. Two types of optimizations are accomplished here simultaneously i) iteration and ii) computation. We have already mentioned 50% or above iterations reduced at every reference pixel of left image except first pixel  $L(x-d_{max})$ . First pixel (black color of left image) traverses from  $R(x-d_{max})$ ,  $R(x-d_{max}+1)$ ,  $R(x-d_{max}+2)$  ...  $R(x+d_{max})$  along the scan line. So the first search occurs from  $-d_{max}$  to  $+d_{max}$  over the scan line. In this case, for first search window cost has been calculated for every candidate's pixels of right image for every window using the mask of size 3x3. Suppose the first reference pixel matches with the third pixel( $x-d_{max}+2$ ) of right image as shown in Fig. 10. After tracking the first matching, the area only. This process will be continued for successive pixels along the scan line.

Let  $w_1 \times w_2$  is the mask size and matching range is  $\lambda$ . So the first searching computational cost  $F_1 = \{(w_1 \times w_2) - 1\} \times (\lambda + 1)$ .

Second optimization cost  $F_2 = \{[(w_1 \times w_2) - 1] \times (\lambda + 1)\} / 2$ . Third, fourth ...  $\lambda^{th}$  searching computational costs be the same i.e.  $F_{\lambda-1} = \{[(w_1 \times w_2) - 1] \times (\lambda + 1)\} / 2$ . Fast area based algorithm [14] requires for every reference pixel computational cost,  $F_{AB} = \{(w_1 \times w_2) - 1\} \times (\lambda + 1)$ .

But in hierarchical disparity [24] or left -right checking algorithm requires for every reference pixel computational cost,  $F_{HD} = 2 \times \{(w_1 \times w_2) - 1\} \times (\lambda + 1)$ .

For the image size is  $M \times N$ , the total computational costs of Fast area based algorithm [14] is  $F_{AB} = (M \times N) \times \{(w_1 \times w_2) - 1\} \times$

$(\lambda + 1)$ . Total computational costs of hierarchical disparity method's [24] is  $F_{HD} = 2 \times [(M \times N) \times \{(w_1 \times w_2) - 1\} \times (\lambda + 1)]$ .

The computational costs of proposed SAA method,  $F_{SAA} = [(M \times N - 1) \times \{(w_1 \times w_2) - 1\} \times (\lambda + 1)] / 2 + \{(w_1 \times w_2) - 1\} \times (\lambda + 1)$ .

Suppose there are  $n$  candidates' pixels within the search range  $\lambda$ .

1) So according to the proposed SAA method the 1<sup>st</sup> search comparison occurs only in  $n+1$  pixels. For  $(-n+1)^{th}$  pixel matching i.e. for 2<sup>nd</sup> search the SAA method sets the reduced match range  $\lambda'$ . proposed method divides the searching space into two regions indicated by 2<sup>nd</sup> and 3<sup>rd</sup> arrow in right image.

Suppose  $flag$  is a tracking pointer, if  $flag < 0$  then SAA algorithm sets  $d_{max1} = -d_{max}$ ; and  $d_{max2} = 0$ ; else  $d_{max1} = 0$ ,  $d_{max2} = +d_{max}$ . So the upper bound and lower bound of matching regions are set according to above criterion and falls into two regions. As per theme of section -IV the second pixel (green color of left image) of reference image searched only 2<sup>nd</sup> search area. Accordingly, the third pixel (violet color of left image) of reference image is matched only 3<sup>rd</sup> search area indicated by the same color in right image. So the subsequent reference pixel matches either 2<sup>nd</sup> search area or 3<sup>rd</sup> search.

This match occurs in  $n/2$  pixels and all forthcoming matching will be occurred within  $n/2$  pixels. So excluding first pixel there are  $(n-1)$  rest of pixels within the search range. Hence the total iterations are for proposed SAA method =  $(n+1) + (n/2) \times (n-1)$ . But for Fast area based algorithm the total iterations are =  $n^2$ . Left -right checking or bidirectional requires the total iterations =  $2 \times n^2$ .

2) In a stereo matching  $n$  varies from  $-10$  to  $+10$ . So let  $n=10$  means  $n/2=5$ . Thus the total iterations are for proposed SAA method =  $(n+1) + (n/2) \times (n-1) = (10+1) + (11/2) \times (11-1) = 66$ .

3) Fast Area Based algorithm requires the total iterations are =  $n^2 = 11 \times 11 = 121$ .

4) Hierarchical disparity method or bidirectional requires the total iterations =  $2 \times n^2 = 2 \times 11^2 = 242$ .

5) So iteration reduction compared to the Fast Area Based algorithm =  $(55/121) \times 100\% = 45.45\%$ .

6) Iteration reduction compared to Hierarchical disparity method or bidirectional method =  $(176/242) \times 100\% = 72.72\%$ .

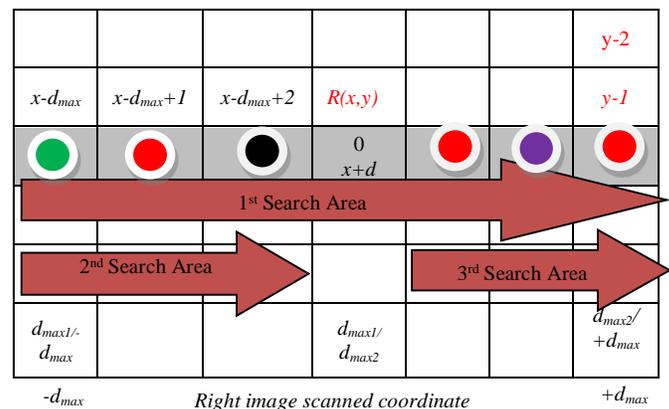
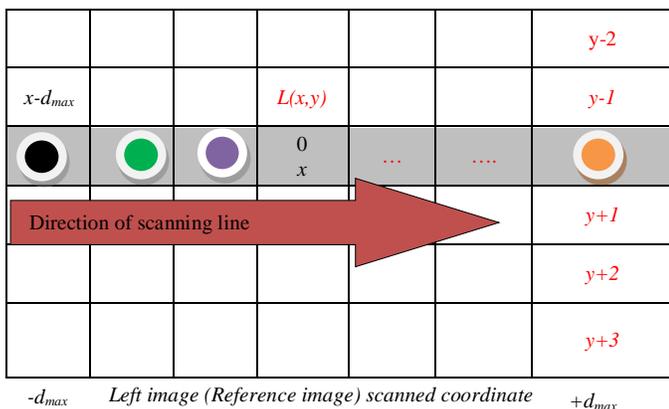


Fig. 10. Scanning Process from  $-d_{max}$  to  $+d_{max}$  on Right Image Mentioning the Co-Ordinate System.

Flag pointer always controls match range  $\lambda'$  by changing the parameter  $d_{max1}$  and  $d_{max2}$  that depends upon the previous matching position of  $x$ . From  $d_{max1}$  to  $d_{max2}$  for every window cost is determined by summing up the product of difference between two pixels by the following pseudo statements.

sum = sum + (v\_left - v\_right)\*(v\_left - v\_right); the parameters v\_left and v\_right of (i,j)th set up by

v\_left = image\_left.pixel[m+i][n+j];

v\_right = image\_right.pixel[m+i+d][n+j];

where  $i$  and  $j$  varies from -1 to +1 for of a mask size  $3 \times 3$ . The window costs are stored in an array Mtemp[ $d + d_{max}$ ].pixel[m][n]. These cost values are gone to minimum function as arguments. The minimum function implements the key idea of the proposed method by assessment of its comparisons among the window costs. The function performs the dual tasks i) it can be able to determine the matching  $x$  axis pixel position " $j$ " by setting  $j=i$  and find the desired disparity with the following pseudo statement-

1) if (temp[i].pixel[x][y] < temp[j].pixel[x][y])

j = i;

and

2) The SAA algorithm selects the upcoming matching position range  $\lambda'$  by setting the flag pointer either -1 or +1 with the following statement.

if(j<10)

flag = -1;

else

flag = 1;

Flag pointer controls the reduced matching position range  $\lambda'$  and thus both the computational costs and number of iterations has been reduced always.

## VI. EXPERIMENTAL RESULTS

The experiments have been performed on two Middlebury standard stereo images: i) Tsukuba stereo pair, ii) Venus stereo pair as shown in Fig. 11. The computational time, speed, accuracy and gain performances of the proposed algorithm have been justified over the said stereo images (Tsukuba Head and Venus stereo pair). Experiments are performed on Intel Core i3, 2.3 GHz processor with 4 GB DDR3 RAM. The algorithms have been implemented using Visual C++ programming language. The size of the left and right images of Tsukuba head is (width×height) = (384×288) pixels and, the ground truth image size is (width×height) = (348×252) pixels. The size of the left and right images of Venus stereo is (width×height) = (434×383) pixels and the ground truth image size of Venus is (width×height) = (348×252) pixels. Mask size of  $11 \times 11$  is used for every operation in this research.

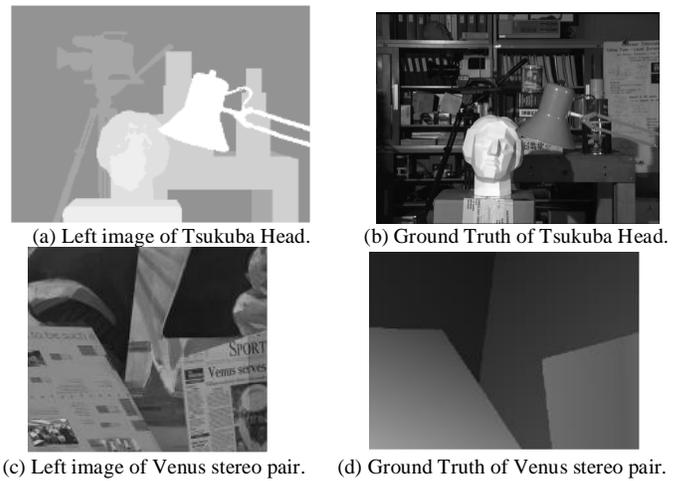


Fig. 11. Standard Stereo Image (Reference Image) and their Ground truth Image.

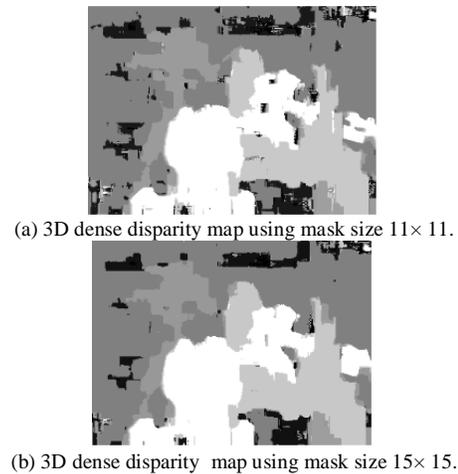


Fig. 12. Estimated 3D Dense Disparity Map of Tsukuba Head using SAA Method.

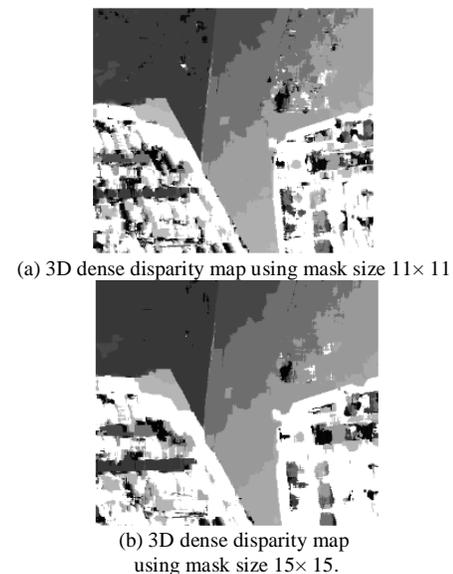


Fig. 13. Estimated 3D Dense Disparity Map of Venus Stereo using SAA Method.

The experimental results state that the SAA algorithm is currently the best window cost method among all existing state-of-the-arts methods. The top performer algorithms are reported in [28], [29], [30], [31] and [32]. All are ranked by Middlebury benchmark [37]. So we have to prove the claim by comparing the time and speed with the top performer algorithms which is demonstrated in Table I and Table III. The disparity maps of the Middlebury data sets for Tsukuba head and Venus stereo pair are estimated by proposed SAA method illustrated in Fig. 12 and Fig. 13. The corresponding numerical time reductions are depicted in Table II and Table IV. From the tables it is evident that in both cases the proposed SAA algorithm outperforms the current and earlier top performer algorithms. Moreover, the proposed method is faster than all others top performer algorithms. The accuracy of the SAA algorithm for Tsukuba head is 93.8% i.e. the bad pixel in percentage with the error threshold is only 6.2% which is almost the same of the top algorithms. Little variation of accuracy occurs due to orientations of pixel redundancy. The performance enhancements by the said SAA method are discussed in Section VII. Table V and VI show the noteworthy gain enhancement of proposed SAA method. The experimental results analyzed in four phases are stated.

#### A. Experiment 1: Observation of 3D Structures of Experimental Output

Both the Tsukuba and Venus stereo pair of input images contain different objects at variable depth of positions. Contextual and forefront objects are positioned at different depth. Four objects are placed at different depth of Venus stereo of input image. Stereo pair also encloses some distinct areas like head of the statue, table lamp, video camera, Venus sport paper, and another paper and background paper wall. Such types of areas are quite challenging to isolate from other objects by stereo matching. So the first work is to differentiate the variable depth of objects by assigning the altered gray level value of output image. Nearby object is presented by deep white color and outermost object is presented by dark or deep black color. It is detecting that the experimental 3D construction of output image is recreated evidently in Fig. 12 and Fig. 13 where the face of the statue, table lamp, video camera, Venus sport paper, another paper as well as remarkable objects are known easily. Comparing the output images of Fig. 12(a) and Fig. 13(a) with ground truth image of Fig.11(b) and Fig.11(d), respectively, the camera and its background objects have been recovered almost correctly and the object closer to the stereo acquisition system. The objects depths are noting that nearby objects are realized by additional white color and outermost objects are realized by dark grey level value or black as shown in Fig. 12 and Fig. 13. Thus, the camera and nearby objects such as face of Tsukuba, table lamp and sports paper, 2<sup>nd</sup> paper (to be such: left side) of Venus stereo are visualized by more white color. On the other hand, the camera and outermost objects such as video camera, book shelf, background wall of Tsukuba stereo pair and background paper wall of Venus stereo are recreated with dark grey or black color. Object borders are evidently detecting in computed dense disparity image, i.e. border localization problem of article [14] are resolved by the SAA method. The experimentally estimated images are tested again by the edge

detection algorithm and the output object's borders are acknowledged which are illustrated in Fig. 14.

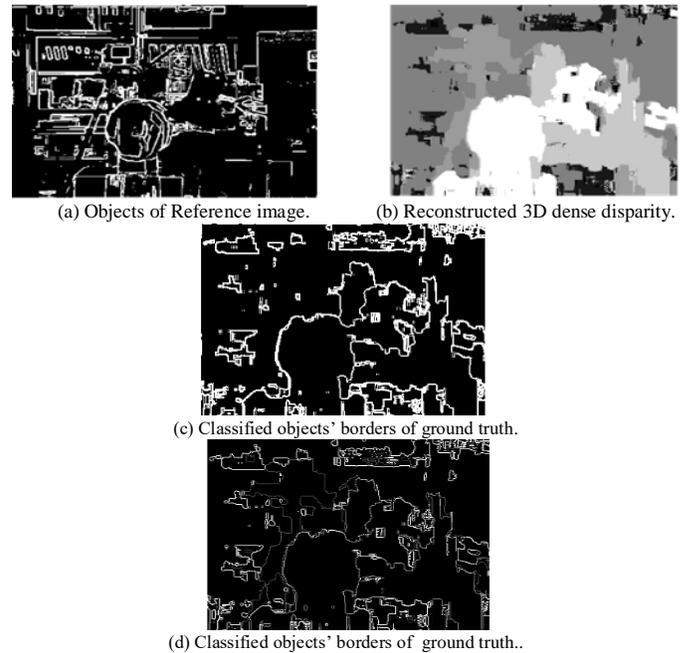


Fig. 14. Localized Objects' Borders.

The estimated dense disparity's 3D structure is recovered and its objects border are correctly identified which are outlined in Fig. 14(c) and (d). So the result ensures that the similar depths are found in estimated dense disparity.

#### B. Experiment 2: Computational Cost Calculation and Comparison with Existing State-of-the-Arts Methods for Middlebury Standard Tsukuba Head Image Pair

Disparities of reference image are estimated by SSD technique using adaptive search algorithm without any pruning. The disparities are estimated with the search range from -10 to +10. The effects of self-adaptive search are investigated with respect to computational costs, speed (in *fps*), gain and accuracy. The computational costs, speed and gain performance results of adaptive method has been compared with previous fastest literatures [14, 24, 25] and the current state-of-the-arts methods [29-32]. Fast Area Based algorithm [14] reports 3229 $\mu$ s required by Intel Core i3, 2.3 GHz processor with 4 GB DDR3 RAM for 384 $\times$ 288 image resolution of Tsukuba head. This experiment also results 310 *fps* and 16 disparity levels. Hierarchical Disparity algorithm [24] reports 4243 $\mu$ s required by Intel Core i3, 2.3 GHz processor with 4 GB DDR3 RAM for 384 $\times$ 288 image resolution of Tsukuba head. This experiment also results 235 *fps* and 88 disparity levels. Fast algorithm [25] reports 4617 $\mu$ s required by Intel Core i3, 2.3 GHz processor with 4 GB DDR3 RAM for 384 $\times$ 288 image resolution of Tsukuba head. This experiment also results 216 *fps* and 11 disparity levels. But the proposed SAA algorithm requires only 1872 $\mu$ s on the same hardware with minimum 66 iterations instead of 121 and 242 iterations as mentioned in Section-V. The proposed method also performs the better speed compared to previous popular methods [14, 24, 25]. It shows the highest speed 535 *fps* while

reporting 11 distinct disparity levels. It is mentioned that the ground truth image of Tsukuba pair contains only 7 disparity levels. Table I illustrates the summary of comparisons among the proposed method and existing state-of-the-arts methods.

The SAA's experimental results have been also compared with the result of methods those are tested on Middlebury standard data set. The ranking results in Table I indicate that the proposed SAA method is ranked 1<sup>st</sup> out of existing top state-of-the-arts methods [29-40]. It shows the highest speed 535 *fps* and lowest computational time 1872 microseconds among the latest methods with lower configuration of machine. So the proposed method outperforms all the state-of-the-arts methods in speed and computational time on Middlebury standard Tsukuba head image pair.

So it is claimed that the proposed method is currently the state-of-the-arts method for Middlebury standard Tsukuba head image pair with 1.7X, 2.2X, 2.4X, 106X, 4.8X, 3.7X, 1602X,

1527X, 293X faster than the methods of [14], [24], [25], [29], [30], [31], [32], [39] and [40], respectively. The proposed method performs 59.45% time reduction compare to window-based method by a fast algorithm [25]. From Table II, we find that a reduction of 55.88% computation time comparing with the proposed SAA method and Hierarchical Disparity Estimation [24]. The SAA method also achieves the reduction of 42.02% computation time comparing with fast Area Based Stereo Matching Algorithm [14].

Finally the SAA method achieves 99.06%, 79.20%, 73.25% and 99.93% computational time reduction against the recent state-of-the-arts methods of [29], [30], [31], [32] and [39], respectively using the lower configurations of hardware. The graphical comparison of computational costs and speed (in *fps*) are illustrated in Fig. 15. So from the Table I, Table II and graph of Fig. 15, the proposed SAA algorithm is the better choice on the basis of trade-off between computational time and speed.

TABLE I. DISPARITY ESTIMATION COMPUTATIONAL TIME (IN  $\mu$ S) AND SPEED (IN FPS) FOR THE MIDDLEBURY STANDARD DATA OF TSUKUBA HEAD IMAGE USING SELF-ADAPTIVE SEARCH METHOD

Methods' Name	Computational Time( $\mu$ s)	Speed ( <i>fps</i> )	Computational Machine	Input image & Resolution	Rank
Self-Adaptive Algorithm [Proposed]	1872	535	Intel Core i3 Speed: 2.3 GHz. RAM: 4GB	Middlebury Standard Tsukuba Head 384x288	1
Fast Area Based [14]	3229	310	Intel Core i3 Speed: 2.3 GHz. RAM: 4GB		2
Hierarchical Disparity [24]	4243	235	Intel Core i3 Speed: 2.3 GHz. RAM: 4GB		3
Fast Algorithm [25]	4617	216	Intel Core i3 Speed: 2.3 GHz. RAM: 4GB		4
Tree filtering [31]	7000 (7ms)	143	Intel Core-i7 Speed: 1.8 GHz. RAM: 4 GB		5
Edge-aware & Geodesic filter[30]	9000 (9ms)	111	Intel Core-i5+Geforce GTX card Speed: 3.0GHz. RAM: 8GB		6
DSI & Adaptive Support[29]	200000 (0.2s)	5	Core Duo Speed: 2.2GHz. RAM:NA		7
Pyramid stereo matching [40]	550000(0.55s)	2	Nvidia GeForce GTX 1080 Ti/PCIe/SSE2	KITTI [2012] Data set.	8
Deep self-guided[39]	2860000(2.86s)	0.35	Intel Core- i7-4770 Speed:3.4GHz. RAM.16GB RAM: 16 GB	Middlebury training data set and KITTI [2015]	9
Energy Minimization[32]	3000000 (3s)	0.33	Intel Core- i5 Speed: 1.9 GHz. RAM:6GB	Middlebury Standard data set	10
Fusing Adaptive Support[38]	40500000(40.5s)	0.025	Intel Core- i5-6500 Speed: 3.2 GHz. RAM: 8GB	Tsukuba, Venus, Teddy, Cones [Middlebury Benchmark]	11

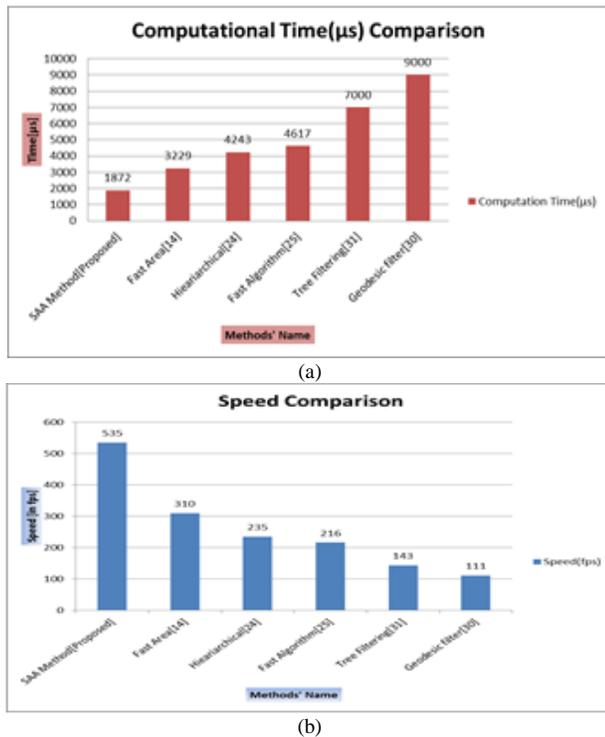


Fig. 15. Upper Side Graph(a) shows the Comparison of Computational Costs and Lower Side Graph(b) shows the Comparison of Speed (in fps) among Proposed and existing State-of-the-Art Methods for Tsukuba Head Image.

C. Experiment 3: Computational Cost Calculation and Comparison with Existing State-of-the-Arts Methods for Middlebury Standard Venus Stereo Image

The computational time and speed performances have been numerically evaluated in Table III for Middlebury standard data of Venus stereo pair. The proposed SAA algorithm also outperforms all other algorithms summarized in Table III. Table III illustrates the comparison of computational costs, speed (fps) of proposed method and other recent state-of-the-arts methods with hardware specifications. The proposed SAA algorithm also obtains the better performance in computational time and speed. It requires only 2652  $\mu s$  instead of 6318  $\mu s$ , 6724  $\mu s$ , 7000  $\mu s$ , 7473  $\mu s$  and 9000  $\mu s$  respectively but with a higher potential speed of 377 fps. The graphical comparisons of time and speed are illustrated in Fig. 16 which depicts the

proposed SAA method runs in lowest time and achieves the highest speed among the current state-of-the-arts methods. Another numerical evaluation and comparisons are depicted in Table IV in which proposed method performs 64.51% time reduction compared to window-based method by a fast algorithm [25] for Venus stereo pair image. From Table IV 60.55% computational time is reduced by proposed SAA method compared to the Hierarchical Disparity [24]. The algorithm also performs the reduction of 58.02% computational time comparing with the Fast Area based method [14].

So it is evident that the proposed method is currently the best state-of-the-arts method for Middlebury standard Venus stereo pair with 2.3X, 2.5X, 2.8X, 3.3X, 2.6X faster than the top five methods of [14], [24], [25], [30] and [31], respectively.

Finally the SAA method achieves 70.53% and 62.11% computational time reduction against the recent state-of-the-arts methods of [30] and [31] respectively using the lower configurations of hardware. So from the Table IV and graph of Fig. 16, the proposed SAA algorithm again proves that it is the best choice on the basis of trade-off between computational time and speed. In both cases (Tsukuba and Venus stereo pair) the SAA algorithm performs the lowest computational costs and highest speed.

D. Experiment 4: Accuracy Measurement and Comparisons with Three Established Methods

The accuracy of this algorithm has been justified over standard stereo images of Tsukuba head. Experiments are performed on Intel Core i3, 2.3 GHz processor with 4 GB DDR3 RAM. The algorithm has been implemented using Visual C++ programming language. To determine the correspondence of a pixel of reference image, the window costs are estimated for the candidates' pixels of right image within the search range -10 to +10 pixels. The proposed SAA method estimates the accuracy in percentage with the error threshold 1. The accuracy of SAA method is 93.8%. The numerical evaluations confirm that the bad pixel in percentage is only 6.2% with error threshold 1 for proposed method. But using the same resolution of image, bad pixels in percentage were 6.33%, 7.88% , and 7.18% reported in [30], [31] and [32] respectively for Tsukuba head with the experiments of Middlebury stereo data set.

TABLE II. COMPUTATIONAL TIME REDUCTION (IN %) OF PROPOSED METHOD FOR TSUKUBA HEAD

Computational Time( $\mu s$ ) for Self-Adaptive Algorithm [Proposed]	Existing state-of-the-art methods		Computational Time Reduction (%) by SAA method compared to the methods of 2 <sup>nd</sup> column
	Methods' Name	Computational Time( $\mu s$ )	
1872	Fast Area Based [14]	3229	42.02
	Hierarchical Disparity [24]	4243	55.88
	Fast Algorithm[25]	4617	59.45
	Tree filtering [31]	7000 (7ms)	73.25
	Edge-aware & Geodesic filter[30]	9000 (9ms)	79.20
	DSI & Adaptive Support[29]	200000 (0.2s)	99.06
	Energy Minimization[32]	3000000 (3s)	99.93
	Pyramid stereo matching[40]	550000(0.55s)	99.65
	Deep self-guided[39]	2860000(2.86s)	99.93

TABLE III. DISPARITY ESTIMATION COMPUTATIONAL TIME (IN  $\mu$ S) AND SPEED (IN FPS) FOR THE VENUS STEREO IMAGE USING SELF -ADAPTIVE SEARCH METHOD

Methods' Name	Computing Time( $\mu$ s)	Speed (fps)	Computing Machine	Input image & Resolution	Rank
Self-Adaptive Algorithm [Proposed]	2652	377	Intel Core i3 Speed: 2.3 GHz. RAM: 4GB	Middle -bury Standard Venus Stereo data set 434x383	1
Fast Area Based [14]	6318	158	Intel Core i3 Speed: 2.3 GHz. RAM: 4GB		2
Hierarchical Disparity [24]	6724	148	Intel Core i3 Speed: 2.3 GHz. RAM: 4GB		3
Tree filtering [31]	7000 (7ms)	143	Intel Core-i7 Speed: 1.8 GHz. RAM:4GB		4
Fast Algorithm [25]	7473	133	Intel Core i3 Speed: 2.3 GHz. RAM: 4GB		5
Edge-aware & Geodesic filter[30]	9000 (9ms)	111	Intel Core-i5+Geforce GTX card Speed: 3.0Ghz RAM: 8GB		6

TABLE IV. COMPUTATIONAL TIME REDUCTION (IN %) OF PROPOSED METHOD FOR VENUS STEREO PAIR

Computational Time( $\mu$ s) for Self-Adaptive Algorithm [Proposed]	Existing state-of-the-art methods		Computational Time Reduction (%) by SAA method compared to the methods of 2 <sup>nd</sup> column
	Methods' Name	Computational Time( $\mu$ s)	
2652	Fast Area Based [14]	6318	58.02
	Hierarchical Disparity [24]	6724	60.55
	Tree filtering [31]	7000 (7ms)	62.11
	Fast Algorithm[25]	7473	64.51
	Edge-aware & Geodesic filter[30]	9000 (9ms)	70.53

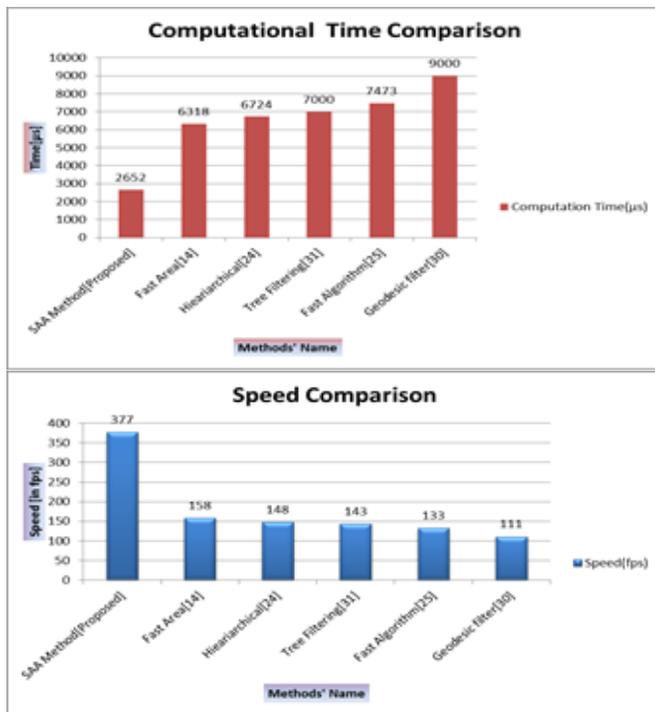


Fig. 16. Upper Graph shows the Comparison of Computational Costs and Lower Graph shows the Comparison of Speed (in fps) among Proposed and Existing State-of-the-Arts Methods for Venus Stereo Images.

## VII. PERFORMANCE ENHANCEMENT ANALYSIS

The performance of the proposed SAA approach has been compared to the state-of-the-arts methods. The comparison tools are computational time, speed and gain. Our target was to speed up computational costs with no degradation of accuracy. Since the state-of-the-arts algorithms yield very similar accuracies and 3D dense disparity maps confirm the effectiveness of the proposed matching algorithm. Table V illustrates the performance enhancement of SAA method compared to the established state-of-the-art methods. Proposed SAA method shows 72% gain enhancement compared to Fast Area Based Method. Enhanced gain is calculated as follows.

Gain Enhancement by proposed SAA method against Fast Area Based Method [14]:

$$= \{(\text{Computational time of Fast Area Based Method} \div \text{Computational time of SAA Method}) \times 100 - 100\} \%$$

$$= \{(3229/1872) \times 100 - 100\} \%$$

$$= 72\%$$

Similarly the gain enhanced by proposed SAA method against the state-of-the-art methods in [24], [25], [30] and [31] is 126%, 146%, 380% and 273% respectively which is illustrated in Fig. 17. Fig. 17 shows the performance enhanced graph of proposed SAA method. Since the speed of proposed method is very high, the estimated gains of this algorithm are automatically high. The numerical measurements of this claim is

strongly supported by experimental data of Table V in which the SAA algorithm is compared with the top five (5) algorithms. The bar diagrams illustrate the increased gain (in %) of SAA method compared to the current methods. The proposed method outperforms the existing state-of-the-art methods with respect to the performance enhancements.

Accordingly, the performance of the SAA method has also been tested on Middlebury standard Venus stereo data set and the estimated experimental data are summarized in Table VI. From this table the proposed SAA method also performs better than the existing top five algorithms for Venus stereo pair data set.

The gains have been estimated following the same procedures as mentioned earlier and the SAA method enhances 138%, 153%, 181%, 239%, 163% gains compared to established state-of-the-art methods of [14], [24], [25], [30] and [31], respectively.

The SAA method enhances 239% gain over Geodesic filter method [30], but it actually increases (239-100)= 139 % gain against Geodesic filter method [30]. Similarly, the proposed

method enhances 163% gain over Tree Filtering [31] method and it actually increases (163-100) = 63% gain against Tree Filtering [31] method. So in the case of Venus data, the proposed SAA algorithm also outperforms the existing top five algorithms. In terms of computational time, speed and gain achievement, the proposed matching approach is faster and better than the existing state-of-the-art methods.

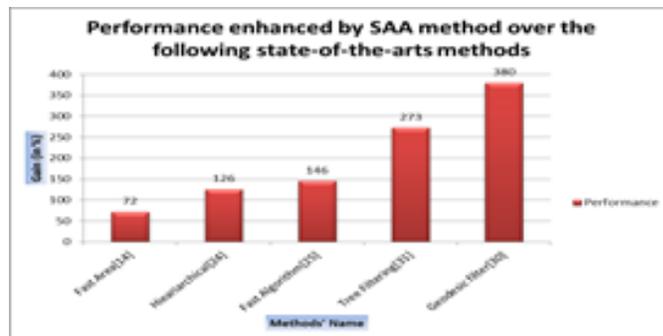


Fig. 17. Graph shows the Performance Enhancement of SAA Method Comparing to Existing State-of-the-Art Methods.

TABLE V. QUANTITATIVE EVALUATION OF PERFORMANCE OF PROPOSED SAA METHOD WITH TOP FIVE (5) ALGORITHMS

Methods' Name	Computational Time( $\mu$ s)	Speed (fps)	Performance Enhanced (%) by SAA method compared to the methods of first column	Input image & Resolution
Self-Adaptive Algorithm [Proposed]	1872	535	x	Middle-bury Standard Tsukuba Head 384x288
Fast Area Based [14]	3229	310	72	
Hierarchical Disparity [24]	4243	235	126	
Fast Algorithm[25]	4617	216	146	
Tree filtering [31]	7000 (7ms)	143	273	
Edge-aware & Geodesic filter[30]	9000 (9ms)	111	380	

TABLE VI. QUANTITATIVE EVALUATION OF PERFORMANCE OF PROPOSED SAA METHOD WITH TOP FIVE (5) ALGORITHMS FOR VENUS STEREO PAIR

Methods' Name	Computational Time( $\mu$ s)	Speed (fps)	Performance enhanced (in %) by SAA method compare to the methods of first column	Input image & Resolution
Self-Adaptive Algorithm [Proposed]	2652	377	x	Middlebury Standard Venus Stereo data set. 434x383
Fast Area Based [14]	6318	158	138	
Hierarchical Disparity [24]	6724	148	153	
Tree filtering [31]	7000 (7ms)	143	163	
Fast Algorithm[25]	7473	133	181	
Edge-aware & Geodesic filter[30]	9000 (9ms)	111	239	

## VIII. CONCLUSION

The main objective of this paper is to speed up the computational time. We have done this by a new technique called Self-Adaptive Algorithm that infers the upcoming matching pixel's position. This algorithm itself readjusts as well as reduces the search range of candidate's pixels based on the previous matched pixel's position. The speed of our algorithm is 535 fps for input images of Tsukuba head pair and 377 fps for input images of Venus stereo pair. Thus it can scan, calculate, process and display output 535 frame/second for the case of standard Tsukuba head image pair and 377 frame/second for the standard Venus stereo pair. The estimated gains of proposed SAA method are 380 and 239 for Tsukuba head and Venus stereo respectively whereas the gain of existing state-of-the-arts- methods is 100. The estimated accuracy is almost same as the accuracies of previous methods. Since the accuracy and 3D dense disparity maps of proposed method are very similar with the existing state-of-the-art algorithms, it confirms the effectiveness of the proposed matching algorithm. Moreover, the SAA algorithm does not require additional programmable 3D hardware like 3D Graphics Processing Unit (GPU). The proposed SAA method demonstrates the state-of-the-arts results and outdoes the existing top methods. In future we will consider fuzzy based technique to extend the research works for better realization of its behaviors.

## REFERENCES

- [1] S. T. Barnard and M. A. Fischler, "Stereo vision," in Encyclopedia of Artificial Intelligence, pp. 1083-1090. John Wiley, New York, 1987.
- [2] W. Hoff and N. Ahuja, "Surfaces from stereo: Integrating feature matching, stereo correspondence estimation and contour detection," IEEE Trans. Pattern Anal. Machine Intell., vol. 11, no. 2, pp.121-136,1989.
- [3] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," IEEE Trans. Pattern Anal. Machine Intell., vol. 16, no. 9, 1994.
- [4] Olga Veksler, "Stereo matching by compact windows via minimum ratio cycle," Proceedings of the IEEE International Conference on Computer Vision (ICCV),2001, pp. 540-547.
- [5] Md.Abdul Mannan Mondal, Md.Al-Amin Bhuiyan "Disparity Estimation By A Two-Stage Approximation Real Time Algorithm" Proceedings of the International Management and Technology Conference (IMT), 2004, pp. 12-17.
- [6] S. S. Intille and A. F. Bobick, "Stereo correspondence-space images and large occlusion stereo," Proceedings of the European Conference on Computer Vision (ECCV), 1994, pp. 179-186.
- [7] A. Fusiello and V. Roberto, "Efficient stereo with multiple windowing," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (ICVPR),1997, pp. 858-863.
- [8] K. Muhlmann, D. Maier, J. Hesser, R. Manner, "Calculating dense stereo correspondence maps from color stereo images, an efficient implementation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (ICVPR), 2001, pp. 30-36.
- [9] Md.Abdul Mannan Mondal, Md. Mustafa Kamal, Md. Al- Amin Bhuiyan,"Stereo correspondence Estimation By Real Time Approximation Algorithm Using Virtual Masking System.," Proceeding of 8th International Conference on Pattern Recognition and Information Processing (PRIP'05),2005.
- [10] Md. Mozammel Hoque Chowdhury, Md.Abdul Mannan Mondal, Md.Al-Amin Bhuiyan, "3D Imaging with stereo vision", Proceeding of 7th International Conference on Computer & Information Technology(ICCIT), 2004, pp. 307-312.
- [11] Md.Abdul Mannan Mondal, Md. Mustafa Kamal, Md.Ariful Hyder Md.Al-Amin Bhuiyan, "Stereo Correspondence Estimation by Diagonal Mask Searching ", Proceeding of 8th International Conference on Computer & Information Technology, (ICCIT), 2005, pp. 379-383.
- [12] Md.Abdul Mannan Mondal, Md. Mustafa Kamal, , Md. Mozammel Hoque Chowdhury , Md.Ariful Hyder ,A.K.M Zaidi Satter ,Md.Al-Amin Bhuiyan, "Stereo Correspondence Estimation: A Two-Stage Approximation Algorithm Using Arbitrary Window Pixel ", Proceeding of 8th International Conference on Computer & Information Technology, (ICCIT), 2005, pp. 384-387.
- [13] D.Schartein ,R.Szeliski, " A taxonomy and evaluation of dense two frame stereo correspondence algorithms", Int. J. of Comp. Vission,Vol. 47, no.1-3,pp.7-42,2002.
- [14] Lugi Di Stefano,Massimiliano Marchionni, Stefano Mattocchia, "A fast Area Based Stereo Matching Algorithm" , Image and vision Computing 22,pp.983-1005,2004.
- [15] Elisabetta Binaghi, Ignazio Gallo, Giuseppe Marino, Mario Raspanti, "Neural adaptive stereo matching", Pattern Recognition Letters, Vol.25,pp. 1743-1758,2004.
- [16] Abijit S.Ogale and Yiannis Aloimonos , "Shape and the Stereo Correspondence Problem", Int. J. of Comp.Vission, Vol.65,no.3,pp.147-167,2005.
- [17] Sukjune Yoon,Sung-Kee Park, Sungeul Kang, Yoon Keun Kwak, "Fast correlation-based stereo matching with the reduction of systematic errors", Pattern Recognition Letters,Vol. 26, pp. 2221- 2231,2005.
- [18] L. Kotoulas, A. Gasteratos,G.Ch.Sirakoulis , C. Georoulas, and I. Andreadis, "Enhancement of Fast Acquired Disparity Maps using a 1-D Cellular Automation Filter", proceedings of the fifth IASTED International Conference on Visualization, Imaging and Image Processing, 2005, pp.7-9.
- [19] Kuk-Jin Yoon and In So Kweon, "Adaptive Support -Weight Approach for Correspondence Search", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.28,no. 4, pp.650-656, April 2006.
- [20] P.H.S. Torra, A. Criminisi, "Dense stereo using pivoted dynamic programming", Image and Vision Computing Vol.22, pp. 795-806, 2004.
- [21] Lazaros Nalpantidis,Georgios Ch. Sirakoulis and Antonios Gasteratos, "Rievview of stereo matching algorithms for 3D vision", 16th International Symposium on Measurement and Control in Robotics ISMCR, 2007, pp.116-124.
- [22] Salvador Gutierrez,Jose Luis Marroquin, "Robust approach for disparity estimation in stereo vision", Image and Vision Computing, Vol.22, pp. 183-195, 2004.
- [23] Michael Bleyer, Margrit Gelautz, " A layered stereo matching algorithm using image segmentation and global visibility constraints", ISPRS Journal of Photogrammetry and Remote Sensing Vol.59,pp.128-150,2005.
- [24] Christopher Zach,Konrad Karner,Horst Bischof,"Hierarchical Disparity Estimation with Programmable 3D Hardware " Proc. of Int. Conf. in Central Europe on Computer Graphics, Visualization and Computer Vision, 2004, pp.275-282.
- [25] Mohammad Shorif Uddin "Stereo correspondence estimation using window-based methods by a fast algorithm", Journal of Electronics and Computer Science,vol.4,pp.5-11,2003.
- [26] D. Min, J. Lu, and M. N. Do, "A revisit to cost aggregation in stereo matching: How far can we reduce its computational redundancy?" Proc.of IEEE ICCV, Nov. 2011. pp. 1567-1574.
- [27] D.Min, J. Lu, and M. N. Do, "Joint histogram-based cost aggregation for stereo matching," IEEE Trans. Pattern Anal. Mach. Intell., vol.35, no. 10, pp. 2539-2545, Oct. 2013.
- [28] A. Hosni, C.Rhemann, M. Bleyer, C. Rother, and M. Gelautz. Fast cost volume filtering for visual correspondence and beyond . PAMI,35,pp.504-511, 2013.
- [29] Nadia Baha & Slimane Larabi "Accurate real-time disparity map computation based on variable support window " International Journal of Artificial Intelligence & Applications (IJAIA), Vol.2, No.3, pp. 22-33, July 2011.

- [30] Xun Sun , Xing Mei , Shaohui Jiao , Mingcai Zhou , Zhihua Liu , Haitao Wang “Real-time local stereo via edge-aware disparity propagation” *Pattern Recognition Letters* , vol. 49,pp. 201-206, 2014.
- [31] Q. Yang, “Stereo matching using tree filtering” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 4,pp. 834-846, 2015.
- [32] Mikhail G. Mozerov & van de Weijer, “Accurate stereo matching by two-step energy minimization”, *IEEE Transactions on Image Processing*,vol. 24,no. 3,pp. 1153-1163, March 2015.
- [33] Zbontar, J., & LeCun, Y. “ Computing the stereo matching cost with a convolutional neural network”, *Proceedings of the IEEE conference on computer vision and pattern recognition*, October 2015, pp.1592-1599.
- [34] Zbontar, J., & LeCun, Y. “Stereo matching by training a convolutional neural network to compare image patches”. *Journal of Machine Learning Research*, vol.17, no.2, pp. 1-32, April 2016.
- [35] Md. Abdul Mannan Mondal & Md. Haider Ali, “Disparity Estimation by a Real Time Approximation Algorithm” *International Journal of Image Processing (IJIP)*, Vol. 10 : Issue -3, pp.126-134, 2016.
- [36] Md. Abdul Mannan Mondal & Md. Haider Ali, “Performance Review of the Stereo Matching Algorithms” *American Journal of Computer Science and Information Engineering*,vol. 4,no.1,pp- 7-15, 2017.
- [37] D. Scharstein, R. Szeliski, *Middlebury stereo evaluation-version 2*, 2010, <http://vision.middlebury.edu/stereo/eval/>.
- [38] Wenhuan Wu,Hong Zhu,Shunyuan Yu and Jing Shi “ Stereo Matching with Fusing Adaptive Support Weights” *IEEE Access*,vol.7 ,pp. 61960-61974, May 2019.
- [39] Williem , In Kyu Park “Deep self-guided cost aggregation for stereo matching” *Pattern Recognition Letters*,vol. 112,pp. 168-175,September 2018.
- [40] Jia-Ren Chang and Yong-Sheng Chen “Pyramid Stereo Matching Network” *Proceeding of IEEE conference Computer vision and pattern recognition*, June 2018, pp. 5410-5418.
- [41] Siti Safwana Abd Razak, Mohd Azlishah Othman, Ahmad Fauzan Kadmin “The effect of Adaptive Weighted Bilateral Filter on Stereo Matching Algorithm” *International Journal of Engineering and Advanced Technology*, vol. 8, Issues 3, pp.284-287, February 2019.

# Natural Language Processing based Anomalous System Call Sequences Detection with Virtual Memory Introspection

Suresh K. Peddoju<sup>1</sup>, Himanshu Upadhyay<sup>2</sup>  
Applied Research Center  
Florida International University  
Florida, USA

Jayesh Soni<sup>3</sup>, Nagarajan Prabakar<sup>4</sup>  
School of Computing and Information Sciences  
Florida International University  
Florida, USA

**Abstract**—Malware has become a significant problem for the security of computers in this scientific era. Nowadays, machine learning techniques are applied to find anomalous activities in computers especially in virtualization environments. Identifying anomalous activities in virtual machines with virtual memory introspector and analyzing data with machine learning techniques are need of current trend. In this paper, an anomaly detection method is implemented using Natural Language Processing (NLP) based on Bags of System Calls (BoSC) for learning the behavior of applications on Windows virtual machines running on Xen hypervisor. During this process, system call traces are extracted from normal applications (benign processes) and malware affected applications (malicious processes) with the help of virtual memory introspection. Preprocessing of extracted system call sequences is done to obtain valid system call sequences through filtering and ordering of redundant system calls. Further, analysis of behavior of system call sequences is carried out with NLP based anomaly detection techniques. During this process, Cosine Similarity Algorithm (Co-Sim) is applied to identify malicious processes running on a VM. Apart from this, Point Detection Algorithm is applied to precisely locate the point of compromise in the system call sequences. The results shown in this paper indicates that both of these algorithms detect anomalies in the running processes with 99% accuracy.

**Keywords**—System call sequence; anomaly detection; natural language processing; memory forensics; cosine similarity

## I. INTRODUCTION

Nowadays, virtualization is playing a vital role in distributed systems. It became popular due to its usage and applicability. The significant advantage of virtualization is to provide vast resource sharing, load balancing, and protecting system resources. With the development of virtualization technologies, hypervisor-based methods have evolved to scan virtual machines (VM) and identify the threats happening on it. In the current market, the latest malware is more sophisticated and robust so that no malware detection techniques are capable of detecting and protecting the virtual machine. Thus, many organizations are facing cyber threats to their data and resources. Hypervisor-based malware detection techniques overcome these problems in comparison to host-based malware detection techniques. Virtual Machine Introspection (VMI) is the most versatile malware detection technique to monitor and

analyze cyber threats on virtual machines [1][2][3]. VMI is a technique to control the virtual machine run-time state at the hypervisor level, and it is used for forensic analysis of VM activities.

In hypervisor-based environment, it is important to observe virtual machine activities through hypervisor to keep track of benign and malicious activities happening on it. Memory forensics is good technique to extract and analyze memory activities. In this paper, we built a memory forensics architecture which uses VMI. All memory data structures are extracted (including system call sequences) to monitor anomalous activity in VM.

One of the techniques to identify the anomalous behavior of VM is to trace system call sequences of all running applications on VM. Hypervisor will extract system call sequences from memory of VM in runtime. Anomaly detection techniques are applied on collected data to find any anomalies in system call sequences by comparing benign and malicious data. This process will help in identifying the compromised VM on hypervisor. One of the efficient approaches for anomaly detection is Bag of System Calls (BoSC). Kang et al. in 2005 [4] introduces it as a frequency-based technique. According to this method, system call sequences  $S_i$  are represented as a list  $\{C_1, C_2, \dots, C_n\}$ , where  $n$  is the number of unique system calls, and  $C_i$  is count of system calls, present in the generated input sequence of system calls.

In this paper, we study the richness of using BoSC technique to detect malicious behavior at the process level in a hypervisor based environment. Further, we also propose an algorithm that detects anomalies at a particular point of time using cosine angle similarity. The results shows that considering the sequence of system call occurrences is powerful for detecting real-time anomalies in running processes on Xen hypervisor.

The outline of the paper is as follows: Section 2 describes state of the art related to proposed techniques. The subsequent section provides a system overview. Section 4 discusses the system call feature extraction and pre-processing. In the next section, we explain the proposed algorithm. Furthermore, we give an in-depth explanation of the environmental setup in section 6. The results of the proposed algorithms are presented in section 7. Finally, we conclude the paper in section 8.

## II. RELATED WORK

Classifying malware in any production system is of crucial importance for the security of its software components. Static analysis and dynamic analysis are two types of different malware analysis methods. Due to an increase in malware threats, there is a substantial increase in research work on malware detection.

In the static analysis method, we directly analyze source files without executing them [5]. Masud et al. [6], extracted 4-gram byte codes with five different static features of assembly instructions and combined them. For malware detection, they used two classification algorithms, namely decision tree algorithm and support vector machine. Ye et al. [7] used an association mining algorithm that generated association rules by developing an Intelligent Malicious code Detection System (IMDS) to obtain import function information. Finally, they used an association rule-based classification algorithm to detect malware.

However, techniques such as encryption, packing of malware, and polymorphism affect static based anomaly detection methods. Analyzing the behavior of an application is known as dynamic analysis. Its basic idea is to analyze the execution of the application [8]. This approach solves many of the problems of static-based analysis.

Many authors have used Hidden Markov Model (HMM) based classifier to detect anomalies in system calls [9][10][11][12][13][14]. However, each author uses a different set of techniques for improving the precision of anomaly detection. Alarifi and Wolthusen [15] took sequences from a virtual machine and then trained them using HMM. Their HMM-based method gave fewer detection rates since it required fewer training samples. The detection rate was 97% by using 780k system calls for training. Wang et al. [11] used the probability score and threshold value of the whole sequence. Cho et al. [13] used HMM by training regular user-level privilege operations. Hoang et al. [14] introduced an anomaly detection technique for multi-layer by using the sliding window approach. Warrender et al. [9] provide a comparison of STIDE [16], RIPPER [17], and HMM-based methods. These methods had different performance characteristics, while HMM performed with good accuracy. However, HMM requires multiple passes through the training data, high computational power, and needs large storage, especially for significant sequence length. Time series based modeling has been performed in [18][19]. The Kernel State Modeling (KSM) technique uses sequences of system call sequences as an individual task of kernel modules [20]. This method calculates the probability of occurrences of the finite number of states in malicious traces of system calls and compares against the expectations of normal traces. The KSM results in higher detection rates in comparison to HMM-based methods for UNM dataset. For feature extraction, neural-net based embedding is used for single dimensions data [21][22][23][24]. Suresh et al. [25][26] introduce machine learning algorithms for feature extraction for multidimensional data.

## III. SYSTEM OVERVIEW

The proposed framework and methodology is described in this section. This framework describes how system call sequences are extracted and analyzed by using a VMI based architecture and machine learning methods. This workflow collects system call traces of running processes and introspects the malicious behavior of processes on guest VM. The following subsections describe the architecture, methodology, and procedure to create custom malware.

### A. Architecture

The architecture of the proposed memory forensic framework, as shown in Fig. 1, consists of four modules: the Virtualization module, the Advanced Cyber Analytics module, the Malware repository module, and the Test Control Center module. The proposed framework acquires smart memory introspection features, analyzes them with advanced cyber analytics algorithms along with a control center for managing the system for visualizing the results.

The following sub-sections describe the functionality of individual modules and their components.

1) *Virtualization*: In this module, smart memory introspection is performed on Virtual Machine (VM) using VMI API to introspect and perform memory forensics. This module consists of different sub-modules such as Introspector and Security Agent.

a) *Introspector*: This module extracts low-level data from the memory of virtual machines running on a hypervisor, and transfers this data to agent listener(s) for anomaly analysis. The Introspector interfaces with hypervisors to ensure that the state of the virtual machines (running, stopped, or shut-down) can be manipulated, and VMs can be added and deleted as needed.

b) *Security Agent*: This sub-module initiates scans on VMs using the LibVMI library to perform introspection. Its primary mechanism is to extract data from a VM and send the data to the agent listener for further analysis. The Security Agent has various features that allow the agent to scan processes, invariant data structures, and to monitor files changes.

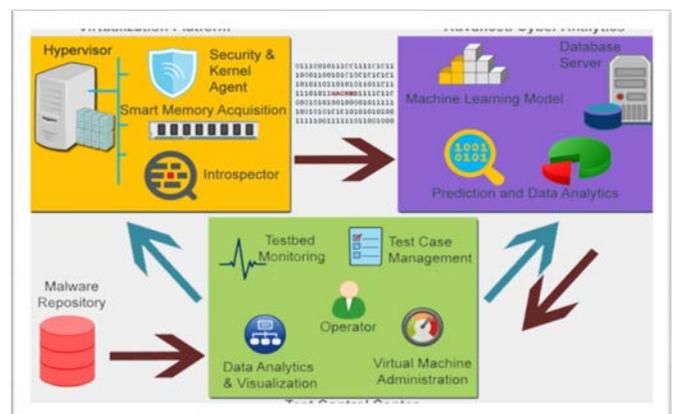


Fig. 1. System Call Traces with Virtual Machine Introspection.

2) *Advanced cyber analytics*: This module comprises of different machine learning and deep learning algorithms to train the model and perform a test on that model for further prediction and analysis of data. The baseline data is considered as benign data, and the test vector injected data is known as malicious data. The data extracted by using the introspection module is stored on a database server and then analyzed using different cutting-edge machine learning techniques.

3) *Malware repository*: This repository consists of a massive set of malware that compromises kernel-level data structures. This repository includes different malware for Windows and Linux. This malware repository also consists of custom malware sets to compromise the specific context of kernel data structures.

4) *Test control center*: With the help of the Test Control Center module, the operator can control and manage the whole framework and its modules with a user interface. The operator can handle the VM operations, such as creation, deletion, stop, start, pause, and view. Also, the operator can control the VMs by installing or running malware and benign applications. The operator visualizes the processed results from the Advanced Cyber Analytics module for further analysis.

### B. Methodology

In the current implementation of this framework, system call traces are collected from live VM using Virtual Memory Introspection method. An Introspector package developed on hypervisor which consists of two modules *introspector* and *security agent*. Among these modules, introspector module gets connected with the VM and initiates the security agent module to extract the system call traces from live memory of VM. Further, security agent sends extracted data to database with the help of other application called *Agent Listener*. This application intern stores information into database. Next step is to pre-process and analyze the collected system call traces using anomaly detection algorithms. In view of these, a custom application is designed to manage the VM, initiate the scanning, view results and many more. For further study, an operator can process these traces.

### C. Custom Malware

A set of custom malware were created to compromise system call sequences by way of DLL injection. This injection hooks into the write function of processes and initiates additional system calls by creating a hidden file on disk. This set of custom malware is used in experiments to compromise system call sequences.

## IV. FEATURE EXTRACTION TECHNIQUE

A process behavior is defined with an approach based on angle similarity. As part of this method, the occurrences of system calls generated by the process are considered, instead of the temporal ordering of system calls. This paper presents a technique called angle similarity which is similar to text classification for anomaly detection, where a sequence of system calls is considered as the document, and individual system calls are viewed as a word. The system-call sequence

are extracted under normal operation are collected from the hypervisor. Fig. 2 shows the sample sequence of system calls.

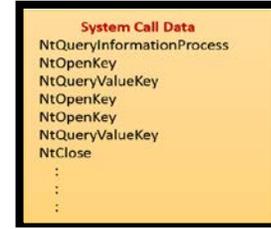


Fig. 2. Sample Sequence.

According to this approach, each and every system call is mapped to a unique number from 0 to 450 to a given sequence of the system calls. The total unique system calls for Windows is 450. A sample mapping of system calls is shown.

System Call Name	NtQueryInformationProcess	NtOpenKey	NtQueryValueKey	NtOpenKey	NtQueryValueKey	NtClose
Mapping Number	25	18	23	18	23	15

We create a Bag of System Calls of 450 dimensions where each cell value designates the frequency of the  $i^{th}$  system call. The following Fig. 3 shows a sample Bag of System calls:

	0	1	2	3	4	5	6	7	8	9	...	440	441	442	443	444	445	446	447	448	449	
0	1	0	0	0	6	10	17	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0

Fig. 3. BoSC of 450 Dimensions.

## V. DETECTION ALGORITHM

The proposed approach computes the cosine similarity between the features from normal processes and malicious processes. Cosine similarity is a similarity measure between two vectors that calculate the cosine angle between them.

The cosine angle between two vectors is calculated using their Euclidean dot product. Equation 1 shows the Euclidean dot product.

$$A \cdot B = \|A\| \|B\| \cos \theta \quad (1)$$

Given two vectors of n dimensions, A and B, the cosine similarity value is calculated as the function of  $\cos(\theta)$  shown in Equation 2:

$$\text{Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (2)$$

Where  $A_i$  and  $B_i$  are the features of vectors A and B respectively in the equation.

### A. Anomaly Detection Algorithm

The following algorithm detects anomalies in the running processes in Windows VM on Xen hypervisor.

For a given set of processes in baseline and test data, use its system-call sequences and mapping table to map system-call name to number. An anomalous system call sequences can be detected by using Algorithm #1, which is shown.

### B. Point Detection Algorithm

A Point detection algorithm detects a particular point in the process execution where the malicious attack has happened.

Sequence length is the number of the system calls taken into consideration. Sequence length of the system call is provided as input to the Point Detection Algorithm as given below in Algorithm #2, BoSC of an anomalous process from the above anomaly detection algorithm #1, and BoSC of a normal process.

For point detection algorithm, we use a sliding window of varying lengths and calculate the cosine similarity for that particular window. If the cosine similarity is less than 0.99, then that process within that window is considered as anomalous. Fig. 4 depicts the point detection method.

Algorithm 1: Anomaly detecting process for system call sequences.

```

Algo #1   Anomaly Detection Algorithm



---


Step 1:  for each process  $B(\mathcal{P}_i)$  in Baseline  $\{ \mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \dots, \mathcal{P}_n \}$  do
    Map  $S_i \rightarrow S_{num}$  using the System Call mapping table.
    Where  $S_i$  is system call name and  $S_{num}$  is System call mapping number
    Create bag-of-words for  $\mathcal{P}_i$  by counting occurrences of each  $S_i$ 
Step 2:  for each process in  $T(\mathcal{P}_i)$  the TestData  $\{ \mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \dots, \mathcal{P}_n \}$  do
    Prepare combined list  $C(\mathcal{P}_i)$  from  $B(\mathcal{P}_i) \cap T(\mathcal{P}_i)$ 
Step 3:  for each process  $C(\mathcal{P}_i)$  in combined list  $\{ \mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \dots, \mathcal{P}_n \}$  do
    if  $C(\mathcal{P}_i) \neq$  Mapping table  $M(t)$  then
        P is Anomalous;
    else
        Map  $S_j \rightarrow S_{num}$  using the System Call mapping table.
        Create bag-of-words for  $\mathcal{P}_j$  by counting occurrences of each  $S_j$ 
        Similarity =  $f(B(\mathcal{P}_i), B(\mathcal{P}_j))$ 
        if Similarity < 0.99 then
            P is Anomalous;
        else
            P is Normal;
        end do
    end do

```

Algorithm 2: Point detection for system call sequences

```

Algo #2   Point Detection Algorithm



---


Step 1:  for each Anomalous Process  $A(\mathcal{P}_i)$  and Baseline process  $B(\mathcal{P}_i)$  do
    for each  $i$  in range  $[\text{length}(\text{BoSC of } A(\mathcal{P}_i)) - (\text{sequence\_length})]$  do
     $\text{Seq}(B(\mathcal{P}_i)) = \text{BoSC\_of\_}B(\mathcal{P}_i)[i:i+\text{sequence\_length}]$ 
     $\text{Seq}(A(\mathcal{P}_i)) = \text{BoSC\_of\_}A(\mathcal{P}_i)[i:i+\text{sequence\_length}]$ 
     $\text{SimVal} = \text{Cosine\_Similarity}(\text{Seq}(B(\mathcal{P}_i)), \text{Seq}(A(\mathcal{P}_i)))$ 
    if  $\text{SimiVal} < 0.99$  then
    Return  $A(\mathcal{P}_i), \text{Seq}(A(\mathcal{P}_i))$ 
    end do
    end do

```

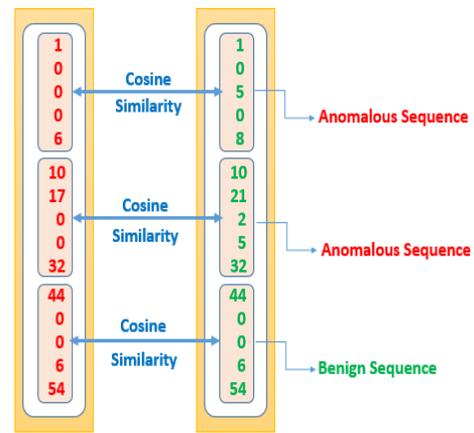


Fig. 4. Point Detection Method.

## VI. ENVIRONMENT SETUP

The proposed framework is developed on Xen 4.12 hypervisor and managed virtual machines (VM) with Libvirt 5.4.0 library. For getting memory addresses of running processes virtual machine introspector method are being imposed with latest version of DRAKVUF library. The current implementation of this framework consists of two modules namely *Introspector* and *Security Agent*. These modules extract system call traces by inspecting the VM called System Under Test (SUT) using the LibVMI library on top of DRAKVUF in combination with a recall profile of Google. This recall profile is files in JSON that comprises of memory mappings and offsets of windows data structures. The above two specified modules, are written in Go Language to process the request and extract the system call traces from VM with LibVMI functions. LibVMI library services and the Libvirt library are used to create, start, or stop virtual machines of windows. An application is designed for operator to extract the system call traces. This Application is written in Microsoft Visual Studio .NET framework and comprises of user-defined API calls for introspector communication and other related function calls. An agent transmits extracted data to the database server. Finally, the stored data is analyzed using different machine learning algorithms. The whole experimental setup is shown in Fig. 5.

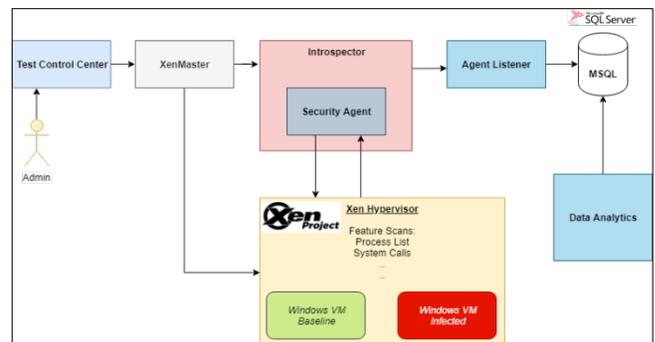


Fig. 5. Experimental Framework for Extracting System Call Traces.

VII. RESULTS

In this section, we present the results of the proposed algorithms.

A. Anomaly Detection Algorithm

We evaluated this algorithm with system-call traces of 1,000,000 system calls with multiple experiments. The total number of unique system calls in Windows operating system is 450. Fig. 6 and 7, display the top 5 system call with their frequencies of a normal SUT application and a malicious SUT application, respectively.

The result shown in Fig. 6 and 7 clearly differentiates between malicious SUT application system call frequencies in comparison with benign SUT application. The following Table I shows the similarity score between malicious and normal SUT application.

From Table I, we can say that the cosine similarity of a normal SUT Applications is higher whereas malicious SUT application is lower in compared with normal SUT application.

Furthermore, the cosine similarity value is independent of the number of records. Fig. 8 demonstrates this characteristic. We observe the same cosine similarity behavior even with the varying number of records.

B. Point Detection Algorithm

For the point detection algorithm, we tested with a sequence length of 3, 5, 10, and 15. From Fig. 9, we observe that sequence length of 5 gives an ideal cosine similarity value for a single scan.

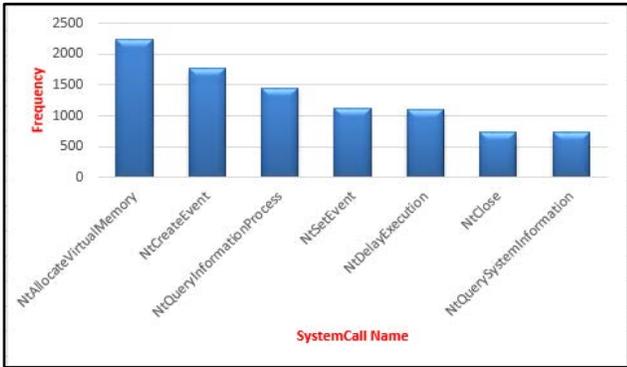


Fig. 6. Top 5 System Calls based on their Frequency of a Normal SUT Application.

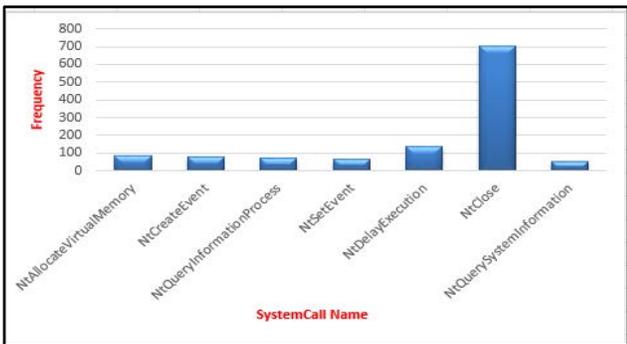


Fig. 7. Top 5 System Calls based on their Frequency of a Malicious SUT Application.

TABLE I. COSINE SIMILARITY BETWEEN NORMAL SUT AND MALICIOUS SUT APPLICATIONS

Applications	Normal SUT Application	Malicious SUT Application
Similarity Score	0.99	0.20

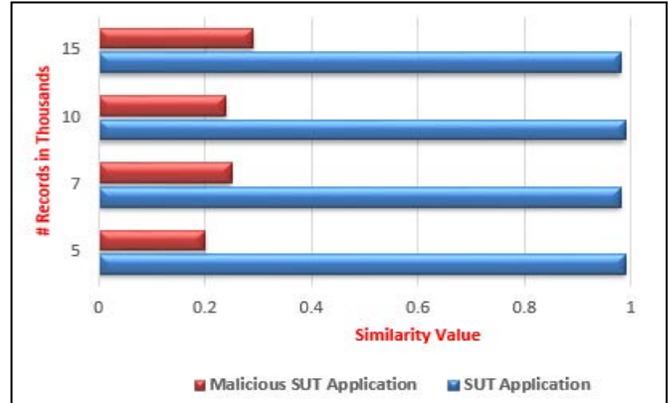


Fig. 8. Cosine Similarity Value for Normal SUT and Malicious SUT w.r.t # Records in Ten Thousand.

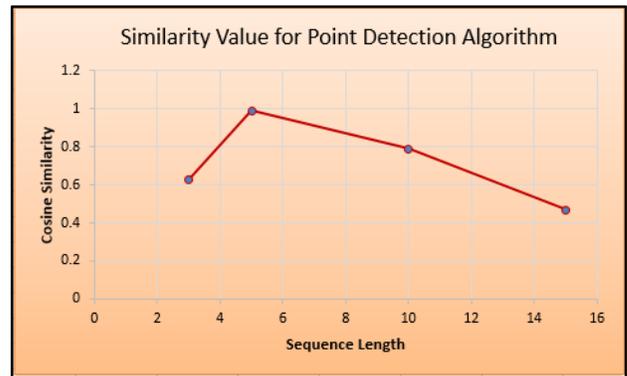


Fig. 9. Cosine Similarity Value w.r.t Sequence Length for a Single Scan.

Furthermore, we evaluated the algorithm with varying scan times. From Fig. 10, we found that with a sequence length of 5, the cosine similarity value is consistently higher in comparison with all other sequence lengths with varying scan times.

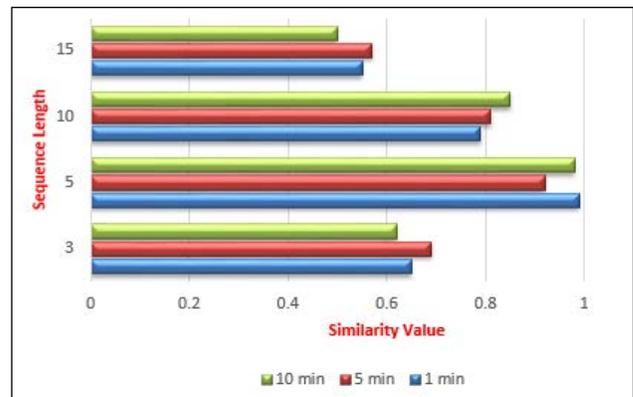


Fig. 10. Cosine Similarity Value w.r.t Sequence Length and Varying Scan Times.

### VIII. CONCLUSIONS

All intrusion-based detection algorithms work on the hypothesis that regular activities differ from irregular events (intrusions). Anomaly detection algorithms learn a program's behavior. The behavior is in the form of the frequency of system calls raised by the processes under evaluation. We presented two anomaly detection algorithms. Both algorithms calculate the cosine similarity between the processes under examination based on the frequency of system calls. Anomaly Detection Algorithm detects anomaly between benign and malicious system call sequences whereas point detection algorithm detects the timeframe of the malicious attack in the anomalous process. With the help of both of these algorithms we can able to detect malicious behavior of system call sequences with 99% accuracy rate.

### ACKNOWLEDGMENT

This work was funded by TRMC of DoD. We are very much thankful for providing facilities and infrastructure to do our experiments. We thank all who directly and indirectly helped us in doing this experiments and results.

### REFERENCES

- [1] M H Ligh, A Case, J Levy, A Walters. "The Art of Memory Forensics," 2014.
- [2] Xen Project available at <https://www.xenproject.org/>, 2013.
- [3] Hizver, Jennia, and Tzi-ckerChiueh. "Real-time deep virtual machine introspection and its applications." ACM SIGPLAN Notices. Vol. 49. No. 7. ACM, 2014.
- [4] D. Fuller and V. Honavar, "Learning classifiers for misuse and anomaly detection using a bag of system calls representation," in Proceedings of the Sixth Annual IEEE Systems, Man and Cybernetics (SMC) Information Assurance Workshop. IEEE, 2005, pp. 118–125.
- [5] Payet E, Spoto F. Static analysis of Android programs [J]. Information and Software Technology, 2012, 54(11): 1192-1201.
- [6] Masud M. M., Khan L, Thuraisingham B. A scalable multi-level feature extraction technique to detect malicious executables[J]. Information Systems Frontiers, 2008, 10(1):33-45.
- [7] Ye Y, Wang D, Li T, et al. IMDS: Intelligent malware detection system[C]//Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2007: 1043-1047.
- [8] Egele M, Scholte T, Kirda E, et al. A survey on automated dynamic malware-analysis techniques and tools[J]. ACM computing surveys (CSUR), 2012, 44(2): 6.
- [9] C. Warrender, S. Forrest, and B. Pearlmutter, "Detecting intrusions using system calls: alternative data models," in Security and Privacy, 1999. Proceedings of the 1999 IEEE Symposium on, 1999, pp. 133–145.
- [10] Soni, J., Prabakar, N., and Upadhyay, H. (2019) "Deep Learning approach to detect malicious attacks at the system level: poster," Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks. ACM, pp. 314-315.
- [11] W. Wang, X.-H. Guan and X.-L. Zhang, "Modeling program behaviors by hidden Markov models for intrusion detection," in Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on, vol. 5. IEEE, 2004, pp. 2830–2835.
- [12] D.-Y. Yeung, and Y. Ding, "Host-based intrusion detection using dynamic and static behavioral models," Pattern Recognition, vol. 36, no. 1, pp. 229–243, 2003.
- [13] S.-B. Cho and H.-J. Park, "Efficient anomaly detection by modeling privilege flows using a hidden Markov model," Computers and Security, vol. 22, no. 1, pp. 45 – 55, 2003.
- [14] X. D. Hoang, J. Hu, and P. Bertok, "A multi-layer model for anomaly intrusion detection using program sequences of system calls," in Proc. 11th IEEE Intl Conf. Networks, 2003, pp. 531–536.
- [15] Alarifi and Wolthusen "Anomaly detection for ephemeral cloud IaaS virtual machines," in Network and system security. Springer, 2013, pp. 321–335.
- [16] S. Forrest, S. Hofmeyr, A. Somayaji, and T. Longstaff, "A sense of self for Unix processes," in Proceedings of the 1996 IEEE Symposium on Security and Privacy, May 1996, pp. 120–128.
- [17] W. Lee and S. J. Stolfo, "Data mining approach for intrusion detection," in Usenix Security, 1998.
- [18] Soni, J., Prabakar, N. and Kim, J-H. (2017) "Prediction of Component Failures of Telepresence Robot with Temporal Data." 30th Florida Conference on Recent Advances in Robotics
- [19] G. S. Thejas, J. Soni, K. Chandna, S. S. Iyengar, N. R. Sunitha, and N. Prabakar. 2019. Learning-Based Model to Fight against Fake Like Clicks on Instagram Posts. In SoutheastCon 2019. Huntsville, Alabama, USA. In press.
- [20] S. S. Murtaza, W. Khreich, A. Hamou-Lhadj, and M. Couture, "A host-based anomaly detection approach by representing system calls as states of kernel modules," in Software Reliability Engineering (ISSRE), 2013 IEEE 24th International Symposium on. IEEE, 2013, pp. 431–440.
- [21] Soni, J., Prabakar, N., and Upadhyay, H. (2019) "Feature Extraction through Deepwalk on Weighted Graph," Proceedings of the 15th International Conference on Data Science (ICDATA'19), Las Vegas, NV, 2019.
- [22] Soni, J., Prabakar, N. (2018) "Effective Machine Learning Approach to Detect Groups of Fake Reviewers," Proceedings of the 14th International Conference on Data Science (ICDATA'18), Las Vegas, NV, 2018.
- [23] P. Suresh Kumar and S. Ramachandram, "Fuzzy based Integration of Security and Trust in Distributed Computing," Proc of Springer 7th International Conference Soft Computing for Problem Solving (SocProS'2017), Indian Institute of Technology, Bhubaneswar, December 2017.
- [24] P. Suresh Kumar, Himanshu Upadhyay and Shekar Bansali, "Health Monitoring with Low Power IoT Devices using Anomaly Detection Algorithm," IEEE conference SLICE-2019, Rome, Italy, June 2019.
- [25] P. Suresh Kumar and Pranavi S, "Performance Analysis of Machine Learning Algorithms on Diabetes Dataset using Big Data Analytics," Proc of IEEE 2017 International Conference on Infocom Technologies and Unmanned Systems (ICTUS'2017), Dubai, United Arab Emirates(UAE), December 2017. pp 580- 585.
- [26] A. Rishika Reddy and P. Suresh Kumar, "Predictive Big Data Analytics in Healthcare," Proc of IEEE 2016 Second International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, 2016, pp. 623-62.

# Parkinson's Disease Diagnosis using Spiral Test on Digital Tablets

Najd Al-Yousef<sup>1</sup>, Raghad Al- Saikhan<sup>2</sup>, Reema Al- Gowaifly<sup>3</sup>  
Reem Al-Abdullatif<sup>4</sup>, Felwa Al-Mutairi<sup>5</sup>, Ouiem Bchir<sup>6</sup>

College of Computer and Information Sciences  
Computer Science Department, King Saud University  
Riyadh, Saudi Arabia

**Abstract**—For a proper diagnosis, Parkinson's disease (PD) requires frequent visits to the doctor for physical tests, causing a huge burden on the patient. As PD impairs the handwriting ability, the handwriting pattern can be used as an indicator for PD diagnosis. More specifically, the Static Spiral Test (SST) and the Dynamic Spiral Test (DST), that consists in retracing spirals using digital pen. Such exam can be self-conducted by the patient, and thus it would be convenient and non-time-consuming for both the patient and the medical staff. In this project, we designed and implemented a system that automatically self-aid-diagnoses PD using SST and DST on digital tablets. The system includes two main components, image processing techniques to pre-process and extract the appropriate visual features and machine learning techniques to recognize PD automatically. The conducted experiment showed that the semi-local Edge Histogram Descriptor extracted from DST drawing, and conveyed to a Gaussian Kernel Support Vector Machine outperforms the other considered systems with an accuracy, specificity and sensitivity around 90%.

**Keywords**—Component; Parkinson's disease (PD); computer-aided diagnosis; pattern recognition

## I. INTRODUCTION

Parkinson's disease (PD) is a disorder that degenerates neurons which yields to failure of motor function because of smaller ratio of dopamine that is produced in brain [1], it affects the patient motor abilities such as speaking, writing, and walking. It has been estimated that around 7 to 10 million people worldwide have PD [2]. Its symptoms often appear gradually without being noticed by the patient. They are classified into those affecting movement (motor symptoms) and those that do not (non-motor symptoms).

Motor ones are easier to detect. One of the primary motor symptoms is the tremor which is an unintentional, rhythmic, slow muscle movement [3]. It occurs when the person is motionless and begins either in one hand, one foot, or one leg [2].

Even though PD cannot be cured, frequent monitoring is of high importance in order to obtain the proper diagnosis and help controlling the symptoms. Diagnosing Parkinson disease (PD) requires running different physical tests on the patient. These tests are performed by the physician in a clinic. Moreover, more than one visit is necessary to follow up with the patient's condition.

Given that PD is an age-related disorder where most patients are over 50 years old [4], who may have other age-related conditions, visiting the clinic may be an inconvenient task for them. Performing the tests remotely and comfortably at home, would decrease the burden of these tests and thus, encourage patients to do them. Moreover, that would decrease the load on the medical staff.

One solution to this problem could be to ask a patient to draw a certain pattern. Then, using image processing techniques extract the visual content of the drawn image and convey it to a machine learning algorithm that decides on the patient diagnosis.

Since PD is a type of movement disorder that impairs handwriting ability, we propose a system that will detect PD using the handwriting pattern. We will use both the traditional Static Spiral Test (SST) and the Dynamic Spiral Test (DST). SST requires the patients to follow a static spiral drawing. For the DST, the spiral that the patient is supposed to follow appears and disappears during the test time, so that the patient has to memorize the pattern and keep drawing it [5].

The proposed approach is intended to diagnose Parkinson's disease using both the static and the dynamic spiral tests independently and jointly. It will use image processing techniques to extract the needed feature from the spiral images and machine learning to automatically recognize Parkinson's disease.

For the feature extraction step, we intend to consider the following visual descriptors:

- Histogram of oriented gradients (HOG) [6],
- Edge Histogram Descriptor (EHD) [6],
- An application designed feature based on acceleration as described in [5], and
- Deep learning Auto-encoder generated feature.

For the classification step, we use Support Vector Machine (SVM) [7].

This paper is organized as follows: Section II presents and discusses the related works. Section III describes the methodology adopted in this research. Section IV sets the experiments that are conducted to assess the performance of the proposed system. Section V. reports the obtained

experimental results and analyze them. Lastly, Section VI concludes and summarizes the paper.

## II. RELATED WORKS

In the following we outline the computer-based approaches that are related to our work. They are either based on gait, voice or handwriting pattern.

### A. Gait based Approaches

Previous studies have shown that how the person walk can be a prediction of developing PD disease. In fact, Bridenbaugh and Kressig (2013) [8] concluded that there is an association between gait and cognition and that elderly people with gait impairments were more likely to develop cognitive impairments and problems in memory in addition to weaknesses in some processing functions.

A gait-based approach to detect PD disease is proposed in [9]. The step length can be estimated using the change in the waist height and the leg angle while walking. In order to record the gait characteristics such as step frequency and stride length, the Pedestrian Dead Reckoning (PDR) system and the smartphones' accelerometer sensor are used. The classification is done through SVM classifier [7].

### B. Voice based Approaches

Recent research has been conducted to study the connection between PD and speech impairment such as dysphonia [10]. The authors in [11], introduced an algorithm for PD diagnosis based on voice analysis. The test consists of pronouncing the letter "A" for 3 seconds. From the recorded sound; 22 features are extracted. They are based on the pitch, the jitter, the shimmer and the noise ratio. Then, feature selection is performed using genetic algorithm (GA) [12]. Afterwards, SVM classifier [7] is used to distinguish between PD and healthy subjects.

The authors in [13], developed a clinical expert system to detect PD from the subject's vocals. The system extracts three voice recordings of each subject pronouncing the letter "A" for 5 seconds and then uses a waveform matching algorithm [14] to extract 44 acoustic features which are based on: noise, pitch perturbation, amplitude of the spectral envelope measures and nonlinear ones. Then, it uses the Bayesian classifier [15] to distinguish PD patients from non-PD ones.

### C. Handwriting based Approaches

Drotár et al. (2016) [16] introduced a system of aided Parkinson's diagnosis based on kinematic characteristics and pressure in handwriting. The dataset of the proposed approach includes the task of drawing an Archimedean spiral. The entropy feature based on the pressure applied on the writing surface is extracted, then the SVM classifier is applied for the recognition task [7].

The authors in [17], suggested an approach to diagnose PD by using the handwriting pattern of the patient by conducting two tests, the spiral and the meander tests. The spiral test requires the subject to draw a spiral while the meander requires the patient to draw more structural lines. The structural Cooccurrence Matrix feature is extracted from both drawings. The features of the spiral test and those from

the meander test are conveyed separately to the classifier and also combined. Then three classifiers are used separately. They are Naïve Bayes [18], Optimum-Path Forrest OPF [19], and Support Vector Machine SVM [7]. The results showed that the highest accuracy rate is obtained when using spiral test and SVM classifier [7].

The authors in [5], proposed a static and dynamic spiral tests diagnosis approach for PD that uses the dynamic spiral test and the static spiral test referred to as DST and SST respectively. These two tests are conducted using an electronic equipment like a computer or a tablet.

SST consists in providing the subject with a static spiral test, the subject is then required to follow the pattern in order to draw a spiral. DST consists in providing the patient with the dynamic spiral. In other words, the spiral appears and disappears periodically. The subject has to continue drawing the pattern even if the model disappears. From both the DST and the SST spiral, the acceleration feature is extracted and then the dissimilarity of acceleration histograms (DAH) is computed. DAH is then used as an indicator, such that a small DAH means that the subject is healthy while a large one means he is a PD patient. We should mention here that the authors in [5] did not use any machine learning techniques to recognize PD patients. Moreover, they did not suggest a way of specifying a threshold allowing to discriminate between healthy and PD subjects.

## III. METHODOLOGY

In this project, we proposed and designed a pattern recognition system that is able to analyze the obtained results automatically. The main components of this system are:

- 1) Feature extraction, which translates the visual content of the drawn spiral image into a numerical vector.
- 2) Feature combination which allows to combine SST and DST tests.
- 3) Machine learning to learn a model that allows discrimination between the drawings of a PD patient and non-PD patient, and thus yields the categorization of the patients as PD or non-PD.

From the SST and DST drawn images, we extract four features which are:

- Histogram of Oriented Gradients [20],
- Edge Histogram Descriptor [21],
- Acceleration based descriptor [5],
- Auto-encoder generated descriptor [22].

In the following, we give a brief description of each component of the system and motivate its choice.

### A. Histogram of Oriented Gradients

Histogram of Oriented Gradients (HOG) is a feature descriptor that is being widely used in many applications [20], [23]. It describes the edges present in an image by using the gradient at each pixel. In fact, the gradient is a mathematical

tool that allows measuring the direction and amplitude of the change in the pixels' values.

The gradient of an image at pixel  $F(x,y)$  is defined as:

$$\nabla F = \left[ \frac{\delta F}{\delta x}, \frac{\delta F}{\delta y} \right]$$

Where,  $\nabla F$  is the gradient of the pixel,  $\frac{\delta F}{\delta x}$  is the partial derivative of the pixel with respect to x, and  $\frac{\delta F}{\delta y}$  is the partial derivative of the pixel with respect to y.

The gradient direction [24], that reflects the direction of the most rapid increase in the pixel intensity, can be expressed as:

$$\theta = \tan^{-1} \left[ \frac{\delta F}{\delta y} / \frac{\delta F}{\delta x} \right]$$

The gradient magnitude reflects the edge strength [24], and it can be expressed as:

$$\|\nabla F\| = \sqrt{\left(\frac{\delta F}{\delta x}\right)^2 + \left(\frac{\delta F}{\delta y}\right)^2}$$

The gradient descriptor is computed using the gradient filter. Each pixel and its neighborhood pixels are correlated using the filter in Fig. 1.

The following steps describe the HOG algorithm:

- 1) Compute the gradient vector using the filter in Fig. 4.
- 2) Compute the 9-bin histogram of the obtained direction ( $20^\circ$  is used for each direction) [20].
- 3) Compute the 9-bin histogram of the obtained amplitudes.
- 4) The HOG vector is then the concatenation of the sub histograms obtained in steps 3 and 4.

HOG describes the edges in an image by computing the gradient at each pixel. This can be achieved by using the HOG filter [24]. The obtained gradient magnitude and direction of each pixel are used to compute two 9-bin histograms.

For both SST and DST, the main difference between the spirals drawn by a PD patient and a non-PD patient is the shape of the spiral. Our main interest is the edges' oscillations of the drawn spirals. Since HOG descriptor reflects the edges in the image it is believed to discriminate between the structure/shape of the spiral between PD and non-PD patient.

0	0	0
$-\frac{1}{2}$	0	$\frac{1}{2}$
0	0	0

Fig 1. The Gradient Filter.

### B. Edge Histogram Descriptor

Edge Histogram Descriptor (EHD) uses histograms to describe local and global edge distribution [21]. It represents the shape content of an image. The edge of each pixel is classified into five types: vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional edges. The pixel edge type is determined by correlating each pixel and its neighboring pixels by edge detection filters. Fig. 2 shows the five filters corresponding to the 5 considered directions.

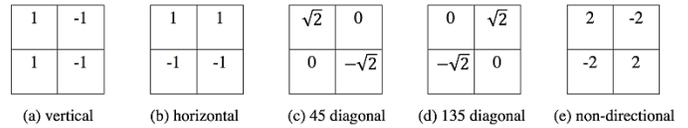


Fig 2. Edge Filters.

After filtering the image with the 5 filters, the 5 results are compared. Each pixel is assigned the direction of the filter type that has the maximum response. Then, three types of histograms are computed: global, local and semi-local histogram.

The global histogram is computed over the whole image and thus it is a 5-length vector, one entry for each direction. For the local histogram, the original image is partitioned into 4\*4 non-overlapping blocks called sub-image which yields a total of 16 sub-images as shown in Fig. 3.

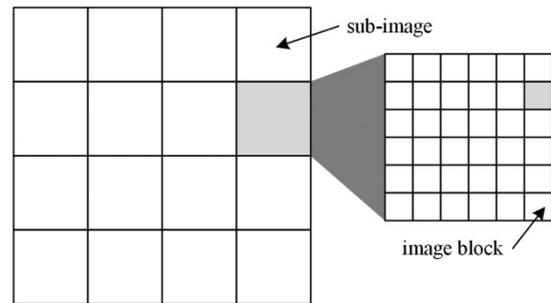


Fig 3. Image Partitioning into 16 blocks [13].

For each sub-image an edge histogram is generated to represent the frequency of occurrence of the different types of edges in each image-block. Generating 5-bin histogram for each of the 16 sub-images yields a total of 80 bin histogram.

The semi-local histogram is obtained by concatenating 13 histograms as follows:

- 4 Histograms are computed by summing the local histograms over each row of blocks (Fig. 7(a)).
- 4 other histograms are generated by summing the local histograms over each column of blocks (Fig. 7(b)).
- One histogram is computed by summing the 4 upper-left sub-blocks' local histograms (Fig. 7(c) block A).
- One histogram is obtained by summing the 4 upper-right sub-blocks' local histograms (Fig. 7(c) block B).
- One histogram is generated by summing the 4 down-left sub-blocks' local histograms (Fig. 7(c) block C).

- One histogram is generated by summing the 4 down-right sub-blocks' local histograms (Fig. 7(c) block D).
- One histogram is generated by summing the 4 central sub-blocks' local histograms (Fig. 7(c) block E).

Fig. 4 displays how the image is partitioned in order to compute the semi-local histograms.

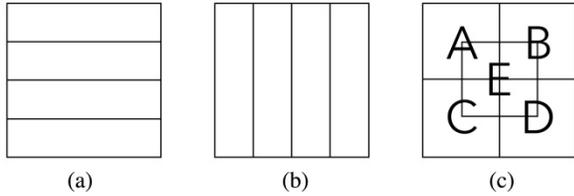


Fig 4. Image Partitioning for Semi-Local Histogram Computation. (a) Into 4 Rows (b) Into 4 Columns (c) Into 4 Sub-Block Sets.

Edge histogram descriptor (EHD) is one of the most common methods for detecting the shape [21]. It describes the global, semi-local and local edges in an image. In fact, the image is partitioned into image blocks. For each image block an edge distribution histogram is computed by categorizing the edges into five types using edge detection filters.

Since EHD captures both local and global edges, it would be able to discriminate between the shape of PD and non-PD spirals through an appropriate description of the edge oscillations.

### C. Acceleration

The authors in [5] propose an application dedicated feature for SST and DST drawing where at each time  $t$ , the coordinates of the drawing  $(x_t, y_t)$  are recorded. The velocity is then defined as the distance between two consecutive samples at  $t-1$  and  $t$  as expressed as follows:

$$V_t = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}$$

Since PD patients are more likely to be confused leading to instantaneous speedup or slowdown, the velocity changes during the drawing for both tests. In other words, instantaneous acceleration changes at time  $t$ . This acceleration is defined as the difference between the velocity at time  $(t)$  and at time  $(t-1)$ . Expressed as in:

$$A_t = V_t - V_{t-1}$$

Acceleration histograms are then computed for SST and DST drawing. Then, the dissimilarity between the two obtained histograms, DAH, is computed as:

$$DAH = \|H_{SST} - H_{DST}\| = \sum_{i=1}^{10} (H_{SST}(i) - H_{DST}(i))^2$$

People who don't suffer from PD are expected to show similar performances in SST and DST and thus would get a DAH close to zero.

We should mention here that the acceleration based on SST and DST test could be used separately as a feature descriptor. It can also be used to combine both tests as described in the previous formula.

Since PD patients are expected to be confused during the drawing of SST and DST leading to instantaneous speedup or slowdown, the velocity and acceleration change for both tests. This feature is used by the authors in [5] suggesting the use of SST and DST.

### D. Auto-Encoder

An auto-encoder is an unsupervised neural network learning algorithm. Auto-encoders take unlabeled data as an input, encode them and then try to reconstruct the encoded data as accurately as possible. An auto-encoder is composed of an input layer, an output layer and several hidden layers as shown in Fig. 5.

In order to compute the neural-net weights for each hidden layer, the previous layer is considered as an input layer and as an output layer. This yields a 3-layer network as shown in Fig. 6.

Then, the network parameters are determined as for any standard 3-layer artificial neural-net. This process is repeated for all the hidden layers starting from the first one where the input layer is the actual input data.

Once the network can accurately reconstruct the input this means that the hidden layer contains enough information to represent the output. Thus, an auto-encoder may act as a feature extraction engine by using the encoding layer as a feature descriptor [22].

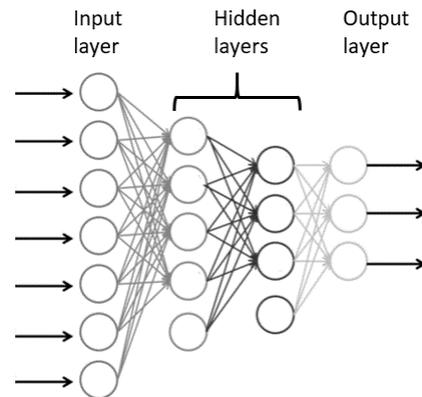


Fig 5. Layers of an Auto-Encoder.

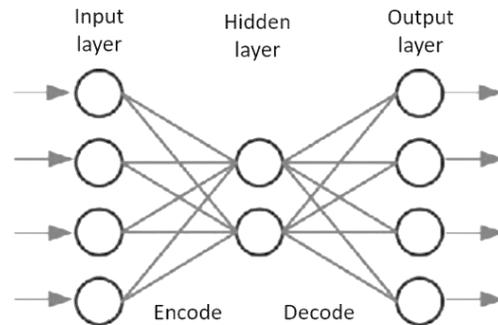


Fig 6. A Three-Layer Auto-Encoder.

An auto-encoder is an unsupervised neural network with multiple hidden layers that is trained with a standard weight

adjustment algorithm to reproduce the input with fewer unites which compels the hidden layer units to become feature detectors and then predict the classes based on the output from the previous layers. As auto-encoder have shown great results for knowledge extraction [22], it would be able to capture the intrinsic characteristic of the drawings and thus, it would be able to distinguish PD and non-PD patients.

#### E. Combinations of SST and DST Features

Consider SST and DST drawings of the same patient  $j$ . Let  $F_{ij}^{(1)}$  be the feature  $i$  extracted from the SST image of patient  $j$  and let  $F_{ij}^{(2)}$  be the feature  $i$  extracted from the DST image of patient  $j$ . We consider to approaches to combine  $F_{ij}^{(1)}$  and  $F_{ij}^{(2)}$  into a single vector  $F_{ij}^c$ .

1) *Concatenation*: In order to obtain the combined feature  $F_{ij}^c$  for patient  $j$  with respect to feature  $i$ , we concatenate  $F_{ij}^{(1)}$  and  $F_{ij}^{(2)}$  as follows:

$$F_{ij}^c = [F_{ij}^{(1)} \ F_{ij}^{(2)}]$$

The dimensionality of  $F_{ij}^c$  will be the double of the dimensionality of feature  $i$ .

2) *Difference*: We can combine  $F_{ij}^{(1)}$  and  $F_{ij}^{(2)}$  by considering their difference. We define  $F_{ij}^c$  as:

$$F_{ij}^c = \sum_{k=1}^d (F_{ijk}^{(1)} - F_{ijk}^{(2)})^2$$

Where  $d$  is the dimensionality of feature  $i$ ,  $F_{ijk}^{(1)}$  is the  $k^{\text{th}}$  entry of feature vector  $F_{ij}^{(1)}$ , and  $F_{ijk}^{(2)}$  is the  $k^{\text{th}}$  entry of feature vector  $F_{ij}^{(2)}$ .

When using the concatenation approach, we assume that both the information from SST and DST are useful and we will convey them both at the same time to the classifier.

On the other hand, using the difference as a method of combining the two tests, assumes that the performance of a PD patient will be better for SST test than for DST test since DST involves testing the short memory of the patient, and thus the difference would be meaningful as a way of discrimination.

#### F. Support Vector Machine

A Support Vector Machine is a supervised machine learning technique for the purpose of binary data classification. It finds the optimal hyperplane that separates two classes. As reported, most of previous works used SVM [7]. In fact, it has been proven to be an affective classifier for 2-class problems [7].

### IV. EXPERIMENTAL DESIGN

#### A. Dataset Description

The dataset used to assess the proposed system is obtained from [5] [25]. It was collected from 15 non-PD patients and 57 PD patients. Knowing that both SST and DST were used, that

gives a total of 30 non-PD and 114 PD recordings. A sample from the dataset is displayed in Fig. 7.

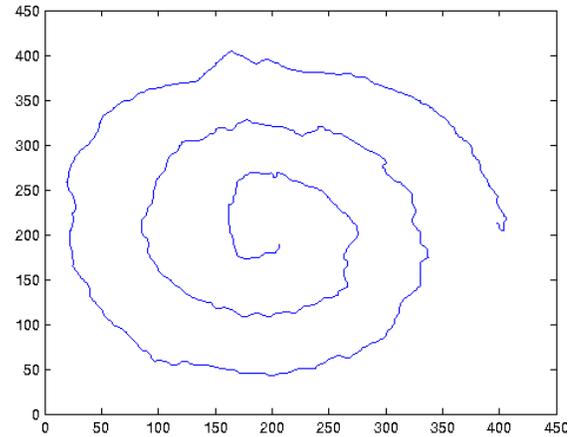


Fig 7. A Sample from the Dataset.

The drawings are recorded using a digitized graphics tablet shown in Fig. 8.



Fig 8. Digitized Graphics Tablet [5].

#### B. Assessment Technique

As an assessment technique, we use the K-fold-cross-validation [26] -with  $k$  equal to 10-. It divides the training set randomly into  $K$  subsets of relatively the same size. The  $K^{\text{th}}$  subset is used for testing while the first  $K-1$  subsets are used as the training set. This process is repeated  $K$  times so that each subset is used as the testing set once. Then the performance of the cross-validation is defined as the average performance of the  $K$  iterations [27].

#### C. Assessment Measures

To assess the proposed approach, we use different statistics. Specifically, accuracy, sensitivity and specificity.

- Accuracy [28], measures the overall correctness of the classifier:
- Sensitivity [28], (also known as Recall or True Positive Rate) measures the correctly classified positive tests over the actual positive.
- Specificity [28] measures the correctly classified negative tests over the actual negative.

We should mention here that since the data is unbalanced (57 PD patients and 15 control), the accuracy maybe misleading. On the other hand, the sensitivity is more appropriate, since it is more important to detect wrongly a PD patient than missing one.

In addition to the previously mentioned statistical performance measures, The Receiver Operating Characteristic (ROC) [29] is also a performance technique that is used to assess the recognition system over all possible thresholds. ROC curves of all the possible approaches are plotted by sketching the 1-Specificity on the x-axis against the Sensitivity on the y-axis.

Area Under the Curve (AUC) is the total area underneath the ROC curve. It measures the performance of a recognition system. If AUC is 1.0, it means that the system can perfectly distinguish between two classes.

#### D. Experiment Description

The first step of the experiment consists in extracting the 4 considered features from SST and DST drawings. This yields 19 features matrices:

- $\Pi 1$ : HOG feature from SST
- $\Pi 2, \Pi 3, \Pi 4, \Pi 5$ : EHD features from SST. We will consider the local EHD, semi-local EHD, the global EHD, and the overall EHD descriptors separately.
- $\Pi 6$ : acceleration feature from SST
- $\Pi 7, \Pi 8, \Pi 9$ : Auto-encoder features from SST. We consider 6 layers with different number of nodes which are 1000, 500, 100, 64, 32, and 16 as shown in Fig. 9. As input to the SVM classifier, we convey the output of the layers of size 64, 32, and 16 separately.
- $\Pi 10$ : HOG feature from DST
- $\Pi 11, \Pi 12, \Pi 13, \Pi 14$ : EHD feature from DST. We will consider the local EHD, semi-local EHD, the global EHD, and the overall EHD descriptors separately.
- $\Pi 15$ : acceleration feature from DST
- $\Pi 16, \Pi 17, \Pi 18, \Pi 19$ : Auto-encoder features from DST. We consider 6 layers with different number of nodes which are 1000, 500, 100, 64, 32, and 16 as shown in Fig. 9. As input to the SVM classifier, we convey the output of the layers of size 64, 32, and 16 separately.

In the second step we combine SST and DST using the concatenation approach. This gives 9 feature matrices:

- $\Pi 20$ : HOG from concatenating SST and DST
- $\Pi 21$ : Local EHD from concatenating SST and DST
- $\Pi 22$ : Semi-local EHD from concatenating SST and DST
- $\Pi 23$ : Global EHD from concatenating SST and DST
- $\Pi 24$ : Overall EHD from concatenating SST and DST
- $\Pi 25$ : acceleration from concatenating SST and DST

- $\Pi 26$ : auto-encoder feature of size 64 from concatenating SST and DST
- $\Pi 27$ : auto-encoder feature of size 32 from concatenating SST and DST
- $\Pi 28$ : auto-encoder feature of size 16 from concatenating SST and DST

Then, we combine SST and DST tests using the difference approach. We get 9 feature matrices:

- $\Pi 29$ : HOG from the difference between SST and DST
- $\Pi 30$ : Local EHD from the difference between SST and DST
- $\Pi 31$ : Semi-local EHD from the difference between SST and DST
- $\Pi 32$ : Global EHD from the difference between SST and DST
- $\Pi 33$ : Overall EHD from the difference between SST and DST
- $\Pi 34$ : acceleration from the difference between SST and DST
- $\Pi 35$ : auto-encoder feature of size 64 from the difference between SST and DST
- $\Pi 36$ : auto-encoder feature of size 32 from the difference between SST and DST
- $\Pi 37$ : auto-encoder feature of size 16 from the difference between SST and DST

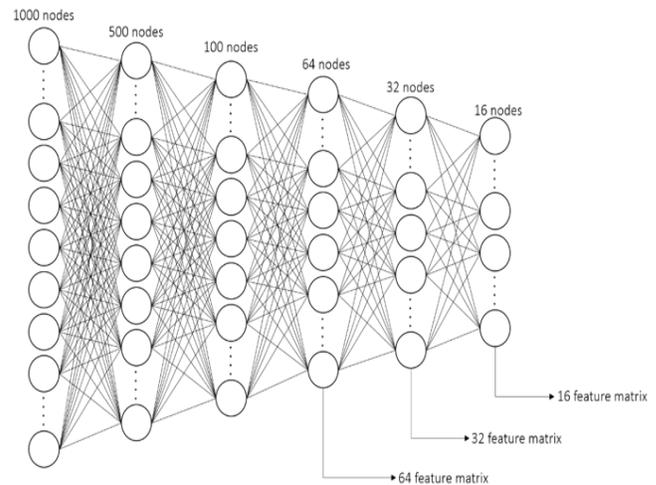


Fig 9. Number of Nodes in each Layer of the Encoding Part in the Auto-Encoder.

Finally, each of the obtained 37 matrices ( $\Pi 1 \dots \Pi 37$ ) will be conveyed separately to an SVM classifier.

Besides, we will consider three variants of the SVM classifier. More specifically, in the first experiment, we assume that the data is well separated and the boundary between the two classes is linear, and thus we use hard-margin SVM [30]. In the second experiment, we assume that the data

is not well separated and that an over-fitting problem may occur and conduct the experiment on a soft-margin SVM [31] to check if this assumption holds. In the third experiment, we assume that the boundary separating the two classes is not linear and that a mapping of the extracted features to a new feature space is necessary. Therefore, we use the Gaussian kernel-SVM [32]. We should mention here that the 37 extracted features are considered in each of the three experiments and conveyed separately to each considered classifier.

Lastly, the performance measures described previously are computed for each considered experiment with respect to the 37 features. The analysis and comparison of the results would allow to conclude on the design of the PD diagnosis system and its effectiveness.

### V. RESULTS AND DISCUSSION

To explore the possibility that the data is not linearly separable, we use Gaussian kernel SVM [32]. The Gaussian kernel happens to be one of the most important algorithms among the kernel based, it maps the data features onto a higher dimensional space using the Gaussian kernel function [32].

Fig. 10-14 show the performance results of using HOG, EHD, Acceleration and Auto-encoder features as input to Gaussian kernel SVM [32].

Fig. 10 displays the performance results when using the HOG feature on SST, on DST, on the concatenation of both tests, and on the difference between them. The best performance is obtained when extracting HOG from DST drawings with a sensitivity of 85.67% and specificity of 45%.

Fig. 11 shows the performance results of the global, semi-local, local and the overall EHD feature on SST, on DST, on the concatenation of both tests, and on the difference between them. As we can see, the semi-local EHD extracted from DST drawings gives the best results which is 91% for sensitivity and 90% for specificity.

Fig. 12 shows the performance results when using Acceleration feature as input for the Gaussian Kernel SVM [32]. We notice that concatenation of the acceleration feature of SST and DST achieves the best results with a sensitivity of to 98.3% and specificity of 25%.

Fig. 13 displays the accuracy, sensitivity and specificity results when conveying the auto-encoder feature as input to the Gaussian kernel SVM [32]. We notice that the concatenation of the 32 encoded layers of SST and DST gives the best results with a sensitivity and specificity of 95% and 10%, respectively.

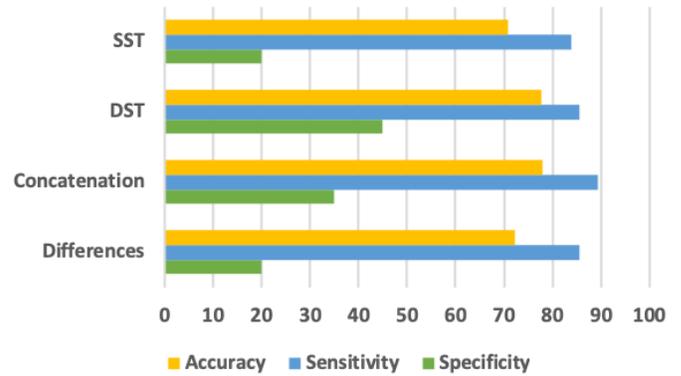


Fig 10. HOG Descriptor Performance using Gaussian kernel SVM.

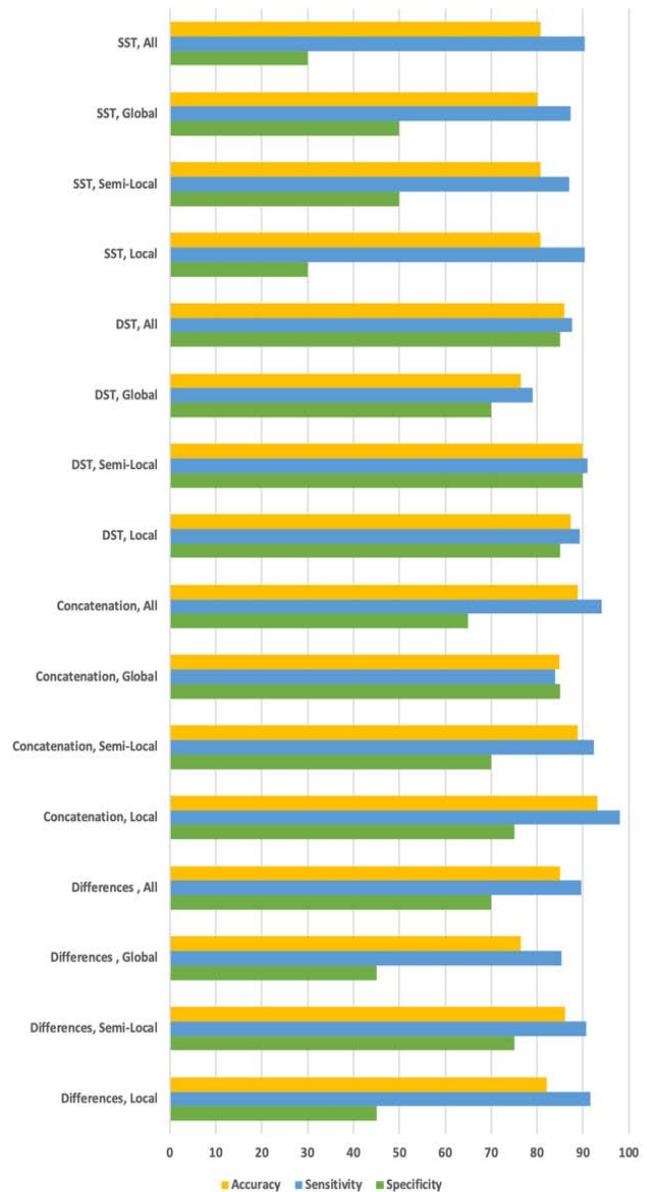


Fig 11. EHD Descriptor Performance using Gaussian kernel SVM.

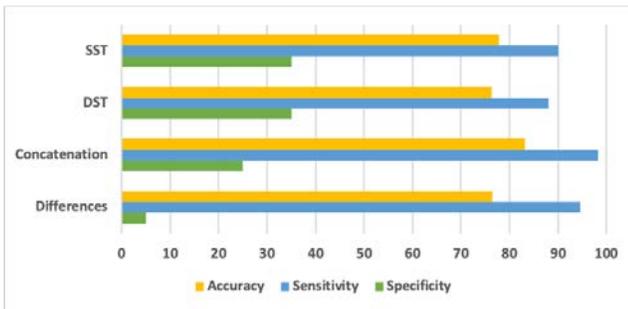


Fig 12. Acceleration Features Performance using Gaussian kernel SVM.

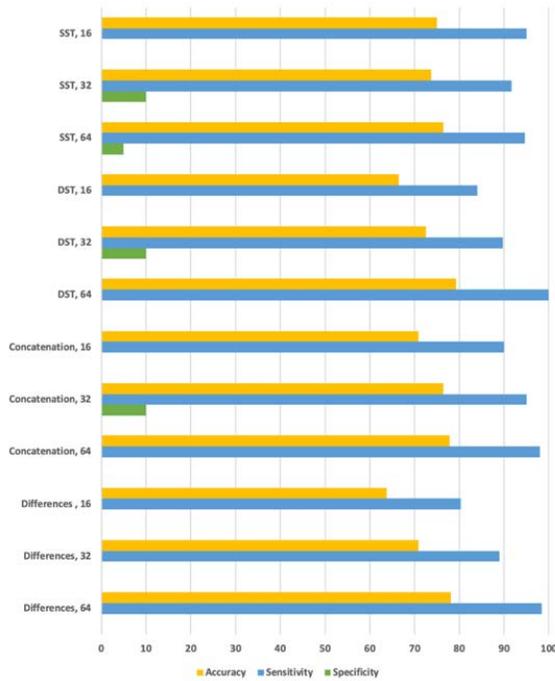


Fig 13. Auto-Encoder Features Performance using Gaussian kernel SVM.

Fig. 14 and Table I show the ROC and the AUC when using Gaussian Kernel SVM [32]. We can see that the semi-local EHD extracted from DST drawings outperforms the other features.

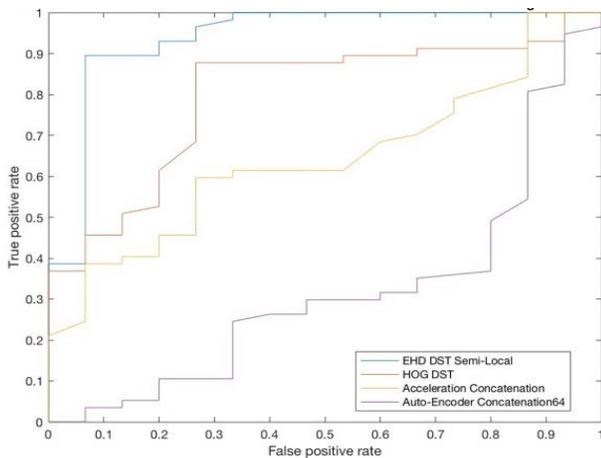


Fig 14. ROC Curves using Gaussian Kernel SVM.

TABLE I. SUMMARY OF THE BEST AUC VALUES USING GAUSSIAN KERNEL SVM

Best AUC Values	Experiment details		
	Size	Type of test	AUC
HOG	18	DST	0.797661
EHD	65	DST	0.938012
Acceleration	32	Concatenation	0.645029
Auto-encoder	128	Concatenation	0.314620

The results of semi-local EHD extracted from DST drawings outperforms all the other descriptors, when using Gaussian kernel SVM [32], it has an accuracy of 90%, sensitivity of 91.3%, and specificity of 90%. Gaussian kernel SVM [32] had an astounding improvement compared to hard and soft margin SVM [30] [31]. This indicates that the data is not linearly separable and mapping the data points using the Gaussian kernel SVM [32] allowed SVM to better classify the data.

In conclusion, the semi-local EHD descriptor extracted from DST drawing, and conveyed to a Gaussian kernel SVM [32] gives the best results. This is due to the fact that semi-local EHD describes 13 overlapping regions of the image separately and then concatenates the obtained histograms. This gives a good local edge information of the spiral drawings and allow distinguishing PD from non-PD. Moreover, the Gaussian kernel SVM [32] was able to classify the data better than hard-margin SVM [30] and soft-margin SVM [31] since it allowed mapping the data to a new feature space in which the data is linearly separable. Therefore, for the PD aided diagnosis system, we propose the system displayed in Fig. 15.

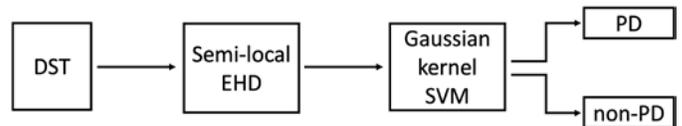


Fig 15. PD Aided Diagnosis System.

TABLE II. FEATURE EXTRACTION RUNNING TIME PER IMAGE

Feature extraction running time		Time (seconds)
HOG		3.72
EHD	All	7.82
	Global	4.47
	Local	4.05
	Semi-Local	4.54
Acceleration		9.3361
Auto-encoder	16	17.940
	32	17.914
	64	17.843

The proposed system contains two main components which are feature extraction and SVM classifier. Table II depicts the time need to extract each considered feature in seconds. We should mention here since the semi-local EHD descriptor gives the best results, the feature extraction phase will take only 4.54 seconds for the obtained system. Moreover, the time complexity for running SVM is linear with respect to the number of support vectors.

## VI. CONCLUSION

Recently, the number of people with PD has augmented considerably. This makes it one of the major health problems.

Since it has no cure, an early detection is very important in order to allow an appropriate treatment. Moreover, it is crucial to monitor regularly the progress of the symptoms. However, this requires the patient to often visit the physician dealing with transportation, waiting, appointments, etc. This is inconvenient, especially that PD affects mostly elderly people. Besides, it involves the physician time and efforts.

In this project, we alleviated the monitoring of PD by designing a self-conducted test that uses recent technology advances along with pattern recognition techniques.

As a typical pattern recognition system includes a feature extraction step and a classification step in this project, we described the features extraction techniques that we investigated. We also outlined the machine learning technique, SVM classifier that will be applied.

A review of computer-based PD detection approaches using new technologies was outlined in this work. These approaches are based either on image or signal data. The latter source of data concerns gait on voice pattern analysis while image data is related either to the analysis of brain images or the analysis of handwriting pattern.

During these experiments, we used SST and DST image data gathered from a tablet device [5]. We investigated several features and conveyed them to an SVM classifier. We also investigated each test separately, SST and DST, and two ways of combining them. We implemented and assessed their performances. After analyzing the results, we conclude that the semi-local EHD [21] extracted from DST drawing, and conveyed to a Gaussian Kernel SVM [32] outperforms the other considered systems with an accuracy, specificity and sensitivity around 90%.

In order to investigate further the use of deep learning in extracting the visual descriptors from SST and DST drawing and its ability to discriminate between PD and non-PD drawing pattern, we plan to collect more data. The large size of the data would allow an effective training of the deep learning network.

Although the Gaussian kernel SVM gave good results, it may be enhanced further. As future work, we plan to investigate the use of Gaussian Mixture Classifier (GMM) [33]. In fact, GMM can model each class with several Gaussian and thus, can deal with variability of the data within each class.

Another way of enhancing the results would be to use fusion techniques. They could be applied on a set of classifiers, a set of visual descriptors, or a set of different data sources. The latter could be done by using the drawing pattern along with the speech pattern to discriminate between PD and non-PD patients.

## REFERENCES

- [1] S. N. Heyn, C. P. Davis and M. C. Stöppler, "Parkinson's Disease Symptoms, Causes, Stages, Treatment, and Life Expectancy," MedicineNet, 28 August 2018. [Online]. Available: [https://www.medicinenet.com/parkinsons\\_disease/article.htm](https://www.medicinenet.com/parkinsons_disease/article.htm). [Accessed 8 September 2018].
- [2] "Tremor in Parkinson's," American Parkinson Disease Association, [Online]. Available: <https://www.apdaparkinson.org/what-is-parkinsons/symptoms/tremor/>. [Accessed 8 September 2018].
- [3] MedicineNet, "Tremors," MedicineNet, [Online]. Available: <https://www.medicinenet.com/tremor/article.htm>. [Accessed 8 September 2018].
- [4] Parkinson's News Today, "Parkinson's Disease Statistics," Parkinson's News Today, [Online]. Available: <https://parkinsonsnewstoday.com/parkinsons-disease-statistics/>. [Accessed 17 September 2018].
- [5] M. E. Isenkul, B. E. Sakar and O. Kursun, "Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease," in The 2nd International Conference on e-Health and Telemedicine, Istanbul, 2014.
- [6] Sikora, Thomas, "The MPEG-7 visual standard for content description-an overview," IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6, pp. 696-702, June 2001.
- [7] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer and D. Haussler, "Article Navigation Support vector machine classification and validation of cancer tissue samples using microarray expression data," Bioinformatics, vol. 16, no. 10, p. 906-914, October 2000.
- [8] S. Bridenbaugh and R. W. Kressig, "Quantitative Gait Disturbances in Older Adults with Cognitive Impairments," Current Pharmaceutical Design, vol. 20, no. 19, pp. 3165-3172, 2014.
- [9] K.-C. Lan and W.-Y. Shih, "Early Diagnosis of Parkinson's Disease using a Smartphone," in The 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC-2014), Tainan, 2014.
- [10] M. Little, P. Mcsharry, E. Hunter, J. Spielman and L. Ramig, "Suitability of Dysphonia Measurements for Telemonitoring of Parkinsons Disease," IEEE Transactions on Biomedical Engineering, vol. 56, no. 4, pp. 1015-1022, 2009.
- [11] M. Shahbakhi, D. T. Far and E. Tahami, "Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine," Journal of Biomedical Science and Engineering, vol. 7, no. 4, pp. 147-156, 2014.
- [12] D. E. Goldberg, Genetic algorithms in search, optimization, and machine learning, vol. 27, New York: Addison-Wesley, 1989.
- [13] L. Naranjo, C. J. Pérez, Y. Campos-Roca and J. Martín, "Addressing voice recording replications for Parkinson's disease detection," Expert Systems with Applications, vol. 46, no. 15, pp. 286-292, 2016.
- [14] I. R. Titze and H. Liang, "Comparison of Fo Extraction Methods for High-Precision Voice Perturbation Measurements," Journal of Speech, Language, and Hearing Research, vol. 36, no. 6, pp. 1120-1133, 1993.
- [15] M. Singh and G. M. Provan, "Efficient learning of selective Bayesian network classifiers," in Thirteenth International Conference on Machine Learning, Bari, Italy, 1996.
- [16] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal and M. Faundez-Zannuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease," Artificial intelligence in medicine, vol. 67, no. 1, pp. 39-46, 2016.
- [17] J. W. M. de Souza, S. S. A. Alves, E. d. S. Rebouças, J. S. Almeida and P. P. R. Filho, "A New Approach to Diagnose Parkinson's Disease Using a Structural Cooccurrence Matrix for a Similarity Analysis," Computational Intelligence and Neuroscience, vol. 2018, pp. 1-8, 2018.

- [18] F. Keller, "Naive Bayes Classifiers," Saarland University, 3 March 2005. [Online]. Available: <http://www.coli.uni-saarland.de/~crocker/courses/learning/lecture10.pdf>. [Accessed 20 October 2018].
- [19] J. P. Papa, A. X. Falcão, V. H. C. de Albuquerque and J. M. R. Tavares, "Efficient supervised optimum-path forest classification for large datasets," *Pattern Recognition*, vol. 45, no. 1, pp. 512-520, 2012.
- [20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE*, San Diego, 2005.
- [21] C. S. Won, S.-J. Park and D. K. Park, "Efficient Use of MPEG-7 Edge Histogram Descriptor," *ETRI Journal*, vol. 24, no. 1, pp. 23-30, February 2002.
- [22] C.-Y. Liou, W.-C. Cheng, J.-W. Liou and D.-R. Liou, "Autoencoder for words," *Neurocomputing*, vol. 139, pp. 84-96, 2014.
- [23] P. H. Kassani and A. B. j. Teoh, "A new sparse model for traffic sign classification using soft histogram of oriented gradients," *Applied Soft Computing*, vol. 52, no. C, pp. 231-246, 2017.
- [24] Udacity, "Computational Photography by Georgia Tech," Udacity, 23 February 2015. [Online]. Available: <https://sa.udacity.com/course/computational-photography--ud955>. [Accessed 22 Septemser 2018].
- [25] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbaş, F. Gurgun, S. Delil, H. Apaydin and O. Kursun, "Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828-834, 2013.
- [26] I. H. Witten, E. Frank, M. A. Hall and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, Cambridge: Elsevier, 2011.
- [27] K. Duan, S. S. Keerthi and A. N. Poo, "Evaluation of simple performance measures for tuning SVM hyperparameters," *Neurocomputing*, vol. 51, pp. 41-59, 2003.
- [28] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283-298, 1978.
- [29] R. Kumar and A. Indrayan, "Receiver operating characteristic (ROC) curve for medical researchers," *Indian Pediatrics*, vol. 48, no. 4, pp. 277-287, 2011.
- [30] C.-W. Hsu, C.-C. Chang and C.-J. Lin, "A Practical Guide to Support Vector Classification," National Taiwan University, Department of Computer Science, Taipei, 2010.
- [31] J. Shawe-Taylor and N. Cristianini, "On the generalization of soft margin algorithms," *IEEE Transactions on Information Theory*, vol. 48, no. 10, pp. 2721 - 2735, 2002.
- [32] Z. Liu and H. Xu, "Kernel Parameter Selection for Support Vector Machine Classification," *Journal of Algorithms & Computational Technology*, vol. 8, no. 2, pp. 163-177, 2013.
- [33] A. M. Patrikar and J. P. Baker, "Improving accuracy of Gaussian mixture model classifiers with additional discriminative training," in *International Joint Conference on Neural Networks (IJCNN)*, Vancouver, 2016.

# Still Image-based Human Activity Recognition with Deep Representations and Residual Learning

Ahsan Raza Siyal<sup>1</sup>, Zuhaibuddin Bhutto\*<sup>2</sup>, Syed Muhammad Shehram Shah<sup>3</sup>, Azhar Iqbal<sup>4</sup>, Faraz Mehmood<sup>5</sup>,  
Ayaz Hussain<sup>6</sup>, Saleem Ahmed<sup>7</sup>

Department of Electronic Engineering, Dawood University of Engg. & Technology, Pakistan<sup>1</sup>

Department of Computer System Engineering, Balochistan University of Engg. & Technology, Pakistan<sup>2</sup>

Department of Software Engineering, Mehran University of Engineering & Technology, Pakistan<sup>3</sup>

Department of Basic Sciences, Dawood University of Engineering & Technology, Pakistan<sup>4,5</sup>

College of Information and Communication Engineering, SungkyunKwan University, Suwon, Republic of Korea<sup>6</sup>

Department of Computer System Engineering, Dawood University of Engg. & Technology, Pakistan<sup>7</sup>

**Abstract**—Iterative Recognizing human activity in a scene is still a challenging and an important research area in the field of computer vision due to its various possible implementations on many fields including autonomous driving, bio medical, machine intelligent vision etc. Recently deep learning techniques have emerged and successfully deployed models for image recognition and classification, object detection, and speech recognition. Due to promising results the state of art deep learning techniques have replaced the traditional techniques. In this paper, a novel method is presented for human activity recognition based on pre-trained Convolutional Neural Network (CNN) model utilized as feature extractor and deep representations are followed by Support Vector Machine (SVM) classifier for action recognition. It has been observed that previously learnt CNN knowledge from large scale data-set could be transferred to activity recognition task with limited training data. The proposed method is evaluated on publicly available stanford40 human action data-set, which includes 40 classes of actions and 9532 images. The comparative experiment results show that proposed method achieves better performance over conventional methods in term of accuracy and computational power.

**Keywords**—Human activity recognition; action recognition; deep learning; transfer learning; residual learning

## I. INTRODUCTION

Over the recent decade, the human activity recognition has been a highlighted topic for researchers because of its various applications which include video surveillance, Human-machine interaction, ambient-assisted living, smart system design and autonomous driving. Automatic classification of an action in a given scene is a challenging and critical task. There are two main approaches for human activity recognition; which includes traditional handcrafted features representations and a deep learning approach or deep representations. The learning-based approach introduce concept of classification by trained feature extractor followed by a state-of-the-art classifier. The deep learning approach have made remarkable growth in activity recognition task.

The deep learning model in [1] has introduced multi-stream 3D CNN for limited learning data. The author in [2] utilizes integrating body pose, part shape, and motion data for

activity recognition task. The process of training the deep neural network from the scratch involves huge amount of trainable data and learning this much of parameters require high computational resources and hours or days for training. Training the model for real-world applications, collecting annotated hug amount of specific task related data is very time consuming and costly [3]. Therefore, accumulation of sufficient task related learning data may not be feasible choice in many cases [4]. It is challenging to produce adequate results by applying deep learning methods. For mitigating this issue researchers reviewed their approach for implementation of deep learning models on smaller data-sets and this makes them relating the problem to human vision system. We humans learn several categories in our lives just from few samples and this capability is achieved by accumulating previously learnt knowledge over the period of time and transferring it for learning the new task [5]. Researchers came to the conclusion that previously learnt knowledge contribution in learning new tasks through connection and similarity between new task and the old one can produce significant improvement in efficiency of methods. By this idea studies suggest pre-trained models for classification can be utilized to classify new classification task [6]. Hence, the CNN models trained to classify certain object can be fine-tuned for new task even in different domain [7].

In this paper, human activity recognition method is proposed which is based on pre-trained Convolutional Neural Network (CNN) model utilized as feature extractor and deep representations are followed by Support Vector Machine (SVM) classifier for action recognition. The results show that proposed method can produce significant performance results for human activity recognition.

The rest of the research paper is outlined as follows. The Learning methods are explained in Section II. Then Section III details about the related work. The Section IV methodology applied for the proposed approach. The experimental results are illustrated and explained in Section V. Finally, paper is concluded in conclusion section.

\*Corresponding Author

## II. RELATED WORK

This section of the paper discusses existing state-of-the-art methods for human activity recognition using both approaches of handcrafted techniques and deep representation learning. The methods that use handcrafted techniques such as extended SURF [14], HOG-3D [15]. The techniques which utilizes motion-based feature descriptors such as exploiting motion in [16], using gaze [17], by improved trajectories [18], from multiple views based on view-invariant feature descriptor [19], have achieved notable performance for classifying human activities. However, these methods have some limitations such as, requirement of proficient intended feature detector and descriptors for feature extraction. This process requires skilled manpower, consumes time and it is a cost inefficient method.

For all these reasons, researchers prefer deep learning approaches for human activity recognition. This approach has been used for various domains in the recent past such as image classification, object detection, speech recognition. Same approach has also been explored for human activity recognition. Some of the contributions are as using Global spatial-temporal attention [20], based on skeleton data [21], 3D ConvNets [22], learning Spatio-Temporal with 3DConvNets [23].

Human activity recognition based on videos has always been highlighted research topic for researchers and they have contributed a lot in the last decade. Image-based action recognition models have been developed and evaluated for efficient action recognition. Researchers have contributed to resolve issues related to accuracy improvement and less computational power requirements.

Some researchers have used transfer learning in cross-domain for human action recognition to improve accuracy and performance of the model such as cross-domain knowledge transfer was carried out in [24]. Based on human poses human part-detectors can be employed to detect different human parts in the scene and then these parts are encoded into poses for human activity classification [25]. In [26], author employs trained neural network to perform pose estimation. Additionally, it transfers previously learnt knowledge to target model in order to perform training of new task has shown improved accuracy of the model and it saves time and money. Furthermore, research work has been reported reported in [28-31].

## III. LEARNING METHODS

### A. Residual Learning

Residual learning is a machine learning technique in which the network has stack of layers. Let us assume  $H(x)$  contained by a neural network block, where  $x$  denotes the input parameter. The difference between the true distribution  $H(x)$  and input sample  $x$  is given as  $x = H(x) - x$ . When we rearrange it we will get  $H(x) = F(x) + x$ . Both equations will be approaching a value or curve which has

certain limit called asymptote line but the way of understanding for both equations could be different. If all the added layers are similar or they return the identical value, then error can occur which will not be greater than shallower part. If the system has nonlinear layers, then it can create issue in order to find the approximate value in identity mapping layers. Another solution for this might be that if identical mappings are good to choose then the weights of nonlinear layers will assume towards value zero to get the result of identical mappings by multilayer mappings.

But in actual scenario the identical mappings are not most favorable, so we have to reformulate the process to get the accurate result. It can also be solved when identical mapping has the value closer to zero than it will be easy to solve. Experiments show that residual functions have small and simple responses to provide reasonable approach to solution.

### B. Identity Mapping using Shortcuts

As we know that residual learning has number of layers. It can be written in equation as:

$$y = F(x, \{w_i\}) + x \quad (1)$$

here  $x$  and  $y$  are the input and output, respectively, and  $F(x, \{w_i\})$  represents residual learning. In Fig. 2, it has been shown that the two layers in which  $F = W2\sigma(w_i x)$  here  $\sigma =$  rectified linear unit.  $F + x$  perform element wise addition. This is also a short cut connection which will neither consider as an extra element in the equation nor it will create a difficulty in calculation process. Plain or residual networks have same number of parameters used for example computational cost, width of depth etc. so we can easily compare them. In above equation  $x$  (input) and  $F$  must have same dimensions if the dimensions are not same then we are able to calculate another kind of projection called as liner a projection

$$y = F(x, \{w_i\}) + W_{sx} \quad (2)$$

Experiments shows that to label the degradation problem the identical mapping is enough, but we can also use square matrix.  $F$  (residual function) has two layers and its value is adaptable but more layers are also possible. If  $F$  (residual function) has single layers, then it's called as linear layer. We have seen that we have fully occupied layers to keep it simple that is applicable to the first layer also called as convolution layer. Multiple convolutional layers can be shown by the function  $(x, \{w_i\})$ . This function will also use to perform element wise addition which includes two featured maps and work as channel by channel one by one.

### C. Transfer Learning

The transfer learning approach tries to utilize the previously learnt knowledge to solve another problem which may have different domain. To initialize the process notation is used which is introduced by Pan and Yang (2010). Notation

has two major components, domain  $D$  which is learnable data-set having given probability distributions, data may be images with different resolutions and pixel values, and other component is task  $T$  which may be defined as target function or labels. Labels are the marking of the image which provides the class or category which assist to learn them accordingly.

Transfer learning is the deep learning technique to set a deep neural network using features learnt for a source problem ( $TS, DS$ ) and the same network can be fine-tuned and employed for target task ( $TT, DT$ ). This approach has been introduced to generate more accurate model than training a model from scratch. Another utilization of the pre-trained neural network for  $TS$  is as a feature extractor for  $TT$  and the extracted features can be utilized on training another machine learning method to optimize accuracy. By utilizing this approach, one will represent  $DT$  data learnt for  $TS$  task by use of pre-trained network representations on small data which is inadequate to train these methods.

Transfer learning is one of the approaches that utilizes pretrained models as feature extractor by replacing fully connected layer and extracting feature data from last pooling layer of the deep neural network which may be followed by a generic state-of-the-art SVM classifier [8], this approach has been employed on many classification and recognition tasks [9]. In this paper our proposed model also falls under the same approach. We evaluate recently benchmark models like GoogleNet [10], VGG-16, VGG-19 [11] and ResNet-18 [12] on stanford40 [13] data-set, based on performance in terms of accuracy and learning validation we selected ResNet-18 as a source model for generating a target model to classify action in a given image. ResNet-18 is used to extract features from input image and deep representations are followed by SVM classifier for action class recognition.

#### IV. METHODOLOGY

Transfer learning in a machine learning approaches utilizes previously learnt knowledge to train a model for new task with less computational requirements as compare to train the model from scratch. Transfer learning based on pre-trained CNN are useful in training the model with smaller data-sets. However, CNN are prone to overfitting with limited data-set but it can be avoided by increasing the training data on the expenses of increased cost and time-consuming process. For these reasons transfer learning is very convenient way to train a model with help of pre-trained deep representations as source architecture for creating new architecture in order to perform a new task. In the proposed method, we have evaluated publicly available popular pre-trained networks including GoogleNet, VGG16, VGG19 and ResNet-18 and selected the pre-trained model ResNet-18 on bases on performance on the stanford40 data-set for human action recognition problem. The ResNet-18 has been trained on a million of images of 1000 different categories of ImageNet data-set [27]. ResNet-18 architecture consists of 18 layers and having input layer of size  $224 \times 224 \times 3$ . This pre-trained CNN has the ability to categorize 1000 different classes like pencil, mouse, cat, dog, keyboard and many more. For the reason of extensive learning of deep

representations of various classes of images, it is very useful to transfer this learnt knowledge to classify human action recognition.

In this paper, pre-trained CNN Resnet-18 model is used as a feature extractor, input image is augmented with a size of ResNet-18 first layer ( $224 \times 224 \times 3$ ). The architecture consists of 17 convolutional layers C1 to C17 and one fully connected layer (fc1000). Features are collected from the last pooling layer in the case of ResNet-18 'Pool\_5' layer, and with these deep representations state-of-the-art classifier is trained to classify action in the given image. Fig. 1 shows the block diagram of proposed method.

##### A. CNN Architecture for Feature Extraction

Deep CNN (DCNN) is one of the general classes of deep neural networks (DNN). In past years, DCNN has contributed in a large scale to the computer vision field by surpassing the performance of machine learning algorithms. However, problems come up whenever network become widen. Many researchers have quoted that the accuracy becomes compromised turning into saturated. There is also a case of test errors at the time of training deep networks consisting several layers. The DCNNs are also vulnerable to the problems of vanishing gradient. In that case, a minor gradient prevents the layer weights to be updated.

Whenever this problem comes, the training of the deeper layers of the network is inefficient due to slow training of deeper layers. Therefore, residual learning can be used to counter this problem.

In case of residual learning, there is a training of the network of features are conducted instead of features. ResNet is a DCNN model which is considered state-of-the-art; it carries a concept of residual learning. In its architecture, ResNet uses an alternative connection through connecting a  $n_h$  layer to the input with a  $n + i$  layer. The architecture of ResNet consists of many residual building blocks. Let's consider the input of residual block to be  $x_{i-1}$  and the output is  $x_i$ , after attempting several operations like, convolution, ReLU activation and batch normalization on input, the output is  $f(x_{i-1})$ . looking residual learning, it can be defining  $f(x_{i-1}) = x_i - x_{i-1}$ . However,  $x_i = f(x_{i-1}) + x_{i-1}$  is obtained. Through this method, the information of previous layer is enabled to be added is the current layer. In Fig. 2, a basic building block of ResNet is shown. Many variations of ResNet can be found. The evaluation of knowing accuracy using ResNet-18 is presented in this work, which consist of 17 layers of convolutional and one fully connected layer. Through residual learning, the layers have been stacked up one after another. Every convolutional layer in the block of residual is followed by its connected ReLU layer and batch normalization layer. Max pooling layer is used at the end of first residual block. An average pooling layer is used after the fifth residual block which is later on connected with the FC layer. The size of the FC layer is equivalent to the class number.

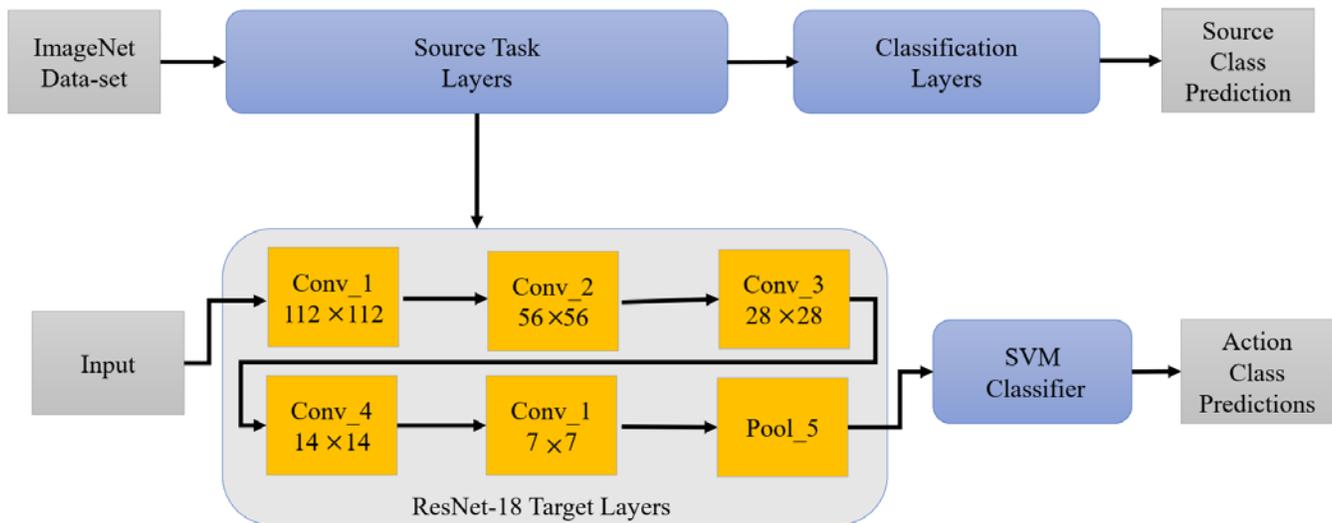


Fig 1. Block Diagram of Proposed Method.

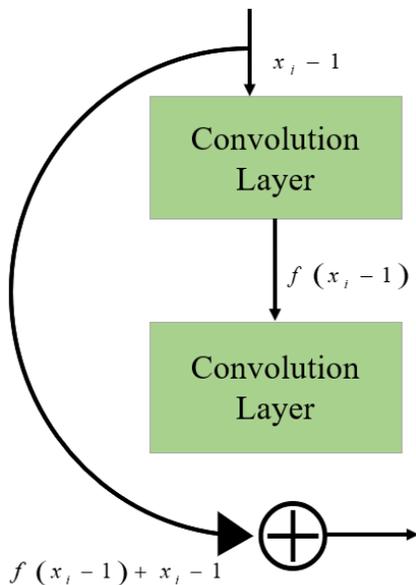


Fig 2. Resnet Architecture: Basic Building Block of Residual Learning.

### B. Classification

Classification is a method of designating an unknown sample to already defined class or it is based on the data which is used to help in understanding the program. When there are different classifiers Support Vector Machine (SVM) is one of the classifiers that uses classification algorithm with two groups of data. Tasks related to pattern recognition SVM is considered as the popular one. Basically, SVM operates on machine learning algorithm. In this method there is a  $n$ -dimensional space. The data items are represented by  $x_i$  and  $y_i$ , each one has its own representation that is  $x_i$  represents the attribute of sample and  $y_i$  denotes the class usually it has positive or negative value. In SVM the classification is found by using a line called as hyper plane which divides the plane

in to two parts and the classes on lie on the either sides. The cost function  $\left(\frac{1}{2}W^T W\right)$  is kept maximum to find out the hyper plane. This can be limited by an equation as:

$$y_i = \begin{cases} +1 & \text{if } w \times x_i + b \geq 1, \\ -1 & \text{otherwise,} \end{cases} \quad (3)$$

In this equation  $w$  represents the weight factor and  $b$  is the distance between the hyper planes to the origin which is called as bias. Since we are interacting with 10 classes of human actions this is definitely a problem which has not only one class but has more than one class. The multi class problems can be used to solve the problems occur in binary class. The equation shown above is having the contents which are made by combining different binary class SVMs to get a multi class SVM. The two other kinds of programming which can be used are one-vs-one and one-vs-all methods.

Now we talk about One-vs-all method. In this method there are  $n$  numbers of binary classifiers each of them recognizing a particular class. The  $c$  is using for  $c$ -class problems and  $i$ th classifier is used to create a boundary between other classifiers. There is another method called as winner-takes-all which is used to assign value to a class the value is unknown sample usually a negative value is also be accepted.

On the other hand one-vs-one method is also called as binary classifier which contains two samples  $i$  and  $j$  they can be assigned when we pick a positive sample of class  $i$  and negative sample of class  $j$ . In this method an approach for classification is used called as max-wins voting in which each classifier gives the value of one or two classes the vote will be in the favor of assigned class and increased by one and an unknown sample is also assigned to a class which has largest vote. one-vs-one method provides better accuracy that's why

SVM is using this problem. However, the initial of one-vs-one method has greater overhead than multi class classifier.

### V. EXPERIMENTATION AND RESULTS

In this section we discuss the experiment setup in terms of preprocessing, learning process and evaluated proposed method observations and results. The proposed method is tested on publicly available Stanford 40 human action data-set. It includes 40 different classes of daily life human actions like phoning, walking, jumping and more. Each action class have approximately 180 to 300 images of bounding box of the action performing person, some sample images from the data-set with their corresponding labels are shown in Fig. 3. For experimental purpose only 10 classes are used to evaluate the performance of four pre-trained models on the data-set , Images in the data-set are different in size to make these images compatible with input layer of pre-trained models, images are gone through augmentation process as a preprocessing followed by learning implementation and based on the performance of Resnet-18 it is selected as source architecture for feature extraction for the proposed method, Table I shows class-wise accuracy of the pre-trained models and Fig. 4 shows overall accuracy on 10 classes of the data-set.

In the proposed method, based on accuracy on the data-set we selected Resnet-18 as source architecture for feature extraction, Resnet-18 has input layer of size 224 by 224 and images in the data-set have different sizes, images are augmented to the size of input layer of pre-trained Resnet-18. The architecture comprises of 17 convolutional layers and only one fully connected layer. The proposed method utilizes the architecture for feature extraction there for deep representations are extracted from last pooling layer which is 'pool5' in case of Resnet-18. These deep representations and then followed by state-of-the-art SVM classifier for predicting action class in a given image, the proposed method achieves 87.22% accuracy on the data-set. Fig. 5 shows the classified actions from the test data-set with predicted labels and confidence score and Fig. 6 shows the confusion matrix.

TABLE I. COMPARISON OF CLASSIFICATION RESULTS ON STANFORD DATA-SET

S. No	Action	ResNet-18 [12]	VGG-16 [11]	VGG-19 [11]	GoogLeNet [10]
1	Applauding	71.19	64.41	64.41	64.41
2	Brushing	70	70	43.33	56.67
3	Cleaning	85.71	93.65	96.83	89.83
4	Climbing	98.31	94.92	91.53	91.53
5	Cutting trees	88.33	95	95	93.33
6	Cooking	92.94	92.94	91.76	96.47
7	Jumping	96.55	96.55	93.1	90.8
8	Phoning	72.73	77.92	79.22	70.13
9	Playing guitar	91.86	80.23	84.88	94.19
10	Riding bike	96.51	91.86	96.51	96.51



Fig 3. Shows Sample Images from Stanford40 Data-Set with Labels.

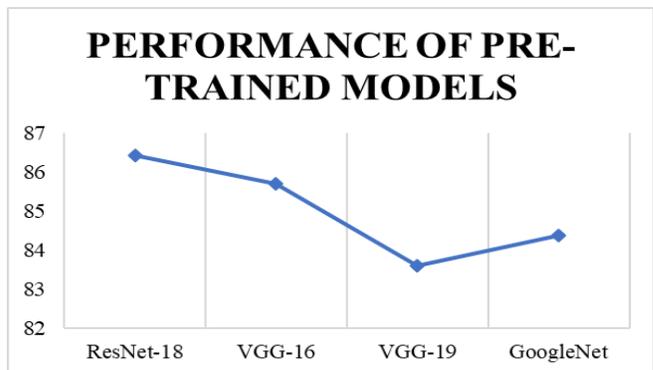


Fig 4. Overall Accuracy of Pre-Trained Models on Stanford40 Data-Set.



Fig 5. Classified Actions on Test Data-Set with Predicted Labels and Confidence Score.

1	120				1			5	1	
2		168		3	3	2		1		
3		2	156		1	8	1	4		
4	1	1	3	126	3	1	1	33	1	
5	1		3	8	81		2	24	1	
6	1		2		1	111	1	3		
7	3	5		3		4	161	1		
8		2	3	10	13	3	4	118	1	1
9		1	1	5	1	1	1	8	155	
10		4					2			169
	1	2	3	4	5	6	7	8	9	10

Fig 6. Confusion Matrix of Predicted Classes.

## VI. CONCLUSION AND FUTURE WORK

In this paper a novel method has evaluated based on transfer learning of deep representations by using pre-trained Resnet-18 convolutional neural network as source architecture for feature extraction and state-of-the-art SVM classifier is trained to classify target data-set. It has been established that by using transfer learning technique previously learnt knowledge can be utilized to learn new task with limited data size. Transfer learning comes very handy when the data-set is not adequate for training the deep model from the scratch and also training the deep model from the scratch on very large amount of data requires computational resources and it is very time costly way which can be avoided using transfer learning. In addition to that it has been noted that SVM as a classifier performs better than a convolutional neural network and moreover, handcrafted representation-based methods require preprocessing and manual feature extraction, the proposed method eliminate these requirements as it directly accepts RGB images as input and extract features from them. The performance of the proposed method is evaluated on open source stanford40 human action data-set, it achieves 87.22% accuracy.

Some future directions in the human activity recognition and classification research are given as:

### A. Utilizing Image-based Models for Other Area of Research

Deep learning network has been emerged and proven its superiority over other traditional methods in many areas of research and similarly in computer vision field. However, video-based models are training with the complexity and difficult implementation, thus benefiting from pre-trained models on images would be better solution to explore. In addition, image-based models have done a better job on capturing spatial relationships of objects which might be utilized in action recognition. These image-based models can be explored for medical image processing, disease detection and classification.

### B. Interpretability on Temporal Extent

All the frames in the video are not equally significant for activity recognition, few of the frames are critical and required to learn deep representations of temporal interpretability of video-based models. Above all else, activities, particularly long-length activities can be considered as a sequence of primitives. It is intriguing to have an interpretability of these primitives for example, how are these primitives sorted out in the temporal area in activities, how would they add to the arrangement task, can we just utilize not many of them without sacrificing recognition performance in order to achieve fast training and less computation.

### C. Complexity Reduction Techniques

Learning deep representations is very complex and specially dealing with dataset having high dimension would require high computational power and time. It is commonly helpful to reduce data dimension not only for reason of computational efficiency but also improve accuracy of analysis. These dimensions reduction technique can be apporioned into two significant manners, they can be separated as techniques that can be utilized for supervised or non-supervised learning and into techniques that either entail feature selection or feature extraction.

## REFERENCES

- [1] V. A. Chenarlogh, F. Razzazi "Multi-stream 3D CNN structure for humanaction recognition trained by limited data", IET Computer Vision, Vol. 13 Issue 3, pp. 338-344, 2019.
- [2] H. El-Ghaish, M. Hossain, A. Shoukry, And R. Onai, "Human Action Recognition Based onIntegrating Body Pose, Part Shape, andMotion", IEEE Access, vol.6, pp. 49040 – 49055, 2018
- [3] Hossein Rahmani, Ajmal Mian and Mubarak Shah, "Learning a DeepModel for Human ActionRecognition fromNovel Viewpoints", IEEE Transactions on Pattern Analysis and Machine Intelligence ,Vol. 40 , Issue: 3 , pp. 667 – 681, 2017.
- [4] Cao, X., Wang, Z., Yan, P., Li, X., "Transfer learning forpedestrian detection", Neurocomputing100, p. 51-57, 2013.
- [5] F. Fei, L. "Knowledge transfer in learning to recognize visual objects classes", International Conferenceon Development and Learning (ICDL). 2006.
- [6] K. Chatfield, K. Simonyan, A. Vedaldi, "Return of the devil inthe details: Delving deep into convolutional nets", arXiv preprintarXiv:1405.3531, 2014.
- [7] H. Nam, and B. Han, "Learning multi-domain convolutional neural networks for visual tracking", arXiv preprintarXiv:1510.07945, 2015.
- [8] M. D. Zeiler, and R. Fergus. "Visualizing and understanding convolutional networks", European Conference on Computer Vision, Springer, 2014.
- [9] H. Azizpour, S. A. Razavian, J. Sullivan, "From generic to specific deep representations for visual recognition", IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2015.
- [10] S. Christian, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. "Going deeper with convolutions", IEEE conference on computer vision and pattern recognition, pp. 1-9. 2015
- [11] S. Karen, and A. Zisserman. "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556 (2014).
- [12] H. Kaiming, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition", IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.
- [13] B. Yao, X. Jiang, A. Khosla, A.L. Lin, L.J. Guibas, and L. Fei-Fei. Human Action Recognition by Learning Bases of Action Attributes and

- Parts. International Conference on Computer Vision (ICCV), Barcelona, Spain, November 6-13, 2011.
- [14] G. Willems, T. Tuytelaars, and L. V. Gool. "An efficient dense and scale-invariant spatio-temporal interest point detector" European conference on computer vision. , Springer 2008.
- [15] A. M. Klaser, Marsza, and C. Schmid. "A spatio-temporal descriptor based on 3d-gradients", BMVC-19th British Machine Vision Conference, 2008.
- [16] M. Jain, H. Jegou, and P. Bouthemy, "Better exploiting motion for better action recognition", IEEE Conference on Computer Vision and Pattern Recognition. 2013.
- [17] A. Fathi, Y. Li, and J. M. Rehg, "Learning to recognize dailyactions using gaze", European Conference on Computer Vision, Springer, 2012.
- [18] H. Wang, and C. Schmid, "Action recognition with improved trajectories", IEEE International Conference on Computer Vision. 2013.
- [19] A. B. Sargano, P. Angelov, and Z. Habib, "Human Action Recognition from Multiple Views Based on View-Invariant Feature Descriptor Using Support Vector Machines", Applied Sciences, vol. 6, issue 10, pp. 309, 2016.
- [20] Y. Han , S. L. Chung , A. M. Ambikapathi ,W.Y. Lin, S.Su, "Robust Human Action Recognition Using Global Spatial-Temporal Attention for Human Skeleton Data", International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8-13 July 2018.
- [21] W. Cho, T. Z. Win, M. T. A. Win. "Human Action Recognition System based on Skeleton Data", IEEE International Conference on Agents (ICA), Singapore, 28-31 July 2018.
- [22] S. Ji, W., Yang, M. Yu, "3D convolutional neural networks for human action recognition", IEEE transactions on pattern analysis and machine intelligence, vol.35 issue 1: p. 221-231.2013
- [23] D. Tran, L. Bourdev, R. Fergus, L. Torresani, "Learning spatio temporal features with 3D convolutional networks", IEEE International Conference on Computer Vision (ICCV).2015.
- [24] L. Cao, Z. Liu, and T. S. Huang, "Cross-dataset action detection", IEEE conference on Computer vision and pattern recognition (CVPR), San Francisco USA, June 2010.
- [25] S. Maji, L. Bourdev, and J. Malik, "Action recognition from a distributed representation of pose and appearance", IEEE Conference on Computer Vision and Pattern Recognition, pp. 3177–3184. 2011.
- [26] J. Tompson, R. Goroshin, A. Jain, Y. L. Cun, and C. Bregler, "Efficient object localization using convolutional networks", IEEE Conference. on Computer Vision and Pattern Recognition, pp. 648–656, 2015.
- [27] J. Deng, W. Dong, *et al.*, "ImageNet: A Large-Scale Hierarchical Image Database", IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, June 2009.
- [28] Z. Bhutto, *et al.*, "Scaling of Color Fusion in Stitching Image", International Journal of Computer Science and Network Security, Vol. 19, No. 4, pp. 61-64. April 2019.
- [29] A. S. Chan, K. Saleem, Z. Bhutto, *et al.*, "Feature Fusion Based Human Action Recognition in Still Images", International Journal of Computer Science and Network Security, Vol. 19, No. 11, pp. 151-155, November 2019.
- [30] N. K. Baloch, Z. Bhutto, *et al.*, "Finger-vein Image Dual Contrast Enhancement and Edge Detection", International Journal of Computer Science and Network Security, Vol. 19, No. 11, pp. 184-192, November 2019.
- [31] A. Siyal, Z. Bhutto, *et al.*, "Ship Detection in Satellite Imagery by Multiple Classifier Network", International Journal of Computer Science and Network Security, Vol. 19, No. 8, pp. 142-148. August 2019.

# Design and Experimental Analysis of Touchless Interactive Mirror using Raspberry Pi

Abdullah Memon<sup>1</sup>

Institute of Information & Communication Technologies  
(IICT), Mehran University of Engineering and Technology,  
Jamshoro, Pakistan

Syed Mudassir Kazmi<sup>2</sup>

Dept. of Electronic Engineering,  
Mehran University of Engineering and Technology,  
Jamshoro, Pakistan

Attiya Baqai<sup>3</sup>

Dept. of Electronic Engineering,  
Mehran University of Engineering and Technology,  
Jamshoro, Pakistan

Fahim Aziz Umrani<sup>4</sup>

Dept. Telecommunication Engineering,  
Mehran University of Engineering and Technology,  
Jamshoro, Pakistan

**Abstract**—A prototype of a smart gesture-controlled mirror with enhanced interactivity is proposed and designed in this paper. With the help of hand gestures, the mirror provides some basic amenities like time, news, weather etc. The designed system uses Pi cam for image acquisition to perform the functions like gesture recognition and an ultrasonic sensor for presence detection. This paper also discusses the experimental analysis of human gesture interaction using parameters like the angle which ranges its horizon from 0 to 37 degree when tilting the forearm up and down and 0 to 15 degrees when the forearm is twisted right to left and otherwise based on yellow, pink and white background colours. Additionally, the range of detection using an ultrasonic sensor is restricted to the active region of 69.6 to 112.5 degrees. Moreover, time delay which takes half a second for a time as retrieve with the system and take 6 to 10 seconds for fetching headlines and weather information from the internet. These analyses are taken into account to subsequently improve the design algorithm of the gesture-controlled smart mirror. The framework developed comprises of three different gesture defects under which mirror will display the mentioned information on its screen.

**Keywords**—Smart Mirror; Raspberry Pi; Pi-camera; Application Programing Interface; Hue Saturation Value; Region of Interest

## I. INTRODUCTION

Human beings have always been striving to create new technology to make their life easy, swift and comfortable. Smart devices [1-5] including mobile phones, TVs, refrigerators, etc. are part of that chain. In today's fast-paced life, one is strained to do multitasking. This research involves making the mirror interactive so that a user can not only beautify him or her but can also perform regular chores like checking the weather, accessing inbox or look the to-do list.

Keeping in view the limitations of the voice-controlled devices the interaction is based on gesture-controlled making the proposed design also useful for speech impaired. The framework proposed utilizes finger count gestures to fetch and display the time, weather and news from the specified internet location. The efficiency of the interactive mirror is analyzed

by observing the effect of different gestures from different angles using different backgrounds and their timely response to display intended information.

Moreover, relative to other touch or voice control smart mirrors this device is also useful for deaf and dumb. With that, it didn't restrict the user to be in the close vicinity like in conventional touch-based smart mirrors. The framework comprises finger count gestures to fetch and display the time, weather and news from the specified internet location. The paper is organized as follows: The work related to the smart mirrors and gestured controlled hardware and software is provided in Section 2 whereas Section 3 illustrates the design steps and development of the prototype. Section 4 illustrates the software algorithm used in the system. Section 5 describes the analysis of the developed system based on and its interpretation of obtaining different results. The paper is concluded in Section 6.

## II. LITERATURE REVIEW

Many researchers proposed various methods and interaction techniques for smart mirrors. The intervention of gestures to interact with devices gave a new direction to smart homes and smart devices. The author in [1] proposed Smart Reflect—a software platform for developing smart mirror applications. The authors in [2-3] uses external computers, additional controllers, sensors and Application Program Interface (API) integration to operate and control the smart mirrors. With the use of a spy mirror foil, [4] uses the Wii Balance Board and Kinect to reflect the user's position and their gestures. The author in [5] shows a survey of smart notice board mirror. Most of the smart mirrors developed so far are thoroughly based on displaying the information with less emphasis on the quick and easy way of doing it. Despite using fewer components additional sensors, external computers, controllers and software platforms are used to get it done. The author in [6] uses face recognition authentication to provide access to data feeds on the smart mirror. The devices in [7,8] take the voice commands to control the mirror. By the time, different interaction mechanisms like touch and voice are introduced to provide a better experience.

People do prefer few efforts to interact with the devices. A survey in [9] illustrates the convenient ways of gesture interaction over touch and voice with different algorithms. To formulate the best technique for mid-air interaction, several methods were considered as [10] uses the dollar method with the best trajectory like multi-finger for processing of short command gestures. The author in [11] proposed finger transformation algorithm for controlling the mobile. Moreover, to select the commands on multi-touch display a finger count gesture technique and analysis is introduced in [12]. The main limitation of these methods is related to the fact that the simple point-to-point trajectory comparison works well with short gestures and few features, but would not probably scale well to long single or multiple trajectories, as required by different application domains. Besides that, the external environment has a vital role in mid-air interaction between human and machine. An algorithm in [13] is proposed to overcome the effect of lighting condition on gesture interface using unit-gradient vector (UGV), background subtraction methods and Hue saturation value (HSV) thresholding. To this end, the gestures are also performed using different objects around users. The author in [14] uses experimental results to formulate a method that helps users to interact and tailored gesture profiles for objects existing in the omnipresent environment. However, the omission of environmental interference in interaction is still a challenge. The authors in [15,16] used a tracking device and reveals temporal division and spatial division gesture techniques to interact with the large public display. To formulate the best technique for hand gesture [17] used the object contour method for fast and accurate hand detection and tracking algorithm. The author in [18] improve gesture recognition by focusing spatial channels on hands. The author in [19] uses Raspberry Pi with the external camera to capture finger gesture using colour markers. Likewise, [20] illustrates a remote free approach to control a led using hand gesture. Besides that, [21,22] proposed Patch-Levy-based Bees Algorithm (PLBA) and Median-Average Otsu's thresholding method to overcome the limitation of traditional Otsu's thresholding method. In contrast to these complex algorithms, we have used the traditional Otsu's method and derived better results.

In this paper, we have critically analyzed the responses of gesture on finger count mechanism using techniques like background or foreground segmentation, and colour information of the hand to subtract hand images. Although several robust techniques and algorithm are proposed to optimally execute these techniques we critically use contour convex hull and convexity defects and traditional techniques to design robust algorithms to provide a better interactive experience.

### III. IMPLEMENTATION OF THE PROPOSED SYSTEM

This section illustrates the hardware components used to design the interactive mirror. The designed interactive mirror is composed of a Raspberry Pi controller, a display module, a pi-camera and an ultrasonic sensor. The mirror displays the information on its display surface when connected to the internet via WIFI. The system conceptual diagram is shown in Fig. 1(a). The proposed mirror support comfort and additional

flexibility by its gesture-controlled mechanism. The detail description of hardware and software is presented as follows.

#### A. Hardware

The mirror is composed of a LED display module of Samsung upon which a 1mm acrylic sheet is pasted such that it reflects the person when there is no light passing through it and displays information when light passes through it. With such a small thickness of the acrylic sheet as shown in Fig. 1(b), it offers high-quality visualization of the information displayed on its screen as shown in Fig. 2. Furthermore, the Pi camera module v2 is used in the system to capture hand gestures which by computer vision techniques analyze a different set of gestures performed by human fingers and interpret them to display useful information. The system also has the capability of detecting the person with an ultrasonic sensor with a separate processing control based on Arduino microcontroller. The power of the display monitor is controlled by the a passive infrared sensor placed right at the bottom of the device. It takes input from an ultrasonic sensor based on the distance measurement of the person from a mirror, it triggers an LED light transmitter to transmit the ON/OFF code to the built-in receiver of the monitor for making the monitor on/off as shown in Fig. 2. The system with all its components is controlled by a Raspberry Pi microcontroller.

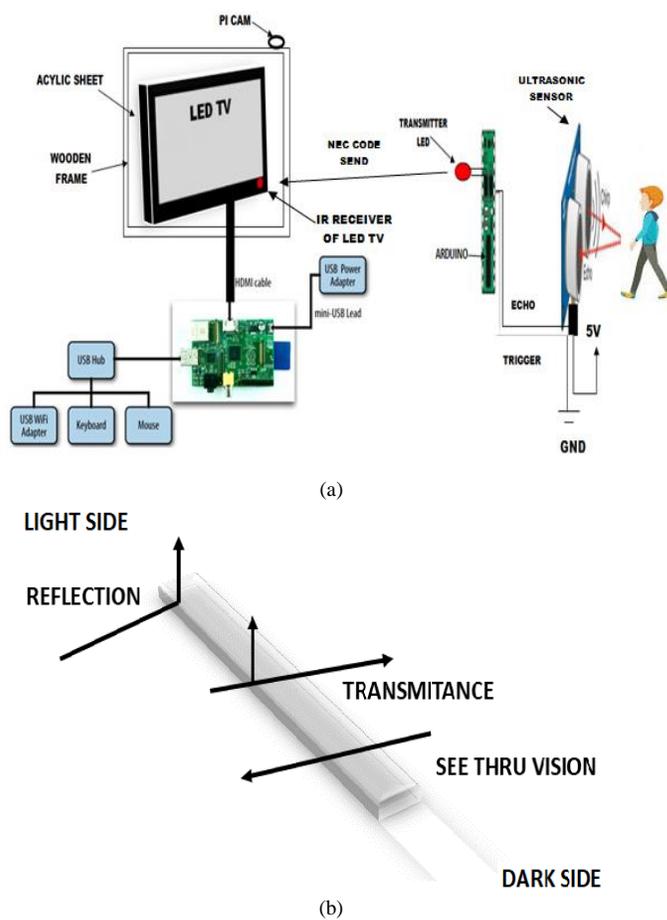


Fig. 1. (a) Conceptual Diagram of the Interactive Mirror, (b) The Interpretation of Acrylic Two-Way Sheet.

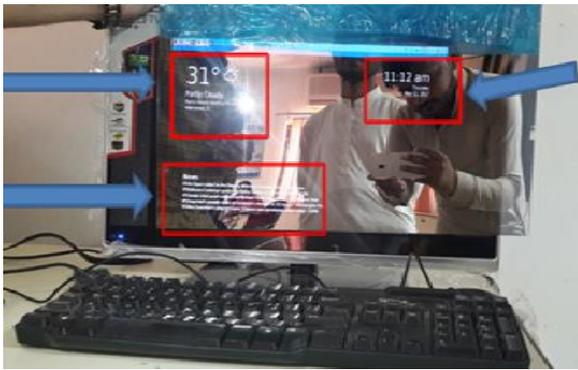


Fig 2. The Information Displayed on Acrylic Sheet (Mirror).

**B. Software**

The section describes the software strategy used to operate the system. For this, Raspberry Pi has all the necessary packages and library such as OpenCV, Numpy, and Matplotlib. Besides that, Python interpreter is installed on it. Therefore, we can run the designed code through Python Idle without using a terminal window. The algorithm is implemented using python 3.0 and OpenCV inside Raspbian desktop OS.

- Gesture Algorithm:

A basic contour detection method in the Region of Interest (ROI) is implemented for gesture recognition. The program algorithm is depicted in Fig. 3, which at first import necessary packages along with Date & Time, Weather and Newsfeed files. Next to the video frames are captured from which for the Hand Region, the image is segmented by cropping captured frame. This cropped image is our Region of Interest. This ROI is further processed to improve quality by converting ROI in Grayscale which subsequently converts an image into binary i.e. high-intensity pixels are treated as one and low-intensity pixels as zero. Based on binary values it decides whether the pixel is of interest or not. After that, Gaussian Blur is applied to reduce noise to acquire the shape of the tracked objects. Besides, the Otsu’s Binarization Thresholding Method using OpenCV is used to highlight the particular colour range and automatically approximate the threshold value of bimodal image from image histogram.

To this end, from the processed image in the ROI, a maximum contour is found in the region and in that contour, convex hull and convexity defects are identified. Convex hull or points are general tips of the fingers and other point and subsequently found the convexity defects, which are the deepest points of deviation on the contour. Later on, cosine rule is applied for finding the angles of all defects (the number of fingers extended). These angles must be greater than 90 degrees as it can’t interpret the angles less than 90. Moreover, it examines the maximum contour area in a range between 2000 to 10000 pixels and decides whether the hand is within the specified range. When the hand is with the area of detection it gives the indication and acknowledges the user to make the gesture. Additionally, the system will then count the number of defects and according to the number of defects in Fig. 4, it will fetch different information as summarized in Table I.

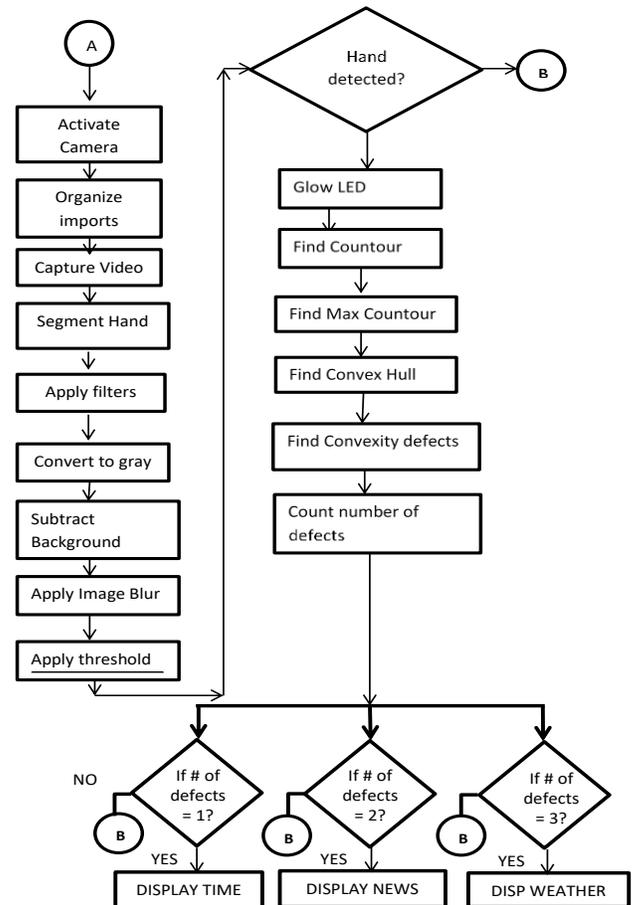


Fig 3. The Algorithm for Gesture Detection and Implementation.

TABLE I. INFORMATION DISPLAYED ON DIFFERENT HAND GESTURES

S. No.	Gestures (No of fingers)	Information Fetched
1	Two	Time and Date
2	Three	News Feed
3	Four	Weather

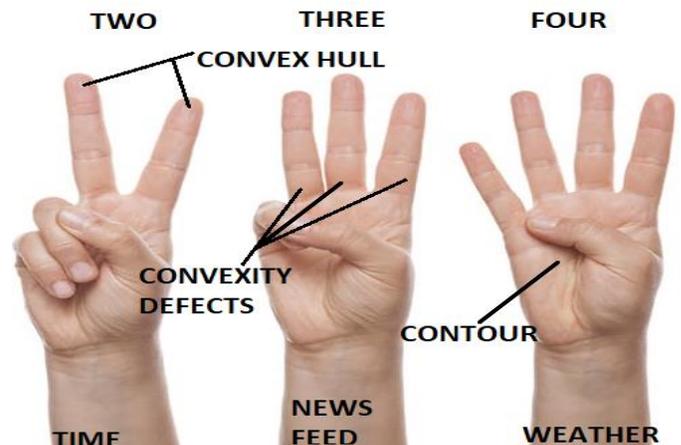


Fig 4. Finger Count Gestures.

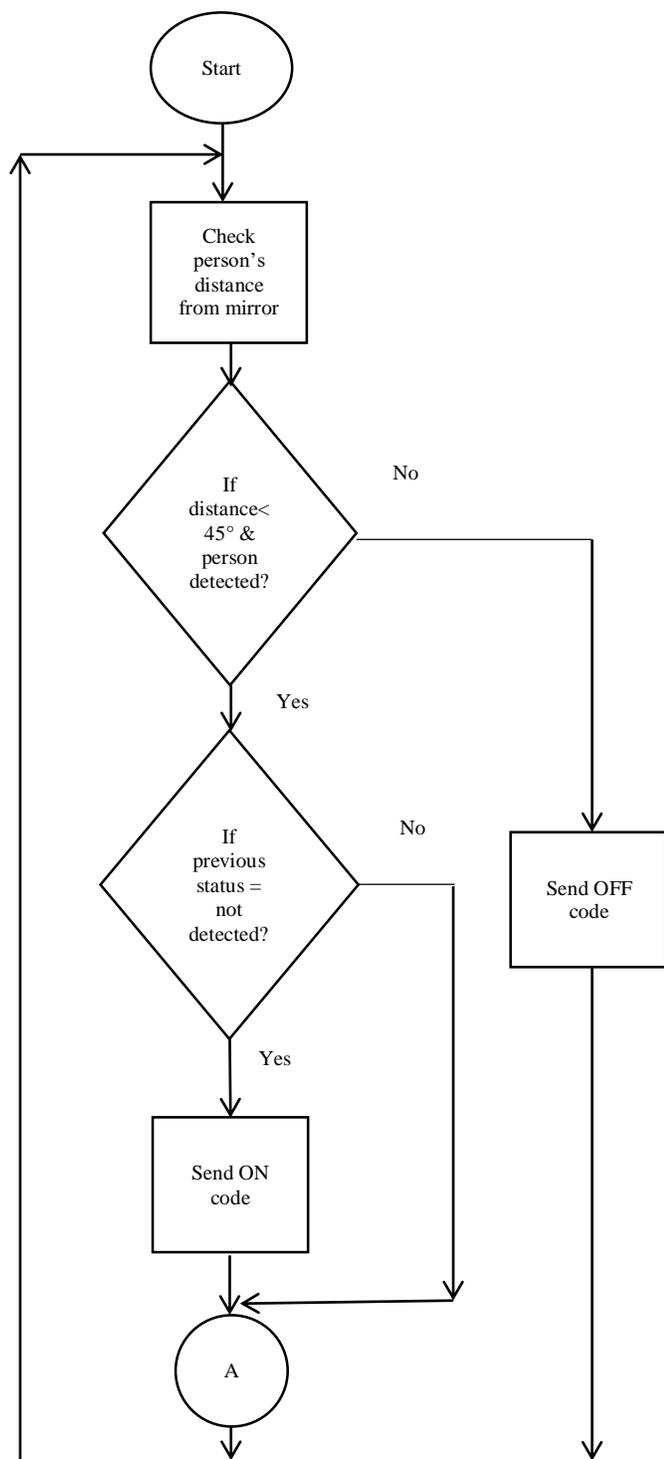


Fig 5. Presence Detection Algorithm.

Finally, for showing the output window (e.g. Weather Display) in full screen we have used the Tkinter module.

- Presence Detection Algorithm

The presence detection algorithm depicted in Fig. 5 works when a person appears within 18 to 36 inches range of distance in front of the mirror. The custom-designed transmitter transmits the signal code to the receiver of the

monitor to turn it on or off. When a person is detected by the ultrasonic sensor the ON or OFF code for Samsung, using the National Electrical Code (NEC) protocol is sent using a built-in IR remote library of Arduino. A check is maintained which is used to resend the code as a result when the object goes away; it automatically turns off the interactive mirror and switches to an ordinary mirror. The presence detection algorithm has separate processing controller which does overlap with the main Raspberry Pi processor. The algorithm works on simple decision-making structures that are based on the distance calculation value from the ultrasonic sensor, with controlling variable declared within the program in C++ that does not let the program enter the same structure repeatedly.

#### IV. EXPERIMENTAL SETUP FOR ANALYSIS

To analyze the smart mirror functionality with the gesture interaction the mirror is placed on the desk under 500 lux light intensity with sheets of different colours pasted in front of the mirror to examine the gesture interpretations and their efficiency concerning change in background colours as shown in Fig. 6(a) & (b). An Android Smartphone Samsung Galaxy S7 Edge is tied on the forearm with a wrist push band as shown in Fig. 7 having clinometer with a bubble's [22] application installed in it to scrutinize the gesture interaction from different angles. The experiments are conducted keeping the hand at approximately 90 degrees and 60 inches away from the pi camera. 18 trials were conducted with different background colour sheets and for each colour; the light intensity is varied at three different levels. A built-in Ambient light sensor AMIS-74980x in Samsung Galaxy S7 Edge smartphone with the resolution of 0.0152 lux/count is used in experiments for measuring light intensity as perceived by the human eye. Now based on the position of the hand an LED is connected to GPIO pin of the Raspberry Pi is made to blink so that the user is notified that a gesture has been detected. For distance range, calculation HC-SR04 ultrasonic sensor is mounted at the bottom center of the smart mirror to incorporate presence detection.



(a)



(b)

Fig. 6. (a) Side View Experimental Setup for Analysis, (b) Front View of the Experimental Setup for Analysis.

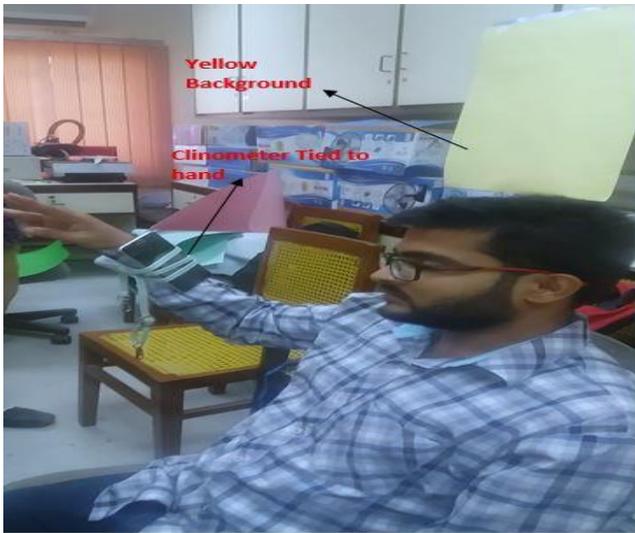


Fig. 7. Experimental Setup for Presence Detection Analysis.

## V. RESULTS AND DISCUSSION

This section discusses the results obtained for the time taken to display information, the effect of finger count gestures on hand recognition accuracy, the effect of deviation of a user from Line of Sight facing the mirror, variation in distance range, variation in ambient light and effects of changing background colours on hand detection accuracy.

From Fig. 8, one can observe the time delay for different information displayed on the surface of the mirror calculated using Time Module in Python. Fig. 8 proposes that the interactive mirror on an average took 0.5 seconds to display date and time to the specified location on the mirror screen, which is the last time as it was taken from the operating date and time. Moreover, it can also be observed that it shows

uniformity in different conducted trails. The time is taken to display the headline news from the internet which illustrates that interactive mirror can take around 8-10 seconds to fetch the news feeds from an internet source and display it to the mirror surface, the readings for time consumption in-display depend upon the number of factors viz-à-viz internet speed, network traffic and server activeness. The readings are taken at an internet speed of 250-280 kbps. Fig. 8 suggests that it normally takes 8-10 second to fetch the information and eventually display it to the surface of the mirror.

**Time Delay Analysis For information to appear With a Gesture**

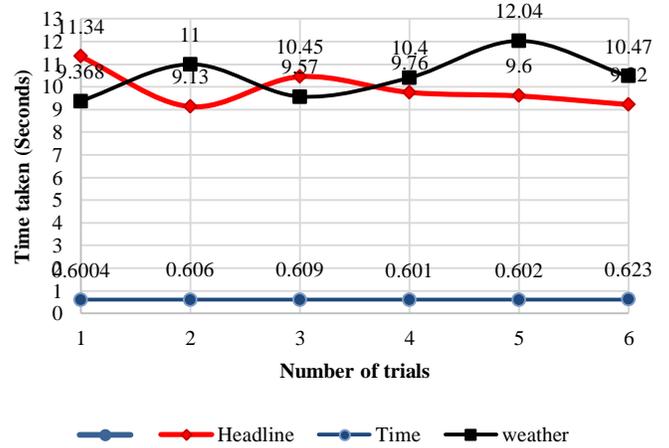


Fig. 8. Time Delay Measurement for Calling Time & Date, Weather and Headlines Information.

**Accuracy For detecting Gesture Type**

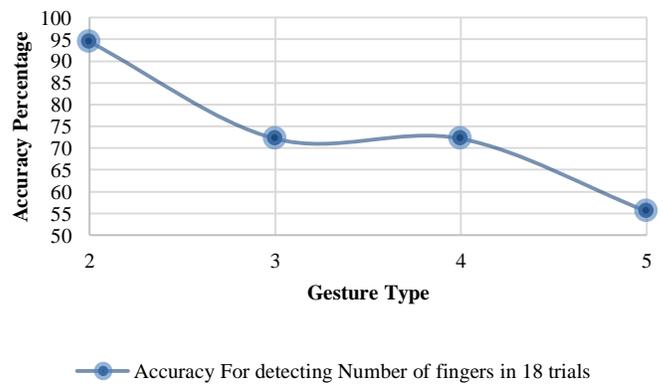


Fig. 9. Performance of Hand Gesture Recognition Accuracy with Different Backgrounds.

In order to analyze the gesture interactivity by contour, convex hull and convexity defect a total of 18 readings were taken for each gesture to examine its accuracy. In Fig. 9, 18 readings are taken on six different colour backgrounds (Dark green, Skin colour, White, Pink, Yellow, Light green) with varying light intensity as mentioned in the experimental setup.

The accuracy is calculated and plotted against gesture type using equation (1):

$$\text{Accuracy} = \frac{\text{Number of times gesture recognized easily}}{100 / \text{total trials}} \quad (1)$$

The term easily refers to the best possible defect identification with almost less than 1 second time delay. From Fig. 9, it is deduced that accuracy decreases from 94% to 55% with the requirement of recognizing more fingers or defects. The background colour and light intensity are interlinked with each other as all the backgrounds are vulnerable for some gesture at some value of light intensity especially when the light intensity is greater than 100 lux. The yellow colour background (Fixed for this setup in Fig. 9) compared to other colours yielded better results.

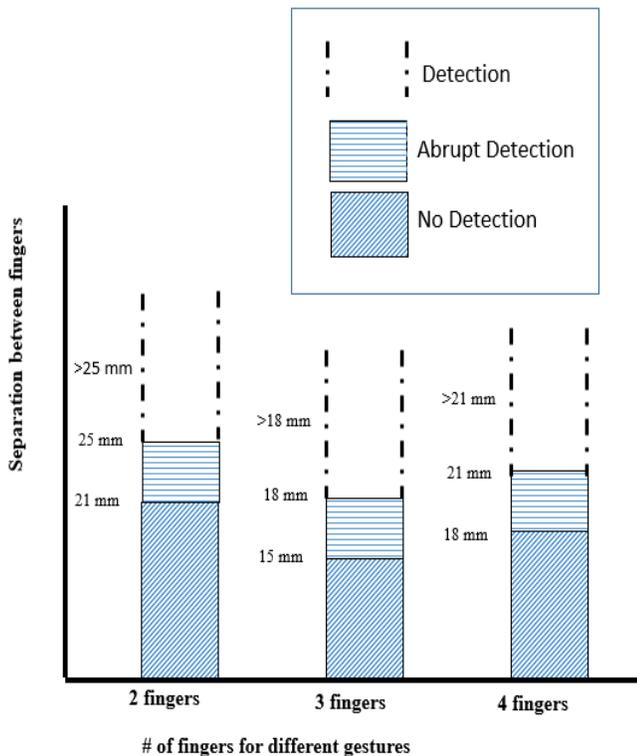


Fig 10. Minimum Distance between the Figures for Finger Count Gesture Detection.

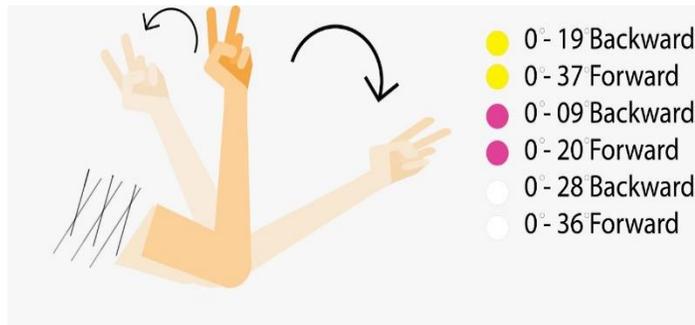


Fig 11. Forward and Backward Tilt Angle with Different Colour Backgrounds.

Fig. 10 narrates the minimum distance between two fingers for 2, 3 and 4 finger count gestures detection. Such that, for 2 fingers, which is index and middle the minimum distance should be 25 millimetres (mm) or greater, as per the algorithm the two fingers didn't detect the gesture up to the distance of 21 mm. However, it gives an abrupt reading from 21 mm to 24 mm, which means within this distance it sometimes detects fingers count gesture. Moreover, for the fingers gesture 3 it didn't accept the gesture defect up to 15 mm. Subsequently, it relies abrupt between 16 mm to 18mm and detect the gesture count 3 for 18 mm or higher. Additionally, for the 4-fingers count gesture, the escape of no detection is from 0 to 18 mm and abrupt is from 19 to 21 mm and the detection zone is 21 or higher. The prime reason of difference of distance between each of the finger count gestures is the angle at which each finger is extended; for a normal human being the index and middle fingers are extended at 90 degrees. In contrast, the ring finger is extended at 70 to 80 degrees when the rest of the finger(s) is held by the thumb. In addition, when four fingers are extended the degree of extension of index and middle is 90 degrees and the angle of extension of ring finger ranges 87 to 90 degrees which differ from person to person. Consequently, the angle at which little finger is extended is 88 to 90 degrees. Additionally, one of the concerns for the limitation is that in this paper we critically analyze the basic contour detection algorithm without subtracting the background so it helps the developers to know the findings of the background interference and design robust contour detection algorithm for the gesture defects.

The finger count gesture is also prone to an inclination of the forearm at yellow, white and pink backgrounds. To take it in consideration, we attached the Samsung S7 edge phone at the forearm and took the readings, Fig. 11 gives the forward and backward angles of the forearm lifted at 3.5 feet (usual placing of a mirror) with the light intensity of 100 lux considered as a normal room light. For the yellow background, the range of detection ranges 0-19 degrees backward, as after the backward angle of inclination increases to 19 it didn't detect the gesture defect. Consequently, for the same background, the forward angle ranges 0-37 degrees as after 37 degrees the system doesn't detect the gesture. Moreover, for the pink background, the backward angle of inclination is 0 to 9 degree and the forward angle is 0 to 20 degrees for the gesture to be detected. However, for the white background, the angle of the forearm for gesture detection ranges 0 to 28 degrees backward and 0 to 36 degrees forward.

It is kept in regard from Fig. 12 that the device functionality is also susceptible to the twist of the wrist as the wrist is curled from its actual position the gesture detection possesses the susceptibility and it also differs with backgrounds. To check this, we took 100 trails, for yellow background, the angle of detection ranges 0 to 15 degrees right as after the right twist increases the 15 degrees the camera didn't detect any gesture detection. Similarly, as the wrist twist at the left side, the detection ranges 0 to 10 degree and offers no detection afterwards. Further, for the pink background, the range of twist angle is 0 to 13 degrees and no detection after 13 degrees at the left curl. Furthermore, for the

right curl the range limit to 0 to 6 degrees. Additionally, for the white background, the left twist angle is restricted to 0 to 23 degree and the right twist angle ranges 0 to 15 degrees.

We took only these three backgrounds in consideration as on the sharp red or the combination of Red-Blue-Green (RGB) if red increases the mark of 190 with the hue and saturation increases the point of 220 it exhibits abrupt or no gesture detection. The similar sort of functionality is observed when the light intensity increases the mark of 120 lux as it's a common consideration the sharp light imposes a great impact on the camera's shutter owing to the fact the existing algorithm is prone to light intensity. Similarly, the width of fingers for swift gesture count detection should be at least 1.8 cm to get the gesture detection.

Fig. 13 narrates the presence detection analysis such that we have connected an ultrasonic sensor at the bottom center of the mirror; the mirror draws cone region of 0-180 degrees. At a cone angle of 0-56.5 degrees, the sensor doesn't detect the person and the mirror remains OFF with the region of 56.6 to 69.5 degrees the mirror posse's abrupt response as at some instant it becomes ON. Consequently, with the same range at a similar instant, it remains OFF. However, when the person comes in the range of 69.6 to 112.5 degrees the mirror remains ON. Additionally, in the range of 112.6 to 123.75 degrees the mirror again gives abrupt readings when it finds a person within this range. Further, in the angle range 123.76 to 180 degree, the mirror doesn't detect the person and remains OFF.

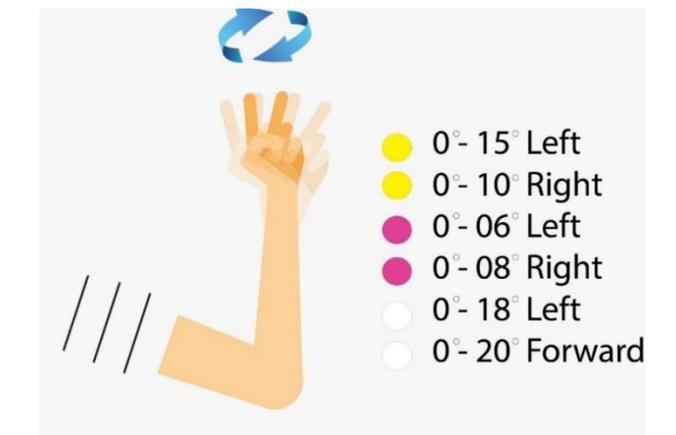


Fig 12. Right and Left Twist Angle of the Wrist with Different Backgrounds.

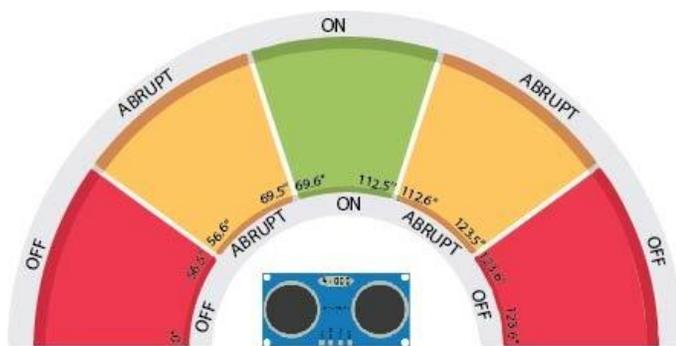


Fig 13. Presence Detection Analysis.

## VI. CONCLUSION

A touch-free reflective interface is designed in this paper that displays informational data. The designed Interactive Mirror acts as a traditional mirror as well as it provides instant access to information on the mirror surface efficiently and effectively to provide time optimization for the user. With that, we also have carried out an analysis to improve interaction mechanism. In addition, we have used a simple and cost-effective method of interaction. The results show that the mirror performs better with light colour backgrounds and the result go worse as the background colour goes sharp and can further be improved for sharp backgrounds by implementing background refinement algorithms to make it robust. This mechanism accurately works if a user appears at 56.6° (degrees) to 69.5° and 69.6° to 112.5° cone of the ultrasonic sensor. But, when a user turns at 112.6° to 123.75° it can produce false detection. This problem can be overcome by using tunnel propagation sensor or using multiple ultrasonic sensors to cater side poses as well. It is recommended to carry out more study on improving gesture recognition algorithm for continuous inputs from the user with that Artificial intelligence can also be added in this interactive mirror. Moreover, the background subtraction method can be used to overcome the gesture predisposition. Besides, with the help of analysis new techniques can be used to overcome gesture vulnerability. Additionally, swipe gestures and 3D Dressing amenities can be designed in the future device.

## ACKNOWLEDGMENTS

This research work has been carried out in Top Quality Lab IIT Building, and Innovation and Entrepreneurship Center (IEC) Mehran University of Engineering and Technology, (MUE) Jamshoro. The authors are thankful to Institute of Information & Communication Technologies (IICT) Mehran University of Engineering and Technology Jamshoro for facilitating this research.

## REFERENCES

- [1] D. Gold, D. Sollinger, and Indratmo, "SmartReflect: A modular smart mirror application platform," 7th IEEE Annu. Inf. Technol. Electron. Mob. Commun. Conf. IEEE IEMCON 2016, 2016
- [2] S. Yong, G. Liqing, and D. Ke, "Design of Smart Mirror Based on Raspberry Pi," 2018 International Conference on Intelligent Transportation, Big Data & Smart City vol. 67, pp. 6-9, 2018.
- [3] V. Mehta, B. Walia, V. Rathod, and M. Kothari, "Smart and Interactive Home Using Raspberry Pi," International Journal of Computer Sciences and Engineering Vol 6 no. 3, pp. 241-245, 2018.
- [4] D. Besserer, J. Bäurle, A. Nikic, F. Honold, F. Schüssel, and M. Weber, "Fitmirror: A Smart Mirror for Positive Affect in Everyday User Morning Routines," Proc. Work. Multimodal Anal. Enabling Artif. Agents Human-Machine Interact., pp. 48-55, 2016.
- [5] A. R. Barse, "A Survey Paper on Smart Mirror Notice Board," International Journal of Science Technology Management and Research vol. 2, no. 10, pp. 1-3, 2017.
- [6] P. K. A. and A. E. S. M. Anwar Hossain, "SMART MIRROR FOR AMBIENT HOME ENVIRONMENTS," J. Can. Dent. Assoc., vol. 70, no. 3, pp. 156-7, Mar. 2004.
- [7] S. Athira, F. Francis, R. Raphel, N. S. Sachin, S. Porinchi, and S. Francis, "Smart mirror: A novel framework for interactive display," Proc. IEEE Int. Conf. Circuit, Power Comput. Technol. ICCPCT 2016, 2016.

- [8] S. Paszkiel and R. Kania, "Interactive, Multifunctional Mirror – Part of a Smart Home," *Informatics, Control. Meas. Econ. Environ. Prot.*, vol. 6, no. 1, pp. 66–68, 2016.
- [9] J. B. Panchal and K. P. Kandoriya, "Survey on Hand Gesture Recognition Methods," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering* vol. 4, no. 11, pp. 271–278, 2014.
- [10] F. M. Caputo, P. Prebianca, A. Carcangiu, L. D. Spano, and A. Giachetti, "Comparing 3D trajectories for simple mid-air gesture recognition," *Comput. Graph.*, vol. 73, pp. 17–25, 2018.
- [11] F. Chen, J. Deng, Z. Pang, M. Baghaei Nejad, H. Yang, and G. Yang, "Finger Angle-Based Hand Gesture Recognition for Smart Infrastructure Using Wearable Wrist-Worn Camera," *Appl. Sci.*, vol. 8, no. 3, p. 369, 2018.
- [12] G. Bailly, J. Müller, and E. Lecolinet, "Design and evaluation of finger-count interaction: Combining multitouch gestures and menus," *Int. J. Hum. Comput. Stud.*, vol. 70, no. 10, pp. 673–689, 2012.
- [13] J. Dulayatrakul, P. Prasertsakul, T. Kondo, and I. Nilkhamhang, "Robust implementation of hand gesture recognition for remote human-machine interaction," *Proc. - 2015 7th Int. Conf. Inf. Technol. Electr. Eng. Envisioning Trend Comput. Inf. Eng. ICITEE 2015*, pp. 247–252, 2015.
- [14] A. Atia, S. Takahashi, and J. Tanaka, "Smart gesture sticker: Smart hand gestures profiles for daily objects interaction," *Proc. - 9th IEEE/ACIS Int. Conf. Comput. Inf. Sci. ICIS 2010*, pp. 482–487, 2010.
- [15] C. Jeffrey and B. Terra, "Touchless Tactile Displays for Digital Signage: Mid-air Haptics meets Large Screens, SIGCHI Conf. Hum. Factors Comput. Syst. 1–4, 2018.
- [16] R. Walter, G. Bailly, and J. Müller, "StrikeAPose: revealing mid-air gestures on public displays," *SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 841–850, 2013.
- [17] D. Lee, J. S. Kim, and H. J. Lee, "Fast hand and finger detection algorithm for interaction on smart display," *Displays*, no. February, 2018.
- [18] P. Narayana, J. R. Beveridge, and B. A. Draper, "Gesture Recognition : Focus on the Hands." *IEEE/CVF Conf on Computer Vision and Pattern Recognition 2018*
- [19] G. S. N and S. Jayanthi, "Virtual Control Hand Gesture Recognition System Using Raspberry Pi," *Asian Res. Publ. Netw. J. Eng. Appl. Sci.*, vol. 10, no. 7, pp. 2989–2993, 2015.
- [20] M. Singh and G. Singh, "Interfacing Raspberry Pi With Led to interact with hand gestures in MATLAB : A Research," vol. 3, no. 1, pp. 414–417, 2015.
- [21] W. A. Hussein, S. Sahran, and S. N. H. S. Abdullah, "A fast scheme for multilevel thresholding based on a modified bees algorithm," *Knowledge-Based Syst.*, vol. 101, pp. 114–134, 2016.
- [22] C. Sha, J. Hou, and H. Cui, "A robust 2D Otsu's thresholding method in image segmentation," *J. Vis. Commun. Image Represent.*, vol. 41, no. October, pp. 339–351, 2016.

# An Enhanced Distance Vector-Hop Algorithm using New Weighted Location Method for Wireless Sensor Networks

Fengrong Han<sup>1\*</sup>, Izzeldin Ibrahim Mohamed Abdelaziz<sup>2</sup>, Xinni Liu<sup>3</sup>, Kamarul Hawari Ghazali<sup>4</sup>

Faculty of Electrical & Electronics Engineering,  
Universiti Malaysia Pahang  
Pahang, Malaysia

**Abstract**—Location is an indispensable segment for Wireless Sensor Network (WSN), since when events happened, we need to know location. The distance vector-hop (DV-Hop) technique is a popular range-free localization algorithm due to its cost efficiency and non-intricate process. Nevertheless, it suffers from poor accuracy, and it is highly influenced by network topology; Especially, more hop counts lead to more errors. In the final phase, least squares are employed to address nonlinear equation, which will gain greater location errors. Aimed at addressing problems mentioned above, an enhanced DV-Hop algorithm based on weighted factor, along with new weighted least squares location technique, is proposed in this paper, and it is called WND-DV-Hop. First, the one hop count of unknown node was corrected by employed received signal strength indication (RSSI) technology. Next, in order to reduce average hop distance error, a weighted coefficient based on beacon node hop count was constructed. A new weighted least squares method was embedded to solve nonlinear equation problem. Finally, considerable experiments were carried out to estimate the performance of WND-DV-Hop, compared the outcomes with state-of-the-art DV-Hop, IDV-Hop, Checkout-DV-Hop, and New-DV-Hop depicted in literature. The empirical findings demonstrated that WND-DV-Hop significantly outperformed other localization algorithms.

**Keywords**—Wireless Sensor Network (WSN); localization algorithm; range-free; distance vector-hop (DV-Hop) localization algorithm

## I. INTRODUCTION

Advanced electronics and cutting-edge wireless communication technologies have fostered large-scale wireless sensor network (WSN). The WSN is composed of great amounts of self-organizing, tiny size, and limited computational sensor nodes, which reflects a multiple hop network [1, 2]. The WSN has been considered as one of the most promising technologies to deal with tough issues. It has been successfully employed in disparate areas for monitoring and tracking purposes, such as military, public industry, agriculture, environment, and health care [3, 4].

Location is a fundamental issue in WSN, as it is crucial to identify where the information is derived from, mainly because the data becomes meaningless with information about location. This is especially true for real-time applications that demand precise location-based services [5]. Some fundamental

techniques developed for WSN require sensor location information, such as geographical routing protocols [6]. Additionally, several principle location-based network services need support from accurate location data, such as network coverage optimization, topology structure, and beacon node clustering [7].

Exact locations may be picked from sensor nodes attached to Global Positioning System (GPS) chips. Yet, this incontrovertibly increases hardware costs despite being equipped as a part of the nodes. The performance of localization accuracy, nonetheless, is not good if GPS is installed in tall buildings or other unsuitable environments surrounded by obstacles. Many scholars have proposed a range of algorithms and models. The localization schemes could be broadly grouped into range-free and range-based localization [5], depending if they need to attach additional hardware device(s). The range-free algorithms only utilize network connectivity data and hop information to calculate the location of the sensor nodes, thus easy to apply and operate. Some of these classic algorithms are Amorphous [8], Centroid [9], Distance Vector-Hop (DV-Hop) [10] and Approximate Point in Triangle Test (APIT) [11]. The range-based localization techniques embed various measuring techniques to collect favorable location data, including Received Signal Strength Indicator (RSSI) [12], Time of Arrival (ToA) [13], Time Difference of Arrival (TDoA) [13], and Angle of Arrival (AoA) [14]. These techniques offer higher localization accuracy with extra hardware cost. Nonetheless, they are sensitive to environmental noises and easily influenced by barriers. Hence, range-free localization algorithm became to most popular location method.

The DV-Hop propagation model has attracted much attention from researchers worldwide for its advantages of simplicity, feasibility, cost-efficiency, and high coverage. It does not rely on measurement error, easy to understand and implement, and has broad popularity within the localization domain. As such, this study investigated the DV-Hop scheme.

The essential function of DV-Hop is to calculate nodes distances by multiplying average hop size and hop counts between nodes. Suppose that the path of minimum hop is nearly beeline, the initial DV-hop algorithm performs better, exclusively when the distribution of nodes is uniform. If otherwise, a wide gap can be expected between the calculated

and the actual location. The literature presents a sea of related approaches to address above issues. Chen et al. [15] proposed a rapid, accurate, and easy DV-Hop localization algorithm called IDVLA. It computes the average of the entire hop size, instead of the initial hop. After that, a new weighted least squares method was used to substitute least squares. Next, an enhanced algorithm, hybrid DV-Hop algorithm, was initiated by Omar et al. [16]. If hop count between unknown and beacon node is 1, the RSSI method was used to estimate distances, or otherwise, the conventional method was applied. Upon localization, the node serves as an assistant beacon node. According to Peng et al. [17], although the GADV-Hop algorithm used genetic algorithm (GA) to minimize the total estimation error, it suffers from high computation complexity. Thus, a trade-off was embedded between localization accuracy and computation intricacy for a new proposed scheme. A comprehensive review of previous work is presented in Section 2.

This study proposes an enhanced DV-Hop localization algorithm based on weighted factor along with a new weighted least squares method for WSN.

The primary contributions of this paper are summarized in the following:

- 1) One hop node and its estimated distances is corrected by using the RSSI technique.
- 2) The average hop size is calibrated by a weighted coefficient to minimize the estimated distance error.
- 3) Finally, a new weighted least square is proposed to address the nonlinear equation to estimate node coordinates.

The remaining segment of this paper is as follows. New-born literature pertaining to DV-Hop is reviewed in Section 2. In Section 3, error analysis of basic DV-Hop is described. Elaboration of proposed WND-DV-Hop algorithm is depicted in Section 4. In Section 5, simulation outcomes and performance evaluation are discussed. Finally, conclusions and future work are presented in Section 6.

## II. RELATED WORKS

This section presents several remarkable studies concerned DV-Hop algorithm. Details pertaining to localization algorithm process are given in the following. Gui et al. [18] proposed two new algorithms based on improved protocol. They merely applied three nearest beacons for unknown node, instead of using all connected beacons to detect the location. It was assumed that two sensor nodes had consistent connectivity and close position. The hypothesis, however, is not always satisfied due to random deployment. Tomic et al. [19] initiated three new localization proposals. The first two proposals applied the geometry method to determine the best beacon node, while the third algorithm substituted it as bounded least squares issue, wherein quadratic programming was used to address. The simulation results proved that the three algorithms gave better performances, but increased computational complexity. A new approach of weighted hop distance is presented in [20]. The average size distance was reduced based on the weighted coefficient value. The simulation outcomes showed improvement in localization accuracy by 10%-15%. To enhance the accuracy, a new localization algorithm was investigated in [21]. First, the estimated distance error was

modified by using the orthogonal polynomial fitting approach in the second stage. Next, square after subtraction was employed in the third stage. Finally, a weighted matrix is used to refine the coordinates of unknown node.

Song et al. [22] introduced two refined algorithms. The first used the mean value of all average hop size, instead of the conventional one that decreased the error by 15%-20%. The IWC-DV-hop algorithm reduced the error by 9%, which enhanced the accuracy by selecting suitable beacon nodes combined with centroid localization. Zhang et al. [23], first, analyzed drawbacks of DV-hop. Next, they proposed a new weighted localization combined with centroid algorithm, which improved the accuracy by 10%. It was concluded that the nearest beacon had a higher impact. Fang et al. [24] presented a compensation coefficient to revise the estimated distances, which could be applied for both random distribution and dynamic topology networks. The simulation results demonstrated that it gave better performance in location accuracy, and reduced error by 18%. However, the proposed algorithm increased both computation overhead and computation time.

In [25], DV-MaxHop was proposed for anisotropic network. Only maximum hop count beacon was selected for location estimation decision, while the rest were omitted. By selecting the optimal MaxHop, it achieved enhancement by 20% in terms of accuracy and efficiency. Wang et al. [26] proposed a hybrid GA with simplex method to gain better accuracy. The weighted coefficient was employed to calculate the estimated distance. Finally, GA combined with simplex method was introduced. Kumar et al. [27] proposed NDV-Hop localization that revised the hop size of unknown nodes by introducing a boost term that greatly eliminated communication. Next, they applied the unconstrained optimization technique to minimize error terms, which significantly minimized time and energy consumption. The simulation results demonstrated that the location error was lower by 18%.

Kaur et al. [28] proposed a nature-inspired algorithm, GWO-DV-Hop. First, Grey-Wolf optimization was employed to enhance hop size by combining grey wolf optimizer. Next, a weighted Grey-Wolf was applied to the weighted average hop distance by considering the impact of all beacons. The proposed algorithm improved the localization accuracy by almost 10%. Kaur et al. [29] also presented a new localization using single mobile beacon node based on advanced path model. It displayed higher accuracy and network coverage, despite under sparse network. Kaur et al. [30] also introduced the Gauss-Newton method to address nonlinear method. This algorithm was implemented under 3D WSN and had managed to reduce error by almost 20%. In [31], three improved DV-hop localization algorithms based on optimization techniques are presented. The average localization errors of the three algorithms had been declined. The ICA-DV-hop resulted in the highest localization accuracy among the three algorithms. Deepak et al. [32] introduced a new metric-based method, in which the author proved that it gave less errors, when compared to the conventional DV-Hop using mathematical analysis.

All proposed algorithms have enhanced localization accuracy to a certain extent, as most of them have several shortcomings, including computational complexity and communication overhead. Hence, this study had attempted to enhance localization accuracy, minimize communication overhead, and reduce computation intricacy.

### III. DV-HOP AND RSSI LOCALIZATION ALGORITHM

#### A. DV-Hop Localization Algorithm

DV-Hop localization scheme was designed by Dragons Niculescu and his team [10] for 2D WSN, which was a classic range-free distributed localization algorithm. Generally, it was divided into three parts.

##### Phase 1: Calculate Minimum Hop Counts

In phase 1, the beacon node,  $A_i$ , broadcasts a packet {ID;  $x_i$ ,  $y_i$ ,  $h_i$ }, ( $x_i$ ,  $y_i$ ), that represents the coordinated of  $A_i$ ,  $h_i$  is hop count, which is increased by 1 from its initial value of 0. If node  $B$  receives a smaller hop count, it will update  $H_i$ , otherwise, the packet will be discarded. Every node gets its minimum hop counts to all beacon nodes by this mechanism.

##### Phase 2: Calculate of Average Hop Size

In the second phase, beacon node,  $A_i$ , calculates its average hop size, expressed as  $AvgHopSize_i$  using Equation (1) given in the following:

$$AvgHopSize_i = \frac{\sum_{i \neq j}^m \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sum_{i \neq j}^m H_{ij}} \quad (1)$$

Where, ( $x_i$ ,  $y_i$ ) and ( $x_j$ ,  $y_j$ ) are the coordinates of beacon nodes  $i$  and  $j$ , respectively. Meanwhile,  $H_{ij}$  is the minimum hop-count value between them, and  $m$  is the number of beacon nodes.

After received the hop size, each beacon node broadcasts its  $AvgHopSize_i$  in the network. The unknown node can calculate distance,  $d_{iu}$ , from the nearest beacon node by multiplying  $AvgHopSize_i$  with hop count using Equation (2) given in the following:

$$d_{iu} = AvgHopSize_i \times H_{iu} \quad (2)$$

Where  $d_{iu}$  and  $H_{iu}$  are estimate distance and the minimum hop count between beacon node  $i$  and unknown node  $u$ , respectively.

##### Phase 3: Calculation of Unknown Nodes Coordinates

In the last phase, the coordinate of each located node is calculated by using least squares or maximum likelihood estimation method.

It is assumed that ( $x_u$ ,  $y_u$ ) are the coordinates of unknown node  $u$ , and let  $d_{iu}$  be distance between  $u$  and  $A_i$ ,  $i \in \{1, 2, 3, \dots, n\}$ , where  $n$  is the number of communicable beacon nodes. Equation (3) is expressed as follows:

$$\begin{aligned} (x_u - x_1)^2 + (y_u - y_1)^2 &= d_{1u}^2 \\ (x_u - x_2)^2 + (y_u - y_2)^2 &= d_{2u}^2 \\ &\vdots \end{aligned} \quad (3)$$

$$(x_u - x_n)^2 + (y_u - y_n)^2 = d_{nu}^2$$

Each equation was subtracted from the last equation since the first one, Equation (3), can be expressed as:

$$\begin{aligned} 2(x_n - x_1)x_u + 2(y_n - y_1)y_u &= d_1^2 - d_n^2 - x_1^2 + x_n^2 - y_1^2 + y_n^2 \\ 2(x_n - x_2)x_u + 2(y_n - y_2)y_u &= d_2^2 - d_n^2 - x_2^2 + x_n^2 - y_2^2 + y_n^2 \end{aligned} \quad (4)$$

$$2(x_{n-1} - x_n)x_u + 2(y_{n-1} - y_n)y_u = d_{n-1}^2 - d_n^2 - x_{n-1}^2 + x_n^2 - y_{n-1}^2 + y_n^2$$

Equation (4) can be transformed into  $AX=B$ ;

$$A = -2 \times \begin{bmatrix} x_1 - x_n & y_1 - y_n \\ x_2 - x_n & y_2 - y_n \\ \vdots & \vdots \\ x_{n-1} - x_n & y_{n-1} - y_n \end{bmatrix} \quad (5)$$

$$B = \begin{bmatrix} d_1^2 - d_n^2 - x_1^2 + x_n^2 - y_1^2 + y_n^2 \\ d_2^2 - d_n^2 - x_2^2 + x_n^2 - y_2^2 + y_n^2 \\ \vdots \\ d_{n-1}^2 - d_n^2 - x_{n-1}^2 + x_n^2 - y_{n-1}^2 + y_n^2 \end{bmatrix} \quad (6)$$

$$X = \begin{bmatrix} x_u \\ y_u \end{bmatrix} \quad (7)$$

The unknown node ( $x_u$ ,  $y_u$ ) can obtain its estimated coordinate based on least square estimations or maximum likelihood estimation method, as follows:

$$X = (A^T A)^{-1} A^T B \quad (8)$$

#### B. RSSI Localization Algorithm

The RSSI was proposed by L. Girod et al. in [8]. The estimate distance between sender and receiver can be obtained by using received power strength under a specific path loss model, as given in Equation (9).

$$p_r(d) = p_r(d_0) + 10n \log_{10} \frac{d}{d_0} + X_\sigma \quad (9)$$

Where,  $d$  is the distance between sender and receiver nodes,  $p_r(d)$  denotes received signal strength at distance  $d$ ,  $d_0$  represents the reference distance, usually taken as 1 m, and  $n$  is the path loss exponent.  $X_\sigma$  refers to power loss due to the shadowing effect on actual environments.

Let  $d_0 = 1$ ,  $RSSI = p_s - p_r(d)$ ,  $A = p_s - p_r(d_0)$ , RSSI can be simplified into Equation (10).

$$RSSI = A - 10n \log_{10} \frac{d}{d_0} + X_\sigma \quad (10)$$

Next, estimated distance,  $d$ , can be determined from Equation (11).

$$d = 10 \frac{A - RSSI}{10n} \quad (11)$$

### C. Error Analysis for Basic DV-Hop Algorithm

- Error by Minimum Hop Count

Problem 1:

The DV-Hop estimates its hop size by hop counts and distances between sensor nodes. Hop-count is discontinuous as the nodes are irregularly deployed within the monitoring region. This suggests a rather major error since several nodes may share similar hop-count with identical beacon node.

In Fig. 1, the red and blue keys denote beacon and unknown nodes, respectively. Fig. 1 depicts that unknown nodes  $j$ ,  $k$ , and  $m$  are within the communication radius of  $R$ , whereby the three of them are one-hop node to beacon node  $i$ . Consequently, the estimated distance between beacon node  $i$  and unknown nodes  $j$ ,  $k$ , and  $m$  is same. Nonetheless, the actual distance between them is apparently different. The estimation scheme of DV-Hop, thus, would lead to a blatant error. This erred trend has motivated this study to formulate an advanced approach that estimates precise minimum hop.

- Accumulated Estimated Error by Average Hop Size

Problem 2:

The basic DV-Hop presumes the minimum hop is nearly beeline distribution, which is in contradictory in the actual scenario. This is especially true when the sensor nodes in WSN are in spares, hence disabling the average hop size to represent the whole network. This demands a modification to be made on the average hop size.

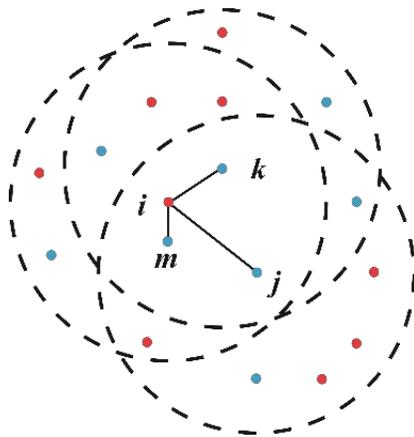


Fig 1. The Relationship of Node Distribution.

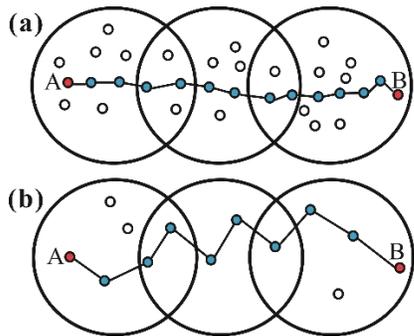


Fig 2. Node Distribution Model; (a), this is a Case Model Node Distribution in Dense; (b), this is a Case in Spares.

Based on Fig. 2(a), since the distance between A and B is on the same line with high network density, the estimated distance is accurate. Nevertheless, in Fig. 2(b), the average hop distance differs, thus resulting in a massive error of estimate distance between A and B to the actual distance. Besides, in phase 3, the unknown node is bound to select the nearest beacon node's average hop size to be multiplied with hop count, hence leading to a grave error.

### IV. PROPOSED ALGORITHM WND-DV-HOP

#### A. Improved DV-Hop based on Weighted Correction

**Phase 1:** The conventional DV Hop algorithm was used in this phase with each node having a minimum hop count.

**Phase 2:** In this phase, the conventional algorithm was improved to determine the correction of one hop count of the unknown node. The average hop size was revised in this phase.

- Correct One Hop Count to Unknown Node

This sub-section describes the improved algorithm to address *problem 1* (see **Section III, C**).

If the hop count of unknown node to neighbour beacon node is 1, the estimated distance is determined using RSSI Equation (11) instead of Equation (2). This enables a comparison of the received wireless signal strength with beacon node at  $(R/2)$ . If  $p_r(d)$  exceeds  $p_r(R/2)$ , its hop value is 0.5, while 1 hop if otherwise. This method introduces the hop hierarchical processing on nodes whose hop is 1, so that the hop value is no longer an integer. This discretisation process can effectively enhance the credibility of hop counts.

- Correct Average Hop-size of Beacon Nodes

This part discusses to overcome *problem 2* (see **Section III, C**).

Beacon and unknown nodes are denoted as  $m$  and  $n$ , respectively. Let coordinates of beacon nodes  $I$  and  $J$  be  $(x_i, y_i)$  and  $(x_j, y_j)$ , respectively. The average hop size of beacon node  $I$  can be determined using Equation (1). The average of  $AvgHopSize_i$  is calculated using Equation (12):

$$\overline{AvgHopSize} = \sum_{i=1}^m AvgHopSize_i \quad (12)$$

Here,  $\overline{AvgHopSize}$  is employed instead of  $AvgHopSize_i$ . The estimated distance between beacon nodes  $I$  and  $J$  can be estimated using Equation (13):

$$d_{ij}^{est} = \overline{AvgHopSize} \times H_{ij} \quad (13)$$

Where,  $H_{ij}$  refers to hop count between beacon nodes  $I$  and  $J$ . Hence, the actual distance between them is as follows:

$$d_{ij}^{act} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (14)$$

Hop error is given in the following:

$$E_i = \frac{1}{m-1} \sum_{\substack{i=1 \\ i \neq j}}^m \frac{(d_{ij}^{act} - d_{ij}^{est})}{H_{ij}} \quad (15)$$

The new average of hop size is given below:

$$AvgHopSize_i^{new} = AvgHopSize_i - E_i \quad (16)$$

- Correct Estimate Distance of Unknown Node

#### 1) One Hop

If the hop between unknown and beacon nodes is 1, the distance is estimated determined using RSSI approach with Equation (11), instead of Equation (2).

#### 2) Multiple Hop Counts

The weighted coefficient was employed to correct the estimated distance for multiple hop counts. The details are given in the following.

If both beacon and unknown nodes are closer, the average error per hop becomes smaller. This offers a more accurate estimated distance of the unknown node hop. Information regarding the location of beacon node obtained by unknown node  $U$  is  $k$ . The unknown node assigns varying weights to each beacon node (e.g. beacon node  $I$ ). Here,  $W_i$  is computed using the following equation:

$$w_i = \frac{\frac{1}{E_i} + \frac{1}{H_{iu}}}{\sum_{i=1}^k (\frac{1}{E_i} + \frac{1}{H_{iu}})} \quad (17)$$

Based on the above calculation, the distance between unknown node  $U$  and beacon node (e.g.  $I$ ) can be obtained by the following:

$$d_{ui} = \left( \sum_{i=1}^k (w_i \times AvgHopSize_i^{new}) \right) \times H_{ui} \quad (18)$$

Phase 3:

Yan et al. [33] proposed an optimal weighted least square for irregular network. Based on this idea, we introduced it to regular network, employed a weighted coefficient matrix  $W$  to address larger error caused by least squares method, see as following.

$$W = \begin{bmatrix} W_1 & 0 & 0 & 0 \\ 0 & W_2 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & W_k \end{bmatrix} \quad (19)$$

Where,  $W_k = 1/H_k^3$ ,  $H_k^3$  is the minimum number hop count between target node  $X$  and anchor node  $A$ . Hence, Equation (8) can be transformed into Equation (20).

$$X = (A^T W^T W A)^{-1} A^T W^T W B \quad (20)$$

### B. Steps of Improved Algorithm WND-DV-Hop

**Step 1.** Minimum hop count is acquired by flooding protocol among nodes, which is like that for DV-Hop localization algorithm in phase 1.

**Step2.** Each node knows its shortest hop count. If the hop count of unknown nodes to neighbour beacon node is 1, RSSI is employed to correct its hop count and to estimate its distance to neighbour beacon node. The information table is updated and the message is forwarded to their neighbours.

**Step 3.** This phase is like basic DV-Hop, whereby each  $AvgHopSize$  of beacon node is calculated by Equation (16) to correct the average hop size.

**Step 4.** The weighted coefficient is applied to correct the estimate distance of unknown node with multiple hop counts.

**Step 5.** The coordinates of unknown nodes are calculated by weighted least square method. The flow diagram of improved algorithm (WND-DV-Hop) is illustrated in Fig. 3.

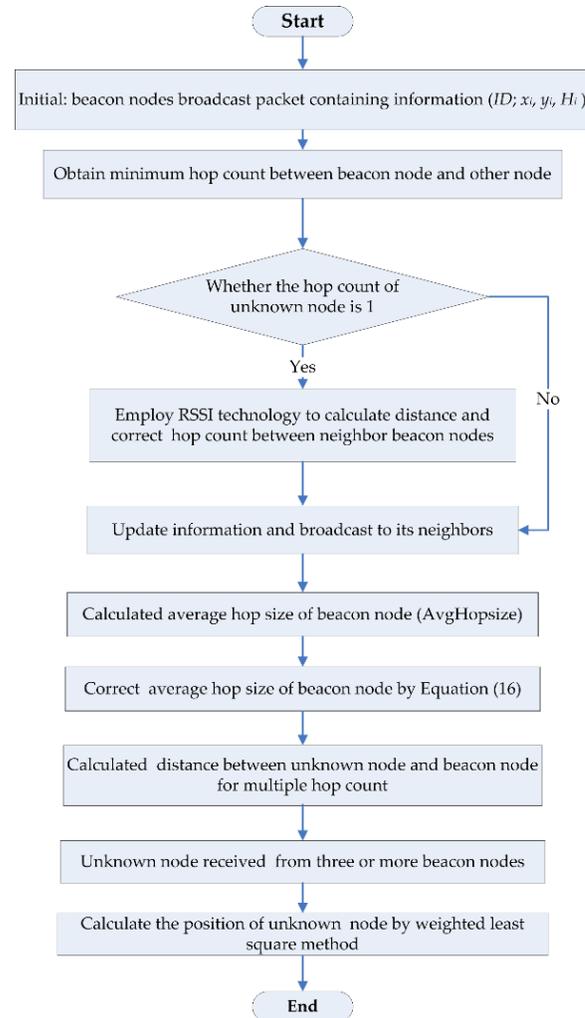


Fig 3. The Flowchart of our Proposed Algorithm (WND-DV-Hop).

## V. EXPERIMENTAL RESULTS AND ANALYSIS

The performance of the proposed WND-DV-Hop algorithm had been assessed by weighing in localization accuracy, localization stability, and computational cost under several conditions, such as node density, beacon node ratio, and communication radius. A classic representation of node distribution in 2D space is portrayed in Fig. 4. A total of 150 nodes are randomly displayed in the  $150 \times 150 \text{ m}^2$  area, including 30 beacon nodes denoted by red pentacles.

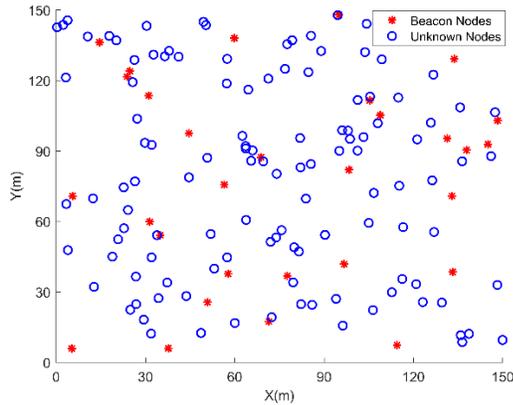


Fig 4. A Typical Example of Node Distribution in 2D Space.

### A. Experimental Environment

In verifying the performance of WND-DV-Hop, comprehensive experiments were conducted in MATLAB 2016a. The experimental outcomes were compared with DV-Hop [10], IDV-Hop [15], CheckOut-DV-Hop [18], and New-IDV-Hop [32] in simulated settings. Table I tabulates the simulation parameters.

TABLE I. SIMULATION PARAMETERS SETTING

Parameters	Value
Network Size	100m×100m
Total nodes	100
Beacon Nodes	30
Communication Range(m)	25

### B. Evaluation Criteria

In evaluating and analysing the achievement of the proposed WND-DV-Hop, localization accuracy, stability, and cost metrics had been considered.

- Accuracy Metrics

#### 1) Localization Error (LE)

The LE refers to error between actual and calculated coordinates of unknown nodes, as defined by Equation (21).

$$LE = \sqrt{(x_u - x_a)^2 + (y_u - y_a)^2} \quad (21)$$

#### 2) Average Localization Error (ALE)

The ALE denotes the sum of localization error for unknown nodes. Its mathematical expression is as follows:

$$ALE = \frac{\sum_{u,a=1}^n \sqrt{(x_u - x_a)^2 + (y_u - y_a)^2}}{n} \quad (22)$$

- Stability Metrics

#### 1) Localization Error Variance (LEV)

The LEV determines the stability of localization algorithm, given as follows:

$$LEV = \sqrt{\frac{\sum_{u,a=1}^n (\sqrt{(x_u - x_a)^2 + (y_u - y_a)^2} - LEV \times R)^2}{n \times R^2}} \quad (23)$$

Here,  $n$  signifies the number of unknown nodes, whereas  $(x_u, y_u)$  and  $(x_a, y_a)$  are the estimated and actual coordinates of unknown nodes, respectively.  $R$  denotes communication range.

- Cost Metrics

#### 1) Computational Cost

Localization time (LT) is employed to evaluate computational cost, which is measured by the time taken to locate an unknown node.

### C. Experimental Results

Simulations for all algorithms were performed as many as 100 times for each result to assess the performance for random deployment. Abbreviations LE, ALE, LEV, and LT are used to represent localization error, average localization error, localization error variance, and localization time, respectively.

- LE for Each Unknown Node

The simulation was performed under the scenario that 100 sensor nodes were irregularly deployed in the area of  $100 \times 100 \text{ m}^2$  with 30% beacon node. The communication range was 25m.

Fig. 5 presents the LE for each unknown node under five algorithms. Apparently, the proposed algorithm (WND-DV-Hop) gave the best outcomes with reduction in localization error at 55%, 35%, 20%, and 25%, respectively, compared with DV-Hop [10], IDV-Hop [15], CheckOut-DV-Hop [18], and New-IDV-Hop [32].

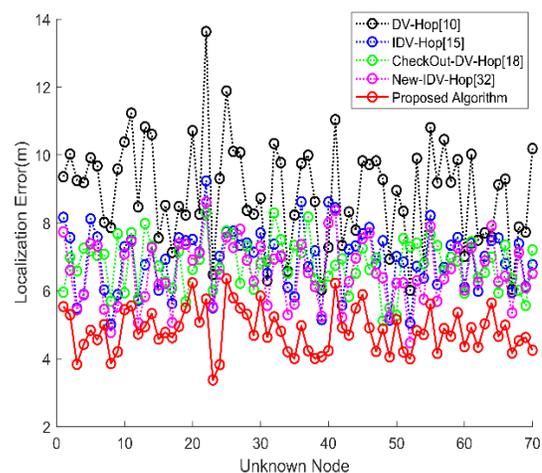


Fig 5. The LE for Each Unknown Node.

TABLE II. LOCALIZATION ERROR AND STANDARD DEVIATION COMPARISONS

Localization Algorithm	Max. LE(m)	Min. LE(m)	Avg. LE(m)	Std. (LE)
DV-Hop [10]	13.6290	6.0095	8.8912	1.5096
IDV-Hop [15]	9.2288	5.0009	6.9184	1.0293
CheckOut-DV-Hop [18]	8.2908	5.1165	6.9184	0.7299
New-DV-Hop [32]	8.643	4.4614	6.5341	1.0991
<b>Proposed algorithm</b>	<b>6.3517</b>	<b>3.2854</b>	<b>4.8309</b>	<b>0.6408</b>

Table II presents the LE and its standard deviation for the five localization algorithms. Upon comparing with the other four localization algorithms, the proposed algorithm yielded the lowest location error in terms of max, min, and average values. The proposed algorithm also recorded the lowest standard deviation, which indicated that WND-DV-Hop had more stability. The proposed algorithm appeared to perform better under average location error with almost 40.95% decrease, when compared with DV-Hop [10].

• Accuracy and Stability Metrics with Variation Factors

Accuracy and stability are two of most critical factors for any algorithm. In this study, ALE and LEV were applied to determine accuracy and stability of the proposed algorithm under the effects of total number of nodes, beacon node ratio, and communication range.

1) Effect of Total Number of Nodes

During this experiment, the sensor nodes were evenly increased from 50 to 350, while communication radius and the proportion of beacon nodes ratio were fixed at 25 m and 10%, respectively. Fig. 6, Fig. 7, Table III and Table IV list the empirical outcomes of ALE and LEV.

In Fig. 6, ALE displayed a declining trend with increment in sensor nodes. This pattern is attributable to the increasing number of nodes within the communication radius that enhanced both hop size and network connectivity. Fig. 6 shows that the proposed algorithm performed better than the other four algorithms. Its average localization error decreased to 70%, 55%, 50%, and 53%, when compared with the algorithm prescribed in [10,15,18,32], respectively.

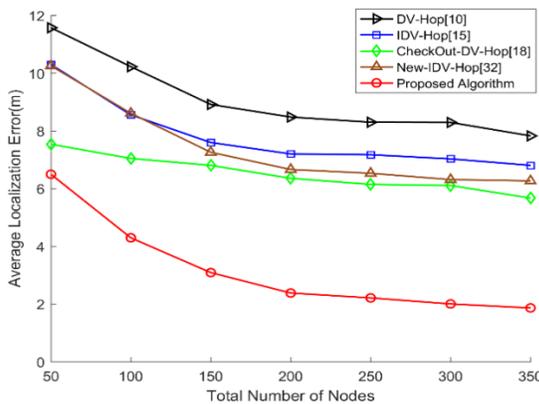


Fig 6. Simulation Results ALE Under Various Total Number of Nodes.

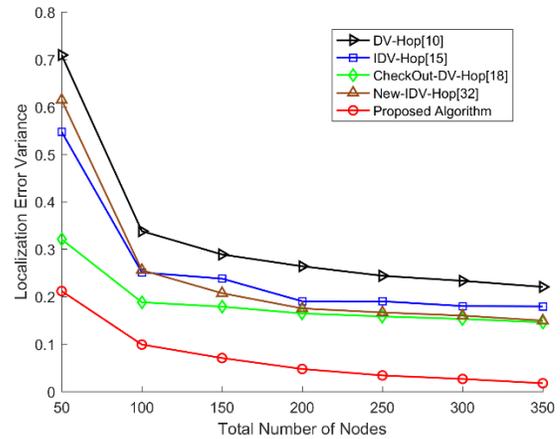


Fig 7. Simulation Results LEV Under Various Total Number of Nodes.

As portrayed in Fig. 7, reduction in LEV was noted with increment in the number of nodes in WSN. Generally, the stability of the proposed algorithm was significantly better than the rest of algorithm. In fact, the proposed algorithm displayed strong stability when sensor nodes exceeded 150. Among all algorithms, the New-IDV-Hop [32] also exerted good performance when the nodes exceeded 100.

As tabulated in Table III, the ALE of the proposed algorithm gave better performance in terms of max, min, and average values. Put simply, upon comparing with the primary DV-Hop [10], a decrease of 64.89% was noted in average term. Both IDV-Hop [15] and New-IDV-Hop [32] displayed superior performance in minimizing localization error.

TABLE III. COMPARISON ALE UNDER VARIOUS TOTAL NUMBER OF NODES

Localization Algorithm	Average Localization Error (ALE)		
	Max.	Min.	Avg.
DV-Hop [10]	11.5730	7.8253	9.0863
IDV-Hop [15]	10.2991	6.8037	7.8076
CheckOut-DV-Hop [18]	7.5354	5.6701	6.5226
New-DV-Hop [32]	10.2543	6.2659	7.4120
<b>Proposed Algorithm</b>	<b>6.4952</b>	<b>1.8630</b>	<b>3.1904</b>

TABLE IV. COMPARISON LEV UNDER VARIOUS TOTAL NUMBER OF NODES

Localization Algorithm	Localization Error Variance (LEV)		
	Max.	Min.	Avg.
DV-Hop [10]	0.7091	0.2205	0.3282
IDV-Hop [15]	0.5472	0.1793	0.2537
CheckOut-DV-Hop [18]	0.3207	0.1456	0.1870
New-DV-Hop [32]	0.4856	0.2512	0.3126
<b>Proposed Algorithm</b>	<b>0.2113</b>	<b>0.0174</b>	<b>0.0723</b>

Table IV shows that the stability performance of improved WND-DV-Hop emerged as the best amidst all algorithms; when compared with DV-HOP [10], IDV-Hop [15], CheckOut-DV-Hop [18], and New-DV-Hop [32], reduction of 77.98%, 71.50%, 61.34%, and 76.87% had been recorded in average term, respectively. This is especially true when the proposed algorithm achieved the lowest location error variance in terms of max, min, and average values.

2) Effect of Beacon Node Ratio

In the experiment, the beacon ration was increased from 10% to 40%, while number of nodes and communication range were fixed at 100 and 25 m, respectively. Fig. 8, Fig. 9, Table V and Table VI present the empirical finding of ALE and LEV under various beacon node ratio.

As illustrated in Fig. 8, increment in beacon nodes ratio led to a declining trend for ALE, stemming from the reduced hop count between the nodes. Increment in beacon nodes led to hop size being approximated to the actual value. The proposed algorithm always gave the lowest values; reduction of 50, 40, 35, and 35% when compared to DV-Hop [10], IDV-Hop [15], CheckOut-DV-Hop [18], and New-IDV-Hop [32], respectively.

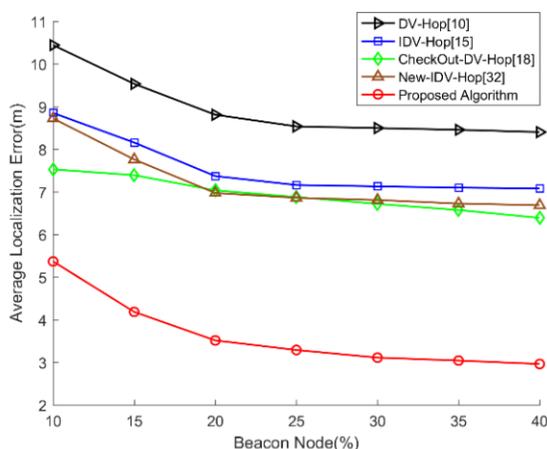


Fig 8. Simulation Results ALE Under Various Beacon Node Ratio.

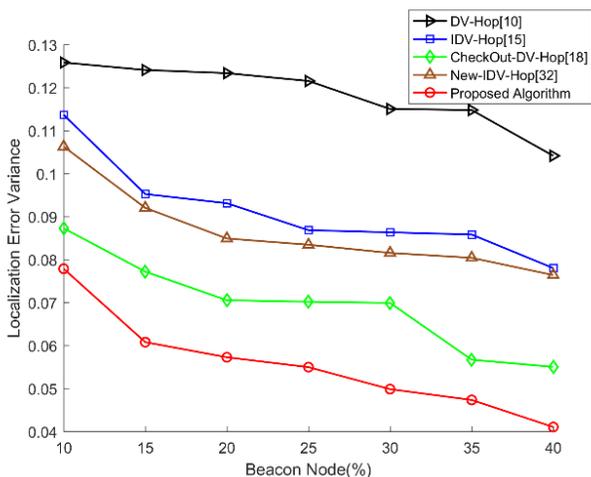


Fig 9. Simulation Results LEV Under Various Beacon Node Ratio.

TABLE V. COMPARISON ALE UNDER VARIOUS BEACON NODE RATIO

Localization Algorithm	Average Localization Error (ALE)		
	Max.	Min.	Avg.
DV-Hop [10]	10.4424	8.4023	8.9537
IDV-Hop [15]	8.8509	7.0817	7.5495
CheckOut-DV-Hop [18]	7.5527	6.3893	6.9306
New-DV-Hop [32]	8.7218	6.6879	7.2195
<b>Proposed Algorithm</b>	<b>5.3675</b>	<b>2.9621</b>	<b>3.6387</b>

TABLE VI. COMPARISON LEV UNDER VARIOUS BEACON NODE RATIO

Localization Algorithm	Localization Error Variance (LEV)		
	Max.	Min.	Avg.
DV-Hop [10]	0.1258	0.1041	0.1184
IDV-Hop [15]	0.1137	0.0780	0.0913
CheckOut-DV-Hop [18]	0.0873	0.0550	0.0695
New-DV-Hop [32]	0.1063	0.0764	0.0864
<b>Proposed Algorithm</b>	<b>0.0779</b>	<b>0.0410</b>	<b>0.0556</b>

As illustrated in Fig. 9, DV-Hop [10] exhibited the worst performance amidst the five algorithms. The proposed algorithm significantly outperformed IDV-Hop [15], and New-IDV-Hop [32]. The Checkout-DV-Hop [32] displayed better improvement, but not as exceptional as the proposed algorithm.

Table V shows that the ALE performance of the original DV-Hop is always the worst, while the proposed algorithm is the best. When compared with DV-Hop [10], IDV-Hop [15], CheckOut-DV-Hop [18], and New-DV-Hop [32], the average localization error decreased under average term by up to 59.36%, 51.80%, 47.50%, and 49.60%, respectively.

As given in Table VI, the LEV for the proposed algorithm emerged as outstanding due to the lowest max, min, and average values. The original DV-Hop is always the poorest from the rest. The LEV of the proposed algorithm decreased to 53.04%, 39.10%, 20.00%, and 35.65%, respectively, when compared with DV-Hop [10], IDV-Hop [15], CheckOut-DV-Hop [18], and New-DV-Hop [32].

3) Effect of Communication Range

In this experiment, the communication range was increased from 20 to 36 m, while sensor and beacon nodes were fixed at 100 and 20, respectively. Fig. 10, Fig. 11, Table VII and Table VIII tabulate the empirical outcomes with varied communication ranges.

Fig. 10 illustrates five algorithms of ALE under varied communication ranges. The empirical findings show that the ALE exerted a declining trend with increment in communication range. The proposed algorithm always had the lowest localization error, which decreased about 50, 35, 30, and 25%, respectively, when compared with DV-Hop [10], IDV-Hop [15], CheckOut-DV-Hop [32], and New-IDV-Hop [32]. Besides, ALE demonstrated only a slight change when the communication range exceeded 32 m.

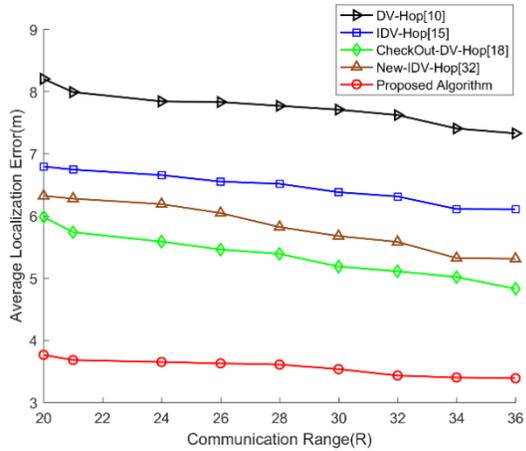


Fig 10. Simulation Results ALE Under Various Communication Range.

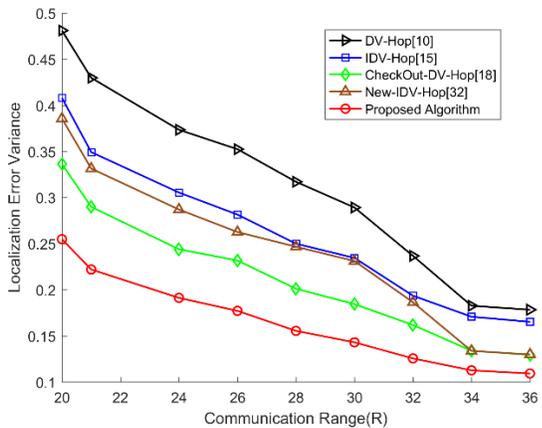


Fig 11. Simulation Results LEV Under Various Communication Range.

Increment in communication range decreased the LEV (see Fig. 11). This result is attributable to the network that is well-connected with the increasing communication range. The proposed localization algorithm always gave the greatest performance, and was followed by CheckOut-DV-Hop [18]. The performances of DV-Hop [10], IDV-Hop [15], and New-IDV-Hop [32] were close. When the communication range exceeded 32 m, the localization error variance for all the algorithms gave low values.

TABLE VII. COMPARISON ALE UNDER VARIOUS COMMUNICATION RANGE

Localization Algorithm	Average Localization Error (ALE)		
	Max.	Min.	Avg.
DV-Hop [10]	8.2024	7.3263	7.7440
IDV-Hop [15]	6.7933	6.1074	6.4633
CheckOut-DV-Hop [18]	5.9815	4.8280	5.3657
New-DV-Hop [32]	6.3221	5.3111	5.8381
<b>Proposed Algorithm</b>	<b>3.7648</b>	<b>3.3880</b>	<b>3.5654</b>

TABLE VIII. COMPARISON LEV UNDER VARIOUS COMMUNICATION RANGE

Localization Algorithm	Localization Error Variance (LEV)		
	Max.	Min.	Avg.
DV-Hop [10]	0.4810	0.1784	0.3156
IDV-Hop [15]	0.4081	0.1653	0.2620
CheckOut-DV-Hop [18]	0.3363	0.1294	0.2125
New-DV-Hop [32]	0.3856	0.1298	0.2438
<b>Proposed Algorithm</b>	<b>0.2548</b>	<b>0.1092</b>	<b>0.1657</b>

The result of both ALE is tabulated in Table VII. It seems that the enhanced algorithm outperformed the others for average localization error. The CheckOut-DV-Hop [32] gave the second-best performance. The average localization error of the proposed algorithm decreased to 53.96%, 44.84%, 33.55%, and 38.93% under average term, respectively, when compared with DV-Hop [10], IDV-Hop [15], CheckOut-DV-Hop [18], and New-DV-Hop [32].

Table VIII shows that the performance of proposed algorithm was outstanding in terms of max, min, and average LEV values. Upon comparing with the proposed algorithm, the best min gap of max and the best average location error variance were displayed by CheckOut-DV-Hop [18] and New-DV-Hop [32]. The original DV-Hop, on the contrary, gave the worst performance.

- Cost Metrics

In order to verify the proposed algorithm under cost metrics, the LT was calculated. This simulation experiment employed 100 nodes that were irregularly deployed in area of 100 × 100 m<sup>2</sup>, the beacon node ratio of 30%, and communication range set at 25 m.

Table IX presents the LT of five algorithms with 100 nodes and 30 beacon nodes. It can be concluded that the LT of the basic DV-Hop [10] algorithm was at minimum, whereas IDV-Hop [15] and New-DV-Hop [32] were close. The localization time of the proposed algorithm was slower than the above three algorithms, but faster than the CheckOut-DV-Hop [18] to some extent. This is ascribed to correct hop and hop count cost relative time. Hence, a pressing need is present to balance the decreasing location accuracy and localization time.

TABLE IX. COMPARISON OF ALGORITHMS WITH LOCALIZATION TIME

Localization Algorithm	Localization Time(s)
DV-Hop [10]	0.6572
IDV-Hop [15]	0.6605
CheckOut-DV-Hop [18]	0.7247
New-DV-Hop [32]	0.6734
<b>Proposed Algorithm</b>	<b>0.7112</b>

## VI. CONCLUSION

Aimed at enhancing localization accuracy, stability, and cost; an enhanced DV-Hop algorithm based on weighted factor is described in this paper. The RSSI scheme was embedded to correct one hop-count of unknown node, which transformed the one hop-count from discrete to continuous values. In order to reduce localization errors, the hop size was modified by using weighted correction factor. The estimated distance between beacon and located node was calculated by weighted hop count, instead of the nearest one. A novel weighted least squares method is proposed to address nonlinear equations in estimating node coordinates. Several parameters were weighed in to determine their effects, including communication range, total number nodes, and beacon nodes ratio. The simulation outcomes revealed that the proposed algorithm was exceptional for localization accuracy and stability, when compared with DV Hop, IDV-Hop, CheckOut-DV-Hop, and New-IDV Hop. In future, we are planning to implement WND-DV-Hop in irregular and sparse network under consider irregular radio patterns. In addition, our proposed algorithm will be assessed under much complicated three-dimensional (3D) wireless network environments with consider specific environmental factors.

## ACKNOWLEDGMENT

This research was supported by Universiti Malaysia Pahang Postgraduate Research Grants, grant number PGRS1903143.

## REFERENCES

- [1] Akyildiz, I. F.; Su, W.; Sankarasubramaniam, Y.; Cayirci, E. J. I. C. m., A survey on sensor networks. 2002, 40, (8), 102-114.
- [2] Yick, J.; Mukherjee, B.; Ghosal, D. J. C. n., Wireless sensor network survey. 2008, 52, (12), 2292-2330.
- [3] Chandanapalli, S. B.; Reddy, E. S.; Lakshmi, D. R. J. I. J. o. M. L.; Cybernetics, DFTDT: distributed functional tangent decision tree for aqua status prediction in wireless sensor networks. 2018, 9, (9), 1419-1434.
- [4] Suo, H.; Wan, J.; Huang, L.; Zou, C. In Issues and challenges of wireless sensor networks localization in emerging applications, 2012 International Conference on Computer Science and Electronics Engineering, 2012; IEEE: 2012; pp 447-451.
- [5] Chowdhury, T. J.; Elkin, C.; Devabhaktuni, V.; Rawat, D. B.; Oluoch, J. J. C. N., Advances on localization techniques for wireless sensor networks: A survey. 2016, 110, 284-305.
- [6] Laoudias, C.; Moreira, A.; Kim, S.; Lee, S.; Wirola, L.; Fischione, C. J. I. C. S.; Tutorials, A survey of enabling technologies for network localization, tracking, and navigation. 2018, 20, (4), 3607-3644.
- [7] Khelifi, M.; Moussaoui, S.; Silmi, S.; Benyahia, I. J. I. J. o. S. N., Localization algorithms for wireless sensor networks: a review. 2015, 19, (2), 114-129.
- [8] Nagpal, R., Organizing a global coordinate system from local information on an amorphous computer. 1999.
- [9] Čapkun, S.; Hamdi, M.; Hubaux, J.-P. J. C. C., GPS-free positioning in mobile ad hoc networks. 2002, 5, (2), 157-167.
- [10] Niculescu, D.; Nath, B. In Ad hoc positioning system (APS), GLOBECOM'01. IEEE Global Telecommunications Conference (Cat. No. 01CH37270), 2001; IEEE: 2001; pp 2926-2931.
- [11] He, T.; Huang, C.; Blum, B. M.; Stankovic, J. A.; Abdelzaher, T. In Range-free localization schemes for large scale sensor networks, Proceedings of the 9th annual international conference on Mobile computing and networking, 2003; ACM: 2003; pp 81-95.
- [12] Girod, L.; Bychkovskiy, V.; Elson, J.; Estrin, D. In Locating tiny sensors in time and space: A case study, Proceedings. IEEE International Conference on Computer Design: VLSI in Computers and Processors, 2002; IEEE: 2002; pp 214-219.
- [13] Harter, A.; Hopper, A.; Steggle, P.; Ward, A.; Webster, P. J. W. N., The anatomy of a context-aware application. 2002, 8, (2/3), 187-197.
- [14] Niculescu, D.; Nath, B. In Ad hoc positioning system (APS) using AOA, IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428), 2003; IEEE: 2003; pp 1734-1743.
- [15] Chen, K.; Wang, Z.-h.; Lin, M.; Yu, M., An improved DV-Hop localization algorithm for wireless sensor networks. 2010.
- [16] Cheikhrouhou, O.; M Bhatti, G.; Alrobaea, R. J. S., A hybrid DV-hop algorithm using RSSI for localization in large-scale wireless sensor networks. 2018, 18, (5), 1469.
- [17] Peng, B.; Li, L. J. C. N., An improved localization algorithm based on genetic algorithm in wireless sensor networks. 2015, 9, (2), 249-256.
- [18] Gui, L.; Val, T.; Wei, A.; Dalce, R. J. A. H. N., Improvement of range-free localization technology by a novel DV-hop protocol in wireless sensor networks. 2015, 24, 55-73.
- [19] Tomic, S.; Mezei, I. J. T. S., Improvements of DV-Hop localization algorithm for wireless sensor networks. 2016, 61, (1), 93-106.
- [20] Tao, Q.; Zhang, L. In Enhancement of DV-Hop by weighted hop distance, 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2016; IEEE: 2016; pp 1577-1580.
- [21] Ma, X.; Liu, W.; Wang, Z. In Node localization of wireless sensor network based on secondary correction error, International symposium on parallel architecture, algorithm and programming, 2017; Springer: 2017; pp 142-151.
- [22] Song, G.; Tam, D. J. I. J. o. D. S. N., Two novel DV-Hop localization algorithms for randomly deployed wireless sensor networks. 2015, 11, (7), 187670.
- [23] Zhang, B.; Ji, M.; Shan, L. In A weighted centroid localization algorithm based on DV-hop for wireless sensor network, 2012 8th International Conference on Wireless Communications, Networking and Mobile Computing, 2012; IEEE: 2012; pp 1-5.
- [24] Fang, X. J. J. o. s. e., Improved DV-Hop positioning algorithm based on compensation coefficient. 2015, 9, (3), 650-657.
- [25] Shahzad, F.; Sheltami, T. R.; Shakshuki, E. M. J. I. T. o. M. C., DV-maxHop: A fast and accurate range-free localization algorithm for anisotropic wireless networks. 2016, 16, (9), 2494-2505.
- [26] Wang, F.; Wang, C.; Wang, Z.; Zhang, X.-y. J. I. J. o. D. S. N., A hybrid algorithm of GA+ simplex method in the WSN localization. 2015, 11, (7), 731894.
- [27] Kumar, S.; Lobiyal, D. J. T. S., Novel DV-Hop localization algorithm for wireless sensor networks. 2017, 64, (3), 509-524.
- [28] Kaur, A.; Kumar, P.; Gupta, G. P. J. W. P. C., Nature inspired algorithm-based improved variants of DV-Hop algorithm for randomly deployed 2D and 3D wireless sensor networks. 2018, 101, (1), 567-582.
- [29] Kaur, A.; Kumar, P.; Gupta, G. P. J. W. N., A new localization using single mobile anchor and mesh-based path planning models. 2019, 25, (5), 2919-2929.
- [30] Kaur, A.; Kumar, P.; Gupta, G. P., Sensor Nodes Localization for 3D Wireless Sensor Networks Using Gauss-Newton Method. In Smart Innovations in Communication and Computational Sciences, Springer: 2019; pp 187-198.
- [31] Kerdabadi, M. S.; Nejad, F. P.; Ghazizadeh, R.; Farrokhi, H. J. I. J. o. U. W. C.; Systems, Wireless sensor network localization using new heuristic optimisation algorithms. 2018, 3, (4), 209-218.
- [32] Prashar, D.; Jyoti, K. J. W. P. C., Distance Error Correction Based Hop Localization Algorithm for Wireless Sensor Network. 2019, 106, (3), 1465-1488.
- [33] Yan, X.; Sun, L.; Zhou J; Song, A., DV-hop localisation algorithm based on optimal weighted least square in irregular areas. 2018, 54, (21), 1243-1245.

# Early Forest Fire Detection System using Wireless Sensor Network and Deep Learning

Wiame Benzekri<sup>1</sup>, Ali El Moussati<sup>2</sup>, Omar Moussaoui<sup>3</sup>, Mohammed Berrajaa<sup>4</sup>

LANOL, Faculty of Sciences, Oujda, Morocco<sup>1,4</sup>

Department of Electronics, Informatics and Telecommunications, ENSAO<sup>2</sup>,  
University Mohammed Premier, Oujda, Morocco

MATSI Lab, ESTO, University Mohamed Premier, Oujda, Morocco<sup>3</sup>

**Abstract**—Due to the global warming, which mechanically increases the risk of starting fires. The number of forest fires is increasing and will increase more and more. To better support the fire soldiers on the ground, we present in this work a system for early detection of forest fires. This system is more precise compared to traditional surveillance approaches such as lookout towers and satellite surveillance. The proposed system is based on collecting environmental wireless sensor network data from the forest and predicting the occurrence of a forest fire using artificial intelligence, more particularly Deep Learning (DL) models. The combination of such a system based on the concept of the Internet of Things (IoT) is made up of a Low Power Wide Area Network (LPWAN), fixed or mobile sensors and a good suitable model of deep learning. That several models derived from deep learning were evaluated and compared enabled us to show the feasibility of an autonomous and real-time environmental monitoring platform for dynamic risk factors of forest fires.

**Keywords**—Forest fire detection; wireless sensor network; deep learning; internet of things; low power wide area network

## I. INTRODUCTION

Forests are indispensable resources to protect the ecological balance on the earth. A Forest fire cause damages to the ecosystems and causes various diseases for the population. In the recent years the number of forest fires has increased considerably throughout the world. Indeed, the 2019–20 Australian bushfire seasons began with several serious uncontrolled fires in June 2019. As of 14 January 2020, this season fires have burnt an estimated 18.6 million hectares [1]. It is estimated that one billion animals have been killed and some endangered species may be driven to extinction.

In addition, the 2019 Amazon rainforest wildfires season saw a year-to-year surge in fires occurring in the Amazon rainforest and Amazon biome within Brazil, Bolivia, Paraguay, and Peru during that year's Amazonian tropical dry season.

As of August 29, 2019, Institute for Space Research (INPE) reported more than 80,000 fires across all of Brazil, a 77% year-to-year increase for the same tracking period, with more than 40,000 in the Brazil's Legal Amazon which contains 60% of the Amazon.

Similar year-to-year increases in fires were subsequently reported in Bolivia, Paraguay and Peru, with the 2019 fire counts within each nation of over 19,000, 11,000 and 6,700,

respectively, as of August 29, 2019. It is estimated that over 906 thousand hectares of forest within the Amazon biome has been lost in fires in 2019.

Considering these examples of damage and for effective management of forest fire, it is necessary to make the: advance warning, prevention and early detection to detect the existence of fire and its location before it spreads to sending notification to the public and concerned authorities, response of mobilization, damage containment and providing medical care as well as relief to damaged citizen.

There have been many innovative techniques proposed in the past to build an accurate fire detection system. The popular forest-fire detection systems based on the satellite imagery monitoring systems like MODIS (Moderate Resolution Imaging Spectroradiometer) used in CANADA [2] and AVHRR (Advanced Very High Resolution Radiometer) used in CHINA [3]. The low resolution of satellites and the long scan period restrict the effectiveness of satellite-based forest fire detection.

The other fire detection system that used camera is proposed in articles [4, 5]; it's developing multi sensor wireless network, which uses IP cameras to detect fire. IP cameras need line-of-sight communication with sensors. Moreover, these cameras should be placed at a high altitude to get good view.

The light detection and Ranging (LIDAR) [6] was one of the famous projects developing in Canada to detect fire. It is system which measures the laser light back scattered by smoke particles. The limitation of this system was high false alert rate as result of climatic conditions for example: proximity of fog, clean particles, shadows etc.

Many researchers carried out work on early fire detection using Wireless Sensor Network (WSN). A system for the identification of forest fires has been developed in South Korea [7], the WSN decides the temperature and humidity after the middleware and the web application examines the collected information sent by sensors [8]. Alper Rifat has deployed on the same WSN principle and proposed temperature-based fire detection algorithm [9]. These systems are referenced for the detection just on the two parameters the temperature and the humidity, however there can exist other parameters which can influence like the wind [9].

This fire surveillance systems do not accomplish timely detection due to long period of scan and not realize constant

surveillance of the forest area. Considering the mentioned limitations of the existing systems presented, so, the most critical issue in a forest fire detection system is immediate response in order to minimize the damage and casualties.

Therefore, is necessary to have a scalable solution that can provide real-time fire detection, detect forest fire in early stage, and reduce false alarms.

The proposed system uses machine learning methods. The functionality of artificial intelligence is widely deployed in all sectors such as agriculture, health and industry. The algorithms of artificial intelligence can detect relations and models that can escape us. So, with using these algorithms, the system can predict fire with high accuracy. This paper present a novel forest fire approach. Firstly, the concept Internet of Things is proposed for the interconnection between the sensors and the different objects. IoT enables monitoring and control of different processes, which optimizes different operations that increase productivity and efficiency. Secondly, we propose models based on Deep Learning to predict triggering and the presence of fire.

This paper is organized as follow: Section 2 covers the background and related work. Section 3 provides the architecture of the proposed forest fire detection system. Section 4 presents results and discussion of the proposed system. Finally, Section 5 concludes.

## II. BACKGROUND AND RELATED WORK

### A. Fire Weather Index

To estimate the risk of occurrence of a forest fire, several approaches exist to calculate an index describing a level of risk for a given region or sub-region. This estimate is calculated according to the weather context. But on a finer and local scale, in a massif recognized as vulnerable to fire, vulnerability to fire presents, with some exceptions, a high spatial variability, linked to nature and structure of the stands, to exposure (to wind, sun), and the soil water reserve capacity. They must also be taken into account in prevention strategies.

The two most used approaches to estimate the risk of fire in a region are the Korean approach [10] and the Canadian approach [11].

Previously, both codes have implemented on the same processor, and we have confirmed that the Canadian approach is more precise, faster and less energy consuming [12]. That's why the Canadian approach is opted in the rest of this work.

The Canadian Method of Forest fire Weather Index (FWI) is made up of six components that take into account the effects of fuel water content and wind on fire behavior.

The diagram below shows the components of the Canadian Forest fire Weather Index Method (Fig. 1). Calculations are based on consecutive daily observations of temperature, relative humidity, wind speed and precipitation in the past 24 hours. The six standard components provide numerical assessments of the relative potential for wildland fire.

Table I provides a classification of fire danger as a function of the FWI index based on the data available from [13].

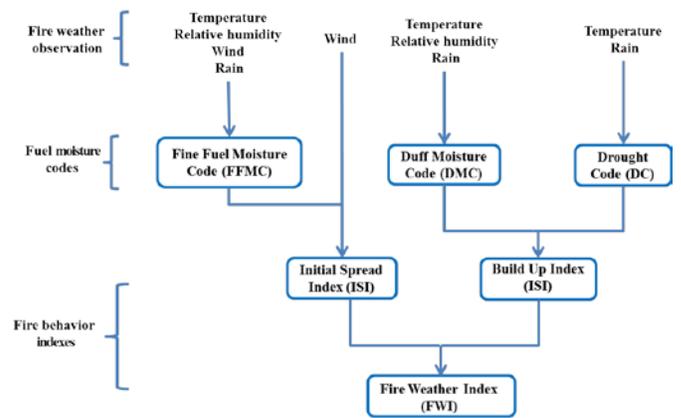


Fig 1. Structure of the Fire Weather Index System.

TABLE I. POTENTIAL FIRE DANGER VERSUS FWI VALUE

FWI class	Value Range	Type of Fire	Potential Danger
<i>Low</i>	0-5	Creeping surface fire	Fire well be self extinguishing
<i>Moderate</i>	5-10	Low vigor surface fire	Easily suppressed with hand tools
<i>High</i>	10-20	Moderately vigorous surface fire	Power pumps and hoses are required
<i>Very High</i>	20-30	Very intense surface fire	Diffucukt to control
<i>Extreme</i>	30+	Developing active fire	Immediate and strong action is critical



Fig 2. The Weather Station VISIOGREEN with Pyranometer.

These data must be collected continuously throughout the year, as the degree of dryness of the deep layers is built over long periods.

The weather station allows us to measure in real time the parameters that need to estimate the FWI (Fig. 2).

### B. IoT System with LoRa

An IoT system brings together many players and technological components. It is made up of connected objects, wireless communication networks, data collection / hosting / processing platforms, applications / services for end users and supervision / securing of the entire chain.

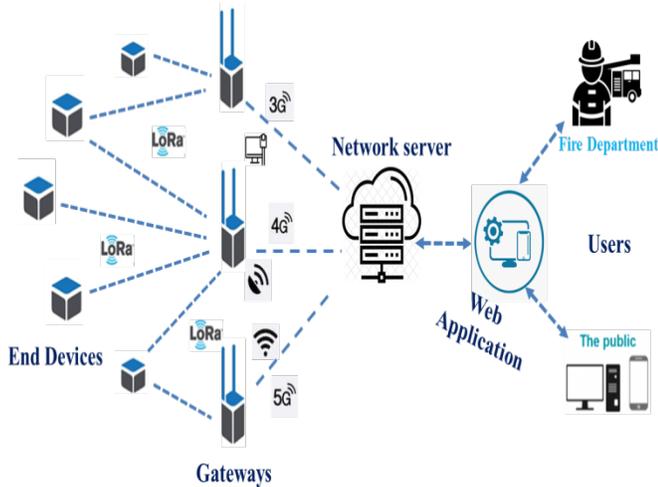


Fig 3. Network Architecture.

The function of the connected object is to collect data from sensors, to process this data and to communicate it using a connectivity function and to receive instructions for carrying out an action. Generally, these functions of the connected object require a power source, especially when the data is preprocessed directly in the object.

The IoT forest fire system aims to provide 24/7 forest fire monitoring and detection (Fig. 3). It reduces the cost and time of human resources and can save lives and reduce loss of property if a fire is detected at an early stage and immediate action is taken. In order to detect a fire, IoT sensors must be distributed around the forest. The forest area is large; it is far from the coverage of the urban network. It is a challenge for the communication between sensors and receivers to exchange data. LoRa, short for Long-Range, is a wireless communication technology suitable for long-range communication. LoRa is a spread spectrum modulation technique derived from CSS (Chirp Spread Spectrum) technology by Semtech. It is a wireless modulation technology with low power consumption and efficient long range capability. It can reach a range of up to 15 km while consuming very little energy (constant RF output of 100 mW at 3.3 V, based on Semtech SX1276), and it is designed specifically for the M2M and IoT network. LoRa transceivers operate in the 860-1000 MHz range. There is a wide range of commercially available LoRa development kits.

After collecting the data, the gateways push the data to the Internet using the MQTT communication protocol via the cellular network). The data is then stored on the Internet server with ThingsBoard [14], a popular open-source IoT platform, and is displayed in an online dashboard. In addition, gateways issue alarms to users via Telegram instantly if they determine that a fire is happening somewhere.

### C. Artificial Intelligence to Deep Learning

Artificial intelligence is the set of theories and techniques used to create machines capable of simulating intelligence. One of the fields of study of artificial intelligence is Machine Learning (ML).

ML is able to reproduce a behavior through algorithms powered by a large amount of data.

As for Deep structured learning, called deep learning, has emerged as a new area of machine learning research. During the past several years, the techniques developed from deep learning research have already been impacting a wide range of information processing work. Deep learning largely covers fields of applications: automatic translation, autonomous vehicles, and medical aid for diagnosis.

The deep neural networks are Multi-Layer Perceptron (MLP) with more than three layers.

Simple Recurrent Neural Network (RNN) is important structure of deep learning, which has made significant breakthroughs especially sequence processing [15]. Gated Recurrent Unit (GRU) and Long Short Term Memory Unit (LSTM) are two other based on RNN model; This two models introduced Memory Unit and additional Gates for solving problems of exploding Gradients and Vanishing Gradients caused by RNN. The success of these models is primarily due to the gating network signals that control how the present input and previous memory are used to update the current activation and produce the current state. These gates have their own sets of weights that are adaptively updated in the learning phase: the training and evaluation process [16].

This part provides a brief overview on recurrent neural networks and its variants, namely the simple RNN, LSTM and GRU.

1) *Simple RNN*: A simple RNN unit is a multiplication of the input ( $x_t$ ) and the previous output ( $h_{t-1}$ ) passed through the activation function.

As shown in Fig. 4, the hidden state of all RNN units at the  $t^{\text{th}}$  time step is determined by the current input  $X_t$  and the previous hidden state  $h_{t-1}$  at the  $(t - 1)^{\text{th}}$  time step.

The RNN model is presented in the form:

$$h_t = f(W_{xh} \cdot X_t + W_{hh} \cdot h_{t-1} + b_h) \tag{1}$$

$$y_t = g(f(W_{hy} \cdot h_t + b_y)) \tag{2}$$

Where  $f$  is a nonlinear activation function,  $W_{xh}$ ,  $W_{hh}$  and  $W_{hy}$  denote weight matrices from the current input layer to hidden layer, the previous hidden layer to current hidden layer and the current hidden layer to output layer, respectively [17]. The  $b_h$  and  $b_y$  are bias vectors.

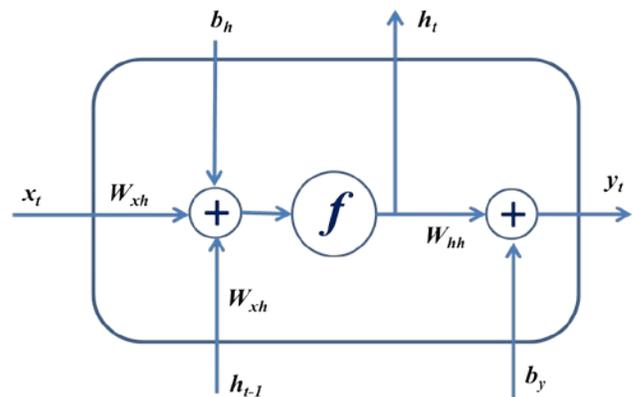


Fig 4. Structure of RNN.

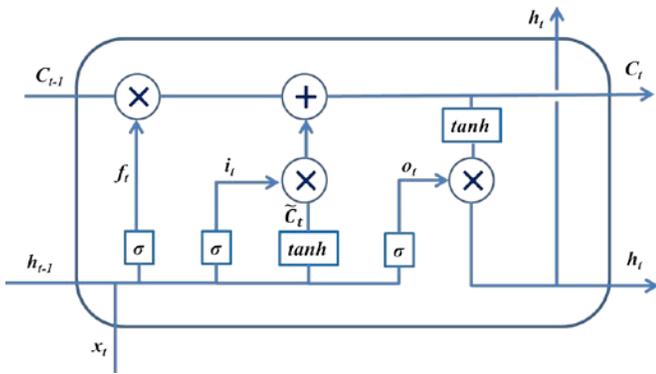


Fig 5. Structure of LSTM Model.

2) *LSTM*: LSTMs are designed to avoid the long-term dependency problem and remembering information for long periods. LSTM is an extension of RNN; it has been proposed to overcome training problems existing in RNN such as exploding and vanishing gradients. As showing in Fig. 5, LSTM Memory block has an input gate, a forget gate and an output gate which regulate the flow of information in and out of the cell [18].

The different equations of the LSTM model are presented and defined in the paper [17]. The main idea of LSTM is that each computational unit is linked not only to a hidden state  $h$  but also to a state  $C$  of the cell which take the role of memory. The transition from  $C_{t-1}$  to the new cell state  $C_t$  is done by transfer with constant gain equal to 1. The cell state can be modified through a gate which authorizes or blocks the update (input gate). Similarly, a gate controls whether the cell status is communicated at the output of the LSTM unit (output gate). LSTM also uses a gate allowing the reset of the cell state (forget gate).

3) *GRU*: GRU networks have performances comparable to LSTMs. A unit requires fewer parameters to learn than an LSTM unit. GRU RNN model is presented in the Fig. 6.

The reset gate  $r_t$  and update  $z_t$  are computed from the two equations as follows:

$$z_t = \sigma(W_z \cdot x_t + U_z \cdot h_{t-1} + b_z) \quad (3)$$

$$r_t = \sigma(W_r \cdot x_t + U_r \cdot h_{t-1} + b_r) \quad (4)$$

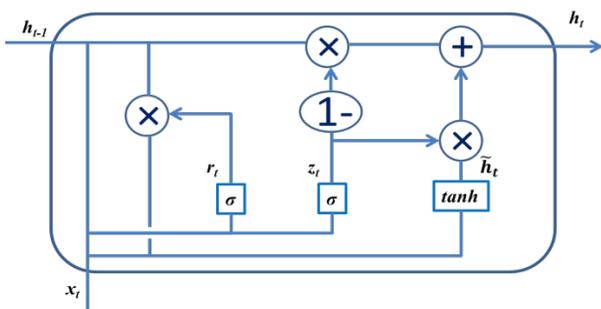


Fig 6. Structure of GRU Model.

Where  $x_t$  is the current time step input and  $h_{t-1}$  is the hidden state of the previous time step.  $W_r, W_z, U_r, U_z$  are weight parameters and  $b_r, b_z$  are biases (Fig. 6).

The GRU-RNN reduce the gating signals to two from the LSTM model. The two gates are called an update gate and a reset gate. GRU's got rid of the cell state and used the hidden state to transfer information and the input and forget gates are merged: update gate. The output gate is replaced by a reset gate. The weights corresponding to these gates are also updated using the back propagation through time.

### III. ARCHITECTURE OF THE PROPOSED FOREST FIRE DETECTION SYSTEM

#### A. Architecture and Management of Proposed System

There are four main components in the IoT system and these are sensor nodes, gateways, Internet servers and end users (for example, firefighters and the public, etc.). A star network topology is adopted in the connection between the sensor nodes and the gateways. The proposed system adopts the cluster topology and hierarchical routing protocols. The nodes members transmit their data to the cluster head, which will transmit to the base station.

Many sensor nodes are distributed evenly over the forest with distances of 1 kilometer and are connected to gateways to transmit data to Internet servers (Fig. 7). To cover the terrain, a long range LoRa communication signal modulation is adopted for the sensor nodes. It offers a wide range of flexibility for the distribution and deployment of sensor nodes. After collecting the data, the gateways send the data via Internet. The data is then stored on the Internet server, and is displayed in an online dashboard.

Each sensor node measures at a given frequency, the values of temperature (T), relative humidity (RH), atmospheric pressure (P), the amount of carbon monoxide (CO), the amount of carbon dioxide (CO<sub>2</sub>), the quantity of Particulate Matter (PM), fine particles. The PM<sub>2.5</sub> and PM<sub>10</sub> fine particle diameters are less than 2.5 and 10  $\mu\text{m}$ , respectively.

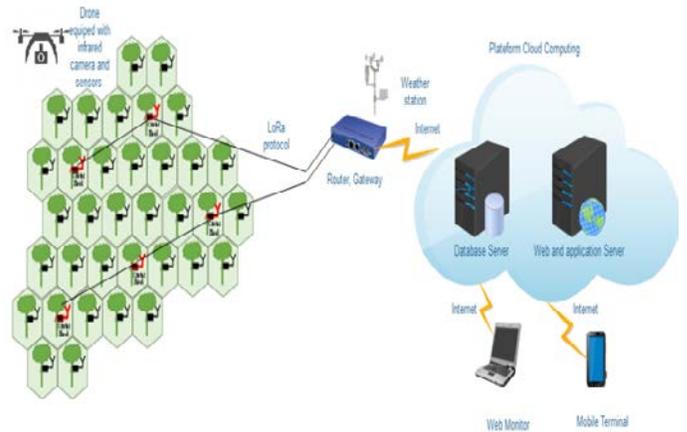


Fig 7. Structure of the Proposed Forest Fire Detection System.

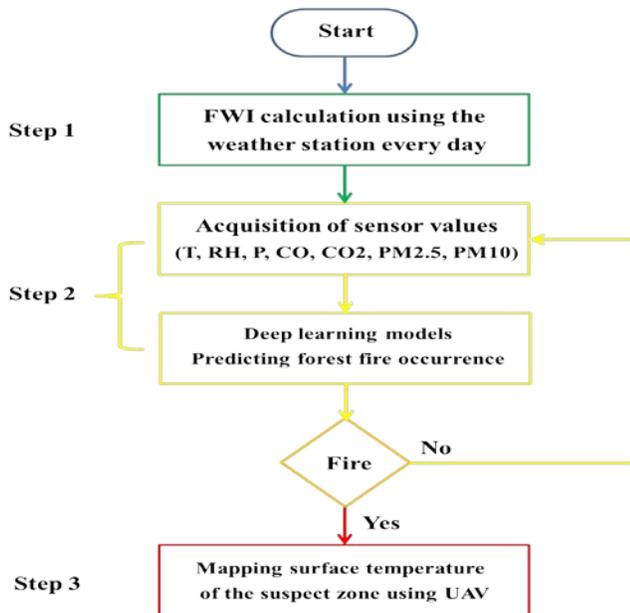


Fig 8. Flowchart of the Forest Fire Detection Algorithm.

The procedure of the proposed fire detection system consists of three steps (Fig. 8): The first step is to determine the FWI from weather data using the weather station located in the forest. The second step is the most important step of the study; it consists of the acquisition of data from each sensor node at a frequency determined according to the danger potential, which is estimated by the calculation of FWI.

The objective of this step is to determine, using the Deep learning algorithm and following the metrics of all the sensor nodes at a given time, whether there is a probability of existence of fires.

If this last function recommends the existence of fire we will trigger the third step which consists in launching an Unmanned Aerial Vehicle (UAV) to measure the values of these sensors in more detail in a mobile way. In addition, the UAV can embark on an infrared camera to have a temperature map. All this information received by the drone will be sent to the control tower to trigger all means of extinguishing the fire if necessary.

### B. Sensor Node

Several questions can be asked when deploying a network of sensor nodes; how many nodes of sensors to put in a forest area, the choice of sensors according to their prices and their qualities. Indeed, all these questions are very important to validate the technical and commercial feasibility study. At this stage, we opted for a reliable and low cost solution to then validate the study on a concrete case and in the field.

The sensor node must essentially consist of a microprocessor, sensors, a long-range, low-consumption transmission part and an energy source optimized for a long duration (Fig. 9).

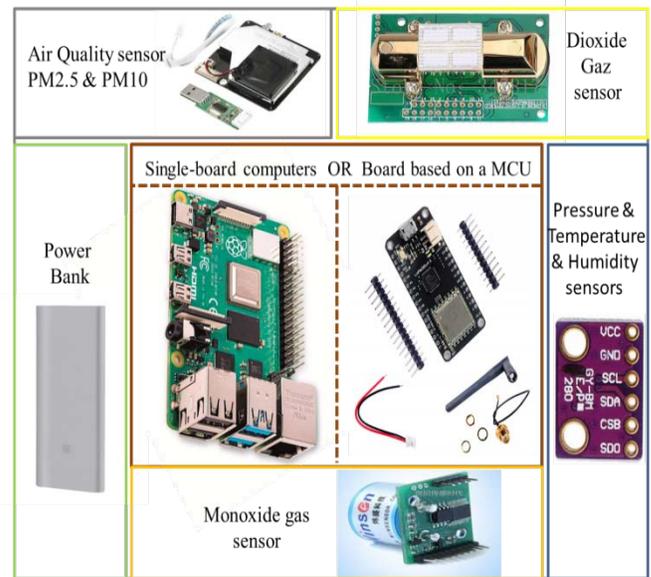


Fig 9. The Main Components of the Sensor Node.

For the sensor test phase and learning the computer models for short-term applications, we opted for the Raspberry single board nano-computer solution, a high-performance board that requires enough energy. On the other hand, for the deployment of the sensor node on the network scale, Lora32u4 is used as a low cost board to acquire sensors data and transfer it to the network. The LoRa32u4 card is a lightweight, low-power board, ideal for creating long-range wireless networks, more flexible than Bluetooth LE and does not require high power unlike WiFi and others. It is an ATmega32u4-based card running at 3.3 V and 8 MHz, with an HPD13 LoRa wireless module, using the SX1276 design of the highly integrated RF transceiver chip. This board is equipped with a Li-Po and Li-ion charging circuit and a standard battery interface, and is compatible with the Arduino development environment.

The choice of sensors for the node is focused on digital components, compact, reliable and inexpensive. To measure temperature, humidity and pressure the BME280 sensor is chosen. For the detection of particles in suspension, the Nova SDS011 sensor using the principle of laser diffusion, allowing to obtain a concentration of particles between 0.3 and 10  $\mu\text{m}$  in air. MH-Z14A-CO2 and ZE07-CO are respectively the two gas sensors used to measure the quantity of dioxide and carbon monoxide in the air (Fig. 10).

The sensor node has a photovoltaic panel to supply the battery in the event of discharge. We have opted for optimal management of electrical energy; we are trying to optimize electrical consumption by implementing a code in C language allowing the lora32u4 board to be put in standby mode between two acquisition operations sends value. This optimization of consumption has allowed us to keep the sensor node autonomous in terms of electrical energy.

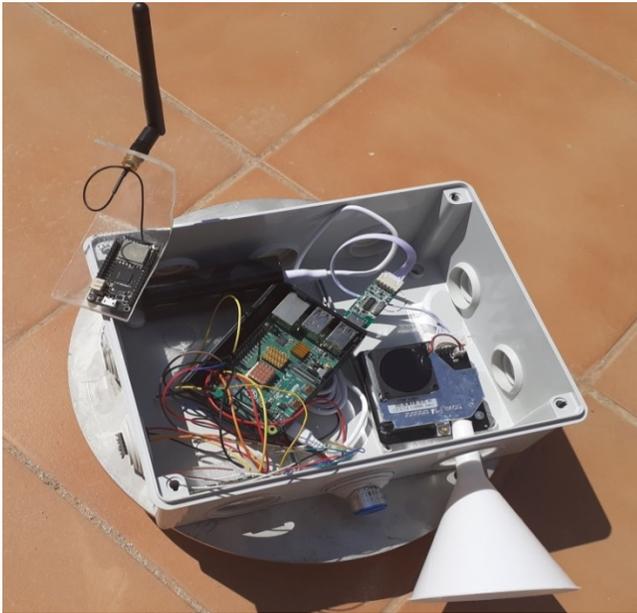


Fig 10. The Main Components of the Sensor Node.

### C. Servers and Applications

Through this project we are trying to build a software and hardware platform capable of monitoring the state of the forest and signaling the risk and the appearance of a fire in the forest. This intelligent system will be able to observe the external environment through sensors, send metrics to a database, and then process this information through deep learning models to make the right decision.

Currently, there is no global standardization of the software and hardware layers of an IoT platform; therefore, developers find it difficult to define an efficient way to transmit, store, manage and provide data. These difficulties increase when you want to manage the data of several applications and devices in an IoT environment or context. For this reason, an IoT platform could very well be suitable, one that supports efficient and appropriate management of devices, implemented with several technologies, protocols, microcontrollers, sensors, actuators and belonging to different entities.

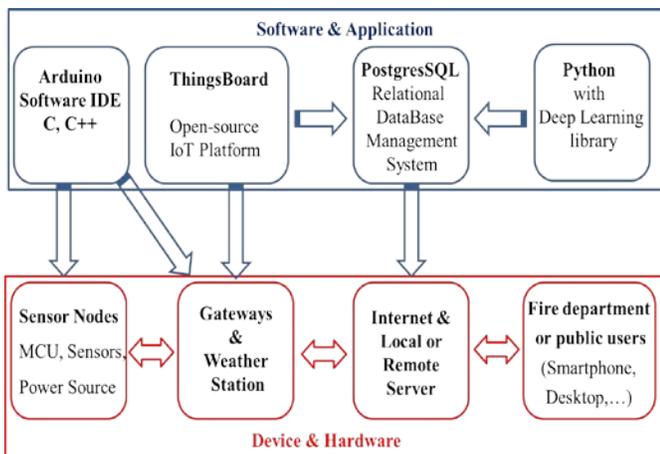


Fig 11. Application-Hardware Relationships.

In this way, the use of a tool capable of centralizing the management of IoT environments, projects and software applications and also capable of providing a good variety of visualization widgets is an interesting solution. There are however several options already available on the market, which is ideal for the context of the work.

Fig. 11 shows the hardware and IT tools chosen for the platform. The most important of these tools is the open-source ThingsBoard tool developed in Java for the management of connected objects [14]. It allows to store and visualize all the data received but also to interact with the objects and manage alerts.

## IV. RESULTS AND DISCUSSION

### A. Setup Simulation

To test the deep learning models we need to have a consistent dataset for better accuracy. These data must be taken under several conditions; with or without fire and for and different meteorological conditions.

To power the data set during a fire, we have placed 3 sensor nodes at a distance  $d$  from a small fire that we have lit. We made measurements every minute and for different distances (Fig. 12). We carried out this operation several times in several days with or without fire and with different meteorological conditions.

### B. Model Supervision

The approach proposed in the document is based on a deep learning model that includes three phases (Fig. 13):

- Dataset pre-processing; after recovering the data, their values are normalized to have the best performance.
- Building the model; after the model is fit on a training dataset that are subdivides to the training data and validation data. The parameters of the model are changed in the training process until having better performance.
- Evaluating the model by prediction; after training the model, this model is evaluated with the test dataset by making prediction of new data.

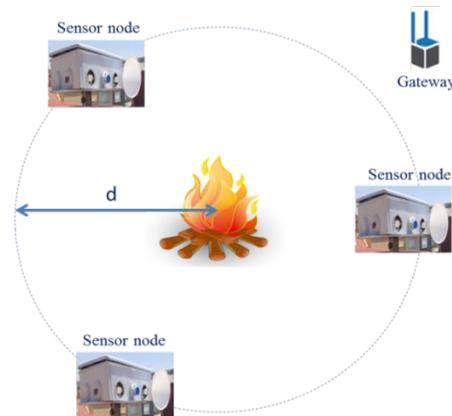


Fig 12. The Sensors Nodes are Distributed in a Circular Fashion Around the Location of the Fire.

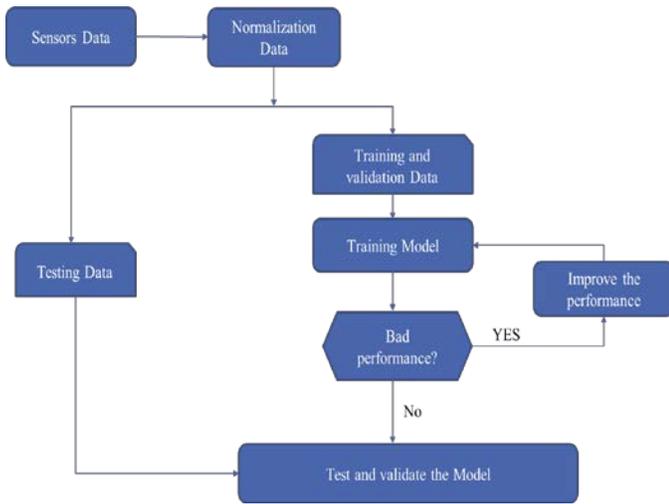


Fig 13. Training Process of Model.

Throughout the process of construction and training model we work with Keras [19] (a Deep Learning library for Theano and TensorFlow) and for generating graphs we use Tensorboard.

Concerning network parameters: the RNN, LSTM and GRU are connected with two hidden layers with 50 neurons and an output layer that makes a single value prediction. For the input and hidden layers, we are using the default activation functions and for the output layer the activation function is ‘Sigmoid’. We set a dropout 0.2 to avoid over-fitting [20]. The networks are trained for 50 epochs and a batch size of 20. In the three networks, for this classification problem, Adam optimizer is the algorithm used to minimize the loss function. For a binary classification such as this approach, the performances of the proposed models are evaluated with the typical loss function: binary cross-entropy (BCE) given in equation 5:

$$BCE = -\frac{1}{N} \sum_{i=1}^N (y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)) \quad (5)$$

Where N is number of measurements,  $y_i$  and  $\hat{y}_i$  are the actual and predict values respectively.

### C. Evaluations and Performances of Model

The experiments are performed on datasets that contains 11 attributes: Temperature, Pressure, Humidity, CO, CO<sub>2</sub>, PM<sub>2.5</sub>, PM<sub>10</sub>, wind speed, wind gust speed, solar radiation and FWI. The dataset have around 10000 examples. For all datasets the target attribute predicts the possibility of fire occurrence (0, 1) “1” indicate the occurrence of fire and “0” indicate his absence. Indeed, each entry has a weight on the prediction of the triggering of the fire, this weight is calculated and updated by the algorithm of the model. We can note that among the different parameters of the algorithm in the proposed method, the parameters CO, PM<sub>2.5</sub>, CO<sub>2</sub> and PM<sub>10</sub> are respectively in descending order of the parameter which has the most influence. Regarding the proposed models Simple RNN, LSTM and GRU can improve the prediction accuracy and stability greatly and effectively as showing in Fig. 14(a).

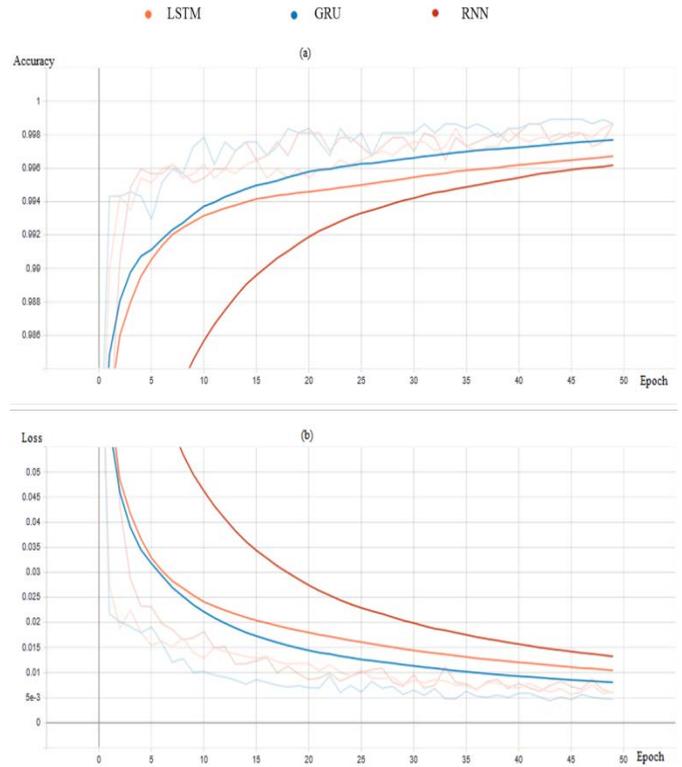


Fig 14. (a) Loss Function (b) Accuracy Graphic.

In result, the data used is divided into two parts. The training datasets with 67% of the observations is used to train the model; the remaining 33% is used to test our model prediction accuracy.

In general, the proposed models can improve the prediction accuracy and stability greatly and effectively.

However, the GRU model generates higher accuracy than the other models during training. GRU also produced significantly smaller loss value compared to the other models as showing in Fig. 14(a). The LSTM model generates less precision than the GRU model but higher compared to the Simple RNN model and it makes longer to train. For training, LSTM ensures less loss compared to Simple RNN as showing in Fig. 14(b). While the Simple RNN model given its simplicity it takes less time to train but it responds to less precision and a relatively more loss than the other models.

We were able to see that the GRU values are also close to the LSTM values with a shorter run time. Indeed, LSTM takes a training time of around 2ms per sample, however GRU puts around 920 us per sample. Simple RNN achieves a training time of around 390 us per sample.

When we evaluated the three models on a number of approximately 2000 test data: the LSTM model gave 4 false predictions, GRU gave only one false prediction and the simple RNN gave 2 false predictions. The GRU model gave 99.89% accuracy and a loss function value of 0.0088, the LSTM model gave an accuracy of 99.82% on the same test data and a loss

value of 0.0298. Finally, the Simple RNN model gave an accuracy of 99.77% and a loss of 0.0062.

Despite that, the results are getting closer for most of them the GRU has provided better results, we could see that GRU model is performs in this case of classification problem and the model could be more easily deployed on small devices.

#### V. CONCLUSION

To limit the damage caused by forest fires and to control the start of fires and its spread, we have presented in this study a method of early detection of forest fires. This method is based on three steps: Estimate the general risk level of the forest, assess and predict in several places the existence or not of fires, and possibly surveyed the place declared to be burning with the help of a UAV.

The originality of this work lies in the use of a wireless sensor network distributed over the entire forest and the deep learning methods to predict in real time a possible start of the fire.

The current system will be implemented on a large scale with multiple sensor nodes to power and augment the data set in order to improve the accuracy and collaboration of data between multiple nodes. The system learning technique can be performed continuously during its operation. We plan in future work to use wind direction sensors to properly estimate and locate the start of the fire.

#### ACKNOWLEDGMENT

This work was supported by University Mohammed Premier. The authors would like to thank the president of University Mohammed Premier for financially supporting this work, as part of the Project PARA.

#### REFERENCES

- [1] Bradstock, A. Ross, R. Nolan, L. Collins, V. Resco de Dios, H. Clarke, E. Jenkins Meaghan, B. Kenny and Matthias M. Boer, "A Broader Perspective on the Causes and Consequences of Eastern Australia's 2019-20 Season of Mega-Fires: A Response to Adams et Al." *Global Change Biology* 0–1.
- [2] Giglio, Louis, W. Schroeder and Christopher O. Justice, "The Collection 6 MODIS Active Fire Detection Algorithm and Fire Products." *Remote Sensing of Environment* 178:31–41, 2016.
- [3] D. Akca, E. Stylianidis, D. Poli, A. Gruen, O. Altan, M. Hofer, K. Smagas, V. S. Martin, A. Walli, E. Jimeno, and A. Garcia, "Pre- and Post-Fire Comparison of Forest Areas in 3D". Springer International Publishing, 2019.
- [4] L. Millan-Garcia, G. Sanchez-Perez, M. Nakano, K. Toscano-Medina, H. Perez-Meana, and L. Rojas-Cardenas, "An early fire detection

- algorithm using IP cameras," *Sensors (Switzerland)*, vol. 12, no. 5, pp. 5670–5686, 2012.
- [5] G. Tsiourlis, S. Andreadakis, and P. Konstantinidis, "SITHON: A wireless network of in situ optical cameras applied to the early detection-notification-monitoring of forest fires," *Sensors (Switzerland)*, vol. 9, no. 6, pp. 4465–4482, 2009.
- [6] B. Koetz, F. Morsdorf, S. van der Linden, T. Curt, and B. Allgöwer, "Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data," *For. Ecol. Manage.*, vol. 256, no. 3, pp. 263–271, 2008.
- [7] B. Son, Y. Her, and J. Kim, "A design and implementation of forest-fires surveillance system based on wireless sensor networks for South Korea mountains," *Int. J. Comput. Sci. Netw. Secur.*, vol. 6, no. 9, pp. 124–130, 2006.
- [8] M. Trinath Basu, R. Karthik, J. Mahitha, and V. Lokesh Reddy, "IoT based forest fire detection system," *Int. J. Eng. Technol.*, vol. 7, no. 2, pp. 124–126, 2018.
- [9] A. R. Ulucinar, I. Korpeoglu, and A. E. Cetin, "A Wi-Fi cluster based wireless sensor network application and deployment for wildfire detection," *Int. J. Distrib. Sens. Networks*, vol. 10, 2014.
- [10] C. H. Lim, Y. S. Kim, M. Won, S. J. Kim, and W. K. Lee, "Can satellite-based data substitute for surveyed data to predict the spatial probability of forest fire? A geostatistical approach to forest fire in the Republic of Korea," *Geomatics, Nat. Hazards Risk*, vol. 10, no. 1, pp. 719–739, 2019.
- [11] P. Jain, X. Wang, and M. D. Flannigan, "Trend analysis of fire season length and extreme fire weather in North America between 1979 and 2015," *Int. J. Wildl. Fire*, vol. 26, no. 12, pp. 1009–1020, 2017.
- [12] W. Benzekri, A. El Moussati, O. Moussaoui, and M. Berrajaa, "Early forest fire detection with low power wireless sensors networks", *LNEE, Advances in Smart Technologies Applications and Case Studies*, in press.
- [13] Grace, Asplund, Ely, Intorf, and Droeg, "The Osborne Fire finder the basic lookout tools: Firemans Guide California Region". U.S. Department of Agriculture Forest Service, 2013.
- [14] ThingsBoard. Thingsboard - open-source IoT platform (2018), <https://thingsboard.io>.
- [15] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent Advances in Recurrent Neural Networks," 2017.
- [16] R. Dey and F. M. Salemt, "Gate-variants of Gated Recurrent Unit (GRU) neural networks," *Midwest Symp. Circuits Syst.*, vol. 2017-August, pp. 1597–1600, 2017.
- [17] Y. Cao, F. Yang, Q. Tang, and X. Lu, "An attention enhanced bidirectional LSTM for early forest fire smoke recognition," *IEEE Access*, vol. 7, pp. 154732–154742, 2019.
- [18] D. Wei, B. Wang, G. Lin and al., "Research on unstructured text data mining and fault classification based on RNN-LSTM with malfunction inspection report," *Energies*, vol. 10, no. 3, 2017.
- [19] G. Aurélien, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019.
- [20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting" *Mach. Learn. Res.*, vol. 15, pp. 1929-195, 2014.

# A Process Model Collection with Structural Variants and Evaluations

Rimsha Anam<sup>1</sup>, Tahir Muhammad<sup>2</sup>

Department of Computer Science  
COMSATS University Islamabad, Lahore Campus  
Lahore, Pakistan

**Abstract**—Today in the era of the latest technologies, Business Process Management Systems (BPMS) have allowed organizations to build process model repositories which help to maintain the flow of operations in the form of various process models. Business process models are virtual models that can imitate the actual activities of an organization. Searching for semantically similar activities between pairs of process models in a repository is known as Process Model Matching (PMM). From the past few years, PMM has been gaining momentum due to its wide range of applications such as integration of process models, process model clone detection, and process model knowledge discovery. Different types of PMM techniques have been applied on available process model repositories but these repositories contained a limited number of process models. Another notable aspect of PMM is that the existing techniques have not achieved the desired results which questions the effectiveness of process model repositories. To address this problem, the authors of this study have developed a substantial, diverse, and carefully developed process model collection. This process model collection is compared with existing SAP collection to highlight its significance and superiority. Furthermore, the proposed process model collection represents structural variations of example process models which are governed by the defined set of rules. To reflect structural variations between process models of our collection, existing structural similarity approaches such as structural metrics and graph edit distance were applied by using a custom-developed tool. Our proposed process model collection is freely available to the research community which can be used to build new PMM techniques and for assessment of existing PMM techniques.

**Keywords**—Business process modeling; process model collection; Process Model Matching (PMM); structural variants

## I. INTRODUCTION

The process model is a conceptual model that represents various dependencies between the activities of an enterprise. Organizations store their business processes into process model repositories which are considered valuable resources to perform various tasks such as business process improvements, software development requirements, and configuration of Enterprise Resource Planning (ERP) systems [1, 2]. The usefulness of process models has encouraged companies to generate huge collections of process models i.e. a Dutch government organization maintains the dataset of more than six hundred process models [3], a multinational company located in Australian holds a collection of more than five thousand process models, another Chinese factory holds even bigger process model collections [4]. However, such process

model repositories are often proprietary and cannot be shared publicly due to privacy issues.

To improve the handling and reliability of process model collections, more process model repositories with extended features are needed. One of the major challenges is searching a process model effectively and efficiently from a collection [5]. To address this challenge various Process Model Matching (PMM) techniques [6-11] have been proposed that can be used to check similarity between process models. Due to the lack of process model collections with diverse features, the performance of these techniques cannot be evaluated rigorously. At this stage, most of the existing PMM techniques lack empirical evaluation due to the limited availability of larger process model collections[4]. This highlights the need for process model collection that contains a large number of process models that are rich in structural features.

Existing process model collections such as PMMC'15 datasets are freely available which contain only 90 models in total (i.e. university admission = 9, birth registrations = 9 and asset management = 72) [12]. Thus, the PMM techniques evaluated on such collections may not perform well when applied to larger process model collections. PMMC'15 datasets also lack feature diversity because each dataset targets a specific domain. Therefore, in this regard more effort is needed to build a diverse process model collection that can be used for effective evaluation of PMM techniques. As a contribution to solving these challenges, authors have developed a standardized process model collection for rigorous evaluation and improvement of PMM techniques.

The proposed process model collection is generated through a systematic approach and is freely available to the research community and can also be used for comparing different PMM techniques. A total of 750 process models were stored in the proposed process model collection. These process models were offered different shapes, sizes and dimensions, etc. This collection is developed by following the constraint in mind that it will object the ability of a PMM technique to check similarity between process models. The rest of the article is structured as follows: Section II provides a brief background of the terms discussed in this study, Section III provides an overview of different studies related to PMM techniques and process model collections. Section IV explains the protocol used for the development of the proposed process model collection. Different variation types used to generate structural variants of the original process models are discussed in Section V. Custom tool developed to

compute structural metrics of the process models is discussed in Section VI. A detailed analysis of the proposed collection is discussed along with the results in Section VII. In Section VIII, a conclusion is made to highlight the significance and strengths of our proposed process model collection. The last section predicts future dimensions for this study.

## II. BACKGROUND

The process model captures the flow of different activities of an organization. Process model structures can be used to represent the order in which activities of the process model are executed, and this is known as control flow [13]. These process models are stored in process model repositories for current and future use. Searching and adding new process models in repositories are crucial tasks for the maintenance of process model collections. Different PMM techniques are used to avoid duplication and efficient searching of process models in a process model collection. PMM techniques are challenged by structural diversities of process models stored in a collection. To evaluate structural similarities between process models, different structural metrics such as size, density, and complexity between similar process models are used.

These structural metrics were proposed since a decade ago and are widely used and accepted for PMM [14]. By computing and comparing the values of these structural metrics, richness, and diversity of process model collections can also be evaluated. Different PMM techniques lack rigorous evaluation due to the unavailability of large and diverse process model collections for the research community. A more rich and structural diverse process model is required which can be used as a benchmark for meticulous evaluation of PMM techniques.

## III. RELATED WORK

Identification of similar activities between two structural variants of a process model is a crucial task. Traditional search engines perform based on text matching which is not enough in the case of PMM [15]. In the study [15], authors have identified important features of process models that can directly affect various aspects of a process model collection. These features are formally categorized into three categories i.e. 'label feature', 'behavioral feature', and 'structural feature' of a process model [16]. These categories play an important role to address various issues of PMM.

Label feature captures the names of activities of the process model, this group of labels used for activities is called a label feature [17]. Each label can be represented differently for corresponding activities of two similar process models [18]. Due to this fact, the benchmark collection must contain process models with labels having similar semantics but are written in different ways. This can help achieve a more standardized and rigorous evaluation of PMM techniques by detecting different labels with similar meaning.

The behavioral feature of the process model demonstrates a causal relationship between various activities of a process model [19], i.e. indirect edges occurring between different nodes of a process model. Behavioral feature is also an important part of PMM techniques as it can help identify

similarities between two different process models. Comprehensive studies [10, 19] have been conducted to address the issues of process model similarities using label features and behavioral features but similarities based on structural features of process models are comparatively less explored and need more attention [20]. In the study [21], authors have produced different structural variants of a process model without making any change to its semantics. In the study [22], the authors also showed different structural variants of process models. However, only a few structural variants were able to maintain the semantics of the original model.

## IV. PROPOSED PROCESS MODEL COLLECTION

A systematic approach is followed for the development of the proposed process model collection with diverse structural features. It is because, in different studies, authors have recommended employing a systematic approach for process design and process reengineering to reducing human bias [23]. Two types of process model developments were involved in the development of collection: a) *collecting and developing original process models*; b) *producing structural variants of original process models developed in the previous phase*. To avoid human error both phases of model generation were supervised by domain experts who had years of research and teaching experience of business process management and process model repository development. Details of both phases of process model collection are as follows:

### A. Collection and Development of Process Models

Different types of process models were collected from various sources such as books, research papers, technical reports, and other online sources such as example models from Object Management Group (OMG®) [24]. No restriction was applied to the selection of sources for original models to target maximum domains. Collection of process models from various domains such as academics, reservation systems, procurement, manufacturing, and payment systems, etc. helped achieve diversity and richness for the proposed process model collection. These collected models were not ready to be stored directly to the process model collection because of the two reasons: 1) these models were in different formats such as images, hand-drawn models, scanned pictures, pdf, videos, etc. 2) these processes were modeled using different modeling languages i.e. Petri-Nets, EPC, BPMN, YAWL, etc. So all of these models were redesigned in Business Process Modeling Notation (BPMN) using a modeling tool called CAMUNDA Modeler [25] which is a lightweight and open-source software that supports BPMN standards. The selection of the tool was inspired by two causes: 1) it comes with various modeling guidelines and generates error-free models; 2) it allows users to export process models in different file formats such as PDF and XML. While modeling these processes, few changes were made to ensure diversity between different models. Furthermore, the process models went through manual testing for logical errors and were improved according to widely accepted modeling guidelines [26]. These modeling guidelines allowed the addition or removal of few elements to ensure that the produced model conformant with existing standards of process modeling. For example, a couple of process models

have *OR-split* with a missing corresponding *OR-join*. It was against BPMN modeling guidelines, according to which each *OR-split* must be accompanied by an *OR-join*. At the end of *Phase I*, a total of 150 models were generated and stored into process model collection.

**B. Generating Process Model Variants**

The goal of this phase was to improve process model collection by adding more process models to the collection that offer distinct structural features. The focus of this study is to create structural diversity between similar types of process models. Considering this fact proposed process model collection was extended by adding 600 structural variants of the original 150 process models. The major cause for the introduction of structural diversity between similar types of process models is that it can help evaluate various PMM techniques rigorously.

Four different types of structural variants were generated from each original process model. These four structural types were inspired by traditional similarity metrics of process models discussed in studies from [21, 22, 27]. From these studies, variation patterns that do not change the semantics of the original process models were chosen for the classification of process model variants of the proposed collection. Classification of structural variants is explained in detail in Section V.

**C. Process Model Collection Dimensions**

The proposed process model collection contains 750 process models out of which 150 are original models and 600 are different structural variants of the original 150 process models. Table I shows the number of process models along with their types. These process models were designed using CAMUNDA Modeler [25] as a modeling tool, and the modeling language used was the BPMN.

TABLE I. STATISTICS OF PROPOSED COLLECTION

1.	Original Models	150
2.	SVT-1	150
3.	SVT-2	150
4.	SVT-3	150
5.	SVT-4	150
Total		750

To validate the strengths of the proposed collection, analysis of the proposed collection was performed. The analysis was conducted by comparing the proposed collection with existing ones and its internal consistency was evaluated by comparing all four variants with original 150 process models.

**V. CLASSIFICATION OF STRUCTURAL VARIATION TYPES**

Classification of structural variations of the original process model is performed as follows: 1) In first variation different structural variants of the original process model were generated by through defined traces. 2) The second variation involves the addition of activities to the original process models. 3) The third variation adds edges to modify control follow of the process model. 4) Structural variants were produced by swapping common activities which are loosely bounded, without changing the semantics of the original process model.

To explain these four classes: structural variation type 1 (SVT-1), structural variation type 2 (SVT-2), structural variation type 3 (SVT-3), and structural variation type 4 (SVT-4), different structural variants of an original process are generated. The sample process model selected to demonstrate the generation of structural variants is a Facebook Co. process model for registration of new candidates into the system. The process model contains 12 activities, 6 gateways, and 2 events as shown in Fig. 1. All possible traces of the Facebook Co. process model are generated which are shown in Fig. 2.

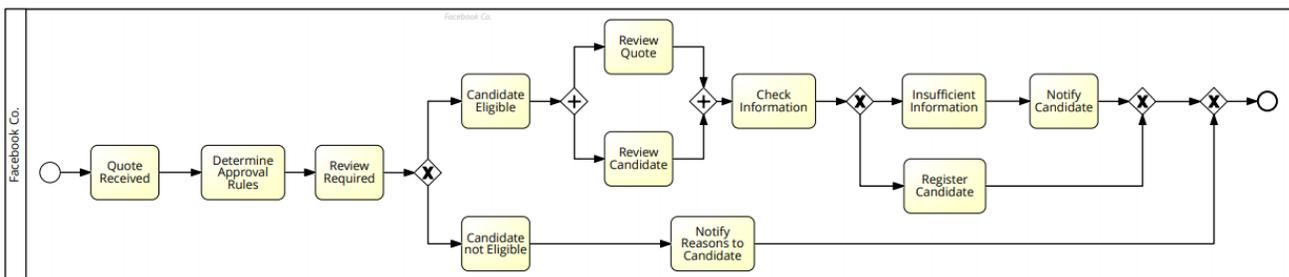


Fig 1. Example of Original Process Model.

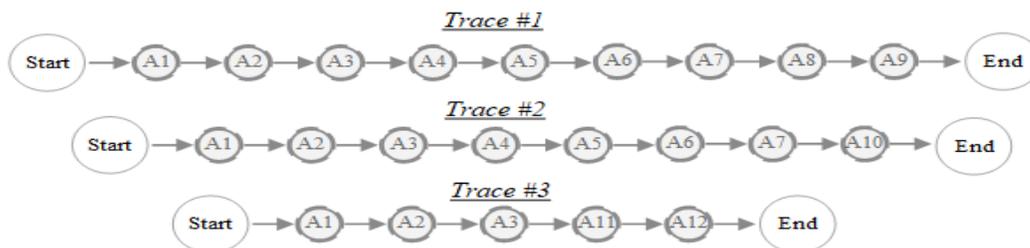


Fig 2. Traces of Original Process Model.

A. SVT-1

In SVT-1, sequential activities can be executed in parallel by defining the traces of the process model. The execution patterns demonstrate the execution order of the activities of the process model, also known as traces of a process model. First, all possible execution patterns of the process model are defined without considering any control block. Activities occurring before and after of a control block are considered as common activities in all traces of the process model. After marking the common activities, the sequential order of activities within any control block is arranged in a way to observe parallel execution through AND gateway.

Each process model has control flow which defines how the execution of activities can be conducted for a process model. The control flow is modified with the help of different types of gateways i.e. AND, OR, and XOR. Through these gateways' activities are executed in parallel or exclusively. In parallel execution of activities, AND-split is used along with AND-join to synchronize the flow of execution. In mutually exclusive execution of activities XOR-split is used along with XOR-join to merge alternative flows. In mutually exclusive flow of execution, the selection of activity that needs to be executed depends upon an external event or data. By making changes to control flow, different structural variants were produced. A structural variant generated by SVT-1 can be transformed back to its original process model by developing a graph of its execution traces. Table II explains design patterns and conditions for the generation of process model variants according to SVT-1.

B. SVT-2

In SVT-2 one or more activities are added to a process model to make a structural change. However, this change in structure will not modify the process model semantically. To add one or more activities first, understand the description of the original model by exploring all of its activities, then the activity is added between two succeeding activities by following the design pattern discussed in Table III.

According to SVT-2 new activity can only be introduced between a pair of activities that are common in traces, and if they are not depending on each other. By following SVT-2 variants of the original models were generated by the addition of new activities, but the original model remained semantically the same. However, if we add new activity between two activities that are closely bounded than it will change the behavior of the original model semantically.

TABLE II. VARIATIONS WITH TRACES (SVT-1)

<b>Description</b>	
Activities without any strict execution pattern can be executed in parallel.	
<b>Design Pattern</b>	
1.	All possible traces of an original process model are defined.
2.	Activities that occur as pre and post activities of the gateway in the traces are called common activities.
3.	Activities within the control block which were executed in a sequence are executed in parallel by introducing AND gateway.
4.	By following steps 2 and 3, all common activities in traces are connected with the activities of the control block by using gateways of the original process model.
5.	All traces should be operated independently.
<b>Condition</b>	
Activities are not tightly bound for sequential execution. Traces are defined without considering any gateways.	
<b>Graphical Representation</b>	
Fig. 3.	

C. SVT-3

Each process model has its control flow which is defined in the form of vertices and edges. SVT-3 deals with the addition of the control edge structure to the original process model by considering pre and post conditions of the activities. The process models observe four basic types of control flow mechanisms i.e. sequential control flow structure, parallel control flow structure (AND-Gateways), choice control flow structure (XOR-Gateways), and looping control flow structure. For sequential control flow all activities are executed in a sequence and succeeding activity is executed after executing its previous one. In parallel execution, AND gateway is used to execute two activities simultaneously while on the other hand XOR gateway is used to offer an exclusive execution among one of the two succeeding activities. In looping structures activities are executed repeatedly in a loop. According to SVT-3 new control edge can be added to all of these four control flow structures by considering pre and post conditions of the activities. According to the study [22], structure variations are identified based on the change primitives such as the addition and modification of nodes and edges in the original process model through the design patterns mentioned in Table IV. However, the addition or deletion of a new control edge does not effects the semantics of the process model [21, 22].

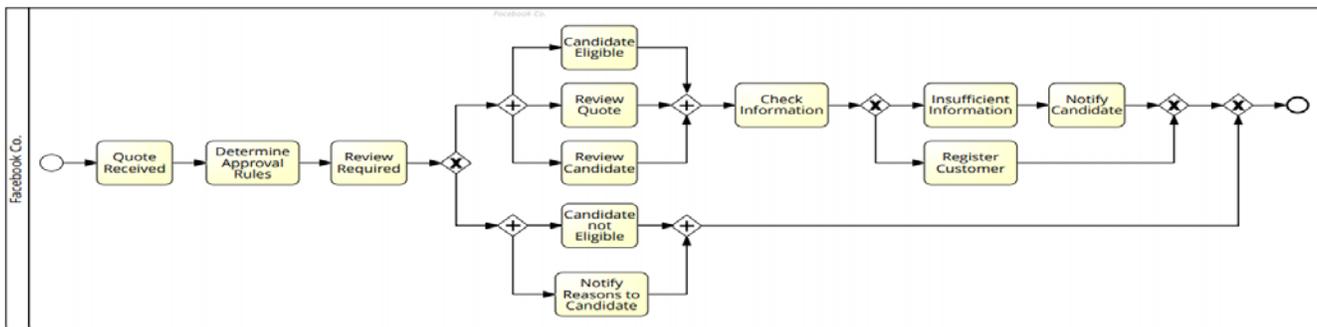


Fig 3. SVT-1 of Original Process Model.

TABLE III. VARIATIONS WITH ADDITION OF ACTIVITIES (SVT-2)

<b>Description</b>
One or more activities are appended with the original process model without effecting its semantics.
<b>Design Pattern</b>
<ol style="list-style-type: none"> <li>1. An activity can be introduced in the middle of two succeeding activities.</li> <li>2. The common activities of the traces defined in (SVT-1), can accommodate one or more new activities.</li> <li>3. New activity could provide a structural variation in the process model, but it will remain semantically the same.</li> </ol>
<b>Condition</b>
One or more activities can only be added when existing activities are not rigorously bound.
<b>Graphical Representation</b>
Fig. 4.

TABLE IV. VARIATIONS WITH MODIFICATION OF CONTROL EDGE (SVT-3)

<b>Description</b>
An addition of control dependency can be added among existing control flow.
<b>Design Pattern</b>
<ol style="list-style-type: none"> <li>1. Control dependency can be added among existing control flow by evaluating the pre and post conditions of the activities.</li> <li>2. The control edge will add up new the condition in the control flow.</li> <li>3. The new condition should be consistent with the existing control block.</li> </ol>
<b>Condition</b>
SVT-3 can only be applied after validating pre and post conditions of control flows in the process model.
<b>Graphical Representation</b>
Fig. 5.

D. SVT-4

According to the SVT-4 order of the common activities in the traces is modified to generate a structural variant of the original process model. First, the execution pattern of the original process model is identified so that activities outside control blocks can be recognized. Another important thing to be considered before doing swapping between activities is that the targeted activities should not be closely bounded. If the target activities are closely bounded than these changes can change the semantics of the variant concerning the original process model. Design pattern rules for SVT-4 are mentioned in Table V. along with the conditions. In Study [21] authors suggest that even after swapping of activities, the semantics will remain the same.

TABLE V. VARIATIONS BY SWAPPING ACTIVITIES (SVT-4)

<b>Description</b>
One or more common activities in traces can be swap with each other.
<b>Design Pattern</b>
<ol style="list-style-type: none"> <li>1. Used the execution traces discussed in SVT-1.</li> <li>2. One or more common activities can only be swapped if they are in a succeeding execution order.</li> <li>3. Swapping can be possible among common activities of the traces.</li> </ol>
<b>Condition</b>
Swapping cannot be applied to activities with strict execution order.
<b>Graphical Representation</b>
Fig. 6.

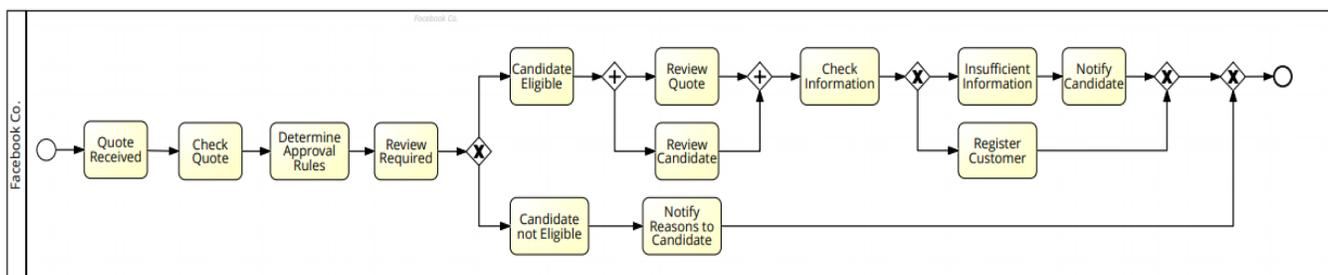


Fig 4. SVT-2 of Original Process Model.

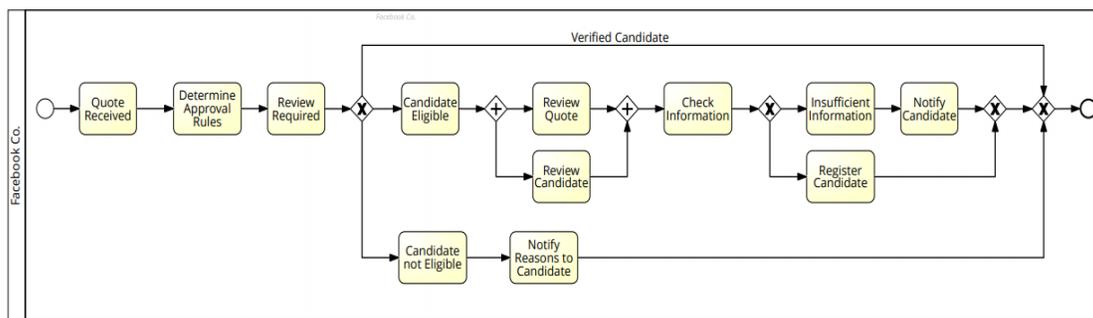


Fig 5. SVT-3 of Original Process Model.

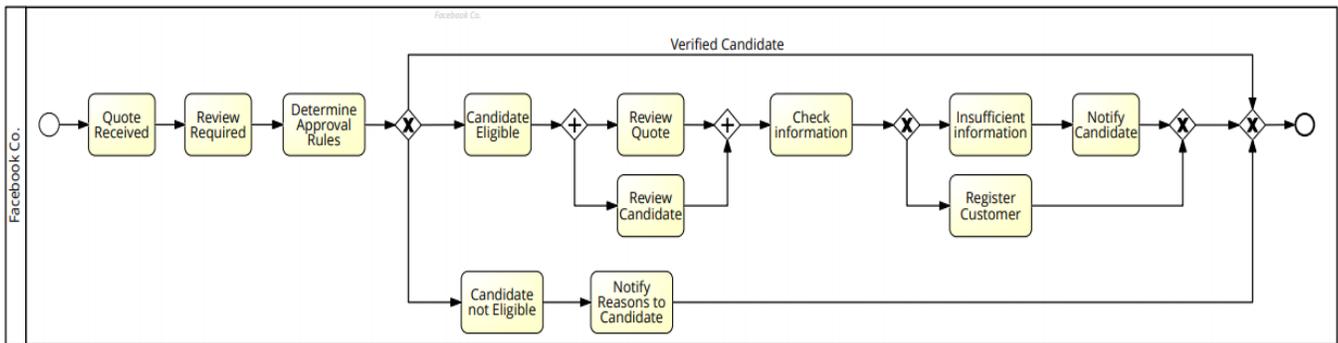


Fig 6. SVT-4 of Original Process Model.

## VI. PROCESS MODEL MATCHING TOOL

A custom tool was developed to compute different structural metrics and GED of the process models stored in the proposed collection. The proposed tool is a desktop application that runs on windows operating system. Java programming language was used for the development of this tool. Fig. 7 shows the functional behavior of the proposed process model matching tool. To perform different types of computations, first all of the process models were stored into XML format by using CAMUNDA modeler. XML files are processed to extract information such as process name and lanes information, Events and task information, control flow information along with their sequence mappings. This information was stored in the form of variables and these

variables were passed to different functions to compute metrics. By mapping various execution sequences business process graphs were produced. Separate functions were written to compute graph edit distance of the process models.

Similarly, BPGs of the process model were generated using extracted information of the process models, and these BPGs were used as input to compute GED. Different GED operations were performed i.e. *Node Insertion or Deletion (SN)*, *Edge Insertion or Deletion (SE)*, *Node Substitution (SB)* on these BPGs, through which values of GED and similarity were computed. Structural metrics and all other results generated by the proposed tool are also uploaded and are available online for the research community [28].

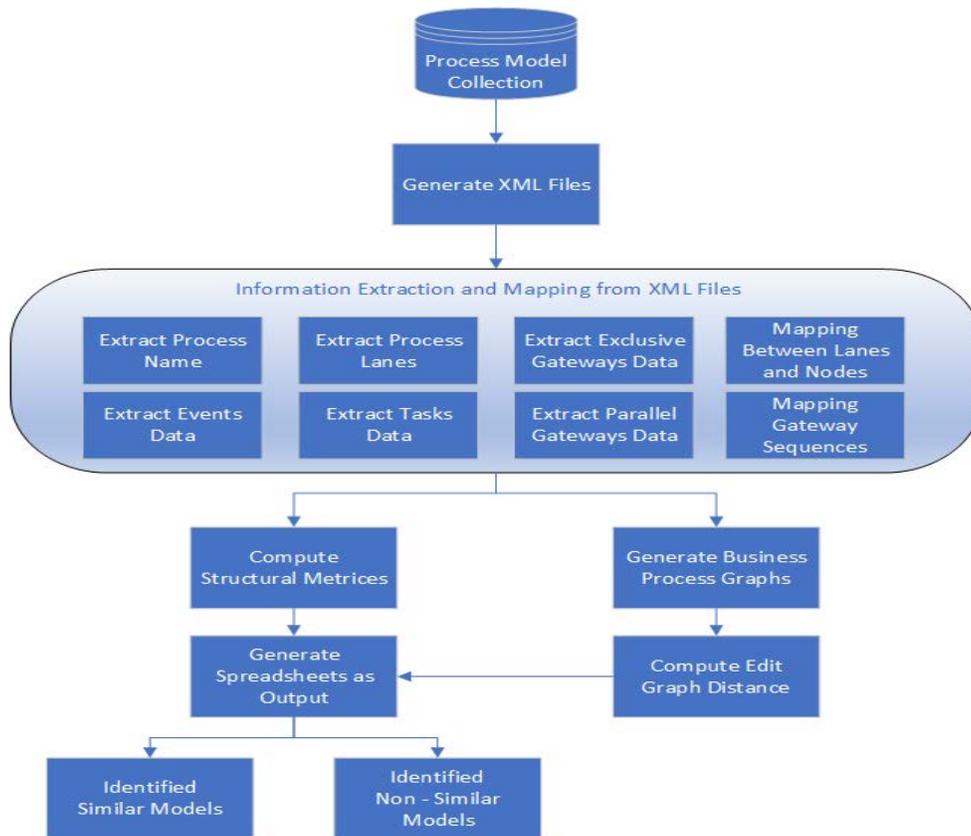


Fig 7. Functional Behavior of Custom Tool for PMM.

## VII. ANALYSIS AND RESULTS

The proposed process model consists of structurally diverse models that are rarely found in existing process model collections. To evaluate the structural diversity of the proposed collection two methods were used first, it was compared with existing SAP's process model collection using widely accepted structural metrics. It should be noted that SAP's process model collection is not openly available to the public so we relied on the results published in previous studies [14, 29].

Secondly, similarity between the original process model and its structural variants is computed by using graph edit distance (GED). The purpose of this computation is to highlight the fact that produced structural variants are semantically similar to their original process model and show less similarity with the rest of the process models in the proposed process model collection.

### A. Comparison of Structural Metrics

Each process model has structural features that can reveal various dimensions of the process model i.e. relationship between various nodes, number of arcs (density), the sequence of execution, etc. There are different types of structural features which can be used for evaluation of our proposed collection. However, to perform a standardized comparison we have chosen widely accepted and most discussed structural metrics proposed by *Mendlings* [14] for process models similarity evaluation. These metrics were designed for various domains such as network analysis, software measurement, and business process models. In this study [14], 26 structural metrics are considered out of which 15 metrics are associated with the size of a process model, and rest are used to express other aspects such as density, sequentiality, separability, and cyclicity of the process model. To compute these structural metrics, a custom tool was developed which is discussed previously in Section VI.

TABLE VI. COMPARISON OF STRUCTURAL METRICS

Structural Metrics	SAP Collection		Proposed Collection	
	Mean	Std.	Mean	Std.
Total Number of Nodes	18.74	20.74	21	8.3
Total Number of Functions	3.81	4.03	19	8.3
Total Number of OR Gateway	1.81	1.08	0	0
Total Number of XOR Gateway	2.78	1.95	2.8	1.5
Number of AND Gateway	2.75	2.18	2.1	2.8
Density of Process Models	0.09	0.08	0.01	0.02
Average Degree of Connectors	1.46	3.3	3.1	0.8

Due to space limitation, we have presented the comparison values of a few structural metrics as shown in Table VI and rest are made available online [28]. It is important to mention here that SAP's process models were modeled in EPC, while for proposed collection BPMN is used as a modeling language. EPC and BPMN are two different modeling languages with different modeling rules. In EPC, each function has one pre-event and one post-event while in BPMN there is no such specific restriction. For instance, a linear process model in EPC with a size of 14 contains 6 functions and 7 events while on the other hand a BPMN process model with size 14 will have 12 activities with 2 events (*Start and End Events*). Considering this fact that there are a smaller number of intermediate events in BPMN than EPC, the process models stored in the proposed collection are more functional as compared to process models in SAP's collection.

In Table VI, the number of nodes that represent the size of the process models has a value of 21 for the proposed collection which is a little bit higher than SAP's collection value i.e. 18.74. This shows that the size of the process models in the proposed collection is almost similar to the ones in SAP's collection. However, considering the richness and functionality of the process models, the proposed collection offers an average value of 19 whereas SAP's collection has a value of 3.81. This concludes that when it comes to functionality the size of process models store in the proposed collection is greater than process models available in SAP's collection. It is also mentioned that the proposed collection does not contain any process model with OR gateway. This was established by following state of the art process modeling guidelines which recommend the use of XOR gateway instead of OR gateway [30]. Due to the absence of OR gateway in the process models its value for the proposed collection is zero. Another structural metric of a process model is density, which indicates the ratio of arcs to the maximum number of arcs that can exist between the same number of nodes. A higher value of density for a process model indicates more arcs between nodes of a process model. It is also a recognized fact, that many of the natural graphs follow a 'power law' according to which models with greater number of nodes have less number of arcs which represent low density [14]. Hence the proposed collection has almost a similar number of arcs as compared to SAP's collection but in terms of functionality and diversity the proposed process model collection is richer and larger (the number of nodes is greater).

Another structural metric is the average degree of connectors, and its higher value for the proposed collection highlights the existence of more logical control flow gateways in process models which indicates structural richness of proposed collection as compared to SAP's collection.

### B. Graph Edit Distance (GED)

In continuance with evaluating strengths of the proposed process model collection, we decided to perform mapping of structural variants to the original process model by using GED [15]. The minimum number of graph-edit operations that are required to get from one graph to another is called GED. It helps to discover the level of structural similarity between original process models and their variants. With the help of similarity values between process models we can verify the

correctness of the proposed collection. To compute the GED of a process model, first the model was converted into Business Process Graph (BPG) [31]. Three types of graph-edit operations that can be performed on these BPGs to compute structural similarity. 1) *Node Insertion or Deletion (SN)*. 2) *Edge Insertion or Deletion (SE)*. 3) *Node Substitution (SB)* [15].

After performing these operations on BPGs, the value of graph-edit distance was computed between each original process model and its four structural variants by using a custom tool discussed in Section VI. A reference process model i.e. P2 was taken to compute overall similarity between all process models.

Higher values of GED indicate minimum structural similarity among a pair of process models. To explain, GED and structural similarity values of proposed collection we have picked 10 random computations as shown in Table VII. The computed values of similarity can occur between 1 and 0 where 1 indicates maximum and 0 means minimum similarity.

After computations, it was observed that each original process model shows maximum similarity with all of its four structural variants. However, when it comes to the general comparison of one process model with other original process models than the similarity values are decreased. This shows that the four structural variants generated against each original process for the proposed collection are semantically similar but structurally different, which makes the proposed collection optimal for the evaluation of PMM techniques.

TABLE VII. COMPUTATIONS OF GED AND SIMILARITY

Query Model	GED	Similarity	Document Model
P2 (Original)	0.00	1.00	P2.0 (Original)
	17	0.89	P2.1 (SVT-1)
	4	0.95	P2.2 (SVT-2)
	1	0.99	P2.3 (SVT-2)
	0	0.98	P2.4 (SVT-4)
	42	0.621	P10.2 (SVT-2)
	35	0.6	P3.4 (SVT-4)
	41	0.65	P8 (Original)
	42	0.67	P30.3 (SVT-3)
	42	0.7	P117.1 (SVT-1)
28	0.76	P4.1 (SVT-1)	

### VIII. CONCLUSIONS

A process model collection is proposed in this study, which consists of 750 process models with distinct structural features. To generate different structural variants of process models, a systematic approach was deployed. To demonstrate strengths of the proposed collection over existing collections a detailed analysis was performed. Two state of the art approaches were employed for rigorous analysis of the proposed process model collection i.e. structural metrics and graph edit distance. First, different structural metrics for both collections were computed through a custom-developed tool. By comparing the values of these metrics, it was found that

the proposed collection contains process models with more diverse and activity-rich. It was also found that process models in the proposed collection are strictly aligned with the modeling guidelines and are comparatively well structured to SAP process models. Secondly the graph edit distance of process models is computed to evaluate the similarity between process models of the proposed collection. The values of similarity between variants in results showed that the proposed collection is consistent with has the potential to be used as a tool to evaluate various process model matching techniques.

### IX. FUTURE WORK

In the future, new PMM techniques can be developed, or existing PMM techniques can be improved by deploying them on the proposed process model collection. Combing structural similarity features along with label and behavioral features can also be performed which may produce some interesting results. It is also mentioned that machine learning could be employed for PMM or evaluation of existing PMM techniques. Another dimension to this study could be the comparison between different similarity features such as understanding the pros and cons of structural similarities with label and behavioral features.

### REFERENCES

- [1] Kuss, E., H. Leopold, C. Meilicke, and H. Stuckenschmidt. Ranking-based evaluation of process model matching. in OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". 2017. Springer.
- [2] Dumas, M., M. La Rosa, J. Mendling, and H.A. Reijers, Fundamentals of business process management. Vol. 1. 2013: Springer.
- [3] Kunze, M., M. Weidlich, and M. Weske, Querying process models by behavior inclusion. *Software & Systems Modeling*, 2015. 14(3): p. 1105-1125.
- [4] Yan, Z., R. Dijkman, and P. Grefen, Generating process model collections. *Software & Systems Modeling*, 2017. 16(4): p. 979-995.
- [5] Tax, N., N. Sidorova, and W.M. van der Aalst, Discovering more precise process models from event logs by filtering out chaotic activities. *Journal of Intelligent Information Systems*, 2019. 52(1): p. 107-139.
- [6] Xue, X., Compact memetic algorithm-based process model matching. *Soft Computing*, 2019. 23(13): p. 5249-5257.
- [7] Huang, Y., W. Li, Z. Liang, Y. Xue, and X. Wang, Efficient business process consolidation: combining topic features with structure matching. *Soft Computing*, 2018. 22(2): p. 645-657.
- [8] Huang, H., Z. Lu, R. Peng, Z. Feng, X. Xuan, P.C. Hung, and S.-C. Huang, Efficiently querying large process model repositories in smart city cloud workflow systems based on quantitative ordering relations. *Information Sciences*, 2019. 495: p. 100-115.
- [9] Shahzad, K., I. Pervaz, and A. Nawab. WordNet based Semantic Similarity Measures for Process Model Matching. in BIR Workshops. 2018.
- [10] Shahzad, K., S. Kanwal, and K. Malik. Process Model Matching with Word Embeddings. in International Conference on Intelligent Technologies and Applications. 2018. Springer.
- [11] Baier, T., C. Di Ciccio, J. Mendling, and M. Weske, Matching events and activities by integrating behavioral aspects and label analysis. *Software & Systems Modeling*, 2018. 17(2): p. 573-598.
- [12] Pergl, R., E. Babkin, R. Lock, P. Malyzhenkov, and V. Merunka, Enterprise and Organizational Modeling and Simulation: 14th International Workshop, EOMAS 2018, Held at CAiSE 2018, Tallinn, Estonia, June 11–12, 2018, Selected Papers. Vol. 332. 2018: Springer.
- [13] Augusto, A., R. Conforti, M. Dumas, M. La Rosa, and A. Polyvyanyy, Split miner: automated discovery of accurate and simple business

- process models from event logs. *Knowledge and Information Systems*, 2019. 59(2): p. 251-284.
- [14] Mendling, J., Metrics for process models: empirical foundations of verification, error prediction, and guidelines for correctness. Vol. 6. 2008: Springer Science & Business Media.
- [15] Dijkman, R., M. Dumas, B. Van Dongen, R. Käärrik, and J. Mendling, Similarity of business process models: Metrics and evaluation. *Information Systems*, 2011. 36(2): p. 498-516.
- [16] Carmona, J., B. van Dongen, A. Solti, and M. Weidlich, *Conformance Checking*. 2018: Springer.
- [17] Leopold, H., H. van der Aa, J. Offenbergh, and H.A. Reijers, Using Hidden Markov Models for the accurate linguistic analysis of process model activity labels. *Information Systems*, 2019. 83: p. 30-39.
- [18] Zeng, Q., X. Tang, W. Ni, H. Duan, C. Li, and N. Xie, Missing Procedural Texts Repairing Based on Process Model and Activity Description Templates. *IEEE Access*, 2020. 8: p. 12999-13010.
- [19] Liu, C., Q. Zeng, H. Duan, S. Gao, and C. Zhou, Towards comprehensive support for business process behavior similarity measure. *IEICE TRANSACTIONS on Information and Systems*, 2019. 102(3): p. 588-597.
- [20] Kim, K.P., Functional Integration with Process Mining and Process Analyzing for Structural and Behavioral Properness Validation of Processes Discovered from Event Log Datasets. *Applied Sciences*, 2020. 10(4): p. 1493.
- [21] Weber, B., M. Reichert, and S. Rinderle-Ma, Change patterns and change support features—enhancing flexibility in process-aware information systems. *Data & knowledge engineering*, 2008. 66(3): p. 438-466.
- [22] Becker, M. and R. Laue, A comparative survey of business process similarity measures. *Computers in Industry*, 2012. 63(2): p. 148-167.
- [23] Banica, B. and L. Patricio. *Service Design for Business Process Reengineering*. in *International Conference on Exploring Services Science*. 2020. Springer.
- [24] “Informative Documents OMG®” October 2019. [Online]. Available: <https://www.omg.org/spec/BPMN/>.
- [25] “Camunda Modeler” October 2019. [Online]. Available: <https://camunda.com/>.
- [26] Corradini, F., A. Ferrari, F. Fornari, S. Gnesi, A. Polini, B. Re, and G.O. Spagnolo, A guidelines framework for understandable BPMN models. *Data & Knowledge Engineering*, 2018. 113: p. 129-154.
- [27] Yan, Z., R. Dijkman, and P. Grefen. Generating synthetic process model collections with properties of labeled real-life models. in *Asia-Pacific Conference on Business Process Management*. 2014. Springer.
- [28] “Proposed Process Model Collection and Structural Metrics,” March 2020. [Online]. Available: <https://bit.ly/3bYyH37>.
- [29] Mendling, J., H. Verbeek, B.F. van Dongen, W.M. van der Aalst, and G. Neumann, Detection and prediction of errors in EPCs of the SAP reference model. *Data & Knowledge Engineering*, 2008. 64(1): p. 312-329.
- [30] Mendling, J., H.A. Reijers, and W.M. van der Aalst, Seven process modeling guidelines (7PMG). *Information and Software Technology*, 2010. 52(2): p. 127-136.
- [31] Dumas, M., L. García-Bañuelos, M. La Rosa, and R. Uba, Fast detection of exact clones in business process model repositories. *Information Systems*, 2013. 38(4): p. 619-633.

# A Gamification Experience and Virtual Reality in Teaching Astronomy in Basic Education

Norka Bedregal-Alpaca<sup>1</sup>, Olha Sharhorodska<sup>2</sup>  
Departamento Académico de Ingeniería de Sistemas e  
Informática  
Universidad Nacional de San Agustín de Arequipa  
Arequipa, Perú

Luis Jiménez-González<sup>3</sup>, Robert Arce-Apaza<sup>4</sup>  
Escuela Profesional de Ingeniería de Sistemas  
Universidad Nacional de San Agustín de Arequipa  
Arequipa, Perú

**Abstract**—Regardless of the country, there is a trend: the world of school and the modern world are two different poles. Young people see school as boring compared to the entertainment of today's technology. Most students prefer to play or surf the internet, but not study. Gamification is projected as a methodological practice that aims to turn classrooms into playful immersion scenarios, using participatory strategies with the incorporation of electronic devices. This article shows the results obtained by applying gamification techniques in the research project aimed at supporting astronomy learning for basic education students. When using the app, the student must overcome challenges to earn different achievements and rewards. Among the results highlights the student's motivation during the learning process and the perception of satisfaction of the personal achievements achieved.

**Keywords**—Gamification; game-based learning; reward system; student motivation

## I. INTRODUCTION

The 21st century is known as the century of creativity, innovation and innovation, as these years face the need to find new ideas and solutions to the many problems that arise in a society of accelerated change (De la Torre, 2006, p.12 cited by [1]). Contrary to this claim, the school creates an artificial environment that does not consider the interests of students and limits their perception and development of the world around them. What's more, teacher and students speak different languages.

Traditionally, the teacher outperforms the student in terms of development and competence, however, in the use of technology the teacher is at a disadvantage because students know and use the new technologies intuitively incorporating them into the way of behavior and socialization. Therefore, an additional problem arises, the students dislike the teachers because they are not in the same tune. Teachers are separated from the modern world, do not use modern educational technologies, have no ICT skills, etc. As a result, students' distrust of the school and loss of learning interest is observed.

Today's children are inundated with entertainment and toys from birth, they perceive that the world is a fabulous place, they consider their parents as magical beings who can be asked or demanded any desire. The Internet and virtual toys create a sense of constant accessibility to anything. Then a question arises how to use technology to motivate learning?

For Marc Prensky, today's young people cannot learn how yesterday's young people cannot learn, so if they learn differently, they must be taught differently, using novel and varied methodologies, forms, methods and means. In this context, gamification arises as a methodology that supports training processes, as it makes learning processes more attractive by enhancing fun, productivity, the ability to retain concepts and the acquisition of skills. Gamification claims the role of games and especially video games, as a vehicle for help for the educational task, allows to involve the student in the practical work inside and outside the classroom, so that it is the students themselves who are the protagonists of their learning processes [2].

According to [3], "*the dynamics of games themselves can increase the attention of students during the learning teaching process by also improving their satisfaction with this process*". In [4], it states that the game in the context of education aims to teach and reinforce different aspects such as knowledge and skills, such as problem solving, collaboration and communication. One of the most important aspects when gamifying in the educational context, according to [4], is the organization, since it directly conditions the expected results.

Students are growing up in a technified society, so educators need to reconfigure their pedagogical strategies so that students can develop different capacities that enable them to understand and address the changing world around them [5]. That is why, in gamification processes ICTs with mobile devices and cutting-edge technologies are being incorporated [6].

## II. GAMIFICATION

Even in the last century, L.N. Tolstoy in his work "General Observations to the Teacher" wrote: "For a student to study well, he must learn eagerly; for you to study at ease, it is necessary that what is taught to the student is understandable and entertaining; that his spiritual strength would be in the most advantageous conditions".

Since ancient times there has been a close relationship between education and play, from its early stages of life, the human being learns by playing. There are several efforts dedicated to analyzing and redesigning the game model, the pioneers in redefining the game concept and its relationship with new technological developments are [7] and [8]; they propose that one of the fundamental elements of the game is

the relationship between the established rules and the player experience with ample room for maneuverability within those rules; that is, the balance between rules and freedom.

Gamification arises as a form of active learning in which the rules of the game are used to achieve real goals. In a game, boring tasks become interesting, complex tasks become simple, and rejected ones become desirable. This happens, for example, when teaching mathematics is done through non-traditional sessions that include dynamics and technology that make mathematics fascinating and more accessible, so that students are encouraged to participate in activities active and creative in which the necessary knowledge, and required skills are acquired and systematized.

It is important to emphasize that gamification is not synonymous with "playing in the classroom" or "learning by playing", nor is it the same to refer to the term for learning through video games, mobile applications (apps) or any other Information Technology Communication (ICT); but it is the use of game design elements in traditionally non-playful contexts [9], these elements are being the mechanics, dynamics and aesthetics.

Gamify is not a question of "designing a game", but of taking advantage of the reward systems that usually have these (points, medals, levels, missions, challenges, achievements, advantages...), as well as the dynamics and aesthetics to create an experience that maintains the interest in the development of educational content [10].

The concept of gamification, according to [11] was presented by Nick Pelling (2002). In [12] contend that until 2010 the use of the term was not widely used. In their gamification by Design work, [13] defines this concept as "a process related to player thinking and gaming techniques to attract users and solve problems". In [14] it presents the work "The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education" and states that gamification is "the use of mechanisms, aesthetics and the use of thought, to attract people, stimulate the action, promote learning and solve problems". The authors mentioned share the same vision on gamification, focused on the influence it has on the psychological and social behavior of the player. They argue that the elements present in the games generate greater disposition and motivation and increase the time the student spends on them, so they can be used more productively to achieve greater learning.

On the other hand, the study [15] "Social Motivations to Use Gamification: An Empirical Study of Gamifying Exercise" postulates that gamification is the production and creation of experiences that bring feelings of control and autonomy to people, in order to influence their behavior, leaving in the background the enjoyment they may experience during the activity.

For [16] gamification "is the application of game metaphors to real-life tasks to influence behavior, improve motivation and encourage involvement in that task". This work assumes that gamification is the use of elements of games in playful environments, that is, the application to an activity of

the aspects that make a game attractive to favor effort, motivation and performance.

In [17] the authors defend the importance of making a good design of the gamified activity, since it is the main action to be carried out and the one that can condition the good or bad result of the activity. Then, you must choose the elements of the game based on pedagogical criteria that allow to analyze the function and the concrete use of all the resources to be created or used.

Specialists also express the need to include an emotional component in gamification.

### III. GAME-BASED LEARNING

Game-Based Learning (GBL) uses the game as a vehicle and tools to support learning, assimilation or knowledge assessment. According to Prensky (2001) this type of learning focuses on those games with educational objectives, which enhance learning outcomes and promote a more fun, interesting and therefore more effective experience.

In [18] and [19], they proposed specifically, the notion of game-based learning has involved supporting teaching, encouraging decision-making, valuing the change in game outcomes based on their actions, and boosting social skills alongside work in team.

In this specific typology a series of commonly applied patterns is set up, highlighting the incorporation of rules and restrictions, dynamic and instant responses to the actions taken by the students, challenges suited to the subject of study that catalyze self-efficacy and progressive learning of difficulty [20].

By using digital educational games (supported by ICT), the GBL represents a substantial improvement that it also brings to digital competition.

Among the advantages of GBL are: (a) Motivates the student, (b) Contributes to reasoning and autonomy, (c) Promotes active learning, (d) Gives the student control of their learning, (e) Provides useful information to the teacher, (f) Empowers the teacher creativity and imagination, (g) Promotes social skills and (h) Contributes to digital literacy.

### IV. CONTEXT

In Peru, astronomy is taught in basic education, and as in other countries, it has a particular problem; teaching is done theoretically, in a few hours of class and at the end of semesters. Students do not pay much attention to the subject, are unmotivated and do not understand the basics of this science. That is why this work describes the experience of designing, implementing and using an application created in order to enhance the learning results of students in Astronomy topics, for which they were proposed as specific objectives:

- Analyzing information about the characteristics of the solar system and planets.
- Differentiate the elements that make up the universe.
- Explain the origin of the universe from the Big Bang theory.

The experience was carried out in 2019 in the field of basic education in Arequipa-Peru, with students of the first degree of secondary education from two national educational institutions. In the I.E. Our Lady of the Assumption has worked with 40 students and in the I.E. Antonio José de Sucre worked with 59 students, in both cases, he was supported by a professor from the I.E. who acted as a pedagogical advisor.

#### A. Moments of Experience

- 1) Coordination of visits by San Agustín National University of Arequipa managers to educational institutions.
- 2) Training session on the use of the application, with assistance of students, teachers in the area of Science Technology and Environment, head and laboratory assistant.
- 3) Meeting for validation of the first prototype on the platform, in which teachers suggested placing content in an introspective way, that is, starting from the general and approaching particular objects, always with the support of information that serve the student to understand the natural structure or phenomena of the astronomical field.
- 4) Meeting for validation of the second prototype with modifications made from the suggestions of the previous meeting.
- 5) Design of the learning session, carried out by teachers of educational institutions considering the skills and capacities to be developed and the respective indicators.
- 6) Programming the learning session with first grade students.
- 7) Evaluation of the results obtained.

GBL was used as a teaching-learning model in order to improve students' motivation and involve them in the different topics to be developed in class, all so that the student can identify, recognize and associate the planets and other celestial bodies. To implement this model, the simulation method was used, which is based on an interplanetary journey using virtual reality. To encourage learning, some training mini-games have been implemented, which allow to reinforce the concepts learned.

#### B. Technical Characteristics

The Unity game engine has been used for system development. The computer application considers two versions.

The first to be used on a computer (2D modules). The second to be used on mobile devices with resolution of 2560 x 1440 pixels, with gyrosopic sensor, Android version 4.1 or higher, which when interacting with Google Cardboard make it possible to display the modules in the virtual reality environment (immersive modules). Using Google Cardboard allows you to achieve greater accessibility through the application, as it is not limited to a physical room.

To use the app, the user must place the mobile device in their VR Box and launch the application.

The decision to use two versions was made because students became dizzy using the VR headset for more than 15 minutes.

All developments have been carried out in constant communication with teachers of educational institutions and under the guidance of astronomy specialists.

#### V. DESCRIPTION OF MODULES

To start the trip the student must register: create their username and password. At the end of registration, you enter the application, where with the help of Astronito (user guide), you are given explanations of how to navigate the modules, the game rules, the prizes to win, etc. The record also gives you the possibility to see your progress, the achievements achieved and the ranking achieved (Fig. 1 and 2).

The home screen displays a set of buttons that allow you to access the relevant information (Fig. 3).

The interstellar journey takes place aboard a spacecraft. The journey begins as a crew member. As the study modules progress and the proposed objectives are met, you can become a Stellar Captain (Fig. 4).



Fig. 1. Admission to the Application.



Fig. 2. Interstellar Travel Home Starting Screen.

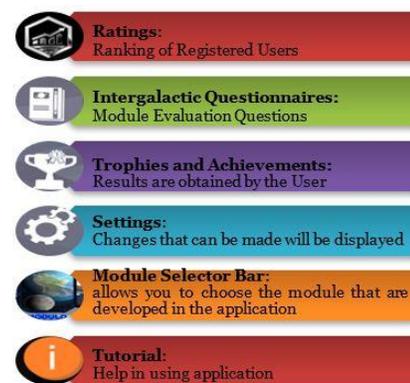


Fig. 3. The Main Screen Buttons of the App.



Fig. 4. Maximum Achievement after Successful Completion of All Modules.

To meet the educational objectives raised and cover the proposed topics, 7 modules were developed. These modules contain the information needed to understand the study topic.

Initially, the development of all modules was planned with the use of virtual reality; however, in the process of developing the application, it was necessary to reduce the time of use of the virtual helmet due to the dizziness experienced by students. As a result, the learning modules were developed in 2D and 3D. As not all students were able to purchase the virtual helmets, they were given the step-by-step instructions to build their own Cardboard with materials that were at their disposal.

When entering any module different options are presented (Fig. 5).

- Videos: educational videos related to the subject in question.
- Virtual game: mini game that supports module learning.
- Questionnaire: set of questions to verify what you have learned.
- Did you know that: interesting facts about the subject covered?

A brief description of each of the modules available in the application is provided below.

**"Planet Earth" module:** This module makes known the relevant data on planet Earth, in an interactive way it is portrayed the different phenomena through which our beloved planet passes and some of the most important characteristics these ones have (Fig. 6).

This module covers the topics:

- Rotation and Translation of the Planet
- Seasons of the year
- Solstice and equinox.

The information is displayed to the user in the form of audiovisual information through a video.

This module is presented as a single sequence of interactive questions, combining visual and auditory information (Fig. 6). Inside the module you have an interactive mini game that allows you to measure the knowledge acquired on Planet Earth. Each success of the student on the proposed question adds 5 points to the ranking and for each error is subtracted 3 points.



Fig. 5. Module Options Menu.



Fig. 6. The Panel View of the "Planet Earth" Module.

With the help of Astronito, if necessary, it is indicated that information is required to improve its performance. The student can view information on the topics to be discussed at the beginning, before questionnaire or minigame, for this action no points are subtracted, but if at the beginning of the sequence of questions the student wants to display the information is subtracted 2 points.

At the end of the module, the accumulated score and the trophy or medal achieved are assigned.

**"Moon" module:** This module explains the lunar phases and the lunar and solar eclipses. (Fig. 7). Three sequences of interactive questions on a given topic are proposed, which do not follow a specific order.

Question sequences are related to the topics:

- Eclipse Solar, where the student is asked to identify the characteristics that lead to this phenomenon and a simulation of the eclipse is made.
- Lunar Eclipse, similar to the solar eclipse, the student must identify the characteristics that lead to this phenomenon and a simulation of the eclipse is made.
- Phases of the Moon, where the student is sought to identify the position in which the moon should be relative to the student's simulated position on earth, so that each of the 4 main phases of the moon can be observed.

**"Solar System" module:** Module comprising the entire solar system where Planet Earth is located (Fig. 8). The features it provides are:

- Display the order of the planets.
- Visualize the translation of each planet.
- Display relevant information about each planet.
- Allow to differentiate between planets through your information.

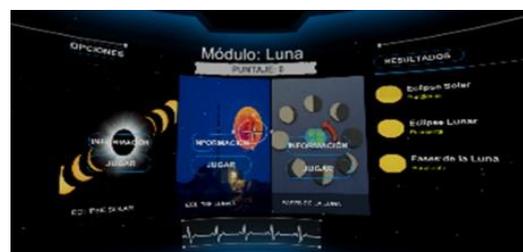


Fig. 7. The Panel View of the "Moon" Module.

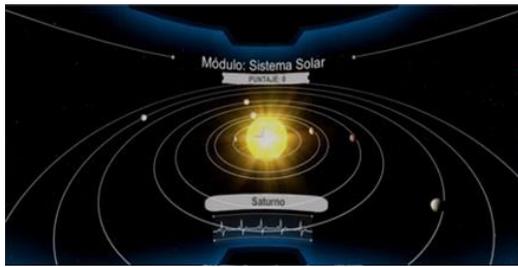


Fig. 8. View of the Minigame of the Module "Solar System".

This module is represented as a mini game where the student must identify the planets and positions that correspond with the help of the information provided by Astronito. At the start of the game, the respective rules and instructions are disclosed. Below is the name of the planet that the user must recognize in the Solar System. In case you hit it is increased by 5 points, otherwise you will be subtracted 3 points. In this module there is no penalty for displaying the information in the middle of the game.

**"Minor Bodies" module:** This module discloses the relevant data to the smaller bodies (Fig. 9).

An evolutionary conceptualization is proposed that goes in the following order:

- Comets: concepts, parts, types, characteristics and curiosities.
- Asteroids: concepts, characteristics and curiosities.
- Meteoroids, Meteors and Meteorites: Explaining them conceptually with a 3D representation.

Relevant aspects of each of them are determined, and a question is revealed at the end of each of the parties, setting the score and type of award. Each correct answer adds 3 points to the module result.

**"Stars" module:** This module explains the life of a star, from its creation to its death. In exploratory form, the different types of stars are disclosed (Fig. 10). Through the tour, by self-discovery, all types of stars must be completed and the proposed questionnaires answered. Upon completion of the game, they are assigned the corresponding score and their respective award.

**"Galaxies" module:** This module explains the different components and different types of Galaxies. In addition, information is given about our galaxy – the Milky Way (Fig. 11).



Fig. 9. Module View "Minor Bodies".



Fig. 10. Module View "Stars".



Fig. 11. Module View "Galaxias".

To discover all this information, the student has to complete three missions (Fig. 12):

- Components of a Galaxy - develops in space, where the student visualizes various bodies of space, the goal is to trap the bodies that are part of a Galaxy. The student gets 3 points when capturing the correct components, otherwise 1 point is subtracted.
- Galaxy Types - the student will need to identify the card, which corresponds to the galaxy that appears randomly on the screen, to click on the card, Astronito gives you additional information and so you can be sure of your choice. If the choice is correct, it wins 10 points, otherwise you lose 2 points for each failed attempt, a maximum of three attempts are allowed.
- Milky Way - relevant data about our galaxy are disclosed. Interactively, their different characteristics are displayed and the student is asked to identify them. Each success of the student on the proposed question adds 5 points to the ranking and for each error is subtracted 2 points.

**Big Bang module:** is the module where one of the theories of Origin of the Universe (Big Bang theory) is released, without touching theories related to the teaching of higher education (quantum and relativity theories) (Fig. 13). After watching an educational video, the learning assessment will be done with the help of a card game, in which the correct position of an event is chosen through a deck, based on the order of its appearance in the Big Bang Theory and according to the question asked.



Fig. 12. Interstellar Missions.



Fig. 13. "Big Bang" Module View.

## VI. REWARD SYSTEM

In order to improve the user's interest in the realization of educational activities, a reward system was established, thus creating periodic moments of satisfaction.

- **Point Accrual:** A system was developed to evaluate the performance of each activity; points are awarded based on the time used to complete an exercise and the number of correct responses.
- **Level scaling:** Several activities were created, which are placed at difficulty levels. The student starts with basic activities until they reach more complex activities.
- **Prizes:** Rewards and virtual prizes have been implemented that the student accesses when completing the levels.
- **Rankings:** The ratings system discloses the progress of each student and promotes competition between them.
- **Missions and Challenges:** There are challenges, within the application, that are cumulative; Completing one creates a new challenge with a higher level of complexity.

At the end of each module, depending on the score acquired, each user is awarded a Trophy or Medal, which are indispensable to obtain a new badge, and consequently, become An Interstellar Captain, as a result a ranking of users is generated (Fig. 14).

The scores required to earn the Trophy according to the module are presented in Table I.



Fig. 14. Users' Ranking List.

TABLE I. TROPHIES

Type	Score
Moon Trophy	The Luna module is required with a score greater than 20 points.
	Having a score greater than 10 points, but less than 20 points you will be given a silver trophy as recognition, but will not count as an official trophy. You will be asked to win the trophy you earn by taking a score greater than 20 to get the official trophy.
	Having a score less than or equal to 10 will prompt you to re-develop the module
Earth Trophy	The Earth module is required with a score greater than 30 points.
	Having a score greater than 20 points, but less than 30 points you will be given a silver trophy as recognition, but will not count as an official trophy. You will be asked to win the trophy you earn by taking a score greater than 30 to get the official trophy.
	Having a score less than or equal to 20 will prompt you to re-develop the module.
Solar System Trophy	It is required to complete the Solar System module with a score greater than 30 points.
	Having a score greater than 20 points, but less than 30 points you will be given a silver trophy as recognition, but will not count as an official trophy. You will be asked to win the trophy you earn by taking a score greater than 30 to get the official trophy.
	Having a score less than or equal to 20 will prompt you to re-develop the module.
Minor Bodies Trophy	Minor Bodies module is required with a score greater than 70 points.
	Having a score greater than 60 points, but less than 70 points you will be given a silver trophy as recognition, but will not count as an official trophy. You will be asked to win the trophy you earn by taking a score greater than 70 to get the official trophy.
	Having a score less than or equal to 60 will prompt you to re-develop the module.
Star Trophy	The Star module is required with a score greater than 60 points.
	Having a score greater than 50 points, but less than 60 points you will be given a silver trophy as recognition, but will not count as an official trophy. You will be asked to win the trophy you earn by taking a score greater than 60 to get the official trophy.
	Having a score less than or equal to 50 will prompt you to re-develop the module.
Galaxy Trophy	Galaxy module is required with a score greater than 70 points.
	Having a score greater than 60 points, but less than 70 points you will be given a silver trophy as recognition, but will not count as an official trophy. You will be asked to win the trophy you earn by taking a score greater than 70 to get the official trophy.
	Having a score less than or equal to 60 will prompt you to re-develop the module.
Big Bang Trophy	The Big Bang module is required with a score greater than 70 points.
	Having a score greater than 60 points, but less than 70 points you will be given a silver trophy as recognition, but will not count as an official trophy. You will be asked to win the trophy you earn by taking a score greater than 70 to get the official trophy.
	Having a score less than or equal to 60 will prompt you to re-develop the module.

The information regarding the Medals is set out in Table II.

TABLE II. MEDALS

Type	Score
"Luna" Medal - Total: 30 points	A maximum score of 30 points is required.
"Earth" Medal - Total: 40 points	A maximum score of 40 points is required.
Solar System Medal - Total: 40 points	A maximum score of 40 points is required.
Minor Bodies Medal - Total: 100 points	A score greater than 75 points is required.
"Stellar" Medal - Total: 100 points	A score greater than 70 points is required.
"Galactic" Medal - Total: 100 points	A score greater than 80 points is required.

TABLE III. PARTICIPANTS' BADGE

Type	Score
Crewer	You get it when you start the adventure.
Captain Solar	Required when obtaining trophies from the Moon, Planet Earth, and Solar System modules.
Star Captain	Required when obtaining trophies from Star, Lesser Corps, and Solar Captain badge.
Galactic Captain	Required when obtaining the Galaxy Module Trophy and having Captain Star badge.
Interstellar Captain	Required when obtaining the Big Bang module trophy and having Captain Galactic badge

As a result of their effort, the student can achieve the badges shown in Table III.

### VII. RESULTS AND DISCUSSION

The application, described in this work, has the appropriate didactic quality to enhance students' astronomy learning, as each stage of software development was accompanied by the pedagogical evaluation and field tests necessary.

Formed by test groups, it was initially necessary to train both students and teachers in the use of digital tools. Students showed interest in learning the correct handling of the application.

During the tests, the opinions and recommendations of both the students and the teachers were collected as pedagogical advisors.

To validate the application, the methodology developed by Abreu was used, and applied in works in which it was intended to evaluate the didactic aspect of educational material [21].

The criteria used for the evaluation of the application were:

- Environment quality and user interface.
- Content and relevance of information.
- Student control.
- Collaborative learning.

- Targeting.
- Applicability.
- Motivation.
- Flexibility.
- Feedback.

Some results of this evaluation are described in article CINAIC – 2019 [22].

The application was first tested and validated by science teachers from two national schools in the city of Arequipa, what were the test venues, and then with the students of these schools. In addition, the validations were made with the early-year students of the San Augustin National University of Arequipa. Opinions on the application were raised on the basis of a survey on the perception of users as to the usefulness of the application and the satisfaction of its use (Fig. 15, 16 and 17).

The application, described in this work, has the appropriate didactic quality to enhance students' astronomy learning, as each stage of software development was accompanied by the pedagogical evaluation and field tests necessary.

The results obtained show the desirability of applying different teaching methodologies that enhance the learning process of students. Results like these can be improved by combining methodologies such as Problem-Based Learning (PBL), Project-Based Learning (PrBL), among others. Results of such experiences can be seen in [23] and [24].

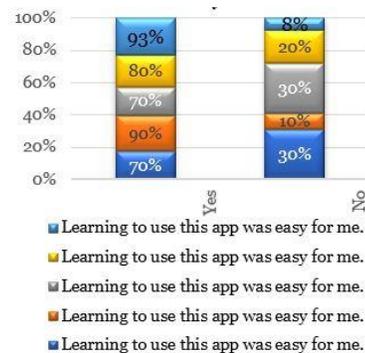


Fig. 15. Some Survey Results.

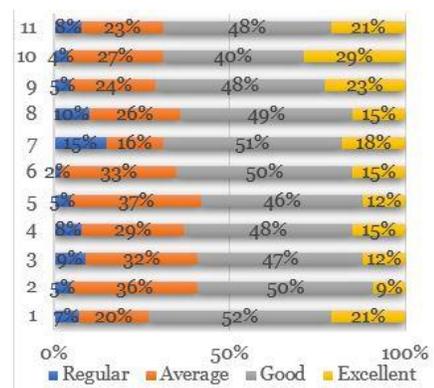


Fig. 16. Utility Assessment Results.

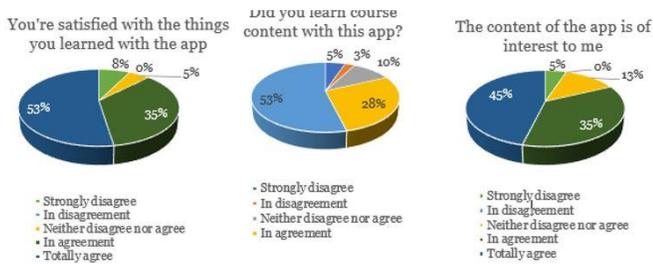


Fig. 17. Some Results Performed.

It is also convenient to measure student satisfaction regarding the use of technology under some model known as the Technology Acceptance Model (TAM), in [25] the TAM model has been applied to measure the satisfaction of virtual platforms in two subjects.

### VIII. CONCLUSIONS

By applying gamification in the educational context, in addition to improving the level of motivation of students, learnings are promoted from interaction activities with the game.

By applying gamification in the educational field integrating learning and play, it is possible to improve the training process, in addition to improving the learning results, generic skills are developed that will serve the student throughout life.

The didactics of the basic sciences must be strengthened and nurtured by novel and varied methods, techniques, means and forms that promote student participation as the center of the training process, among them the study of Astronomy is underlined.

Gamification of the training process should be considered systematically and inescapable in teaching practice. This is the case that this work is intended to be a boost to continue research processes and application of gamification in different educational contexts.

### ACKNOWLEDGMENT

We do want to express our thanks to the San Augustin National University of Arequipa for the support provided in the development and dissemination of this work.

### REFERENCES

- [1] O. Klimenko, "La creatividad como un desafío para la educación del siglo XXI", *Educación y Educadores*, 11(2), 191-210, 2008. <http://www.redalyc.org/articulo.oa?id=83411213>
- [2] C. Corchuelo, "Gamificación en educación superior: experiencia innovadora para motivar estudiantes y dinamizar contenidos en el aula", *EDUTEC, Revista Electrónica de Tecnología Educativa*, N° 63, 29-41, 2018. Recuperado de: <http://dx.doi.org/10.21556/edutec.2018.63.927>
- [3] J. Cuevas and A. Andrade, "Abordajes metodológicos para problemas educativos emergentes", 2016. [https://www.academia.edu/25838315/Hacia\\_la\\_gamificaci%C3%B3n\\_educativa](https://www.academia.edu/25838315/Hacia_la_gamificaci%C3%B3n_educativa)
- [4] R. Contreras and J. Eguía, "Gamificación en aulas universitarias", *Bellaterra: Institut de la Comunicació, Universitat Autònoma de Barcelona*, 2016. Recuperado de [https://www.academia.edu/22834718/Gamificaci%C3%B3n\\_en\\_aulas\\_universitarias](https://www.academia.edu/22834718/Gamificaci%C3%B3n_en_aulas_universitarias)

- [5] S. Peñalva, I. Aguaded and A. Torres-Toukoumidis, "La gamificación en la universidad española. Una perspectiva educomunicativa", *Revista Mediterránea de Comunicación*, 10(1),245-256, 2019. <https://www.doi.org/10.14198/MEDCOM2019.10.1.6>
- [6] D. Vergara and A. Gómez, "Origen de la gamificación educativa", 2017. Recuperado de <http://espacioeniac.com/origen-de-la-gamificacion-educativa-pordiego-vergara-rodriguez-y-ana-isabel-gomez-vallecillo-universidad-catolica-deavila>.
- [7] J. Huizinga, "Homo Ludens". Madrid: Alianza Ed, 1990.
- [8] R. Caillois, "Man, play, and games", University of Illinois Press, 2001.
- [9] S. Deterding, D. Dixon, R. Khaled and L. Nacke, "From Game Design Elements to Gamefulness: Defining Gamification", En: 15th International Academic MindTrek Conference: Envisioning Future Media Environments (9-15), 2011. Recuperado de <https://goo.gl/2n6hOC>
- [10] L. Romero-Rodríguez, A. Torres-Toukoumidis and I. Aguaded, "Ludificación y educación para la ciudadanía", *Revisión de las experiencias significativas. Educar*, 53(1), 109-128, 2017. doi: <http://dx.doi.org/10.5565/rev/educar.846>
- [11] F. Rodríguez and R. Santiago, "Gamificación: Cómo motivar a tu alumnado y mejorar el clima en el aula", México: Editorial Océano, 2015.
- [12] H. Romero and E. Rojas, "La Gamificación como participante en el desarrollo del B- learning: Su percepción en la Universidad Nacional, Sede Regional Brunca", In Eleventh Latin American and Caribbean Conference for Engineering and Technology (LACCEI' 2013) "Innovation in Engineering, Technology and Education for Competitiveness and Prosperity", Cancun, Mexico, 2013.
- [13] G. Zichermann and C. Cunningham, "Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps", Cambridge, MA: O'Reilly Media, 2011.
- [14] K. Kapp, "The gamification of learning and instruction: game-based methods and strategies for training and education", EEUU: John Wiley & Sons, 2012.
- [15] J. Hamari and J. Koivisto, "Social Motivations To Use Gamification. An Empirical Study Of Gamifying Exercise", *ECIS*, 2013.
- [16] A. Marczewski, "Gamification: a simple introduction", E-Book, 2013.
- [17] J. Foncubiarta and C. Rodríguez, "Didáctica de la gamificación en la clase de español", Madrid: Editorial Edinumen, 2014.
- [18] B. Kim, H. Park and K. Baek, "Not just fun, but serious strategy: using Meta cognitive strategies in game-based learning", *Computers and Education*, 52(4), 800-810.
- [19] E. Rincón-Flores, M. Ramírez-Montoya and J. Mena, "Challenge-based gamification and its impact in teaching mathematical modeling", In *Proceedings of the fourth International Conference on Technological Ecosystems for Enhancing Multiculturality, TEEM 2016*.
- [20] S. Erhel and E. Jamet, "Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness", *Computers & Education*, 67, pp. 156-167, 2013. DOI 10.1016/j.compedu.2013.02.019
- [21] R. Aceituno and S. Bruschi, "Aplicação da metodologia AIM-CID nos conceitos da disciplina sistemas operacionais, no domínio de gerenciamento de processos", *Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação*, 2013.
- [22] N. Bedregal-Alpaca, O. Sharhorodska, L. Jiménez-González and R. Arce-Apaza, "Gamificación como estrategia para potenciar el aprendizaje de la astronomía en la educación secundaria", V Congreso Internacional sobre Aprendizaje, Innovación y Competitividad (CINAIC 2019), pp. 710-715, 2019. DOI: 10.26754/CINAIC.2019.0150.
- [23] N. Bedregal-Alpaca, D. Tupacyupanqui-Jaen, M. Rodríguez-Quiroz, L. Delgado-Barra, K. Guevara-Puente and O. Sharhorodska, "Problem-Based Learning with ICT Support: An experience in teaching-learning the concept of derivative," 2019 38th International Conference of the Chilean Computer Science Society (SCCC), Concepcion, Chile, 2019, pp. 1-7. DOI: 10.1109/SCCC49216.2019.8966396
- [24] Victor Cornejo-Aparicio ; Sidanelia Flores-Silva ; Norka Bedregal-Alpaca ; Doris Tupacyupanqui-Jaén. "Capstone courses under the PBL

- methodology approach, for engineering)". 2019 IEEE World Conference on Engineering Education (EDUNINE). DOI: 10.1109/EDUNINE.2019.8875803
- [25] N. Bedregal-Alpaca, V. Cornejo-Aparicio, D. Tupacyupanqui-Jaén and S. Flores-Silva, "Evaluación de la percepción estudiantil en relación al uso de la plataforma Moodle desde la perspectiva del TAM", *Ingeniare. Revista chilena de ingeniería versión On-line* ISSN 0718-3305, *Ingeniare. Rev. chil. ing.* vol.27 no.4 Arica dic. 2019. DOI 10.4067/S0718-33052019000400707.

# Customer Churn Prediction Model and Identifying Features to Increase Customer Retention based on User Generated Content

Essam Abou el Kassem<sup>1</sup>

Mathematics and Statistics Department  
Faculty of Commerce and Business Administration  
Helwan University  
Cairo, Egypt

Shereen Ali Hussein<sup>2</sup>

Management Department  
Faculty of Commerce and Business Administration  
Helwan University  
Cairo, Egypt

Alaa Mostafa Abdelrahman<sup>3</sup>

Business Information Systems Department  
Faculty of Commerce and Business Administration  
Helwan University  
Cairo, Egypt

Fahad Kamal Alsheref<sup>4</sup>

Information Systems Department  
Faculty of Computers and Information  
Beni-Suef University  
Beni-Suef, Egypt

**Abstract**—Customer churn is a problem for most companies because it affects the revenues of the company when a customer switch from a service provider company to another in the telecom sector. For solving this problem we put two main approaches: the first one is identifying the main factors that affect customers churn, the second one is detecting the customers that have a high probability to churn through analyzing social media. For the first approach we build a dataset through practical questionnaires and analyzing them by using machine learning algorithms like Deep Learning, Logistic Regression, and Naïve Bayes algorithms. The second approach is customer churn prediction model through analyzing their opinions through their user-generated content (UGC) like comments, posts, messages, and products or services' reviews. For analyzing the UGC we used Sentiment analysis for finding the text polarity (negative/positive). The results show that the used algorithms had the same accuracy but differ in arrangement of attributes according to their weights in the decision.

**Keywords**—Customer churn; telecom sector; churn prediction; sentiment analysis; machine learning; customer retention

## I. INTRODUCTION

During the enormous increase in numbers of customers who are using the communication sector and in numbers of companies [1], the competitive level between companies raised [2, 3]. Each company tries to survive in this competition through many strategies [4], The Main strategies are: 1) upsell existing customer, 2) increase duration of retention of their customers, 3) acquire new customers. Companies are concerned about seeking to keep or retain their customers as they are considered that as a profit, and it is cheaper to keep them than to earn a new one. Each company tries to keep its customers, by make them more loyal. Customers are great ambassadors in the market [5] as the company can use them for making advertising of the company's product or service.

This free advertising will cost nothing except the high quality and service after the sale.

Each company should be concerned with customer churn prediction and predict customers who are likely to leave the company to preserve its revenue, but companies must do it early [6].

Many types of research assured that machine learning is efficient in prediction through learning from past situations or previous data [7, 8, 9].

A customer churn happens when customers are not satisfied with a service provided by a company. It results in customers switching to another service provider. Customers have different reasons for churn, and all of them should not be treated in the same way. There is a need for a prediction model to predict churn customers and provide a strategy of retention depends on their churn factors [7].

According to vast numbers of people who use social media to show their opinions, whether by text, emotion, picture or video, we use sentiment analysis to analyze and classify every comment into positive, neutral, or negative [10]. And then, we track negative comments to the customer retention department to retain the churned customer.

In this paper, we proposed two approaches for helping companies to keep their customers by identifying the top reasons for churn and predicting the customer before the churn action is taken.

The rest of this paper is organized as: Section 2 presents the literature review. The background of the technique is presented in Section 3. The proposed model for churn detection from social media is briefly explained in Section 4. Experimental results and analysis are discussed in Section 5. We conclude the proposed study in Section 6.

## II. LITERATURE REVIEW

This section presents the literature related to customers' retention and its prediction methods, identifying many factors related to customer retention, and using social media to get users' opinions to enhance retention.

Don Jyh-Fu Jeng, Thomas Bailey, has used hybrid, multiple criteria decision-making (MCDM) method to inspect customer retention framework and they found that the most common response was to look at pricing and customer service [11]. Ali Tamaddoni Jahromi, Stanislav Stakhovych, Michael Ewing used models for churn predictions in a B2B context, and to increase the profitability of retention campaigns; he found that boosting model, logistic regression, cost-sensitive and straightforward decision tree is applied on tests [12]. While Nitish Varshney and S.K. Gupta used social media analytics to get users' opinions through Twitter. The tweets were classified into three categories using a lexicon-based classifier and applying the association rule mining to find the dominant churn factor [13]. J. Vijaya, E. Sivasankar and S. Gayathri have proposed that ensemble classification techniques with hybrid fuzzy clustering provide more accuracy and better performance than single classifiers and clustering [14]. Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., & Anwar, S used a distance factor in classifier decision [5]. Hossain, M.A, Chowdhury, M .R., & Jahan, N. have supported the importance of customer satisfaction in building a relationship between buyer and seller, using a model that was constrained only four constructs (price, network, customer care, and brand image) to explain customer satisfaction [15]. Adnan, Sajid, Awais, M.Nawaz, K.Alawfi, Amir, and Kaizhu have proposed a practical approach to classify, predict and extract important decision rules related to customer churn or not according to an intelligent rule-based decision-making technique, this technique based on rough set theory (RST). Experiments are carried out to evaluate the performance of RST using Exhaustive Algorithm (EA), Genetic Algorithm (GA), Covering Algorithm (CA), and the LEM2 algorithm (LA). Results show that RST based on GA is the most efficient technique for extracting implicit knowledge in the form of decision rules [16]. However, J. Vijaya and E. Sivasankar used RST with other techniques such as Bagging, Random Subspace, and Boosting. Boosting has achieved the highest accuracy of 93.73%. They found that ensemble classification techniques work better with a classification accuracy of 95.13% compared to any single model [17]. Despite Abhishek and Ratnesh have trained four machine learning models which are Logistic Regression, Support Vector Machine, Random Forest and Gradients boosted tree, and they found that Gradient boosted tree is best among other models, Both Random forest and Logistic regression are an average while SVM is underperforming between these models [18].

Most of the previous researches used classical machine learning techniques, and because of that, we tried to use the recent techniques like deep learning for identifying churn factors accurately. Also, we used the power of social media for early churn detection of customers through using sentiment analysis that used in many areas of business analytics models.

In Section 3, we will discuss the algorithms we used in our study and sentiment analysis technique.

## III. BACKGROUND OF TECHNIQUE

We have many algorithms in machine learning; we discuss only three algorithms that we used in our study and sentiment analysis technique.

### A. Machine Learning Techniques-Classification Methods

Many approaches were applied to predict churn in the telecom sector; most of them have used machine learning technology and data mining. Techniques supposed for use in customer churn prediction:

#### a) Deep Learning Algorithm

Deep learning is a subset of machine learning based on neural networks that permit a machine to train itself to perform a task [19].

#### b) Naïve Bayes

Naïve Bayes classifier, also known as simple Bayes or independence Bayes, is a simple probabilistic classifier. This method builds on independence between the input variables, but it performs well even under conditions that might be considered suboptimal for algorithms [20, 21].

#### c) Logistic Regression

Logistic Regression is a traditional machine learning algorithm developed by a statistician. It is used for classification problems as it works through predictions of the relationship between the predictor variable and the output variable [21, 22].

### B. Sentiment Analysis

Sentiment analysis, also known as opinion analysis, is one of the most important techniques used in social media. This technique is used to extract expressions, opinions of internet users, which are expressed in several forms (such as emotions, texts, pictures, videos), analyzing the opinion-oriented, and then classify it into positive, neutral, or negative sentiments [23, 24].

## IV. RESEARCH CONTRIBUTION

### A. Customer Churn Prediction Model

This paper proposed two main contributions; the first one is a model for customer Churn prediction by analyzing user-generated content, and the second model is identifying main attributes that help the retention department to keep their customers and prevent them from the churn.

Customer churn prediction model using UGC proposed in Fig. 1, the proposed model consists of multiple processes, as shown in Fig. 1; these steps are:

Step 1: User creates his user-generated content; this content could be post, opinion, or comments.

Step 2: English treebank applies text preprocessing, stemming, and lemmatization on English text to extract essential words in their basic form.

TABLE I. EXAMPLE\_SENTENCE\_CLASSIFICATION

Example 1	I	hate	this	Company	Polarity Percentage
Classification	Neutral	Negative	Neutral	Neutral	Negative 99.8%
Example 2	Vodafone	is	a good	network	
Classification	Neutral	Neutral	Positive	Neutral	Positive 94%
Example 3	First	improve	your	Customer service	
Classification	Negative	Positive	Positive	Negative	Negative 92.7%

So, in the first example we have one-word negative and do not have any positive word, so the comment classified as a negative comment. While in the second example we have one-word positive and do not have any negative word, so the comment classified as a positive comment. Then in the last example we two positive and two negative words, the comment classified as negative. The classification happens for both the words and the sentence to identify the polarity to be more accurate.

*B. Identifying Customer Churn Factors*

We created a questionnaire and distributed it among customers for building a dataset that can be analyzed by machine learning algorithms. This dataset is built within different telecom companies and put almost known factors in it that later affect their decision, whether churn or not.

Description of survey's attributes

Gender: Gender of the customer.

Age: age of the customer.

The company of cellular communication service: name of the company which the customer-related.

For how many years a customer in the company: no. of years that customer relates to the company.

Customer's line type: line type, whether personal, business, or corporate.

Use to make a complaint: customers have used in displaying a complaint.

A problem happened with the company: mention a problem that happened.

The company tried to retain the customer: mention the response of the company.

Churn: customer ports out.

In the next section, we will present experimental results and analysis of proposed solutions.

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we have explored the experiments and results of the proposed study, our experiments are divided into two experiments. The first experiment analyzed users' comments from social media through sentiment analysis. The second experiment analyzed the dataset by using Naïve Bayes, Logistic Regression, and Deep learning. Then we used the correlation coefficient for finding the most effective factor with churn decision.

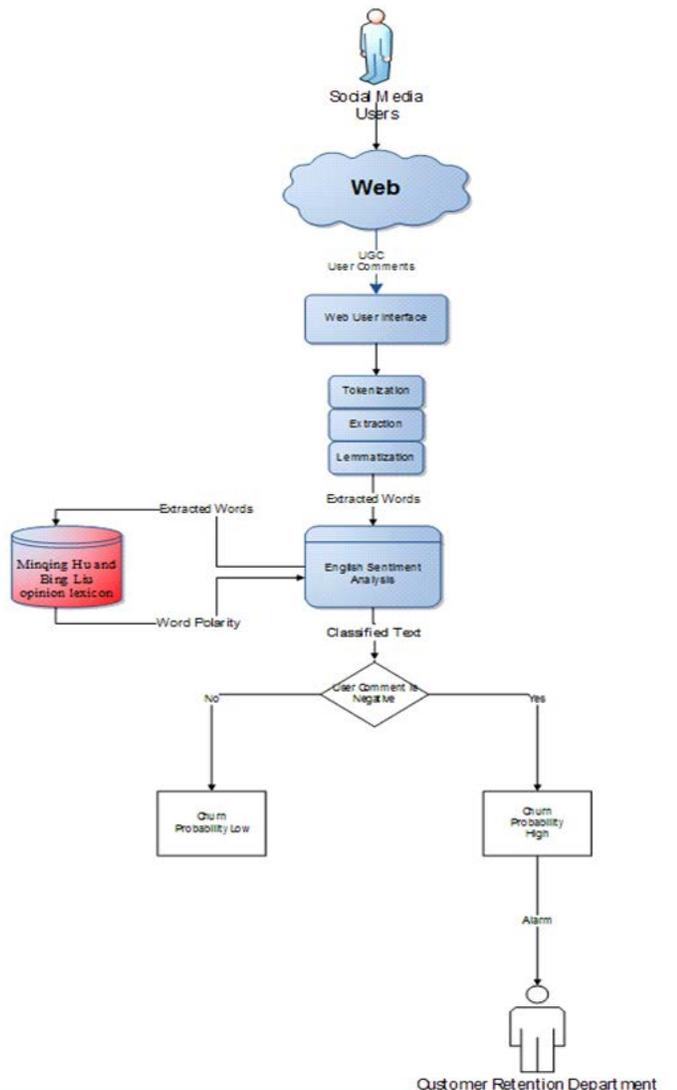


Fig 1. Proposed Sentiment Analysis Model.

Step 3: Sentiment analysis classification: The extracted English words are entered for the classification process by measuring the polarity of each word then each text is classified according to its similarity with each class (positive, negative, neutral).

Step 4: Classify a user's comment, whether positive or negative; if the comment is adverse and churn probability is high, there is an alarm sent to the customer retention department with the user's id to communicate with the customer and try to retain the customer.

Step 5: The output of the proposed model is the sentimentally classified English text.

For example, I hate this company.

The model divides this sentence into four parts then classifies each word whether positive, negative, or neutral. Table I is shown an examples for sentence classification.

A. Experiments

Firstly, we targeted online social media users' comments from the Vodafone UK page. Then we collected many comments and used sentiment analysis to analyze and classify their comments according to positive, negative, or neutral sentiment. We collected 352 comments over six months, the comments from both males and females, most of them within 20-35 years.

Table II is shown a screenshot of the comments used in sentiment analysis.

TABLE II. SOCIAL\_MEDIA\_COMMENTS

Post_ID	Comment_ID	Comment
1	1	I'd love to see how 3g works in my home
1	2	5G is not safe. So far nearly 250 EMF scientists from over 40 nations have signed this urgent appeal.
1	3	When is this coming to np13, south Wales?

Secondly, we have built our survey then distributed it among customers within four telecom companies. Each survey contains many questions in Fig. 2, such as:



Fig 2. Screenshot for the Survey.

Then, we analyzed our survey through three algorithms (DL, NB, LR) to find the dominant churn factor.

B. Results of the Experiment

Here are the results of both customers' comments and dataset analysis.

a) Results of Comments

There are some comments from what we collected from social media (i.e. Vodafone UK Facebook page) and they are shown in Table III.

TABLE III. COMMENT SENTIMENT EXAMPLE

Post_ID	User_ID	Comment	Proposed Model Classification
1	1	I'd love to see how 3g works in my home	Neutral 77.9%
1	2	5G is not safe. So far nearly 250 EMF scientists from over 40 nations have signed this urgent appeal.	Negative 51.8%
5	4	NEVER EVER use Vodafone. Liars, obstructive, and impossible to deal with.	Negative 99.7%
7	3	Vodafone is the best! 5G around the corner. Our best network provider ☐	Positive 96.0%
13	1	First, improve your customer service	Negative 92.7%
19	5	Vodafone is the best☐♥☐ network provider. Love it since 2002♥☐☐	Positive 98.4%
112	2	Great company. I've had faultless home broadband and mobile service from them.	Positive 95.4%

b) Result of Survey

And here is the result of the classified survey. Table IV is shown result by using NB algorithm.

TABLE IV. NAIVE BAYES

Attribute	Weight
Your line type	0.152981
The company tried to retain you	0.080407
age	0.067216
A problem happened for you with the company	0.056266
Gender	0.040690
use to make a complaint	0.035092
For how many years are you a customer of the company	0.034172

Table V shows result by using LR algorithm.

TABLE V. LOGISTIC REGRESSION

Attribute	Weight
Gender	0.090266
use to make a complaint	0.082144
The company tried to retain you	0.075134
For how many years are you a customer of the company	0.072974
A problem happened for you with the company	0.057011
age	0.041539
Your line type	0.005700

Table VI is shown result by using DL algorithm.

TABLE VI. DEEP LEARNING

Attribute	Weight
age	0.1332652
Your line type	0.1158231
For how many years are you a customer of the company	0.0798538
use to make complaint	0.0781901
Gender	0.0721258
A problem happened for you with the company	0.0662030
The company tried to retain you	0.0394301

*c) Result of Correlation*

After we experimented with the algorithms, we used correlation for each algorithm in order to calculate the weight of attributes with the label attribute (churn decision whether yes or no) to know the value with the highest weight which affects customers' decision to be more accurate.

$$r = \frac{(X(i)-X\prime).(Y(i)-Y\prime)}{(n-1).S(X).S(Y)} \quad (1)$$

Table VII shows a screenshot for correlation Naive Bayes table, the rest of the table in Appendix B.

TABLE VII. COR\_NAIVE\_BAYES

Attribute	Weight
A problem happened for you with the company = Withdraw from the credit without using it	0.1673900
The company tried to retain you = Offered minutes for free	0.1556922
A problem happened for you with the company = The internet was cut	0.1321921
The company tried to retain you = Offered more internet megabits	0.1321921
A problem happened for you with the company = Problem in payment plan	0.1096181
A problem happened for you with the company = The internet was very slowly	0.1096181
The company tried to retain you = Offered free gigabits	0.1096181
The company tried to retain you = Reduce the invoice	0.1096181
The company tried to retain you = By offers	0.1068224
age = 36-45	0.1027885

TABLE VIII. COR\_DEEP\_LEARNING

Attribute	Weight
A problem happened for you with the company = Withdraw from the credit without using it	0.167390
The company tried to retain you = Offered minutes for free	0.155692
A problem happened for you with the company = The internet was cut	0.132192
The company tried to retain you = Offered more internet megabits	0.132192
A problem happened for you with the company = Problem in payment plan	0.109618
A problem happened for you with the company = The internet was very slowly	0.109618
The company tried to retain you = Offered free gigabits	0.109618
The company tried to retain you = Reduce the invoice	0.109618
The company tried to retain you = By offers	0.106822
age = 36-45	0.102789

TABLE IX. COR\_LOGISTIC\_REGRESSION

Attribute	Weight
A problem happened for you with the company = Withdraw from the credit without using it	0.16739
The company tried to retain you = Offered minutes for free	0.15569
A problem happened for you with the company = The internet was cut	0.13219
The company tried to retain you = Offered more internet megabits	0.13219
A problem happened for you with the company = Problem in payment plan	0.10962
A problem happened for you with the company = The internet was very slowly	0.10962
The company tried to retain you = Offered free gigabits	0.10962
The company tried to retain you = Reduce the invoice	0.10962
The company tried to retain you = By offers	0.10682
age = 36-45	0.10279

In Table VIII, is a screenshot for correlation Deep Learning table, the rest of the table is in Appendix C.

Screenshot for the correlation Logistic Regression table is shown in Table IX, the rest of the table is in Appendix D.

*C. Experiment Results Summary*

Three main points summarize our results:

*a) UGC Analysis*

The result showed the sentiment of each comment; it divided each sentence into parts then classified the kind of sentence whether positive, negative, or neutral with percentage.

### b) Dataset Analysis

Here we can notice that the weights of attributes have differed from one algorithm to another. In Naïve Bayes, the line type of users has the highest weight while in Logistic Regression, gender has the highest weight, and in Deep Learning, the customer's age has the highest weight.

Whereas the years the customer has joined, the company has the lowest weight in Naïve Bayes. In contrast, line type of the customer has the lowest weight in the Logistic Regression, but in Deep Learning the trial of the company to retain its customer has the lowest weight.

### c) Correlation Analysis

The results of correlation shown the values affect each attribute, which has a high rate on the customer's decision, whether churn or not churns from the company. We can notice that the most effective value in 'A problem happened for you with the company' attribute was 'Withdraw from the credit without using it' in Correlation Naïve Bayes, Correlation Deep Learning, and Correlation Logistic Regression, which means it has the highest weight in customers' decision. And after it came the best way to retain customers was offered minutes for free. And as we see all algorithms have the same order of values.

## VI. CONCLUSION

The importance of this research paper comes from the importance of customer churn in the telecom sector. It helps companies to make more profit as customer churn is considered as one of the most important sources of income for the telecom sector. Hence, we build a model to analyze the behavior of customers and predict whom customers want to churn. In this study, we used Deep Learning, Naïve Bayes, and Logistic Regression algorithms for identifying the most important factors that affect the churning process. And we found the correlation for these algorithms to get the value of attribute which has the highest weight in the decision. Moreover, we analyzed the UGC by using sentiment analysis to analyze and classify customers' opinions. According to the survey and comments, we can track the churned customer to the customer retention department to retain him with the customer's id.

### REFERENCES

- [1] DOAA, M. E.; MOHAMED, H. A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*, 2016, 4.
- [2] Gerpott TJ, Rams W, Schindler A. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommun Policy*. 2001; 25:249–69.
- [3] AHMAD, Abdelrahim Kasem; JAFAR, Assef; ALJOUAAA, Kadan. Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 2019, 6.1: 28.
- [4] Wei CP, Chiu IT. Turning telecommunications call details to churn prediction: a data mining approach. *Expert Syst Appl*. 2002; 23(2):103–12.
- [5] AMIN, Adnan, et al. Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, 2019, 94: 290-301.

- [6] Qureshii SA, Rehman AS, Qamar AM, Kamal A, Rehman A. Telecommunication subscribers churn prediction model using machine learning. In: *Eighth international conference on digital information management*. 2013. P.131–6.
- [7] ULLAH, Irfan, et al. A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*, 2019, 7: 60134-60149.
- [8] Umayaparvathi V, Iyakutti K. A survey on customer churn prediction in telecom industry: datasets, methods and metric. *Int Res J Eng Technol*. 2016; 3(4):1065–70.
- [9] Yu W, Jutla DN, Sivakumar SC. A churn-strategy alignment model for managers in mobile telecom. In: *Communication networks and services research conference*, vol. 3. 2005. p. 48–53.
- [10] SALEENA, Nabizath, et al. An ensemble classification system for twitter sentiment analysis. *Procedia computer science*, 2018, 132: 937-946.
- [11] SORIANO, Domingo Ribeiro; JENG, Don Jyh-Fu; BAILEY, Thomas. Assessing customer retention strategies in mobile telecommunications. *Management Decision*, 2012.
- [12] JAHROMI, Ali Tamaddoni; STAKHOVYCH, Stanislav; EWING, Michael. Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*, 2014, 43.7: 1258-1268.
- [13] VARSHNEY, Nitish; GUPTA, S. K. Mining Churning Factors in Indian Telecommunication Sector Using Social Media Analytics. In: *International Conference on Data Warehousing and Knowledge Discovery*. Springer, Cham, 2014. p. 405-413.
- [14] VIJAYA, J.; SIVASANKAR, E.; GAYATHRI, S. Fuzzy Clustering with Ensemble Classification Techniques to Improve the Customer Churn Prediction in Telecommunication Sector. In: *Recent Developments in Machine Learning and Data Analytics*. Springer, Singapore, 2019. p. 261-274.
- [15] HOSSAIN, Md Alamgir, et al. Customer Retention and Telecommunications Services in Bangladesh. *International Journal of Asian Social Science*, 2017, 7.11: 921-930.
- [16] AMIN, Adnan, et al. Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*, 2017, 237: 242-254.
- [17] VIJAYA, J.; SIVASANKAR, E. Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector. *Computing*, 2018, 100.8: 839-860.
- [18] GAUR, Abhishek; DUBEY, Ratnesh. Predicting Customer Churn Prediction In Telecom Sector Using Various Machine Learning Techniques. In: *2018 International Conference on Advanced Computation and Telecommunication (ICACAT)*. IEEE, 2018. p. 1-5.
- [19] Abdalla, Rifaat, and Marwa Esmail. *WebGIS for Disaster Management and Emergency Response*. Springer, 2018. p. 59.
- [20] VAN DER HEIDE, E. M. M., et al. Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle. *Journal of dairy science*, 2019, 102.10: 9409-9421.
- [21] VAFEIADIS, Thanasis, et al. A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 2015, 55: 1-9.
- [22] AL IMRAN, Abdullah; AMIN, Md Nur; JOHORA, Fatema Tuj. Classification of Chronic Kidney Disease using Logistic Regression, Feedforward Neural Network and Wide & Deep Learning. In: *2018 International Conference on Innovation in Engineering and Technology (ICIET)*. IEEE, 2018. p. 1-6.
- [23] ALHARBI, Ahmed Sulaiman M.; DE DONCKER, Elise. Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. *Cognitive Systems Research*, 2019, 54: 50-61.
- [24] KOWCIKA, A., et al. Sentiment analysis for social media. *International journal of advanced research in computer science and software engineering*, 2013.

APPENDIX A

Gender \*

Female

Male

---

age \*

15-20

21-25

26-35

36-45

Above 45

---

The company of your cellular communication service \*

Orange

Vodafone

Etisalat

We

---

For how many years are you a customer of the company \*

Less than one year

1-5

5-10

More than 10

---

Your line type \*

Personal

Business

Corporate

---

you use to make your complaint \*

Social Media

Customer Service Representative

Call Center Representative

---

A problem happened for you with the company \*

The package was ended without inform me

Withdraw from the credit without using it

Rooming issues

The internet was very slowly

The internet was cut

Problem in mobile data

Stopped the service of calls or cellular data

Bad network

Cash service, the money late to send

cut the network

Way of payment for internet service

Extra money charged

Problem in the Sim card

Problem in recharging

Errors in the invoice

Anghami bundle or call tone without my acceptance

No problem happened

Problem in payment plan

---

The company tried to retain you \*

Offered more internet megabits

Return back the credit

Offered a free bundle

Offered minutes for free

I didn't mention the problem

Nothing happened

Justified the reason and apologized

Regular maintenance

Offered an internet modem for free  
 Reduce the invoice  
 Changed the Sim card from 3G to 4G  
 Went to another store  
 Offered to pay only half of my bill for 6 month  
 Not enough time  
 Corrected the invoice  
 Brought my money back  
 Offered a better plan of payment  
 Made a discount on the new package and return the amount  
 By offers  
 Come back my flex  
 Gave me new Sim card  
 Solve the problem  
 Offered free gigabits

churn (port out) \*

Yes  
 No

APPENDIX B

Attribute	Weight
A problem happened for you with the company = Withdraw from the credit without using it	0.1673900
The company tried to retain you = Offered minutes for free	0.1556922
A problem happened for you with the company = The internet was cut	0.1321921
The company tried to retain you = Offered more internet megabits	0.1321921
A problem happened for you with the company = Problem in payment plan	0.1096181
A problem happened for you with the company = The internet was very slowly	0.1096181
The company tried to retain you = Offered free gigabits	0.1096181
The company tried to retain you = Reduce the invoice	0.1096181
The company tried to retain you = By offers	0.1068224
age = 36-45	0.1027885
The company tried to retain you = Solve the problem	0.0924705
The company tried to retain you = Return back the credit	0.0850286
A problem happened for you with the company = Anghami bundle or call tone without my acceptance	0.0831868
For how many years are you a customer of the company = More than 10	0.0792394
age = 21-25	0.0790395
A problem happened for you with the company = Bad	0.0768866

network	
age = Above 45	0.0761141
A problem happened for you with the company = Rooming issues	0.0740389
A problem happened for you with the company = Stopped the service of calls or cellular data	0.0740389
For how many years are you a customer of the company = Less than 5 years	0.0692804
use to make complaint = Call Center Representative	0.0676756
age = 26-35	0.0650416
A problem happened for you with the company = Problem in recharging	0.0638076
The company tried to retain you = Justified the reason and apologized	0.0638076
The company tried to retain you = Offered a free bundle	0.0638076
The company tried to retain you = Offered to pay only half of my bill for 6 month	0.0638076
A problem happened for you with the company = Errors in the invoice	0.0524717
A problem happened for you with the company = Extra money charged	0.0518476
A problem happened for you with the company = Problem in the Sim card	0.0518476
A problem happened for you with the company = The package was ended without inform me	0.0518476
The company tried to retain you = Brought my money back	0.0518476
The company tried to retain you = Gave me new Sim card	0.0518476
The company tried to retain you = Offered a better plan of payment	0.0518476
The company tried to retain you = Offered an internet modem for free	0.0518476
A problem happened for you with the company = Way of payment for internet service	0.0364868
The company tried to retain you = Come back my flex	0.0364868
The company tried to retain you = Corrected the invoice	0.0364868
The company tried to retain you = I didn't mention the problem	0.0364868
The company tried to retain you = Made a discount on the new package and return the amount	0.0364868
The company tried to retain you = Not enough time	0.0364868
The company tried to retain you = Regular maintenance	0.0364868
The company tried to retain you = Went to another store	0.0364868
Your line type = Business	0.0328753
use to make complaint = Customer Service Representative	0.0279953

The company tried to retain you = Nothing happened	0.0125065
Your line type = Personal	0.0107071
A problem happened for you with the company = Problem in mobile data	0.0107071
For how many years are you a customer of the company = From 5 to 10 years	0.0104629
Gender = Female	0.0065408

APPENDIX C

Attribute	Weight
A problem happened for you with the company = Withdraw from the credit without using it	0.167390
The company tried to retain you = Offered minutes for free	0.155692
A problem happened for you with the company = The internet was cut	0.132192
The company tried to retain you = Offered more internet megabits	0.132192
A problem happened for you with the company = Problem in payment plan	0.109618
A problem happened for you with the company = The internet was very slowly	0.109618
The company tried to retain you = Offered free gigabits	0.109618
The company tried to retain you = Reduce the invoice	0.109618
The company tried to retain you = By offers	0.106822
age = 36-45	0.102789
The company tried to retain you = Solve the problem	0.092471
The company tried to retain you = Return back the credit	0.085029
A problem happened for you with the company = Anghami bundle or call tone without my acceptance	0.083187
For how many years are you a customer of the company = More than 10	0.079239
age = 21-25	0.079040
A problem happened for you with the company = Bad network	0.076887
age = Above 45	0.076114
A problem happened for you with the company = Roaming issues	0.074039
A problem happened for you with the company = Stopped the service of calls or cellular data	0.074039
For how many years are you a customer of the company = Less than 5 years	0.069280
use to make complaint = Call Center Representative	0.067676
age = 26-35	0.065042

A problem happened for you with the company = Problem in recharging	0.063808
The company tried to retain you = Justified the reason and apologized	0.063808
The company tried to retain you = Offered a free bundle	0.063808
The company tried to retain you = Offered to pay only half of my bill for 6 month	0.063808
A problem happened for you with the company = Errors in the invoice	0.052472
A problem happened for you with the company = Extra money charged	0.051848
A problem happened for you with the company = Problem in the Sim card	0.051848
A problem happened for you with the company = The package was ended without inform me	0.051848
The company tried to retain you = Brought my money back	0.051848
The company tried to retain you = Gave me new Sim card	0.051848
The company tried to retain you = Offered a better plan of payment	0.051848
The company tried to retain you = Offered an internet modem for free	0.051848
A problem happened for you with the company = Way of payment for internet service	0.036487
The company tried to retain you = Come back my flex	0.036487
The company tried to retain you = Corrected the invoice	0.036487
The company tried to retain you = I didn't mention the problem	0.036487
The company tried to retain you = Made a discount on the new package and return the amount	0.036487
The company tried to retain you = Not enough time	0.036487
The company tried to retain you = Regular maintenance	0.036487
The company tried to retain you = Went to another store	0.036487
Your line type = Business	0.032875
use to make complaint = Customer Service Representative	0.027995
The company tried to retain you = Nothing happened	0.012507
Your line type = Personal	0.010707
A problem happened for you with the company = Problem in mobile data	0.010707
For how many years are you a customer of the company = From 5 to 10 years	0.010463
Gender = Female	0.006541

APPENDIX D

Attribute	Weight
A problem happened for you with the company = Withdraw from the credit without using it	0.16739
The company tried to retain you = Offered minutes for free	0.15569
A problem happened for you with the company = The internet was cut	0.13219
The company tried to retain you = Offered more internet megabits	0.13219
A problem happened for you with the company = Problem in payment plan	0.10962
A problem happened for you with the company = The internet was very slowly	0.10962
The company tried to retain you = Offered free gigabits	0.10962
The company tried to retain you = Reduce the invoice	0.10962
The company tried to retain you = By offers	0.10682
age = 36-45	0.10279
The company tried to retain you = Solve the problem	0.09247
The company tried to retain you = Return back the credit	0.08503
A problem happened for you with the company = Anghami bundle or call tone without my acceptance	0.08319
For how many years are you a customer of the company = More than 10	0.07924
age = 21-25	0.07904
A problem happened for you with the company = Bad network	0.07689
age = Above 45	0.07611
A problem happened for you with the company = Roaming issues	0.07404
A problem happened for you with the company = Stopped the service of calls or cellular data	0.07404
For how many years are you a customer of the company = Less than 5 years	0.06928
use to make complaint = Call Center Representative	0.06768
age = 26-35	0.06504
A problem happened for you with the company = Problem in recharging	0.06381
The company tried to retain you = Justified the reason and apologized	0.06381

The company tried to retain you = Offered a free bundle	0.06381
The company tried to retain you = Offered to pay only half of my bill for 6 month	0.06381
A problem happened for you with the company = Errors in the invoice	0.05247
A problem happened for you with the company = Extra money charged	0.05185
A problem happened for you with the company = Problem in the Sim card	0.05185
A problem happened for you with the company = The package was ended without inform me	0.05185
The company tried to retain you = Brought my money back	0.05185
The company tried to retain you = Gave me new Sim card	0.05185
The company tried to retain you = Offered a better plan of payment	0.05185
The company tried to retain you = Offered an internet modem for free	0.05185
A problem happened for you with the company = Way of payment for internet service	0.03649
The company tried to retain you = Come back my flex	0.03649
The company tried to retain you = Corrected the invoice	0.03649
The company tried to retain you = I didn't mention the problem	0.03649
The company tried to retain you = Made a discount on the new package and return the amount	0.03649
The company tried to retain you = Not enough time	0.03649
The company tried to retain you = Regular maintenance	0.03649
The company tried to retain you = Went to another store	0.03649
Your line type = Business	0.03288
use to make complaint = Customer Service Representative	0.02800
The company tried to retain you = Nothing happened	0.01251
Your line type = Personal	0.01071
A problem happened for you with the company = Problem in mobile data	0.01071
For how many years are you a customer of the company = From 5 to 10 years	0.01046
Gender = Female	0.00654

# Oscillation Preventing Closed-Loop Controllers via Genetic Algorithm for Biped Walking on Flat and Inclined Surfaces

Sabri Yılmaz<sup>1</sup>, Metin Gokasan<sup>2</sup>, Seta Bogosyan<sup>3</sup>

Department of Control and  
Automation Engineering  
Istanbul Technical University  
Istanbul, Turkey 34398

**Abstract**—In this study, a closed-loop controller is designed to overcome the dynamical insufficiency of the 3D Linear Inverted Pendulum Model (LIPM) via the Genetic Algorithm (GA). The main idea is to still use the 3D LIPM with a closed-loop controller because of its ease at modeling. While suppressing the dynamical flaws only the legs are used, in other words a robot is used which does not have any upper body elements to have a more modular robot. For this purpose, a biped is modeled with the 3D LIPM which is one of the most famous modeling methods of humanoid robots for the ease of modeling and fast calculations during the trajectory planning. After obtaining the simple model, Model Predictive Control (MPC) is applied to the 3D LIPM to find the reference trajectories for the biped while satisfying the Zero Moment Point (ZMP) criteria. The found reference trajectories applied to the full dynamical model on Matlab Simulink and the real biped in the laboratory at Istanbul Technical University. From the simulation results on the flat and inclined surfaces and real-time experiments on a flat surface some dynamical flaws are observed due to the simple modeling. To overcome these flaws a Proportional-Integral (PI) controller is designed, and the optimal value of the controller gains are found by the GA. The results assert that the designed controller can overcome the observed flaws and makes biped move more stable, smoother, and move without steady-state error.

**Keywords**—Humanoid robot; biped walking; Model Predictive Control (MPC); Genetic Algorithm (GA); trajectory planning; Zero Moment Point (ZMP); linear inverted pendulum

## I. INTRODUCTION

In recent years, with the growing interest in humanoid robots, they have been started replacing humans in hazardous environments. Although the use is growing with the interest, there are also lots of difficulties to imitate human-like movements [1, 2]. One of the most interesting movements of the humans is locomoting on two legs since this movement can adapt itself for flat or inclined surfaces, even if uneven surfaces. A stable walking can be described as walking without falling on the ground.

To maintain stability during walking, the Visual Simultaneous Localization and Mapping, also known as Visual SLAM, has been used in [3] to estimate the Center of Mass (CoM), while Zero-Moment Point (ZMP) is measured via force/torque sensors. The increased sensing and computational

load results in a promising performance, also tested under push and perturbations. Moving the torso to maintain stability [4] is another popular approach, extending the stability control problem to that of the humanoid body [5]. Those studies do address the dynamical flaws of control approaches using the simple 3D Linear Inverted Pendulum (LIPM) but at the cost of an increased number of sensors, computational load, and system complexity.

In this study, we aim to improve the walking performance of a biped further exhausting the capabilities of the 3D LIPM based simpler control approaches. The proposed method combines the approach in [6] developed for uneven and inclined surfaces, and the method in [7] based on the kinematic resolution of CoM, and also compensated for some of the dynamical deficiencies of the 3D LIPM. The trajectory generation is performed with the ZMP, but unlike other studies, such as [8] and [9], our objective function takes the ZMP into account as a constraint and aims to minimize the hip tracking error, but not the ZMP error. The justification of this approach is that the derivation of ZMP uses approximations, while the hip point can be derived more accurately by conventional Jacobian kinematics. Simulation results show that the objective function defined at [8] and [9] causes biped to oscillate without a closed-loop controller. Another novelty of this study is the consideration of a biped system alone in the development of improved walking performance and stability, without any compensation coming from the increased number of DoFs of the torso or rest of the humanoid body as in the above-mentioned studies. The use of the simplified model still gives rise to some oscillations at the hip during walking, and these oscillations are eliminated with the use of a simple feedback control, the coefficients of which were determined by Genetic Algorithm (GA) which is a benchmark optimization algorithm used at various areas such as redundant robots [10], tuning of controller parameters [11]. The success of the proposed method is shown both on simulation results on the flat and inclined surfaces and real-time experimental results on a flat surface.

This paper is organized as follows. The second section provides background. In the third section modeling of the biped as a 3D LIPM is explained and the relation between the pendulum and the biped with 12 DoF at Istanbul Technical University is given. In the fourth section, the concept of MPC

is explained, and the results of the simulation with MPC are presented with discussions. In the fifth section, the proposed closed-loop control method, and the developed GA is described and tested with simulation results. In the sixth section, real-time experimental results are provided, with final discussions in the conclusion section.

## II. BACKGROUND

In order to maintain a stable walking, humanoids must meet some stability criteria. One of the most popular concepts of stable walking is ZMP [12]. ZMP is the point, where the normal forces caused by the movement of the humanoid do not produce any moment, hence, this point is concurrent with the center of mass when the robot is inactive. Consequently, keeping the ZMP inside the Support Polygon (SP) of the humanoid during the locomotion guarantees the balance of the humanoid [13].

On the other hand, the exact derivation of the ZMP is a complex task. In order to obtain ZMP easily, approximate dynamical models have been investigated. Because of its simplicity in the representation of ZMP, the most commonly used approximate model is the 3D LIPM [14]. This model provides a reasonable ZMP position while the humanoid is walking and can be used for ZMP trajectory planning, but it is too simple to reflect all the dynamical properties of a humanoid; e.g. it does not contain any relation between the foot and the ground. In some studies, the actuator positions are used for the calculation of CoM [15], but this function does not take contact forces into consideration, hence, it cannot suppress the disturbances associated with contact forces. By position feedback of the joints only joints' reference tracking control can be done without any reaction with the floor, e.g. if the biped is on an inclined surface the position feedback of the joints cannot reflect the inclination of the surface.

The simplicity of the 3D LIPM model has opened the path for the use of Model Predictive Control as the 'trajectory planner' for walking in several studies. MPC allows for real-time implementation of optimal control principles and has gained increasing popularity in many areas, from numerous automotive applications [16-18] to pH neutralization process [19]. It is also a suitable control method for trajectory planning of humanoid robots since the ZMP can be defined by a simple model and the online optimization problem can be solved sufficiently fast, while ensuring the constraints on ZMP. This aspect of MPC is recognized by many researchers in humanoids, starting with the design of a preview for the ZMP to generate the walking patterns [8]. After preview controllers, [9] redefined the trajectory generation problem with the constraints on the ZMP using simulations on a humanoid model. The study also considers push recovery similar to several other studies, such as [20, 21].

## III. MODELING OF THE BIPED

In this section, the modeling of the biped with the 3D LIPM will be discussed and the expressions of the 3D LIPM will be given.

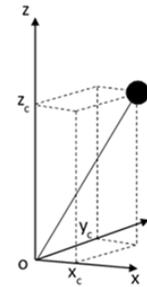


Fig. 1. The 3D LIPM.

3D LIPM can be used to compute the ZMP simply. There is a point mass is accepted to be concentrated on the tip of the 3D LIPM and the pendulum is accepted to be massless. Since the pendulum is 3D, the same equations of motion can be used for modeling the pendulum both on the x-axis and y-axis. In this study 3D LIPM is used for modeling the biped for calculating the ZMP. Fig. 1 shows the isometric view of the 3D LIPM.

The dynamic model of 3D LIPM can be stated in matrix form by taking jerks of the mass that is concentrated on the tip of the 3D LIPM  $\ddot{x}_c$  and  $\ddot{y}_c$  as control inputs.

$$\frac{d}{dt} \begin{bmatrix} x_c \\ \dot{x}_c \\ \ddot{x}_c \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ \dot{x}_c \\ \ddot{x}_c \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \ddot{x}_c \quad (1)$$

$$x_{zmp} = \begin{bmatrix} 1 & 0 & -\frac{z_c}{g} \end{bmatrix} \begin{bmatrix} x_c \\ \dot{x}_c \\ \ddot{x}_c \end{bmatrix}$$

$$\frac{d}{dt} \begin{bmatrix} y_c \\ \dot{y}_c \\ \ddot{y}_c \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_c \\ \dot{y}_c \\ \ddot{y}_c \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \ddot{y}_c \quad (2)$$

$y_{zmp} = \begin{bmatrix} 1 & 0 & -\frac{z_c}{g} \end{bmatrix} \begin{bmatrix} y_c \\ \dot{y}_c \\ \ddot{y}_c \end{bmatrix}$

Here  $x_c$  and  $y_c$  are the linear positions,  $\dot{x}_c$  and  $\dot{y}_c$  are the linear velocities and  $\ddot{x}_c$  and  $\ddot{y}_c$  are the linear accelerations of the 3D LIPM. As stated, before  $\ddot{x}_c$  and  $\ddot{y}_c$  are the jerks of the 3D LIPM.  $x_{zmp}$  and  $y_{zmp}$  which are the linear positions of the ZMP on the x-axis and y-axis respectively, can be written in terms of the three states. Here  $z_c$  is the linear position of the 3D LIPM on the z-axis, which is the height of the 3D LIPM and  $g$  is the acceleration due to the gravitational forces. From now on the height of the 3D LIPM  $z_c$  will be taken as constant  $h$ .



Fig. 2. Solidworks Drawing of the Biped with Pendulum.

Fig. 2 shows Solidworks drawing of the biped with the pendulum. The point mass of the pendulum is located on the hip of the biped so that the CoM of the biped is accepted to be located on the hip.

#### IV. MODEL PREDICTIVE CONTROL

##### A. Preview of the Model Predictive Control

MPC is an optimal control algorithm that uses a model to make predictions about future outputs of a process while satisfying inequality constraints on the input and output variables. MPC can be used for controlling the multi-input multi-output systems. In order to control a system, MPC needs a reasonable accurate model of the system and MPC solves an online optimization problem to find the best control action that makes the output follow the reference.

Fig. 3 shows the block diagram of the 3D LIPM that is controlled by a Linear MPC during this study. By using the MPC, the reference trajectories for the hip of the biped can be produced while ensuring the constraints on ZMP. As stated in Section 3 3D LIPM is transformed into a linear system by taking the height of the pendulum as constant.

In this section all the equations for MPC will be derived according to equations, derived by Wieber [9] for only x-axis as can be seen from the Equations (1) and (2) in Section 3, the derivations of the position of ZMP for x and y axes are analogical to each other. By taking the height of the 3D LIPM as constant, the output equation turns into a linear equation from a nonlinear equation. Equations (1) and (2) can be discretized by trivial integration. With trivial integration, the relation of the next states with the current states and control signal can be written as follows:

$$\hat{X}_{c_{k+1}} = \underbrace{\begin{bmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix}}_A \hat{X}_{c_k} + \underbrace{\begin{bmatrix} \frac{T^3}{6} \\ \frac{T^2}{2} \\ \frac{T}{2} \end{bmatrix}}_B \ddot{x}_{c_k} + u_k \quad (3)$$

$$y_{c_k} = \underbrace{[1 \ 0 \ 0]}_C \hat{X}_{c_k}$$

$$x_{zmp_k} = \underbrace{\begin{bmatrix} 1 & 0 & -\frac{h}{g} \end{bmatrix}}_{C_{zmp}} \hat{X}_{c_k} \quad (4)$$

Here  $\hat{X}_{c_k}$  is the state vector at  $k^{th}$  step,  $\ddot{x}_{c_k}$  is the control signal at  $k^{th}$  step and  $x_{zmp_k}$  is the ZMP position at  $k^{th}$  step.

The constraint on the position of ZMP for a stable walking at  $k^{th}$  step can be defined as follows, where  $x_{zmp_k}^{min}$  the minimum allowed value of the ZMP and  $x_{zmp_k}^{max}$  the maximum allowed value of the ZMP at  $k^{th}$  step:

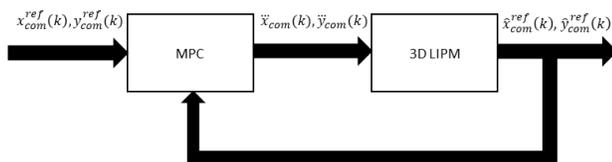


Fig. 3. The 3D LIPM Controlled by a Linear MPC.

$$x_{zmp_k}^{min} \leq x_{zmp_k} \leq x_{zmp_k}^{max} \quad (5)$$

The main purpose of the optimization problem is to find all  $\ddot{x}_{c_i}$ , those minimize the cost function  $J$  stated below. Here iterating the Quadratic Program (QP) by finite  $N$  times which is the prediction horizon, allows solving the Optimal Control problem analytically through some simple matrix manipulations instead of having to solve a more complex algebraic Riccati Equation. With a difference to Kajita's and Wieber's proposed objective functions, here the aim is to minimize the tracking error of the CoM while minimizing the jerks, instead of minimizing the tracking error of the ZMP [8, 9]. Because as seen from the Equation (5) optimization problem guarantees that the ZMP will stay inside the SP. Since the CoM of the 3D LIPM is accepted to be on the hip of the biped, the kinematic relation between the feet soles and the hip can be expressed directly with Jacobian Kinematics. Equation (6) shows the objective function on both the x-axis and y-axis defined in this study. Here  $q_x$  and  $q_y$  are the weight values of the tracking error and  $r_x$  and  $r_y$  are the weight values of the control signal.

$$J = \frac{1}{2} \min_{\ddot{x}_{c_k}, \dots, \ddot{x}_{c_{k+N}}, \ddot{y}_{c_k}, \dots, \ddot{y}_{c_{k+N}}} \sum_{i=k}^{k+N-1} \left[ q_x (x_{c_{i+1}}^{ref} - x_{c_{i+1}})^2 + \frac{1}{2} r_x \ddot{x}_{c_i}^2 \right] + \left[ q_y (y_{c_{i+1}}^{ref} - y_{c_{i+1}})^2 + \frac{1}{2} r_y \ddot{y}_{c_i}^2 \right] \quad (6)$$

By using the recursive relation iterated  $N$  times, all the relation between the jerks and coordinates of ZMP can be defined as follows:

$$\begin{bmatrix} x_{zmp_{k+1}} \\ \vdots \\ x_{zmp_{k+N}} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & T & \frac{T^2}{2} - \frac{z_c}{g} \\ \vdots & \vdots & \vdots \\ 1 & NT & \frac{N^2 T^2}{2} - \frac{z_c}{g} \end{bmatrix}}_{P_x} \hat{X}_{c_k} + \underbrace{\begin{bmatrix} \frac{T^3}{6} - \frac{Tz_c}{g} & 0 & 0 \\ \vdots & \ddots & 0 \\ \frac{(1+3N+3N^2)T^3}{6} - \frac{Tz_c}{g} & \dots & \frac{T^3}{6} - \frac{Tz_c}{g} \end{bmatrix}}_{P_u} \begin{bmatrix} \ddot{x}_{c_k} \\ \vdots \\ \ddot{x}_{c_{k+N-1}} \end{bmatrix} \quad (7)$$

Equation 7 can be shown in compact form as follows:

$$x_{zmp_{k+1}} = P_x \hat{X}_{c_k} + P_u U_k \quad (8)$$

##### B. Simulation Results of the Linear Model Predictive Control

In this section, Matlab Simulink simulation results of the biped are given and discussed. In order to use the Linear MPC some parameters need to be defined before simulations. These are the number of states, number of outputs, number of the control signals, sampling period, prediction horizon, control horizon, initial conditions of the states and the control signals, and weights at the optimization problem. There are 3 states, 1 output and 1 control signal for each axis. The sampling period is chosen to be 0.01 s. The prediction horizon is 150 and the control horizon is 16 steps. The weights' ratios are selected as

1000 both on x and y axes. All the initial conditions are set to 0.

Fig. 4 shows the general block diagram of the system. As stated, before MPC produces jerk inputs for the 3D LIPM while satisfying constraints on the ZMP locations both on the x-axis and y-axis. While 3D LIPM is tracking the reference trajectories, it also produces reference positions for the biped's CoM, which is accepted to be on the hip of the biped as the point mass of the pendulum. The reference positions will be evaluated at the swing leg and stance leg selector block. Mainly it evaluates the reference positions for the swing leg and stance leg with respect to phases of cyclic walking and foot positions. Finally, this block derives reference velocities for the biped's hip point. The evaluated reference velocities will be applied to inverse kinematics function and this function produces angular velocity references for 12 DoF of the biped. These angular velocities are derived and integrated to find the angular accelerations and angular positions. The angular accelerations, velocities, and positions are applied to the biped.

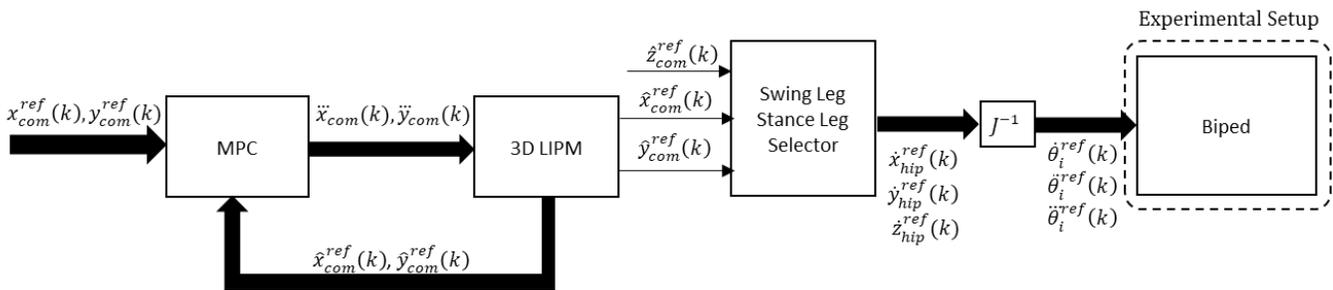


Fig. 4. The 3D LIPM Controlled by a Linear MPC.

Fig. 5 shows the biped's final position after 20 steps. Here the magenta line shows the trajectory of the CoM and the green line shows the projection of the CoM. When the figure is examined, it can be seen that the biped is moving in positive y direction also, although the only position change is expected at x-direction when walking linearly.

Fig. 6 shows the biped's final position and the trajectory of the CoM on xy-plane. The projection of the trajectory of the CoM overlaps with itself, so there is only one line that can be seen in the figure. Also, the movement on positive y-direction can be seen clearly as a result of the oscillation of the biped during this movement. When one leg gets off the ground and becomes the swinging leg, so the biped is at the Single Support Phase (SSP). At this time the support polygon reaches its smallest surface area and the walking becomes less robust as stated before. Also, the glitches on the trajectory of the CoM can be seen from the figure.

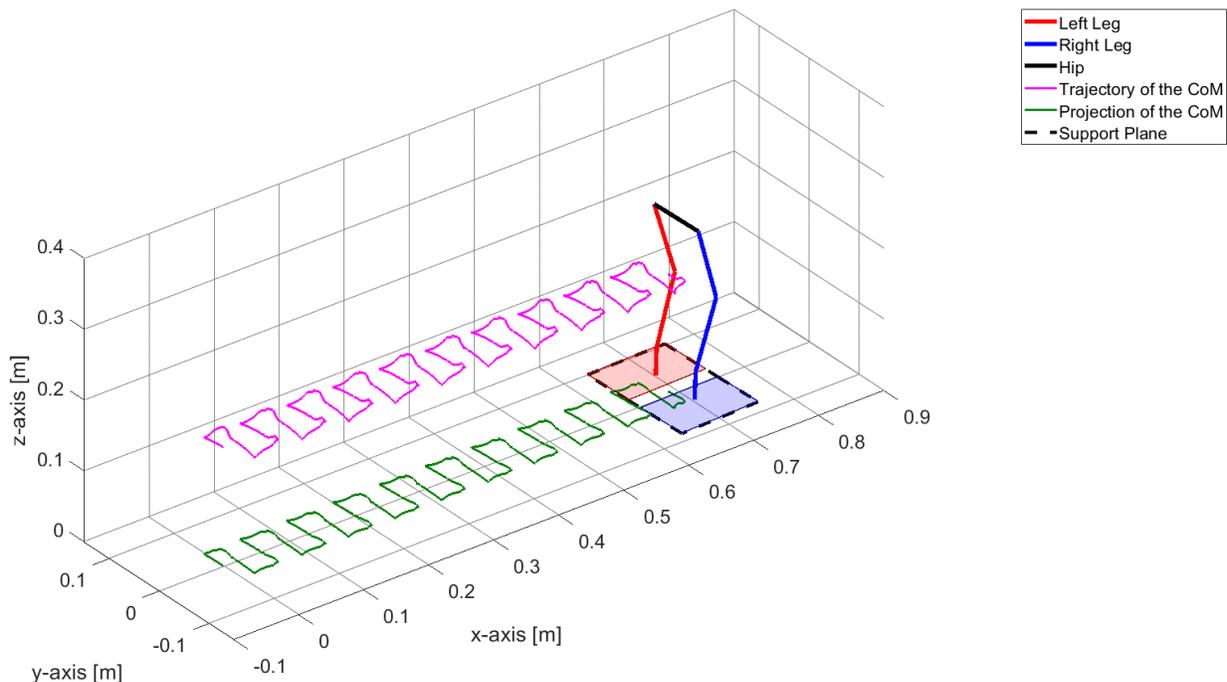


Fig. 5. The Final Position of the Stick Model and CoM Trajectories.

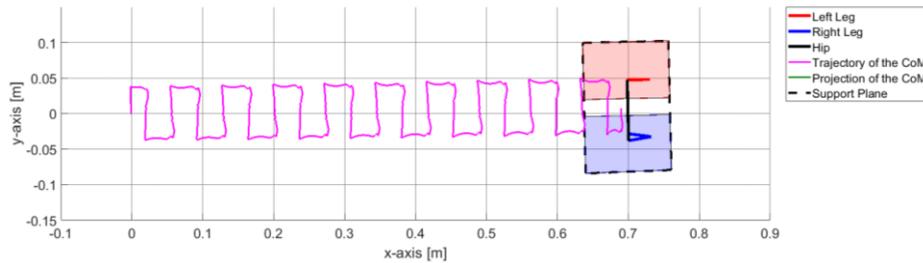


Fig. 6. The Final Position of the Stick Model and CoM Trajectories on xy Plane.

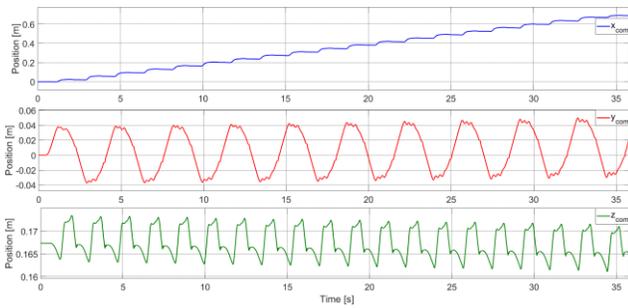


Fig. 7. Trajectories of CoM.

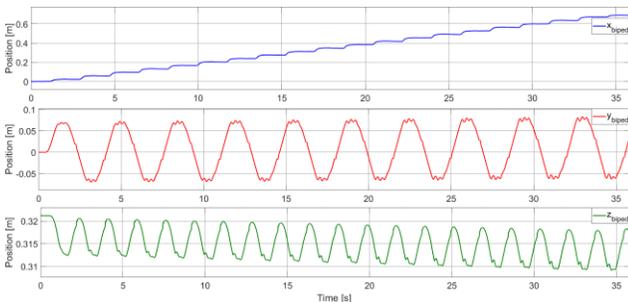


Fig. 8. Trajectories of the Hip.

Fig. 7 shows the trajectory of the CoM on all x, y, and z axes. The first one is the trajectory on the x-axis, the second one is the trajectory on the y-axis and the third one is the trajectory on the z-axis. From the graph of the trajectory on the y-axis, it can be seen that when the biped is at its maximum distance from the middle position on the y-axis, it starts oscillating. In addition to this, as time progresses the offset of the trajectory on the y-axis shifts to positive values.

Fig. 8 shows the trajectory of the hip, -in other words the 3D LIPM-. The figure is arranged the same as the previous figure and again it shows the offset change in the y-axis.

Fig. 9 shows the trajectories on the x-axis. The blue line is the reference trajectory that is defined for the 3D LIPM. The red line is the trajectory of the 3D LIPM or in other words output of the closed-loop system with Linear MPC. The green line is the trajectory of the biped's hip and the magenta line is the trajectory of the calculated ZMP. It can be seen that the 3D LIPM can follow the reference trajectory without any steady-state error, however, the biped cannot follow this trajectory. ZMP has peaks at the beginnings and the endings of the steps because of the inertia of the biped.

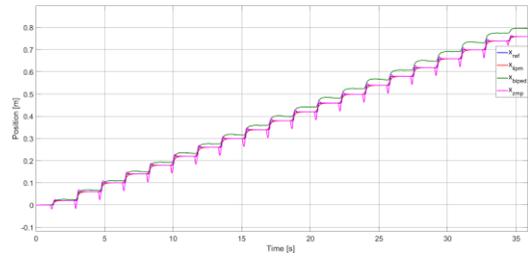


Fig. 9. The Trajectories of the Biped on the x-axis.

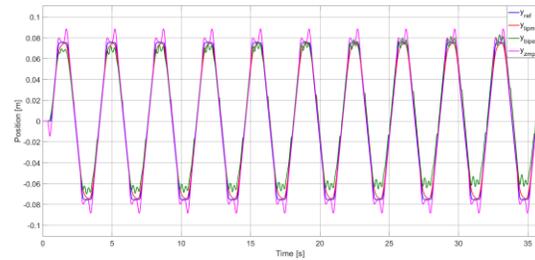


Fig. 10. The Trajectories of the Biped on the y-axis.

Fig. 10 shows the trajectories on the y-axis. The specifications of the figure are the same as the previous figure, the blue line is the reference trajectory, the red line is the trajectory of the 3D LIPM, the green line is the trajectory of the biped's hip and the magenta line is the trajectory of the ZMP. From the figure it can be seen that the 3D LIPM cannot track the reference well or in other words tracks the reference slowly, but it has a smooth movement. The biped has a steady-state error that can be seen from the offsets and also oscillates when it is on the limits on the y-axis. The ZMP has peaks, nevertheless has no error because it is a calculated value, not a measured value as stated before.

Fig. 11 and Fig. 12 show the trajectories of the biped with objective functions defined by Kajita [8] and Wieber [9] on the x-axis and y-axis respectively. The blue lines are the reference trajectories, red lines are the 3D LIPM trajectories, green lines are the biped's hip trajectories and magenta lines are the ZMP trajectories. The main difference is the ZMP trajectories if compared to the objective function that is defined in this study. The oscillations of the biped on SSP can be seen from the figures although the optimization problem tries to minimize the ZMP tracking error. These oscillations cause biped to have steady-state errors on both axes. Figures show that whether the optimization problem tries to minimize the CoM trajectory error or the ZMP trajectory error, the biped oscillates.

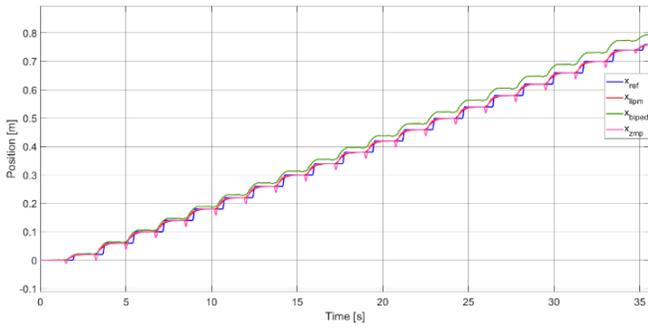


Fig. 11. The Trajectories of the Biped on the x-axis.

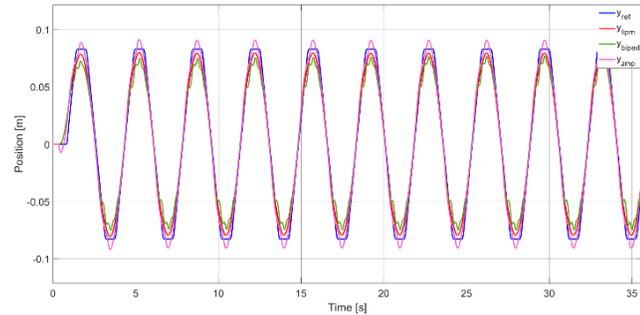


Fig. 12. The Trajectories of the Biped on the y-axis.

## V. CLOSED-LOOP CONTROLLER VIA GENETIC ALGORITHM

In this section the proposed closed-loop control method and the search for the optimal controller gains are explained.

### A. Closed-loop Controller

3D LIPM is one of the most used models for modeling a biped and deriving the ZMP definition as stated before. However, this ease at modeling and less calculation load, the 3D LIPM has some dynamical flaws. The first one is the concentrated mass at the tip of the inverted pendulum, which does not reflect the change at the position of real CoM of the biped during the movement. The biped, which used in this study has 12 motors, and these motors' weights and inertia tensors are much bigger than the weights and inertia tensors of the links, so the links can be negligible during the dynamical analysis. The real CoM of the biped and ZMP of the biped are the functions of these motors' positions and accelerations so the results of these functions change during the movement with respect to positions and accelerations of each motor. These changes cannot be expressed by a simple model. So, although the 3D LIPM tracks the reference trajectory well during the movement, the biped cannot track its reference as successful as the 3D LIPM.

SSP is the least robust phase of the walking because the support polygon has its smallest surface area, which is the equivalent area of the support foot's projection on the ground. And also, while the support polygon has its smallest area, one leg of the biped is swinging and the dynamics of the biped are changing roughly, which cannot be expressed by the 3D LIPM.

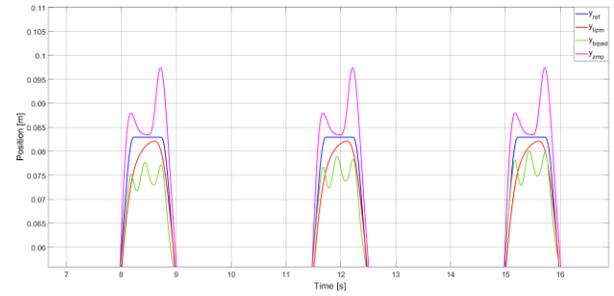


Fig. 13. Detailed view of Trajectories on the y-axis.

From Fig. 13 it can be clearly seen that the biped is oscillating on the y-axis. These oscillations do not cause the biped to fall during this study, but at higher walking speeds they can cause the biped to fall to the ground. Even if biped can still walk it has steady-state errors on both x-axis and y-axis as stated before. These errors can also cause biped to walk into undesired locations.

In order to get rid of these oscillations a closed-loop control method based on a PI controller is searched. The tracking error on y-axis stated as follows:

$$e_y = \hat{y}_{com}^{ref} - y_{hip} \quad (9)$$

Here  $\hat{y}_{com}^{ref}$  is the position of the 3D LIPM on y-axis and  $y_{hip}$  is the position of the biped on the y-axis. The position reference of the biped is the position output of the 3D LIPM. The results of the simulations are examined, and the tracking error is observed at SSP as expected, so this correction must be applied during the SSP. Because of the discontinuity applying this correction as a square wave makes the system unstable. To get rid of instability, the correction must be applied as a sine waveform, so the error must be modulated. In order to modulate the error signal, it is multiplied by the movement of the swinging leg on the z-axis. The proposed PI-controller can be expressed as follows:

$$\hat{y}_{com}^{comp} = K_p \hat{z}_{com}^{ref} e_y + K_i \int \hat{z}_{com}^{ref} e_y dt \quad (10)$$

Here  $\hat{z}_{com}^{ref}$  is the position of the swinging foot on z-axis and  $e_y$  is the tracking error on the y-axis. Fig. 14 shows the detailed block diagram of the biped system with PI-controller and 3D LIPM with Linear MPC. The red blocks are the added blocks to make the biped system as a closed-loop system. An accelerometer is added on the hip of the biped to measure the accelerations both on three axes. The measured accelerations of the hip are processed through the position estimator block in order to estimate the position of the hip on the y-axis. This block mainly filters the accelerations and integrates the positions from accelerations. After finding the position of the hip on the y-axis the Equations (9) and (10) implemented by the added blocks.

The optimal values of the controller gains  $K_p$  and  $K_i$  will be searched by the GA in the next section.

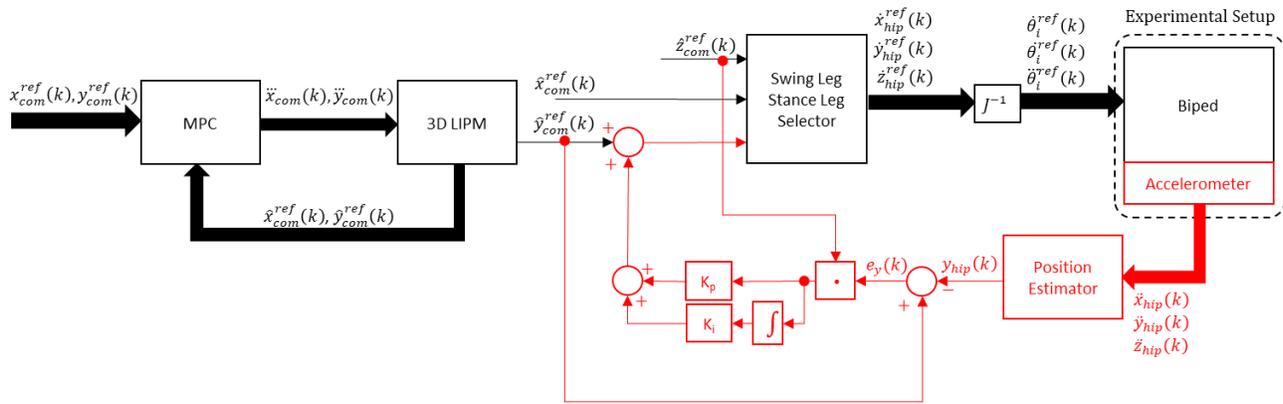


Fig. 14. Detailed Block Diagram of the Biped with PI-Controller.

### B. Genetic Algorithm

The GA is a random based search algorithm based on the theory of natural evolution. It can be used on both constrained and unconstrained optimization problems. Like most of the other optimization methods, GA starts with a population size which is the number of solutions. At every step, GA selects individual solutions randomly from the population and as chromosomes the selected individuals produce a new generation. These newly produced generations inherit the characteristics of the parents as expected. As in evolution, the generations produced from high-quality parents, expected to have better characteristics. In an optimization problem these characteristics can be named as a fitness function. After iterations with successive generations, the population ends up with an optimal solution.

The fitness function is identified as the integral square error on the y-axis and defined as follows:

$$J = \int_0^t (y_{com}^{ref} - y_{hip})^2 dt \quad (11)$$

Here the error is defined as the tracking error of the biped on the y-axis.  $y_{com}^{ref}$  is the position of the 3D LIPM on y-axis

and  $y_{hip}$  is the position of the biped on the y-axis. The population size is selected as 50, the crossover is selected as 0.8. The optimal gain values  $K_p$  and  $K_i$  for 20 steps of walking are found as 265.31 and 115.33 respectively. The cost function value is found as 0.0000801 after 34 iterations.

### C. Simulation Results of the Linear MPC with PI Controller

Fig. 15 shows the biped's final position after 20 steps. Here the magenta line shows the trajectory of the CoM and the green line shows the projection of the CoM. As a result, no anomaly can be seen in the figure. For example, without the PI controller, biped was changing position's offset on the y-axis and resulting as an anomaly in the figure. It seems that the PI controller can overcome this steady-state error on the y-axis.

Fig. 16 shows the biped's final position and the trajectory of the CoM on xy-plane. Again, the trajectory of the CoM and the trajectory of the projection of the CoM overlap with each other and there is only one trajectory line that can be seen in the figure. It can be clearly observed that there is no offset change on the y-axis in this figure.

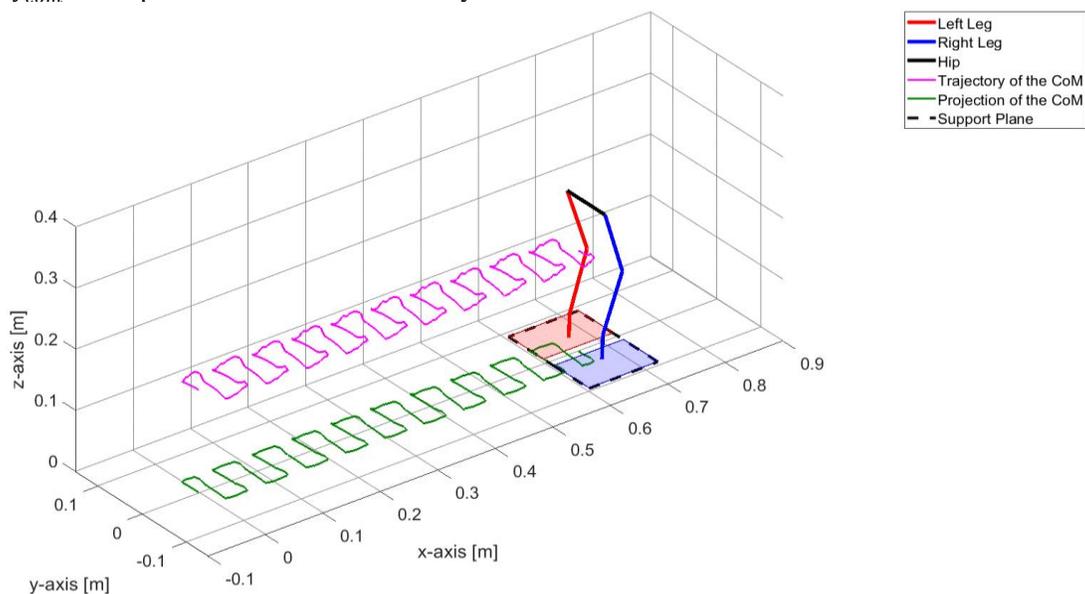


Fig. 15. The Final Position of the Stick Model and CoM Trajectories.

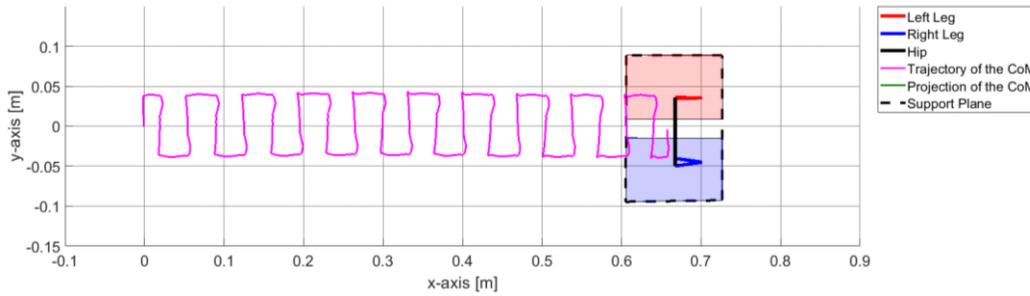


Fig. 16. The Final Position of the Stick Model and CoM Trajectories on xy plane.

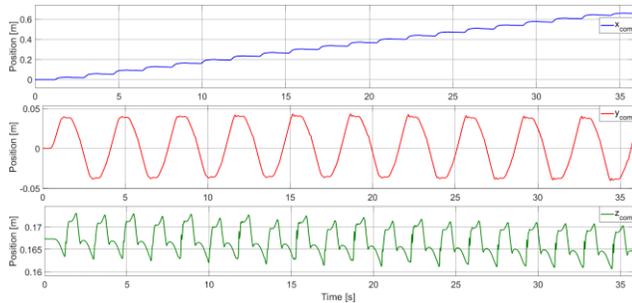


Fig. 17. Trajectories of CoM.

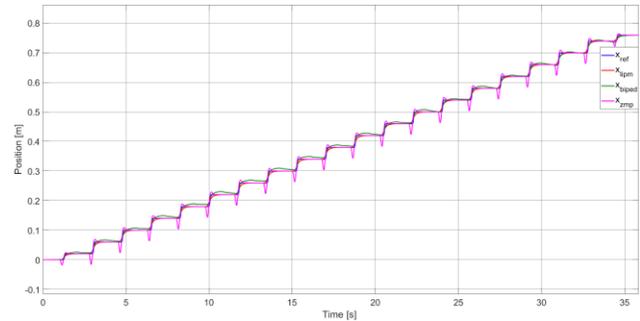


Fig. 19. The Trajectories of the Biped on the x-axis.

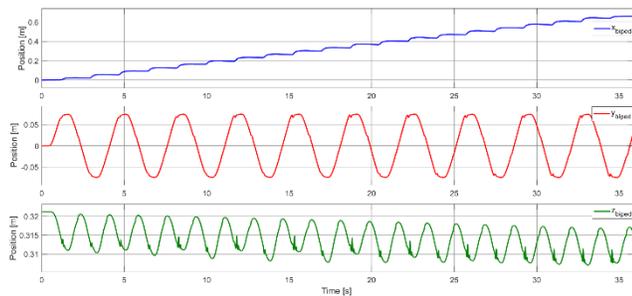


Fig. 18. Trajectories of the Hip.

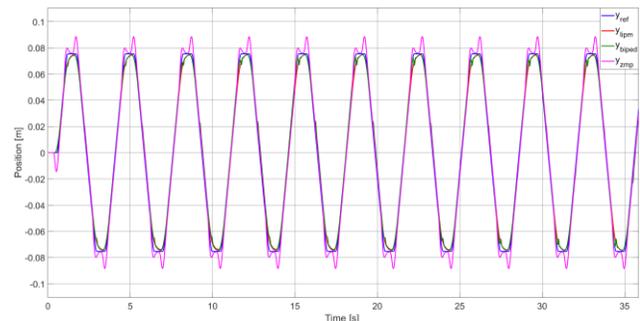


Fig. 20. The Trajectories of the Biped on the y-axis.

Fig. 17 shows the trajectory of the CoM on all x, y, and z axes, from up to down. When the y-axis is examined, it can be seen that there is no offset change in both directions and also there are fewer oscillations on the figure.

Fig. 18 shows the trajectories of the hip of the biped. The arrangement of the figure is the same as the previous one. Again, there is no offset change in y-direction for the CoM, so it can be explained that the biped is walking linearly on the x-axis without any slipping in the y-axis.

Fig. 19 shows the trajectories on the x-axis. The blue line is the reference trajectory to the 3D LIPM, the red line is the trajectory output of the 3D LIPM, the green line is the hip trajectory of the biped and the magenta line is the calculated ZMP trajectory. 3D LIPM can follow the reference trajectory since there is no addition to the Linear MPC used in the previous section. Additionally, the biped tracks the output of the 3D LIPM without any error if compared with the previous section. ZMP has peaks at the beginnings and the endings of the steps because of the inertia of the biped.

Fig. 20 shows the trajectories on the y-axis with the same order as the previous figure. Again, the 3D LIPM tracks the reference trajectory as in the previous section, but here the biped can track the reference not without any, but with a little oscillation. The ZMP has peaks on the x-axis.

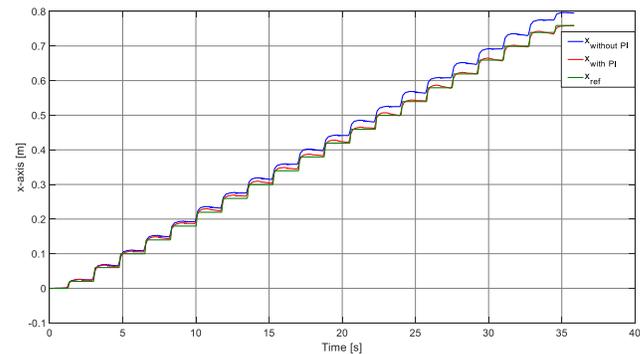


Fig. 21. Comparison of the Open and Closed-Loop System on the x-axis.

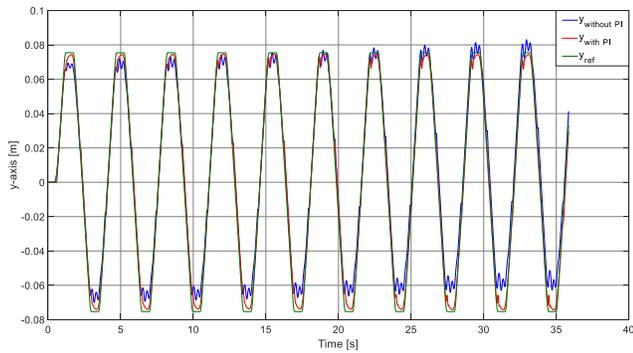


Fig. 22. Comparison of the Open and Closed-Loop System on the y-axis.

Fig. 21 shows the comparison between the open-loop biped system and the closed-loop biped system on the x-axis. As mentioned before there is a PI controller has been added to the system mainly to overcome the oscillations of the biped when it is on SSP. However, it can be seen that the biped also slips on positive x-axis and positive y-axis, because of these oscillations. By inspecting the figure on the x-axis, the red line which is the biped's trajectory with the closed-loop controller can track the green line, which is the reference trajectory without an error, but the blue line which is the biped's trajectory without the PI controller cannot track the reference. It can be understood that these oscillations when the biped is on a less robust phase can cause the biped slip and by overcoming these oscillations biped can follow the trajectory on the sagittal plane without an error.

Fig. 22 shows the effects of the PI controller on the y-axis. The blue line is the biped's trajectory without the PI controller, the red line is the biped's trajectory with the PI controller and the green line is the reference trajectory of the biped. It can be seen that biped cannot track the reference without the PI controller and the biped is oscillating when it's on the SSP. Additionally, the amplitude of the movement on the y-axis is not equal to the amplitude of the reference. As time progresses, the biped is moving on the positive y-axis resulting in an offset

error. When the red line is examined, there are very few oscillations compared to the blue line and it can also be observed that there is neither an offset error nor an amplitude error on the y-axis with the PI controller.

As a final result of all these progression steps, it can be said that the PI controller is suggested for preventing the biped from oscillating when it is on SSP. Unless the model overcomes the oscillations, these will result in slipping on both the x-axis and the y-axis at normal walking speeds. Overcoming oscillations also avoids slipping which can cause the biped to enter an unstable region instead of a stable walking.

Fig. 23 shows all the phases of cyclic walking. The biped starts from the Double Support Phase (DSP) and reaches SSP and DSP consecutively while walking. The shape and the surface area of the support polygon change at every phase with an order. The support polygon is a rectangle at the beginning and then becomes a convex hull which is followed by a smaller rectangle. This two-shape transition goes until the last step which ends with the same rectangle as at the beginning. Red lines are showing the left leg, right lines are showing the right leg, black lines are showing the hip, black circles are showing the position of the CoM, black crosses are showing the projection of the CoM on the walking plane and dashed lines are showing the support planes.

#### D. Simulation Results of the Closed-Loop Controller on Inclined Surfaces

The suggested PI-controller is examined on inclined surfaces to show its success. Firstly, simulation is done on the ascending surface, the slope of the surface is selected as  $10^\circ$  and both the kinetic and viscous friction selected as same as the previous simulations. Secondly, simulation is done on a descending surface, the slope of the surface is selected as  $10^\circ$  as the ascending surface, only the slope direction is changed and the friction coefficients are selected the same as the ascending surface. Fig. 24 shows the ascending and descending surfaces from left to right.

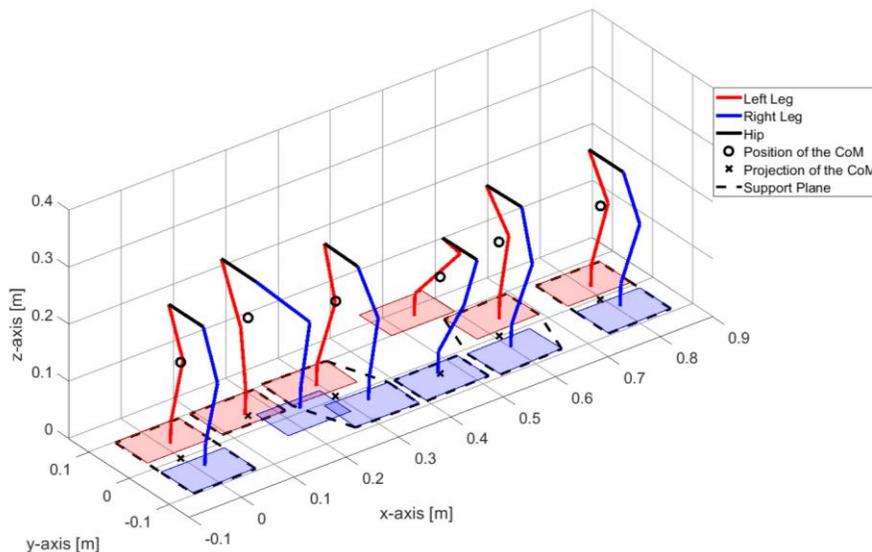


Fig. 23. Phases of Cyclic Walking.

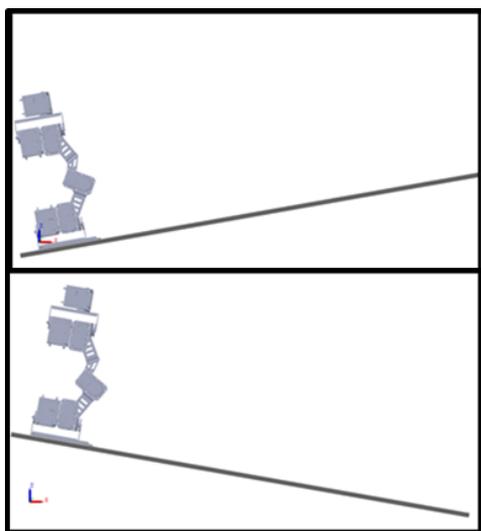


Fig. 24. The Inclined Surfaces.

Fig. 25 and Fig. 26 show the reference trajectories, open-loop trajectories, and closed-loop trajectories on the x-axis and y-axis of the ascending surface, respectively. The green lines are the reference trajectories, blue lines are the trajectories of the biped without the PI controller and blue lines are the trajectories of the biped with the PI controller. It can be seen that biped slips on both situations on the x-axis, the closed-loop system does nothing on the x-axis because biped mainly slips backward because of the slope. But from the second figure, it can be seen that the PI controller suppresses the oscillations on the y-axis as it does on a flat surface.

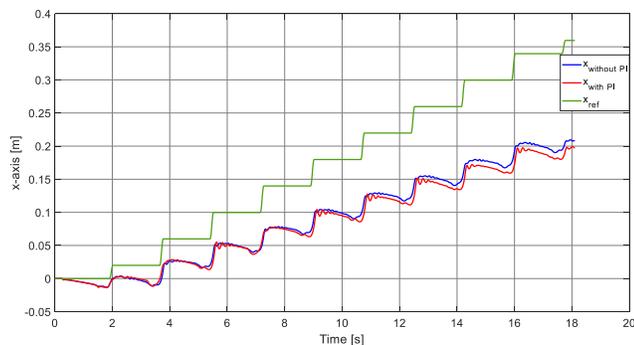


Fig. 25. Comparison of the Open and Closed-Loop System on the x-axis of the Ascending Surface.

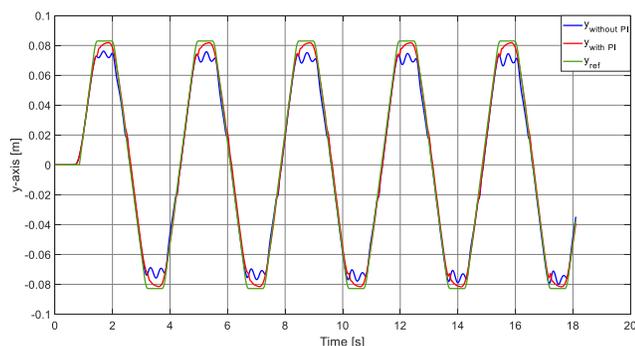


Fig. 26. Comparison of the Open and Closed-Loop System on the y-axis of the Ascending Surface.

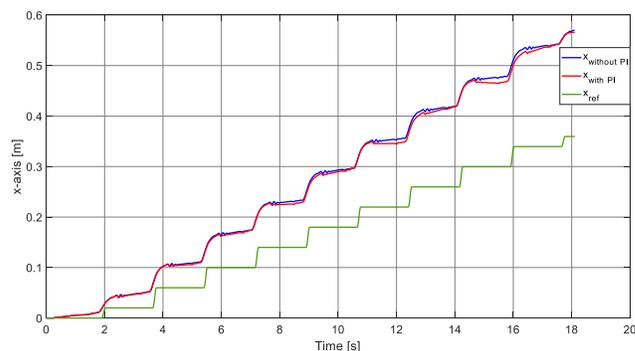


Fig. 27. Comparison of the Open and Closed-Loop System on the x-axis of the Descending Surface.

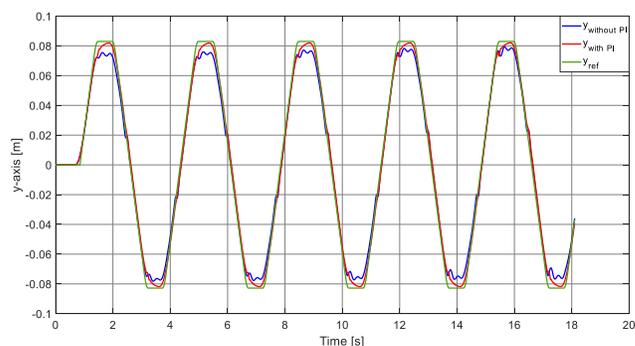


Fig. 28. Comparison of the Open and Closed-Loop System on the y-axis of the Descending Surface.

Fig. 27 and Fig. 28 show the reference trajectories, open-loop trajectories, and closed-loop trajectories on the x-axis and y-axis of the descending surface respectively. As previous figures, the green lines are the reference trajectories, blue lines are the open-loop trajectories and the red lines are the closed-loop trajectories. It can be seen that, biped slips on both situations on the x-axis, the closed-loop system does nothing on the x-axis, because biped mainly slips forward because of the slope. But from the second figure, it can be seen that the PI controller suppresses the oscillations on the y-axis as it does on the flat and ascending surfaces.

## VI. EXPERIMENTAL RESULTS

In this section experimental results of the biped both without the PI controller and with PI controller will be given. The biped consists of 12 Dynamixel servo motors. The main controller is Microautobox 2 from Dspace. All the motors are connected to Microautobox 2 from the serial port and the calculations are done real-time on Microautobox 2.

Fig. 29 shows the initial position of the biped before walking for 20 steps.

Fig. 30 shows the scenes from the initial to the final position of the biped during walking for 20 steps without the suggested PI controller. With the initial and final position of the biped two alternating step scenes are snapped and combined to show the scenes of walking at one figure. It can be seen that the biped also moved on the y-axis, although there is no translation reference is applied. Because of the slipping, the biped nearly moved out of the walking surface.

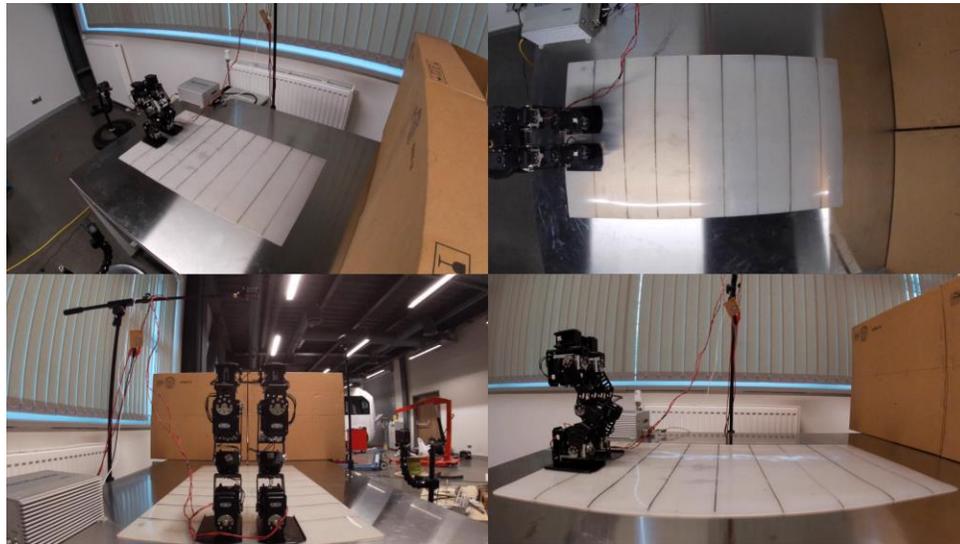


Fig. 29. The Initial Position of the Biped.

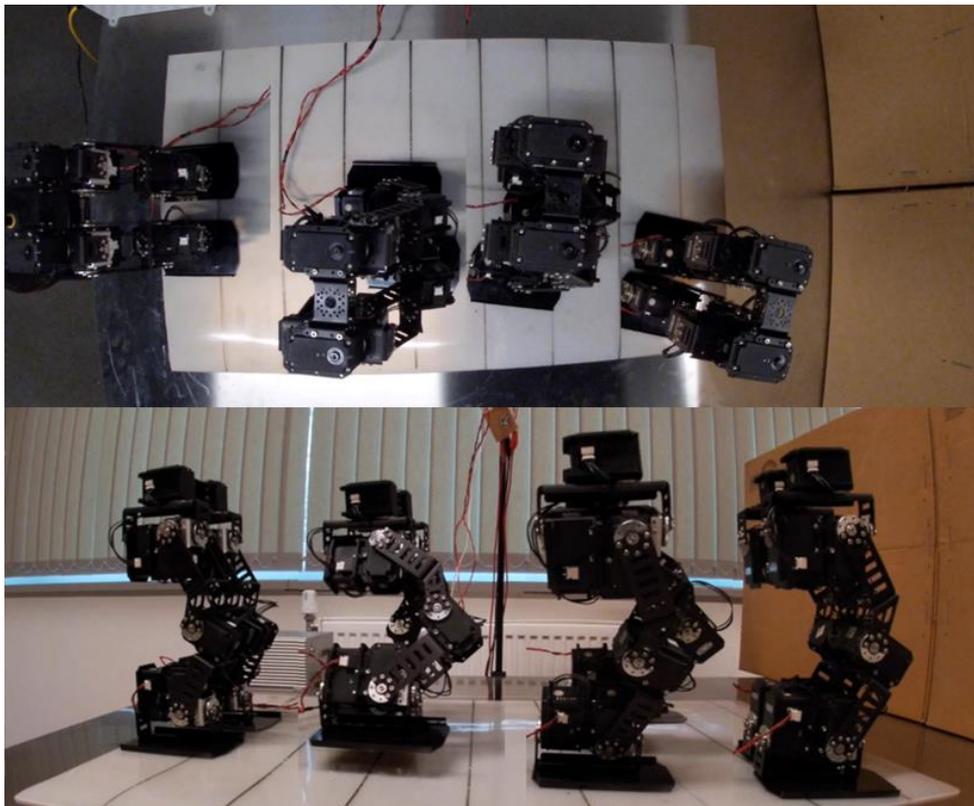


Fig. 30. Walking Scenes of the Open-Loop Biped System.

Fig. 31 shows the scenes from the initial to the final position of the biped during walking for 20 steps with the suggested PI controller. It can be seen that the biped has moved too less on the y-axis if compared with Fig. 29. Again, there is no translation reference is applied but the reason for this little

translation is the backlash effects on motors, so it can be said that the suggested PI controller is also successful in the real-time experiment. And also, the steady-state error on the x-axis is prevented by the closed-loop controller.

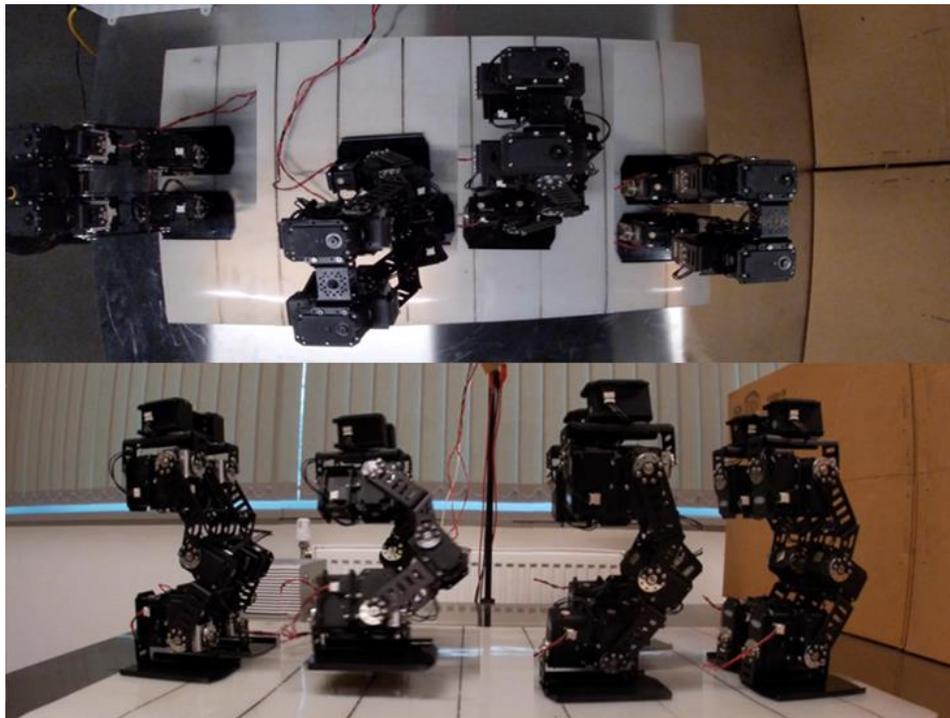


Fig. 31. Walking Scenes of the Closed-Loop Biped System.

## VII. CONCLUSIONS

The effectiveness of the MPC and suitability of the 3D LIPM for biped modeling have explained and proved in literature. But the simple 3D LIPM cannot reflect all the dynamical properties during the walking as expected, it can be used for trajectory planning and making calculations faster.

In this paper a PI controller is suggested, in order to not have a battle with the highly complicated dynamical model of the 12 DoF biped. While using the simple model because of its ease at use, overcoming the dynamical flaws of the simple model is aimed. For this purpose, a biped is modeled with a conventional 3D LIPM model, reference trajectories are created with Linear MPC. The least robust phase of the walking, which is the SSP is examined and PI-controller is added during this phase. The optimal values of the controller gains are searched by the GA, which is a well-known optimization method, to minimize the tracking error during the SSP. After finding the optimal gain values, the suggested method is firstly examined on the flat and inclined surfaces on Matlab Simulink simulations and then applied to the biped in the laboratory on a flat surface.

The success of the suggested method can be both seen in simulation results and real-time experiment results. With the help of the suggested method, all the dynamical flaws of the simple model which cause oscillations and steady-state error on the moving surface can be compensated during the walking. It can be said that, without facing a highly complicated dynamical model of a biped, the dynamical flaws of the simple 3D LIPM can be suppressed by the suggested method. And by this suggested method it is also clear that, this robot does not need any upper body element to overcome the oscillations and establish a stable walking.

Furthermore, the proposed closed-loop controller algorithm will be tried on two-dimensional walking in order to access all the points on a surface. This walking includes rotation of a biped which is not included in this study, this two-dimensional walking will also include an angular position tracking added to the linear position tracking. Another future study will be push recovery as it is a very important task for bipeds to maintain stable walking. And also the control algorithm will be tried on both climbing and descending ladders as future work.

## REFERENCES

- [1] Krotkov, E.; Hackett, D.; Jackel, L.; Perschbacher, M.; Pippine, J.; Strauss, J.; Pratt, G.; Orłowski, C.: "The Darpa robotics challenge finals: Results and perspectives". Springer Tracts in Advanced Robotics 2017, vol. 34, no. 2, pp. 229–240.
- [2] Ishihara, K.; Itoh, T. D.; Morimoto, J.: "Full-Body Optimal Control Toward Versatile and Agile Behaviors in a Humanoid Robot". IEEE Robotics and Automation Letters 2020, vol. 5, no. 1, pp. 119-126.
- [3] Tanguy, A.; De Simone, D.; Comport, A. I.; Oriol G.; Kheddar, "A. Closed-loop MPC with Dense Visual SLAM - Stability through Reactive Stepping". 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 2019, pp. 1397-1403.
- [4] Yong-Duk, K.; Bum-Joo, L.; Jeong-Ki, T.; Seung-Hwan, C.; Jong-Hwa K. "Humanoid Robot HanSaRam: Yawing Moment Cancellation and ZMP Compensation". 2005.
- [5] Ishihara, K.; Itoh, T. D.; Morimoto, J. "Full-Body Optimal Control Toward Versatile and Agile Behaviors in a Humanoid Robot". IEEE Robotics and Automation Letters 2020, vol. 5, no. 1, pp. 119-126.
- [6] Jung-Yup, K.; Ill-Woo, P.; Jun-Ho, O. "Walking Control Algorithm of Biped Humanoid Robot on Uneven and Inclined Floor". Journal of Intelligent and Robotic Systems 2007. 48. 457-484.
- [7] Choi, Y.; Kim, D.; Oh, Y.; You, B. "Posture/Walking Control for Humanoid Robot Based on Kinematic Resolution of CoM Jacobian With Embedded Motion". IEEE Trans. Robot. 2007, 23, 1285–1293.
- [8] Kajita, S.; Kanehiro, F.; Kaneko, K.; Fujiwara, K.; Harada, K.; Yokoi, K.; Hirukawa, H. "Biped Walking Pattern Generation by Using Preview

- Control of Zero-Moment Point". IEEE International Conference on Robotics and Automation, 2003, pp. 1620–1626.
- [9] Wieber, P.-B. "Trajectory Free Linear Model Predictive Control for Stable Walking in the Presence of Strong Perturbations". 6th IEEE/RAS International Conference on Humanoid Robots, 2006, pp. 137–142.
- [10] Parker, J. K.; Khoogar, A. R.; Goldberg, D. E.; "Inverse Kinematics of Redundant Robots Using Genetic Algorithms". Proceedings, International Conference on Robotics and Automation, Scottsdale, AZ, 1989, pp. 271-276 vol.1.
- [11] Pillai, R. P.; Jadhav, S. P.; Patil, M. D. "Tuning of PID Controllers using Advanced Genetic Algorithm". International Journal of Advanced Computer Science and Application Special Issue on Selected Papers from International Conference Workshop On Advance Computing 2013.
- [12] Vukobratovic, M.; Juricic, D. "Contribution to the synthesis of biped gait". Proceedings of the IFAC Symposium Technical and Biological Problem on Control, Erevan, USSR, 1968.
- [13] Scianca, N.; De Simone, D.; Lanari, L.; Oriolo, G. "MPC for Humanoid Gait Generation: Stability and Feasibility". IEEE Transactions on Robotics 2020. PP. 1-18.
- [14] Kajita, S.; Tani, K. "Study of dynamic biped locomotion on rugged terrain-derivation and application of the linear inverted pendulum mode". IEEE International Conference on Robotics and Automation, 1991, pp. 1405–1411.
- [15] Ferrari, P.; Scianca, N.; Lanari, L.; Oriolo, G. "An Integrated Motion Planner/Controller for Humanoid Robots on Uneven Ground". 2019 18th European Control Conference (ECC), Naples, Italy, 2019, pp. 1598-1603.
- [16] Castiglione, T.; Morrone, P.; Falbo, L.; Perrone, D.; Bova, S. "Application of a Model-Based Controller for Improving Internal Combustion Engines Fuel Economy". Energies 2020, 13, 1148.
- [17] Traussnig, A.; Stolz, M.; Horn, M. "MPC based Fan Control for Automotive Application"s. IFAC-PapersOnLine. 2016, 49.
- [18] Guo, L.; Ge, P.; Sun, D.; Qiao, Y. "Adaptive Cruise Control Based on Model Predictive Control with Constraints Softening". Appl. Sci. 2020, 10, 1635.
- [19] Degachi, H.; Chagra, Wm; Ksouri, M. "Nonlinear Model Predictive Control for pH Neutralization Process based on SOMA Algorithm". International Journal of Advanced Computer Science and Application Vol. 9, No. 1, 2018.
- [20] Winkler A. W.; Farshidian F.; Pardo D.; Neunert M; Buchli J. "Fast Trajectory Optimization for Legged Robots Using Vertex-Based ZMP Constraints". IEEE Robotics and Automation Letters, 2017, vol. 2, no. 4, pp. 2201-2208.
- [21] Urata J.; Nshiwaki K.; Nakanishi Y.; Okada K.; Kagami S.; Inaba M. "Online Walking Pattern Generation for Push Recovery and Minimum Delay to Commanded Change of Direction and Speed". IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012, pp. 3411-3416.

# Using Fuzzy c-Means for Weighting Different Fuzzy Cognitive Maps

Mamoon Obiedat<sup>1</sup>  
Computer Information Systems Department  
Hashemite University  
Zarqa, Jordan

Ali Al-yousef<sup>2</sup>  
Computer Science Department  
Jerash University  
Jarash, Jordan

Ahmad Khasawneh<sup>3</sup>  
Prisednt of Irbid National University  
Irbid National University  
Irbid, Jordan

Nabhan Hamadneh<sup>4</sup>, Ashraf Aljammal<sup>5</sup>  
Computer Science Department  
Hashemite University  
Zarqa, Jordan

**Abstract**—Currently, complex socio-ecological problems have increasingly prevailed with uncertainty that often dominates these domains. In order to better represent these problems, there is an urgent need to engage a wide range of different stakeholders' perspectives, regardless of their levels of expertise and knowledge. Then, these perspectives should be combined in an appropriate manner for a comprehensive and reasonable problem representation. Fuzzy cognitive map (FCM) has proven to be powerful and useful as a soft computing approach in addressing and representing such problem domains. By the FCM approach, the relevant stakeholders can represent their perspectives in the form of FCM system. Normally, relevant stakeholders have different levels of knowledge, and hence produce different representations (FCMs). Therefore, these FCMs should be weighted appropriately before the combination process. This paper uses fuzzy c-means clustering technique to assign different weights for different FCMs according to their importance in representing the problem. First, fuzzy c-means is used to compute the membership values of belonging of FCMs to the selected clusters based on the FCMs similarities that show how convergent and consistent they are. According to these membership values, the importance clusters' values are calculated, in which a cluster with a high membership value from all FCMs is the cluster with the high importance value, and vice versa. Next, the importance values for FCMs are derived from the importance values of the clusters by looking at the amount of contributions of FCMs memberships to the clusters. Finally, FCMs importance values are used to assign weight values to these FCMs, which are used when they are combined. The suitability of the proposed method is investigated using a real dataset that includes an appropriate number of FCMs collected from different stakeholders.

**Keywords**—Complex problems; uncertainty; fuzzy cognitive map (FCM); fuzzy c-Means; FCM weight values

## I. INTRODUCTION

The world is facing a large number of various real problems in all aspects of life. On one hand, real-world problems have become complex in nature and are usually multidisciplinary. On the other hand, solving these problems has become an urgent necessity and attracted the attention of

decision makers. However, modelling and capturing the knowledge of these problems face several key challenges, such as a lack of structural/quantitative data, domain complexity, and a lack of sufficient data (comprehensive view) representing adequately the domain knowledge. Moreover, such problems are often characterized by uncertainty, ill-defined, and qualitative imprecise data [1]. In such a case, hard computing approaches may not provide problem modelling or solving [2]. These problems can also include multiple dimensions, and hence, this leads to increase their complexity. For example, environmental problems usually include social/human and ecological dimensions [3]. This highlights the need to share the knowledge of all these dimensions. Here, a comprehensive view of such problem domains can embrace complexity, yet it requires an understanding of their dimensions and components on one side, and the players who interact between these dimensions and components on the other side [4-9].

In fact, soft computing approaches can deal with complex unstructured real-world problems including uncertainty and imprecise data with partial truth [2]. One helpful soft computing approaches that appropriately addresses uncertainty usually inherited in the real-world problems is a Fuzzy Cognitive Map (FCM). FCM approach incorporates the cognitive map concept with the fuzzy logic concept for expressively representing the knowledge of the problem. The purpose of the cognitive map is to represent the domain knowledge as map nodes representing the domain variables and directed arrows representing the connections between these variables [10]. On the other side, fuzzy logic has tolerance in modelling such problems and provides reasoning decision making [11]. It also can mathematically represent the imprecise data in a degree of truth (i.e. degree of membership in the range [0, 1]) [12]. FCM can address the multiple dimensionality of the problem domains including different related perspectives. It represents any complex problem in the form of variables and negative/positive connections between them [1, 7, 13]. The values of connections can be in the form of fuzzy numerical values (i.e. [-1, 1]) or fuzzy linguistic values (i.e. weak, moderate, strong etc.). Moreover, FCM is easily converted to

an adjacency matrix containing the connection values for mathematically handling.

To overcome the challenges of insufficient data and inherited uncertainty of large complex problems, several studies have suggested engaging a large sample of relevant stakeholders in representing the domain knowledge and capturing their perspectives/FCMs [1, 14-17]. Gathering many perspectives would lead to explore many relevant variables and connections with more certainty. Then, the perspectives of the stakeholders are combined to form a comprehensive perspective/FCM with more reliability and relevance in knowledge representation and problem solving [7, 14, 18]. In fact, humans are a vital component of real-world problems and they play a major role in shaping other components. Therefore, their perceptions are often necessary to solve problems. However, stakeholders may have different specializations and positions with different levels of knowledge and experience, and their perspectives are generally dominated by their preferences. Dealing with the stakeholders' perceptions/FCMs equally, especially when they are combined into a holistic view, will lead to an inaccurate solution. Thus, FCMs of the stakeholders should be weighted based on suitable criteria that reveal their importance in effectively representing the domain knowledge. To do so, the goal of this paper is to use the fuzzy c-means clustering algorithm in order to obtain different weight values for different stakeholder perspectives (FCMs). These weights are then used to weight the connections in the corresponding FCMs when aggregated into a whole FCM system that reflects an acceptable consensus perception for a given problem. The reason behind the use of this method is that it reveals the extent of convergence and similarity of the views of stakeholders. It first defines a suitable number of clusters. Then, it assigns a membership degree of each FCM belonging to every defined cluster. Finally, FCMs belonging to the same cluster are as similar as possible, and therefore should take high weights, and vice versa.

To achieve this goal, this paper is structured as follows: Section II overviews FCM approach and presents some related works. The proposed methodology is introduced in Section III, including a brief description of the fuzzy c-means clustering technique. Section IV presents the application of the proposed method on a suitable dataset and discusses the results. Finally, Section V presents the conclusion of the paper.

## II. FCM OVERVIEW AND RELATED WORKS

FCM approach mimics human cognition of a problem domain (system) without looking at their levels of knowledge. Each human can easily develop their system by exploring uncertain concepts of a given domain and then link these concepts by directed arrows labelled with fuzzy signed values representing cause-effect relationships between concepts, see Fig. 1. The concepts represent the relevant elements of the system such as its variables, factors, attributes, etc. The relationships represent the degree of negative or positive influences between the concepts, where they are encoded into adjacency matrix, Table I.

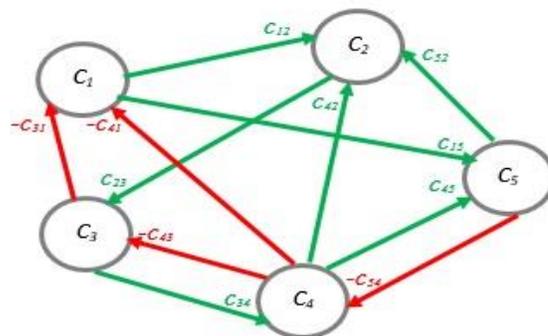


Fig 1. FCM Representing 5 Concepts and the Positive and Negative Relationships between them.

TABLE I. THE ADJACENCY MATRIX OF CONNECTION VALUES OF FCM IN FIG. 1

	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$
$C_1$	0	$c_{12}$	0	0	$c_{15}$
$C_2$	0	0	$c_{23}$	0	0
$C_3$	$-c_{31}$	0	0	$c_{34}$	0
$C_4$	$-c_{41}$	$c_{42}$	$-c_{43}$	0	$c_{45}$
$C_5$	0	$c_{52}$	0	$-c_{54}$	0

FCM approach is a nonlinear dynamic system that allows feedback loops. It uses artificial neural network (ANN) to run the system based on its adjacency matrix and initial state values of concept in order to find the system outcomes, also called FCM inference. The system outcomes represent new state values of concepts and are used to analyse the influence of variables on each other. They are also used to conduct and analyse what-if scenarios in order to explore results that would help in decision making. This process can be applied on each individual FCM system. To address the knowledge of the problem domain as a whole system, the individual FCMs are combined into a collective FCM.

FCM approach has been extensively used in many research domains to model various complex real-world problems. Several studies have considered any relevant stakeholder has an important role in solving such problems [1, 8, 19-21]. The authors in [1] participated different stakeholders to extract their knowledge about environmental problems. They gathered the stakeholders' perspectives of experts and local people as well. Each individual perception was represented in an individual cognitive map which in turn was encoded into adjacency matrix to include the connection values. The authors then combined all perceptions to form a social perception by augmenting each matrix to include all variables identified by all stakeholders. Then, the matrices were added to each other after weighting them using the weighted average method. Each connection value in the resulting combined matrix is the sum of all corresponding values in the individual matrices divided by the number of matrices (FCMs). It is worth noting here that all stakeholders' perceptions have the same weight regardless of their knowledge of the domain. A study in [6] addressed a sustainable socio-economic domain using different knowledge

resources. The study applied the ordered weighted averaging (OWA) aggregation operators introduced in [22] to aggregate the connection values of FCMs. The aggregation operators are numeric operators such as *weighted average*, *arithmetic mean*, *weighted harmonic averaging*, *max*, *min*, etc. The above study enhanced OWA method by learning the connection values of relevant stakeholders associated with the operator used for aggregation.

An advance attempt for aggregating different FCMs developed by different stakeholders was in [7]. The above study developed many FCMs from different stakeholders of different groups in solving a complex water scarcity problem. Toward weighting the FCMs before aggregating, the authors used a novel aggregation method to combine FCMs into an overall FCM [7, 23-24]. They assigned a credibility weight for each individual FCM before combining them to each other. The credibility weights of FCMs were calculated based on a novel measure, called centrality consensus measure CCM, proposed in [25]. The aforementioned study benefited from the 2-tuple fuzzy linguistic representation model [26] to represent the connection values in the fuzzy  $\beta$  values. It is worth mentioning here that this aggregation method can combine fuzzy numerical and linguistic values simultaneously.

Nevertheless, the above studies have not taken into account the consistency and convergence between stakeholders' perspectives (FCMs). When many stakeholders with different knowledge contribute to a given system, this naturally appears inconsistency between them and the holistic system as well. Therefore, the methodology of this paper uses the fuzzy c-means technique to identify consistent and converging FCMs based on their similarities. Then, it benefits from this process to assign weight values to FCMs. The next section details the methodology proposed in this paper.

### III. METHODOLOGY

Clustering techniques, either hard or soft clustering, have been proven to be effective in grouping similar data into clusters. In hard clustering, the data is grouped into crisp clusters, where each data point either belongs to exactly specific cluster or not, and the clusters cannot overlap, see Fig. 2(A). Unlike the hard clustering, the data point in the soft or fuzzy clustering can belong to multiple clusters with different membership values between 0 and 1, and the clusters may overlap, see Fig. 2(B).

In this paper, a fuzzy c-means clustering method is used to know the similarities between FCMs by determining how they belong to a defined number of clusters. For each FCM, a membership degree of its belonging to every cluster will be defined, where FCMs belonging to the same cluster are as similar as possible. The benefit of this is to know the degree of similar FCMs memberships in each cluster. Similarities between FCMs demonstrate the consistency and convergence among the perspectives of FCMs developers. This of course indicates that the more converged and consistent perspectives, the more important they are.

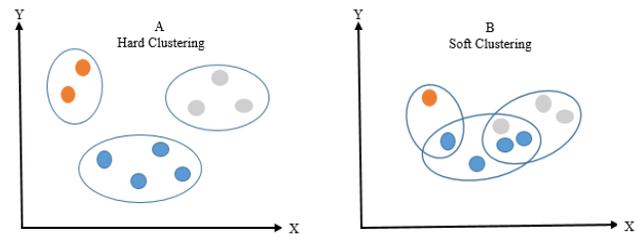


Fig 2. Hard vs Soft Clustering.

Fuzzy c-means, first coined by Dunn [27] and generalized by Bezdek [28, 29], is a fuzzy clustering method in which each element (data point) in a dataset has a degree membership value within the range [0, 1] for each cluster. The total of the element membership values for all clusters equals 1. Fuzzy c-means clustering have been used in many application areas such as clustering, neural network, image analysis, and classification [30].

The structure of the fuzzy c-means algorithm is based on the concept of fuzzy c-partition, proposed by Ruspini [31]. The objective of the fuzzy c-means is to investigate the degree of membership to each sample of data corresponding to each defined cluster centers (centroids). This is done based on the distance between the centroid and the sample of data, in which the degree of membership takes a high value (close to 1) when the sample is near to the centroid. This reflects the similarities of the data samples for clustering.

To put the fuzzy c-means clustering method in the context of the paper methodology, the following steps are performed:

*Step1.* Prepare the data/samples (FCM matrices) to be an appropriate input for the fuzzy c-means as follows:

a) Each sample (FCM) is represented by  $C \times C$  adjacency matrix containing connections between concepts, where  $C$  is the number of concepts in FCM.

b) Convert the resulting matrix into vector of  $C^2$  elements. The goal of this conversion process is to create a suitable sample of data for clustering. The data points of the first column in the matrix will be the [1 ..  $C$ ] data points in the new vector, the data points of the second column in the matrix will be [ $C + 1$  ..  $C * 2$ ], the elements of the third column in the matrix will be [ $C * 2 + 1$  ..  $C * 3$ ], and so on, until the data points of the last column in the matrix will be [ $C^2 - C + 1$  ..  $C^2$ ].

c) Create a matrix of  $N \times C^2$ , where  $N$  is the number of samples (FCMs).

- The output of *Step1* is a matrix that includes all connection values between concepts in all FCMs which is then used as data points (samples) input to the fuzzy c-means clustering as shown in *Step2*.

*Step2.* Apply the fuzzy c-means algorithm to the data points resulting from *Step1* as follows:

a) Select the number of clusters ( $K$ ), and the fuzzification parameter ( $m$ ),  $m$  typically in the range [1.25, 2].

b) Initialize randomly the membership matrix  $U^0$  that includes all membership values  $u_{ij}$  of belonging of  $FCM_i$  to clusters  $K_j$ .

c) At iteration  $S$ , where  $S = 1$  to number of iterations, calculate centroids  $K_i$ , using the following Equation:

$$K_j = \frac{\sum_{i=1}^N u_{ij}^m * X_i}{\sum_{i=1}^N u_{ij}^m} \quad (1)$$

where  $N$  is the number of samples (FCMs), and  $X_i$  is the  $i^{th}$  data point.

Calculate the new membership values  $u_{ij}$ , using the following Equation:

$$u_{ij} = \frac{1}{\sum_{S=1}^K \left( \frac{\|x_i - K_j\|}{\|x_i - K_S\|} \right)^{\frac{2}{m-1}}} \quad (2)$$

where  $\|$  means Euclidian Distance.

a) Update  $U^{S+1} \leftarrow U^S$ .

b) Repeat steps c to e until a maximum number of iterations is reached or a very small change of the membership values, typically  $U^{S+1} - U^S \leq 0.01$ .

- The output of Step2 is a two dimensional matrix; the matrix rows represent the FCMs, the matrix columns represent the clusters, and each row includes the degree membership values of belonging of  $i^{th}$  FCM to all clusters, where  $i = 1$  to the number of FCMs. This matrix is used to obtain weight values for FCMs as presented in the next Step3.

Step3. Assign weights ( $W$ ), where  $i = 1$  to the number of FCMs, to FCM matrices as follows:

a) Calculate the importance values of the clusters as follows

- 1) For  $k^{th}$  cluster, where  $k = 1$  to the number of clusters, calculate the total of its degree membership values derived from all FCM matrices.
- 2) Normalize the calculated values for all clusters to be in the range [0, 1]. This gives an importance value for each cluster.

b) Assign weight values for FCMs based on the importance values of the clusters as follows:

1) Calculate the importance values for the FCMs as follows:

- $i$ . For  $i^{th}$  FCM matrix, where  $i = 1$  to the number of FCMs, then
  - For  $k^{th}$  cluster, where  $k = 1$  to the number of clusters, multiply the degree membership value of belonging of the  $i^{th}$  FCM to the  $k^{th}$  cluster by the importance of the  $k^{th}$  cluster.

- Sum all values of the  $i^{th}$  FCM calculated from step  $i$ .

2) Use the calculated importance values of FCMs to assign weight values  $W$  for these FCMs (matrices).

where  $\sum_i^N W_i = 1$ , and  $N$  is the number of FCMs.

Step4. Use the weight values of FCMs in combining the FCMs to obtain consensus, credible and comprehensive FCM system using the following equation:

$$FCM = \sum_i^N W_i * FCM_i \quad (3)$$

where FCM is the group weighted FCM matrix,  $W_i$  is the weight of  $i^{th}$  FCM matrix.

Next section demonstrates the application of our proposed method on a real dataset.

#### IV. TESTING THE PROPOSED METHOD

The proposed method was tested using the dataset collected in the study [23]. The dataset included 35 samples (FCMs) developed from 35 different stakeholders representing a major real world problem "water scarcity problem". Each FCM contained 13 same concepts. Thus, the adjacency matrix of each sample was 13 x 13 dimensions and included the connection values between these concepts.

Based on Step 1.b in the proposed methodology, a vector of  $1 \times 13^2$  was created for each adjacency matrix of each FCM. Then, the input matrix for fuzzy c-means clustering was of  $35 \times 169$  and included all connection values between concepts in all FCMs, (Step 1.c in the proposed methodology). We used the *silhouette* criteria introduced in [32] to select the suitable number of clusters and centroids, accordingly. This number was 10 cluster. After that, the steps of the fuzzy c-means clustering algorithm (Step2.b-f in the proposed methodology) were implemented to calculate the final membership values  $u_{ij}$  of belonging of FCMs to all clusters. The outcomes (membership values) of the above steps are shown in Table II.

Then, we calculated the importance of the clusters according to the Step3.a in the proposed methodology. Table III shows these importance values. We then used Step3.b to find the importance values of FCMs by multiplying the membership values of their belongings to the cluster by calculated importance values of the clusters, see Table IV.

Finally, we utilized the calculated importance values of FCMs to assign weight values for these FCMs. This process was performed using a ratio method, which was applied to the importance values to make the resulting weight values within the range [0, 1], and the total of these weight values equal to 1. The last column in the Table IV shows the weights assigned to the FCMs.

TABLE II. THE DEGREE MEMBERSHIP VALUES,  $u_{ij}$ , OF BELONGING OF FCMs TO ALL CLUSTERS

Cluster ID FCM ID	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
FCM 1	0.01532	0.00009	0.98080	0.00001	0.00004	0.00000	0.00109	0.00038	0.00008	0.00218
FCM 2	0.00020	0.99077	0.00073	0.00027	0.00016	0.00015	0.00065	0.00135	0.00064	0.00508
FCM 3	0.00504	0.00079	0.99114	0.00002	0.00006	0.00001	0.00032	0.00030	0.00004	0.00229
FCM 4	0.00020	0.00039	0.99905	0.00000	0.00000	0.00000	0.00015	0.00001	0.00003	0.00016
FCM 5	0.00012	0.99312	0.00032	0.00010	0.00005	0.00007	0.00481	0.00058	0.00020	0.00063
FCM 6	0.00048	0.99387	0.00095	0.00005	0.00002	0.00001	0.00062	0.00107	0.00061	0.00231
FCM 7	0.00002	0.00002	0.00000	0.99964	0.00000	0.00000	0.00004	0.00021	0.00000	0.00007
FCM 8	0.00014	0.00006	0.00047	0.00000	0.00004	0.00001	0.00007	0.00016	0.99874	0.00030
FCM 9	0.00027	0.00059	0.00006	0.00001	0.00002	0.00000	0.00012	0.00010	0.99859	0.00023
FCM 10	0.00005	0.00083	0.00006	0.00001	0.00001	0.00002	0.00012	0.00002	0.99858	0.00029
FCM 11	0.00006	0.00024	0.00011	0.00003	0.00006	0.00001	0.00038	0.00035	0.00008	0.99869
FCM 12	0.00346	0.00063	0.00409	0.00013	0.00006	0.00001	0.00268	0.00240	0.00010	0.98644
FCM 13	0.00079	0.00369	0.00225	0.00036	0.00002	0.00002	0.00640	0.00136	0.00016	0.98495
FCM 14	0.00035	0.00010	0.00008	0.00001	0.00002	0.00000	0.00033	0.00016	0.00001	0.99896
FCM 15	0.00087	0.00150	0.00075	0.00004	0.00002	0.00002	0.00153	0.00059	0.00010	0.99458
FCM 16	0.01067	0.00328	0.00119	0.00085	0.00377	0.00025	0.00711	0.00355	0.00139	0.96794
FCM 17	0.99510	0.00009	0.00052	0.00001	0.00002	0.00000	0.00240	0.00039	0.00002	0.00145
FCM 18	0.99928	0.00001	0.00026	0.00000	0.00001	0.00000	0.00023	0.00003	0.00000	0.00018
FCM 19	0.00117	0.00006	0.00021	0.00003	0.00004	0.00000	0.99758	0.00030	0.00000	0.00060
FCM 20	0.00028	0.00066	0.00009	0.00027	0.00003	0.00000	0.00101	0.99726	0.00001	0.00039
FCM 21	0.00001	0.00001	0.00000	0.99984	0.00000	0.00000	0.00004	0.00004	0.00000	0.00005
FCM 22	0.00005	0.00005	0.00004	0.00001	0.99903	0.00005	0.00036	0.00035	0.00001	0.00006
FCM 23	0.00003	0.00005	0.00002	0.00000	0.00000	0.00000	0.99965	0.00013	0.00000	0.00013
FCM 24	0.00031	0.00035	0.00013	0.00004	0.00167	0.00001	0.99447	0.00164	0.00018	0.00121
FCM 25	0.00314	0.00127	0.00231	0.00019	0.97391	0.00026	0.00779	0.00260	0.00285	0.00569
FCM 26	0.00006	0.00027	0.00006	0.00012	0.00017	0.00001	0.00205	0.99595	0.00005	0.00124
FCM 27	0.00108	0.00012	0.00012	0.00003	0.99677	0.00015	0.00071	0.00037	0.00004	0.00062
FCM 28	0.99897	0.00001	0.00085	0.00000	0.00001	0.00000	0.00003	0.00001	0.00001	0.00011
FCM 29	0.00000	0.00000	0.00000	0.00000	0.00000	0.99999	0.00000	0.00000	0.00000	0.00000
FCM 30	0.00269	0.00075	0.99019	0.00003	0.00009	0.00003	0.00252	0.00194	0.00076	0.00100
FCM 31	0.00000	0.00000	0.00000	0.00000	0.00001	0.99998	0.00000	0.00000	0.00000	0.00000
FCM 32	0.00021	0.99403	0.00063	0.00001	0.00008	0.00008	0.00142	0.00148	0.00030	0.00176
FCM 33	0.00021	0.00026	0.00023	0.00006	0.00010	0.00004	0.00150	0.99659	0.00001	0.00100
FCM 34	0.00024	0.00075	0.00021	0.00003	0.00002	0.00001	0.99577	0.00167	0.00002	0.00128
FCM 35	0.00001	0.00001	0.00000	0.00000	0.99990	0.00000	0.00002	0.00003	0.00000	0.00002

TABLE III. THE IMPORTANCE VALUES OF THE CLUSTERS

Cluster ID	1	2	3	4	5	6	7	8	9	10
Cluster importance	0.30408	0.39887	0.39779	0.20022	0.39762	0.20012	0.40340	0.30134	0.30036	0.59619

TABLE IV. FCMs IMPORTANCE VALUES AND THEIR CALCULATED WEIGHT VALUES, ACCORDINGLY

Cluster ID FCM ID	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	FCM Importance	FCM Weight
FCM 1	0.00466	0.00004	0.39015	0.00000	0.00002	0.00000	0.00044	0.00012	0.00002	0.00130	0.39675	0.02945
FCM 2	0.00006	0.39519	0.00029	0.00005	0.00006	0.00003	0.00026	0.00041	0.00019	0.00303	0.39958	0.02966
FCM 3	0.00153	0.00032	0.39427	0.00000	0.00002	0.00000	0.00013	0.00009	0.00001	0.00137	0.39774	0.02952
FCM 4	0.00006	0.00016	0.39741	0.00000	0.00000	0.00000	0.00006	0.00000	0.00001	0.00010	0.39780	0.02953
FCM 5	0.00004	0.39613	0.00013	0.00002	0.00002	0.00001	0.00194	0.00017	0.00006	0.00038	0.39890	0.02961
FCM 6	0.00015	0.39643	0.00038	0.00001	0.00001	0.00000	0.00025	0.00032	0.00018	0.00137	0.39910	0.02962
FCM 7	0.00001	0.00001	0.00000	0.20015	0.00000	0.00000	0.00002	0.00006	0.00000	0.00004	0.20028	0.01487
FCM 8	0.00004	0.00003	0.00019	0.00000	0.00002	0.00000	0.00003	0.00005	0.29999	0.00018	0.30051	0.02231
FCM 9	0.00008	0.00024	0.00002	0.00000	0.00001	0.00000	0.00005	0.00003	0.29994	0.00014	0.30051	0.02231
FCM 10	0.00002	0.00033	0.00003	0.00000	0.00001	0.00000	0.00005	0.00001	0.29994	0.00017	0.30055	0.02231
FCM 11	0.00002	0.00009	0.00004	0.00001	0.00002	0.00000	0.00015	0.00011	0.00002	0.59541	0.59588	0.04423
FCM 12	0.00105	0.00025	0.00163	0.00003	0.00003	0.00000	0.00108	0.00072	0.00003	0.58810	0.59292	0.04401
FCM 13	0.00024	0.00147	0.00089	0.00007	0.00001	0.00000	0.00258	0.00041	0.00005	0.58721	0.59295	0.04401
FCM 14	0.00011	0.00004	0.00003	0.00000	0.00001	0.00000	0.00013	0.00005	0.00000	0.59557	0.59593	0.04423
FCM 15	0.00026	0.00060	0.00030	0.00001	0.00001	0.00000	0.00062	0.00018	0.00003	0.59295	0.59496	0.04416
FCM 16	0.00325	0.00131	0.00047	0.00017	0.00150	0.00005	0.00287	0.00107	0.00042	0.57708	0.58818	0.04366
FCM 17	0.30260	0.00003	0.00021	0.00000	0.00001	0.00000	0.00097	0.00012	0.00000	0.00087	0.30480	0.02262
FCM 18	0.30387	0.00000	0.00010	0.00000	0.00000	0.00000	0.00009	0.00001	0.00000	0.00011	0.30419	0.02258
FCM 19	0.00036	0.00002	0.00008	0.00001	0.00002	0.00000	0.40242	0.00009	0.00000	0.00036	0.40336	0.02994
FCM 20	0.00008	0.00026	0.00004	0.00005	0.00001	0.00000	0.00041	0.30051	0.00000	0.00023	0.30161	0.02239
FCM 21	0.00000	0.00001	0.00000	0.20019	0.00000	0.00000	0.00001	0.00001	0.00000	0.00003	0.20026	0.01486
FCM 22	0.00001	0.00002	0.00002	0.00000	0.39724	0.00001	0.00015	0.00011	0.00000	0.00004	0.39759	0.02951
FCM 23	0.00001	0.00002	0.00001	0.00000	0.00000	0.00000	0.40326	0.00004	0.00000	0.00008	0.40341	0.02994
FCM 24	0.00009	0.00014	0.00005	0.00001	0.00066	0.00000	0.40117	0.00049	0.00005	0.00072	0.40339	0.02994
FCM 25	0.00096	0.00050	0.00092	0.00004	0.38725	0.00005	0.00314	0.00078	0.00086	0.00339	0.39789	0.02953
FCM 26	0.00002	0.00011	0.00002	0.00002	0.00007	0.00000	0.00083	0.30012	0.00002	0.00074	0.30195	0.02241
FCM 27	0.00033	0.00005	0.00005	0.00001	0.39634	0.00003	0.00029	0.00011	0.00001	0.00037	0.39757	0.02951
FCM 28	0.30377	0.00000	0.00034	0.00000	0.00000	0.00000	0.00001	0.00000	0.00000	0.00006	0.30420	0.02258
FCM 29	0.00000	0.00000	0.00000	0.00000	0.00000	0.20012	0.00000	0.00000	0.00000	0.00000	0.20012	0.01485
FCM 30	0.00082	0.00030	0.39389	0.00001	0.00004	0.00001	0.00102	0.00058	0.00023	0.00059	0.39748	0.02950
FCM 31	0.00000	0.00000	0.00000	0.00000	0.00000	0.20012	0.00000	0.00000	0.00000	0.00000	0.20013	0.01485
FCM 32	0.00006	0.39649	0.00025	0.00000	0.00003	0.00002	0.00057	0.00044	0.00009	0.00105	0.39901	0.02962
FCM 33	0.00006	0.00011	0.00009	0.00001	0.00004	0.00001	0.00061	0.30031	0.00000	0.00059	0.30183	0.02240
FCM 34	0.00007	0.00030	0.00008	0.00001	0.00001	0.00000	0.40169	0.00050	0.00001	0.00076	0.40344	0.02995
FCM 35	0.00000	0.00000	0.00000	0.00000	0.39758	0.00000	0.00001	0.00001	0.00000	0.00001	0.39762	0.02951

As shown in Table IV, the weight value of both *FCM 14* and *FCM 11*, (0.04423), was the highest of the weight values. This means that *FCM 14* and *FCM 11* are the most important FCMs which are well perceive the domain knowledge of the problem and that their developers are the most credible stakeholders. Thus, when multiplying the matrices of these FCMs by their credible weights during combining them to other FCMs (*Step4* in the methodology), this will enhance the

compatibility and reliability of the resulting comprehensive FCM system.

On the other hand and as shwon in Table IV, the weight value of both *FCM 29* and *FCM 31*, (0.01485), was the lowest of FCMs weight values. This indicates that these FCMs are less important in representing domain knowledge, and therefore, the stakeholders who developed these FCMs are less reliable (less experienced or knowledgeable). Therefore, the matrices of these FCMs should be weighted with lower weights

in the combination process. Other FCMs with different weight values are treated with the same steps above. It should be noted here that any participation of all relevant stakeholders should not be neglected as mentioned above, but that their different levels of knowledge and preferences should be taken into consideration in any further process.

## V. CONCLUSION

This paper introduces a new method to assign weights for fuzzy cognitive maps (FCMs) developed by different stakeholders representing a given complex problem domain. The main goal of this is to use these weight values in order to reach an accurate consensus representation of the problem after combining the weighted FCMs. To achieve this, important FCMs are given high weights, and vice versa. This paper uses fuzzy c-means clustering algorithm to reveal the importance of FCMs. By this algorithm, the membership values of belonging of FCMs to the clusters are calculated. The membership values are computed based on the similarities between FCMs. Based on this, the importance values of cluster were extracted. Then, importance values of FCMs are identified based on their contributions to the clusters taken into accounts the importance values of the clusters. Consequently, weight values are assigned to FCMs using FCMs importance values. Finally, experimenting with the proposed method on a real dataset has proved its suitability and efficiency in assigning credible weights of for FCMs. As a future work, we intend to combine this proposed method with other effective techniques for more reliable weights either for FCMs or the connections in the FCMs.

## REFERENCES

- [1] Ozesmi, U. and S. Ozesmi, Ecological models based on people's knowledge: a multi-step fuzzy cognitive mapping approach. *Ecological Modelling*, 2004. 176(1-2): p. 43-64.
- [2] Zadeh, L.A., Fuzzy logic, neural networks, and soft computing. *Acm Transactions on Information Systems*, 1994. 37 p. 77 - 84.
- [3] Virapongse, A., et al., A social-ecological systems approach for environmental management. *Journal of Environmental Management*, 2016. 178: p. 83-91.
- [4] Stojanovic, T., et al., The "social" aspect of social-ecological systems a critique of analytical frameworks and findings from a multisite study of coastal sustainability. *Ecology and Society*, 2016. 21(3).
- [5] Gray, S., et al., Modeling the integration of stakeholder knowledge in social-ecological decision-making: Benefits and limitations to knowledge diversity. *Ecological Modelling*, 2012. 229(0): p. 88-96.
- [6] Papageorgiou, K., et al., Fuzzy Cognitive Map-Based Sustainable Socio-Economic Development Planning for Rural Communities. *Sustainability*, 2019. 12(1).
- [7] Obiedat, M. and S. Samarasinghe, A novel semi-quantitative Fuzzy Cognitive Map model for complex systems for addressing challenging participatory real life problems. *Applied Soft Computing*, 2016. 48: p. 91-110.
- [8] Ozesmi, U. and S. Ozesmi, A participatory approach to ecosystem conservation: Fuzzy cognitive maps and stakeholder group analysis in Uluabat Lake, Turkey. *Environmental Management*, 2003. 31(4): p. 518-531.
- [9] Gray, S., et al., Using fuzzy cognitive mapping as a participatory approach to analyze change, preferred states, and perceived resilience of social-ecological systems. *Ecology and Society*, 2015. 20.
- [10] Axelrod, R., *Structure of Decision: The Cognitive Maps of Political Elites*. 1976, Englewood Cliffs, New Jersey: Prentice-Hall.
- [11] Zadeh, L.A., Fuzzy Sets\*. *Information and Control* 1965. 8(3): p. 338-353
- [12] Ross, T.J., *Fuzzy Logic with Engineering Applications* 2004: John Wiley & Sons, Ltd.
- [13] Kosko, B., Fuzzy cognitive maps. *International Journal of Man-Machine Studies*, 1986. 24(1): p. 65-75.
- [14] Mobolurin, A., Generating consensus fuzzy cognitive maps, in *Proceedings of the 1997 IASTED International Conference on Intelligent Information Systems (IIS '97)*. 1997, IEEE Computer Society.
- [15] Nair, A., D. Reckien, and M.F.A.M. Maarseveen, A generalised fuzzy cognitive mapping approach for modelling complex systems. *Applied Soft Computing*, 2019. 84(3): p. 1-13.
- [16] Tan, C., H. Teo, and K. Wei, Promoting consensus in small decision making groups. *Information & Management*, 1995. 28(4): p. 251-259.
- [17] Papageorgiou, E.I., A.T. Markinos, and T.A. Gemtos, Soft Computing Technique of Fuzzy Cognitive Maps to Connect Yield Defining Parameters with Yield in Cotton Crop Production in Central Greece as a Basis for a Decision Support System for Precision Agriculture Application, in *Fuzzy Cognitive Maps*. 2010, Springer Berlin Heidelberg. p. 325-362.
- [18] Kumar, A. and G. Karmakar, Aggregation of opinions using fuzzy numbers. *International Journal of Systems Science*, 2001. 32(12): p. 1399-1411.
- [19] Mehryar, S., et al. Making Use of Fuzzy Cognitive Maps in Agent-Based Modeling. in *Advances in Social Simulation*. 2020, Cham: Springer International Publishing.
- [20] Wu, W., et al., Including stakeholder input in formulating and solving real-world optimisation problems: Generic framework and case study. *Environmental Modelling & Software*, 2016. 79: p. 197-213.
- [21] Gregory, A.J., et al., Stakeholder identification and engagement in problem structuring interventions. *European Journal of Operational Research*, 2020. 283(1): p. 321-340.
- [22] Yager, R.R., On ordered weighted averaging aggregation operators in multicriteria decision-making. *Ieee Transactions on Systems Man and Cybernetics*, 1988. 18(1): p. 183-190.
- [23] Obiedat, M., Robust semi-quantitative fuzzy cognitive map model for complex systems and case study: mitigating the water scarcity problem in Jordan, in *Department of Environmental Management*. 2013, Lincoln University: Lincoln University. p. 277.
- [24] Obiedat, M. and S. Samarasinghe, Fuzzy Representation and Aggregation of Fuzzy Cognitive Maps, in *20th International Congress on Modelling and Simulation*. 2013: Adelaide, Australia.
- [25] Obiedat, M., S. Samarasinghe, and G. Strickert, A New Method for Identifying the Central Nodes in Fuzzy Cognitive Maps using Consensus Centrality Measure, in *19th International Congress on Modelling and Simulation*. 2011: Perth, Australia.
- [26] Herrera, F. and L. Martinez, A 2-tuple fuzzy linguistic representation model for computing with words. *Ieee Transactions on Fuzzy Systems*, 2000. 8(6): p. 746-752.
- [27] Dunn, J.C., A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 1973. 3(3): p. 32-57.
- [28] Bezdek, J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*. 1981: Kluwer Academic Publishers.
- [29] Hathaway, R.J. and J.C. Bezdek, Recent convergence results for the fuzzy c-means clustering algorithms. *Journal of Classification*, 1988. 5(2): p. 237-247.
- [30] Nayak, J., B. Naik, and D.H. Behera, Fuzzy C-Means (FCM) Clustering Algorithm: A Decade Review from 2000 to 2014. 2015. p. 133-149.
- [31] Ruspini, E.H., A new approach to clustering. *Information and Control*, 1969. 15(1): p. 22-32.
- [32] Rousseeuw, P.J., Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 1987. 20: p. 53-65.

# Wind Power Integration with Smart Grid and Storage System: Prospects and Limitations

Saad Bin Abul Kashem<sup>1</sup>  
Robotics and Advanced Computing  
faculty  
Qatar Armed Forces – Academic  
Bridge Program, Qatar Foundation  
Doha, Qatar

Muhammad E. H. Chowdhury<sup>2</sup>,  
Amith Khandakar<sup>3</sup>  
Department of Electrical Engineering  
Qatar University  
Doha, Qatar

Jubaer Ahmed<sup>4</sup>  
Faculty of Engineering, Computing  
and Science  
Swinburne University of Technology  
Sarawak  
Sarawak, Malaysia

Azad Ashraf<sup>5</sup>  
Dept. of Chemical Engineering and Process Technology  
College of North Atlantic  
Qatar

Nushrat Shabrin<sup>6</sup>  
Business Faculty  
University Malaysia Sarawak  
Sarawak, Malaysia

**Abstract**—Wind power generation is playing a pivotal role in adopting renewable energy sources in many countries. Over the past decades, we have seen steady growth in wind power generation throughout the world. This article aims to summarize the operation, conversion and integration of the wind power with conventional grid and local microgrids so that it can be a one-stop reference for early career researchers. The study is carried out primarily based on the horizontal axis wind turbine and the vertical axis wind turbine. Afterward, the types and methods of storing this electric power generated are discussed elaborately. On top of that, this paper summarizes the ways of connecting the wind farms with conventional grid and microgrid to portray a clear picture of existing technologies. Section-wise, the prospects and limitations are discussed and opportunities for future technologies are highlighted. It is envisaged that, this paper will help researchers and engineering professionals to grasp the fundamental concepts related to wind power generation concisely and effectively.

**Keywords**—Wind power system; wind turbines; energy storage system; microgrids; nation grids

## I. INTRODUCTION

Wind energy has been widely used as a renewable resource to generate electricity in some countries such as the United States. This is because wind energy is relatively low cost and environmentally- friendly compared to other non- renewable energy resources such as coal, natural gas &etc. Due to its advantages, many types of research had been carried out for further development of this energy source especially for the integration of wind energy with the smart grid. Hence, the aim of this research is an attempt to focus on the study of prospects and limitations of wind power integration with its power storage system and grid system. In this research, there is no simulation tool or experimental tool is used. This research is focusing on the peer review of the latest research papers that related to our topic. In Section 2 and 3, the use of vertical and horizontal axis wind turbines for a wind power system is reviewed. The energy storage system will be discussed in

Section 4. The integration of wind power with the microgrid and nation grid will be discussed in Sections 5 and 6, respectively.

## II. VERTICAL AXIS WIND TURBINE (VAWT)

The increment of renewable energy is very promising due to the rapid depletion of fossil fuels [1]. For the past decades, horizontal axis wind turbine (HAWT) has been widely used in commercials to generate electricity with higher capacity. However, a wind turbine with a vertical axis has been introduced to the world in the 1900s [2]. Vertical axis wind turbine (VAWT) has been first developed and studied in North America [3]. For instance, Cheng, et al. [3] said that the world's largest vertical axis wind turbine had been built in Canada in the year 1986. VAWT was proposed to encounter the problem of power losses of HAWT. It is the major drawback of HAWT. The use of gear mechanism on HAWT contribute a lot of power losses and high maintenance cost is needed for HAWT. All these factors lead to an increase in the total cost of HAWT. However, VAWT can overcome these problems by replacing the gear box to direct drive generator system which reduces power losses and maintenance costs [4]. Fig. 1 shows the VAWT commonly used in an urban area [8].

### A. Types of VAWT

The design of VAWT is different from HAWT [1, 3]. VAWT is surrounded by its shaft along the vertical axis. The size of the VAWT is smaller than HAWT. The rotation of the turbine designed to be perpendicular to the wind direction to generate electricity. The generator and battery are connected at the bottom of the VAWT. Different type of rotor blade design determined the types of VAWT [5, 6]. Drag-type and lift-type are the most common types of VAWT around the world.

### B. Drag Type

Drag type VAWT generates electricity by converting the torque created by the rotation of the wind turbine. A plurality of flat or concave cup-shaped blade is the most common design

for drag type VAWT. From Fig. 2, normally the drag type devices consist of two or three scoops of the s-shape aerodynamic plate [7]. The blade would experience more drag when it is moving with the wind by less drag when it is against the wind direction. The spin occurs when there is differential drag created by the scoops on the VAWT. According to Wong, et al. [5], the most commonly used rotor for drag type is a Savonius rotor and Sistan rotor. It uses the blades as the obstacle to the wind which utilizes the force to rotate the wind turbine. Nevertheless, it has the better self-start ability compare to lift type VAWT but lower power coefficient on the system [5].

### C. Lift Type

Lift type VAWT has better efficiency compared to drag type. The blade of lift type VAWT is designed to aerofoil-shaped blades. The blades rotate by the aerodynamic lift forces that created when the air passes through the wind turbines. There is always positive static torque that can be observed in helical blades regardless of the incident angles shown in Fig. 3. Toja-Silva, et al. [8] had proved that the helical Savonius wind turbine has a higher power coefficient compare to traditional Sanonius rotor wind turbine. However, it lacks a 'self-initiate' force compare to drag type VAWT [5, 7]. Therefore, a hybrid combination of both drag and lift type of wind turbine are proposed to encounter the weaknesses [5, 8, 9].



Fig 1. Hybrid VAWT.

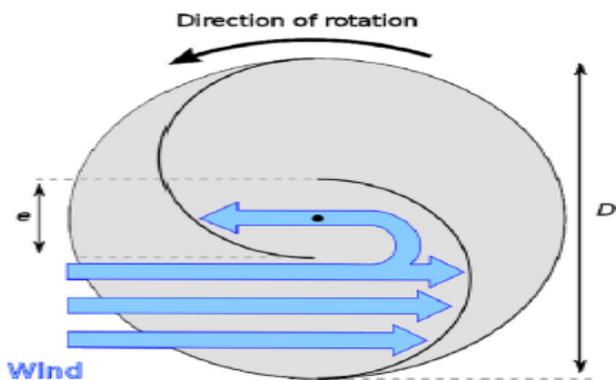


Fig 2. Working Mechanism of Drag-Type VAWT.

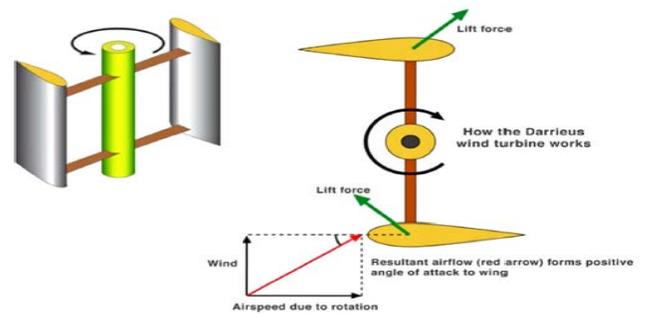


Fig 3. Working Mechanism of Lift-Type VAWT.

### D. Working Mechanism

According to Shahariar et al., [10], the blade on the VAWT is designed to be symmetrical and the angle of the blade is set corresponding to the turbine. By doing this, VAWT can capture the wind in all directions. Fig. 4 shows the various components of the VAWT [10]. First of all, the rotor in VAWT is used to convert the wind energy into mechanical. The rotor will be installed in the hub. The hub is connected to all the rotor blades due to the magnetic levitation. The hub is suspended on the magnets to increase the efficiency of the output. The most common material for the hub is cast iron and cast steel [10]. Besides, the shaft below the hub gets turned by the rotation of rotor blades. Then the generator below the shaft converts the mechanical energy into electrical energy.

1) *Magnetic levitation:* To maximize the rotor efficiency, magnetic levitation was proposed in the design of VAWT [1]. The aims of proposing magnetic levitation are to nullify the weight of the rotor in the system. Levitation magnetic system is done by installing two sets of magnetized ring magnets in such a way that repelling with each other. For instance, there is a magnet that will be installed at the shaft and another magnet under the supporting hub. With this, the repulsion of the magnet can partially nullify the hub and other rotating parts. Therefore, it can reduce the friction between the rotor and mechanical parts in the hub.

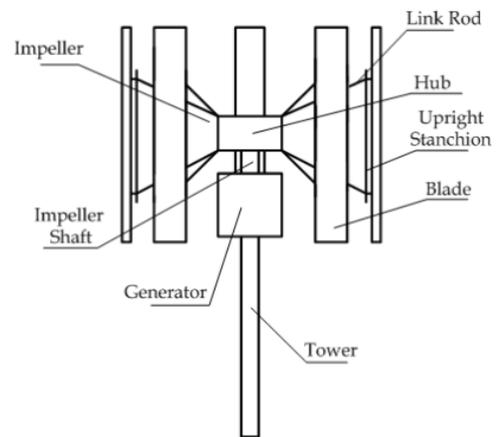


Fig 4. Various Component of the Typical VAWT.

2) *The airfoil of the Turbine*: According to Jin, et al. [7], lift-type VAWT will gradually become the representative of the VAWT because it has a higher efficiency compared to drag-type VAWT. The design of aero-foil is essential to improve the Darrieus type turbine [1]. Nevertheless, the shape and thickness of the aerofoil are determined by the specific data. The National Advisory Committee for Aeronautics (NACA) and Sandia National Laboratory had co-operated to come out with most of the review and data for the different types of aerodynamic characteristics for aerofoil. For instance, the lift and drag coefficient of different type of aerofoil with Reynolds number and ranging angle between 0 to 180° are provided by NACA.

#### E. The Solidity of the Turbine

The number of blades is depending on the smoothness of the rotor operation [11]. However, the number of blades correlated with the solidity. The solidity of the turbine is defined by the ratio of the total swap area and the radius of the turbine. The number of blades and length of the chord is influencing the total area. Nevertheless, all these parameters are affecting the efficiency of the self-starting of the wind turbine. Solidity ( $\sigma$ ) can be computed by using (1) [11].

$$\sigma = \frac{Nc}{R} \quad (1)$$

Where N is the number of blades, C is the length of the chord (m) and R is the radius of the rotor (m). Nonetheless, the number of blades is also affecting the power coefficient of the wind turbine.

#### F. Power Coefficient

Furthermore, the Betz limit is an essential and powerful performance indicator for VAWT [10]. The upstream and downstream stream wind for VAWT is different. Therefore, it is important to study the available wind kinetic energy can be extracted from a wind generator. Furthermore, the theoretical maximum rotor power coefficient, that achieved by the past research is 0.59. However, Shahariar and Hasan [10] had proved that the two blades VAWT has the 0.46 of practical limit for under high wind and the three-blade VAWT has 0.50 of . Nevertheless, the drag-type VAWT only can operate at 1/3 of the Betz limit [12]. There are (2) and (3) used to determine the power co-efficient or efficiency of the rotor [1].

$$C_p = \frac{\text{captured mechanical power by blade}}{\text{Available power in the wind}} \quad (2)$$

$$C_p = \frac{T\omega}{0.5\rho AV^2} \quad (3)$$

#### G. Efficiency of VAWT

According to Shahariar et al., [10] and Zhuga et al., [13], the wind turbine efficiency can be computed by using (4).

$$P_t = 0.5\rho AV^3 \quad (4)$$

The power produced by the generator and the amount of wind energy that can be captured can be determined by the (5).

$$P_a = T\omega \quad (5)$$

Whereas the torque, T can be determined by (6).

$$T = 0.5\rho AC_t V^3 \quad (6)$$

The angular velocity or RPM of the VAWT  $\omega$  can be computed by (7).

$$\omega = \frac{\lambda V}{R} \quad (7)$$

According to Ragheb [14], aerodynamic turbine (lift-type) has a higher efficiency compared to impulse devices (drag-type). For instance, lift-type has an efficiency of approximately 60percent whereas the drag-type only 19-40 percent. This is because there is the possibility of the zero wind speed behind the turbine which clogging any flow through the rotor.

Power efficiency and rotor efficiency of VAWT are determined by the blades design and the rotor design [1, 15, 16]. The aerofoil shape design is essential for the blade to promote or ensure the efficiency of the wind turbine [15, 17]. Besides, the number of blades also affects the smoothness of the rotor operation of the wind turbine. Cheng, et al. [3] had performed the changes in wind turbine efficiency by using a different number of blades. The number of blades is the main factor that affects the aerodynamic load on the wind turbine which includes thrust and side force. According to the researcher Cheng, et al. [3], increasing the blades can reduce the fatigue and bending moment of the wind turbine. Nevertheless, three to four blades are reaching the limitation of additional effect [18].

Furthermore, flexible blades are proposed to increase the efficiency of VAWT [7, 8]. The main intention of proposing the flexible blade is to reduce the vibrations transmitted to the VAWT. Nevertheless, the deflector sheet on the VAWT can increase the efficiency of the system as well [8]. Cheng, et al. [18] had proved that the deflector sheet on VAWT shows the higher power coefficient compares to the rotor with two blades.

#### H. Advantages of VAWT

VAWT is workable under omnidirectional flow [14]. It can catch the wind in all directions. Therefore, VAWT does not require a yaw mechanism for operation [5, 15, 19]. It can sustain under much harsher environment namely higher wind turbulent condition compare to HAWT due to its omnidirectional characteristic. According to Ragheb [14], HAWT does not need to be shut down and it can sustain up to 50 m/s.

Besides, Wong, et al. [5] claims that the blade of VAWT can be easily manufactured compare to HAWT. The uniform section and untwisted characteristic of blades make them easy to fabricate. For instance, the blade of the VAWT does not require twisting or tapering to optimize the performance.

Furthermore, VAWT does not require to be built as high as HAWT. The size of the VAWT is much smaller than HAWT. According to Ragheb [14], a small VAWT only require 10m height to capture the wind. Therefore, they are less visible and can be widely used in the urban area.

Apart from that, there is lower rotational speed can be observed on the VAWT [8]. Therefore, the noise contributed by VAWT is lower than HAWT. The lower rotational speed also results in lower vibrations transmitted to the structure.

Besides, VAWT does not require a yaw-control mechanism. With this, the lower maintenance is required for VAWT thus lower cost is needed for VAWT.

### III. HORIZONTAL AXIS WIND TURBINE (HAWT)

Horizontal axis wind turbine (HAWT) was one of the most common types of wind turbines used for generating wind power [20]. It was being used in both large and small scale. Large-scale utilization of HAWT was usually adopted in wind farms or offshore areas; while small-scale usage of HAWT usually occurred within the residential area, normally integrated with solar panel system [21, 22]. The rotor blades of this particular type of wind turbines are rotating about the horizontal axis in a vertical plane[23]. The main components of horizontal axis wind turbines were the rotor blades, gear-box, electrical generators and the tower-like supporting structure. One of the distinct advantages of HAWT was that it was equipped with a tower, which was able to expose the wind turbine to higher wind speed at an elevated height [23, 24]. Fig. 5 and 6 show the typical 3 blade HAWT and components of horizontal axis wind turbine respectively [21, 35].



Fig 5. Typical 3 Blade Horizontal Axis Wind Turbine.

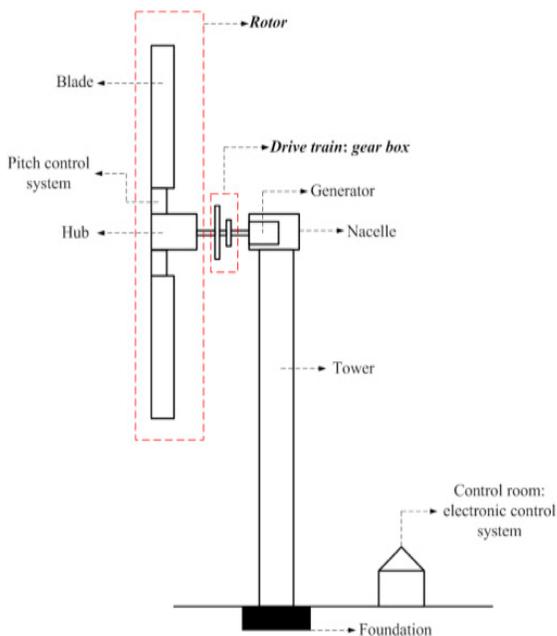


Fig 6. Components of Horizontal Axis Wind Turbine.

#### A. Working Mechanism of HAWT

HAWT had to face the wind direction for it to work with the maximum efficiency. It uses the aerodynamics force of lift to convert the wind energy into electrical power [25]. Most of the rotor blades of HAWT was designed to be airfoil shape. Fig. 7 shows the typical airfoil-shaped blades for HAWT [26].

When the wind blows through the airfoil-shaped rotor blades with the configuration in Fig. 7, the wind velocity that passes through the upper part of the blades would be faster compared to the wind velocity at the bottom of the blade [26]. As higher velocity would result in lower pressure, a pressure difference would occur between the upper and lower zone of the blades. This pressure difference would create a lift force to initiate the movement of the blades in rotational motion and cause the driveshaft to spin. Although the outer blades were rotating at a high speed, the driving shaft was not spinning as fast as the rotor blades. For most of the HAWT, a gearbox was installed within it to convert the low rotational speed of the drive shaft into a higher rotational speed [27]. This was to ensure that the rotation speed was fast enough for the electrical generator to generate power.

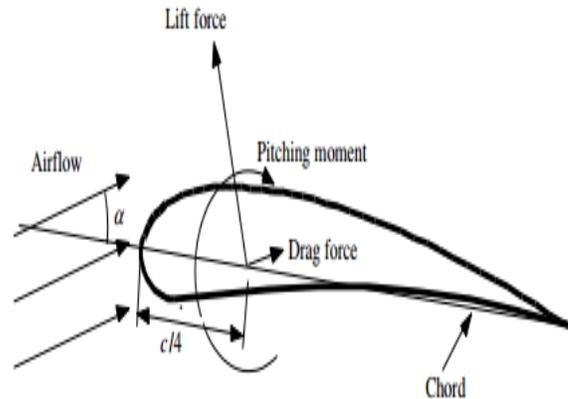


Fig 7. Airfoil-Shaped Rotor Blades for HAWT.

#### B. Power Generation in HAWT

The power that generated in HAWT can be computed through the equation:

$$P = \frac{1}{2} C_p \rho A v^3 \quad (8)$$

Where,  $C_p$  = Power Coefficient

$\rho$  = density of air

$A$  = area by the wind blades

$v$  = wind speed

#### C. Factors Influencing Power Generation in HAWT

A few factors affect the power generation in HAWT. These factors include the wind speed, power coefficient and the area swept by the rotor blades of HAWT.

1) *Wind Speed*: The power generation in the HAWT is strongly influenced by wind speed. The wind speed was classified into three different types, namely cut-in wind speed,

rated output wind speed and cut-out wind speed. The cut-in wind speed referred to the minimal wind speed required for the power generation to initiate [28, 29]. As mentioned in the equation, the power produced was governed by the wind speed, thus a lower wind speed results in a lower power being produced. As the rotor blades continued to rotate, a continuous increment in the wind speed would occur until it reaches rated output wind speed. At the rated output wind speed, the maximum output power could be produced by the electrical generator [30]. In cases where the wind speed went beyond the rated output wind speed, there was no increment in the power output because it was beyond the limit of the electrical generator. As the wind speed kept on increasing, a cut-out wind speed would be reached and the rotating of the rotor blade would be forced to stop to prevent overloading of the wind turbine [31]. Thus, normally in HAWT, the power-regulating system is required as it is often exposed to high wind speed [21, 23, 29]. One of the common power regulating system adopted in HAWT was the pitch angle control system [31]. Whenever the wind speed exceeded the rated wind speed value, the pitch controller would increase the pitch angle to reduce the rotation of the wind turbine [31].

2) *Power Coefficient,  $C_p$* : The output power is always lower than the actual power within the wind due to the energy lost during the conversion of wind energy into mechanical power. This conversion efficiency was known as Power Coefficient,  $C_p$  [32, 33].

$$C_p = \frac{\text{output power}}{\text{wind power}} \quad (9)$$

The theoretical maximum value of the power coefficient was found to be 0.593, known as the Betz limit [34, 35]. However, the practical power coefficient was not able to reach 0.593 due to different configurations of a wind turbine which leads to inefficiencies and losses of energy [36]. According to Sedaghat and Mirhosseini [35], modern HAWT worked with a power coefficient of 0.5, which is very close to the Betz limit. The power coefficient was closely related to the tip speed ratio. Tip speed ratio,  $\lambda$  was the ratio between the speed at the tip of the blade and the wind speed.

$$\lambda = \frac{\text{Speed at the rotor tip}}{\text{Wind Speed}} \quad (10)$$

An optimum tip speed ratio was required for maximizing the efficiency of the conversion [36]. Aside from that, HAWT utilized the lift force instead of drag force for rotating the blades hence the conversion of energy was more efficient [25]. This would result in a higher power coefficient.

3) *Swept Area*: The swept area of the wind blades had a significant impact on the power generated. The swept area referred to the areas being swept by the blades when rotating Fig. 8 shows a detailed view of the swept area by HAWT [25].

The equation used to calculate the swept area of HAWT was the same as calculating the area of a circle.

$$\text{Swept Area, } S = \pi \left(\frac{d}{2}\right)^2 \quad (11)$$

where the  $d$  value referred to the rotor diameter [37]. An increment in the rotor blades length would increase the swept area of the wind turbine. HAWT normally had a large swept area, therefore the installation of the HAWT had to take account of the area swept when operating.

#### D. Yaw Mechanism for HAWT

As wind direction often changes, HAWT was required to install the yaw system [20]. The yaw system was able to control the rotor blades and ensure that it faced the wind direction whenever the wind direction changes [38]. However, one of the limitations of this yaw control system was that when the wind turbine was subjected to a sudden change in wind direction, it would result in the wind turbine to run with yaw errors. Yaw error was a term used to describe the situation where the wind direction was not perpendicular to the rotor blades; normally occurred when the turbine was changing its direction [20]. When the yaw error occurred, the power generated would drop significantly.

#### E. Other Considerations

1) *Noise*: HAWT usually causes a lot of noise when the rotor blades were spinning. This was because the blades of HAWT was usually exposed to higher wind speed as it was installed at the top of the tower [25]. Apart from that, HAWT usually operates at a higher tip speed ratio, result in creating a lot of noise from the wind turbine [20, 36].

2) *Cost*: The cost analysis of HAWT had to take into consideration a few different aspects. As HAWT had been widely used in the past decades, cheaper solutions in manufacturing HAWT had been developed in these years [20]. Despite having cheaper manufacturing costs, the installation and maintenance cost of the HAWT was expensive. This was because most of the parts of HAWT was located at the higher part of the tower which complicates the installation and the maintenance process [20].

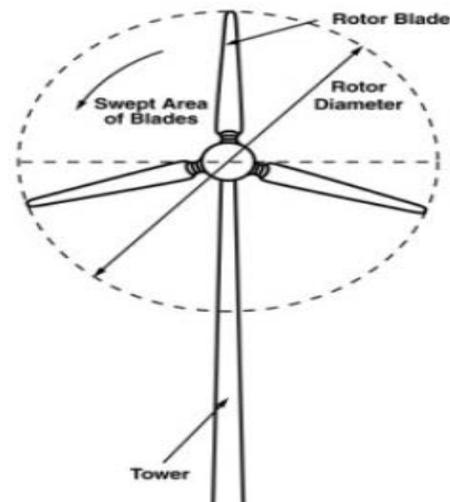


Fig 8. Area Swept by Horizontal Axis Wind Turbine.

#### IV. WIND ENERGY STORAGE SYSTEM

##### A. Energy Conversion in Wind Energy System

The generation part of a wind energy system is done by a wind turbine which comes in different sizes and a different number of blades. The kinetic energy in the flowing wind has been converted into mechanical energy by the turbine to spin the generator for the generation of electricity. The generated electrical power will either be in the form of Alternating Current (AC) or Direct Current (DC) depends on the design of the wind turbine. Electrical cables will be used to transmit the generated electricity to other electrical components in the system for different applications. If the output of the wind generator is DC, a charge controller or voltage regulator will be used to convert the DC to a suitable range of voltage level according to the DC appliances, to charge the batteries, or to be injected into the local DC microgrids. Besides, DC can also be transformed into AC with a voltage level of 240 V (single-phase) or 415V (three-phase) and frequency of 50Hz by an inverter. The AC will be used to supply the local AC appliances or its voltage will be stepped up further by the power transformer for transmission purposes to reduce the power loss [39]. The AC can be injected into either the local microgrids or the national grids to supply the power demand in other areas. If the generated power by the generator is already in AC, the inverter will not be required for the system but a rectifier will be needed to convert the AC into DC so that it can be stored in the battery bank of the system. A sample diagram of the flow of a wind energy system has been shown in Fig. 9 [40].

##### B. Rectifier

If a wind turbine's generator uses an alternator, the output of it will be in three-phase AC. A three-phase rectifier is needed to convert the AC output from the generator into DC before connecting to the voltage regulating device to charge the battery bank [41]. A rectifier is a Power Electronics Device (PED) which comprises several diodes/thyristors. However, the output of a rectifier is always not plain DC but pulsating DC which will be regulated before it has been used to charge the battery bank. The battery bank will provide a smooth and continuous DC output for the circuit in the DC appliances to work properly. Fig. 10 shows the schematic diagram of a grid-tied wind-mill system with the application of a rectifier [42].

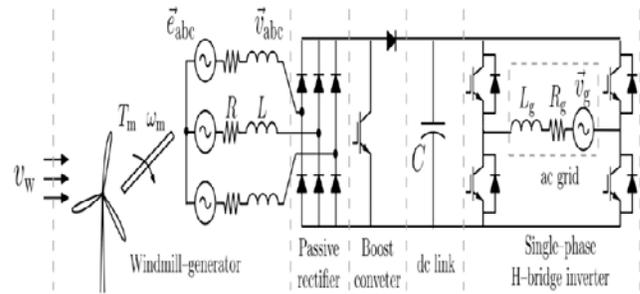


Fig 10. The Application of Rectifier in Wind Energy System.

##### C. Regulator and Charge Controller

A charge controller or a regulator is the device that limits and controls the amount of current that will be used to charge the battery bank. Unlike solar charge controller which will disconnect the connection between the source (solar panels) and charge controller itself when the battery is fully charged, a wind turbine charge controller needs to monitor the battery voltage constantly and able to turn on a dump load (also known as a shunt) if that voltage approaches the float charging voltage of the battery instead of shorting the output of the wind generator as shown in Fig. 9. A dump load, in this case, could a supplementary water heater or underfloor heating device instead of a braking resistor to optimize the generated energy. While a wind generator is rotating at high speed, sudden disconnection of the load will create a huge current spike that might destroy the charge controller and leads to over-speeding of the turbine in which will damage itself. There are two major types of charge controller technologies that can be found on the market where they are Pulse Width Modulation (PWM) and Maximum Point Power Tracking (MPPT). Fig. 11 is showing an example of the implementation of an MPPT charge controller in a wind energy system [43].

##### D. Grid Tie Inverter and Transformer

An inverter is a device that injects electrical power generated by any distributed power generation source in a stand-alone system, micro-grid or the nation grid [44]. When the output from the wind generator is in DC, a power inverter will be needed to convert it to AC before it has been supplied to any AC appliances or injected into the grid. A Step-up transformer will be needed to boost the voltage level which can reduce power loss during the transmission of electricity through the grid. The transformer is normally located at the base of the wind turbine. However, grounding transformer will be located at the critical points through the wind farm which provides a neutral point where all the electricity is then interconnected with a major collector step-up transformer that is connected to the grid.

##### E. Energy Storage System

As mentioned above, an energy storage system will be used to store the excess energy that is generated by the system to be used when the energy demand exceeds the output power of the wind energy system. The electrical energy generated from a wind turbine can be stored in different forms of energy.

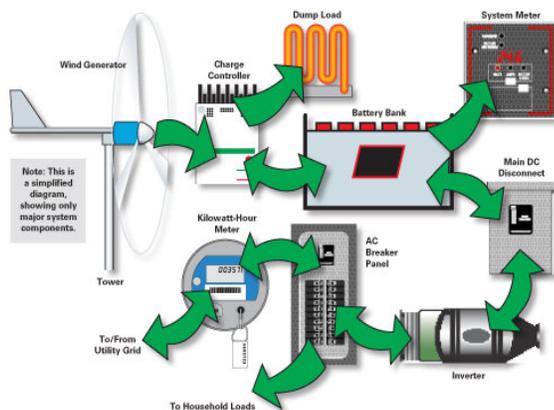


Fig 9. Grid-Tied Wind Energy System with Battery Storage.

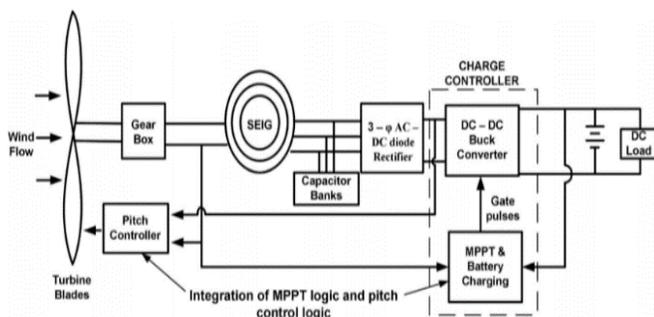


Fig 11. Integration of MPPT Charge Controller with Wind Energy System.

1) **Lead Acid Battery:** There are two major lead-acid batteries that can be found in the current market where they are Flooded Lead Acid (FLA) and Valve Regulated Lead Acid (VRLA). FLA batteries are built from two lead plates which are immersed in the electrolyte which is sulphuric acid in this case. Meanwhile, VRLA batteries operate in the same way while having an extra sealed pressure-regulating valve that works to prevent the hydrogen from venting and release the air in the battery cell [45]. VRLA batteries are having a higher price and lower life-cycle compared to FLA batteries but it comes in smaller size and weight while being maintenance-free. Lead-acid battery often being considered as a support of the wind power system compared to other types of the battery due to its small power density, lower depth of discharge and the shorter number of its life cycle.

2) **Nickel Cadmium Battery:** Nickel Cadmium battery is using nickel hydroxide and metallic cadmium as its electrode while both electrodes are separated by nylon divider and having aqueous potassium hydroxide as the electrolyte for the battery. Nickel Cadmium Battery can operate at a wider temperature range compared to a lead-acid battery. It also offers excellent cycle life and exceptional tolerance of high discharge rates. However, the toxicity of Cadmium and direct competition from other batteries like Lithium-ion leads to its decreased popularity since the 1990s [46]. It has also been banned by some European countries due to environmental issues.

3) **Lithium-Ion Battery:** Lithium-Ion battery is using litigated metal oxide as its cathode and graphite carbon with layer structure as its anode [47]. It is widely used in consumer electronics, automotive and energy storage systems especially for large scale applications like wind farm storage. Lithium-Ion battery comes in small weight, high cell voltage, high overall cell efficiency and power density. Lithium-Ion batteries are also having long life deep cycles which means significantly more to deep discharging. It is also one of the cleanest battery technology that can be found on the market that is having up 0.1 small self-discharge per month [45]. However, a lithium-ion battery with all these features and high-quality performance as shown in Fig. 12 also comes with an extremely high initial cost compared to other types of battery like lead-acid and Nickel Cadmium [47].

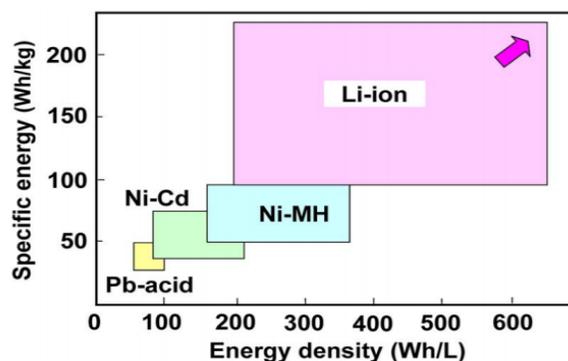


Fig 12. Comparison of Batteries.

4) **Sodium Sulphur Battery:** Sodium Sulphur battery is a kind of high-temperature battery that is usually constructed in a tall cylindrical configuration. The anode and cathode of the battery are made by molten Sulphur and molten sodium respectively while the electrolyte, in this case, is normally  $\beta$ -alumina. A diesel Genset is often installed along with a system that is using Sodium Sulphur batteries because the batteries will suffer some serious damage if it is cooled down without being fully charged. The Sodium Sulphur battery can become a serious player in large-scale energy storage application due to its high energy density and deep cycles in comparison to the Lead-Acid battery. The initial cost for Sodium Sulphur battery is relatively high for now because it is a new battery technology where there is only one manufacturer is producing it, NGK Insulators (Japan) [45]. However, the price is considered to drop with a mass production since the materials of the battery are inexpensive, abundant and recyclable. A schematic diagram of a central sodium electrode tubular Sodium Sulphur cell has been shown in Fig. 13 and Fig. 14 shows the discharge reaction scheme of Sodium Nickel Chloride Battery [48].

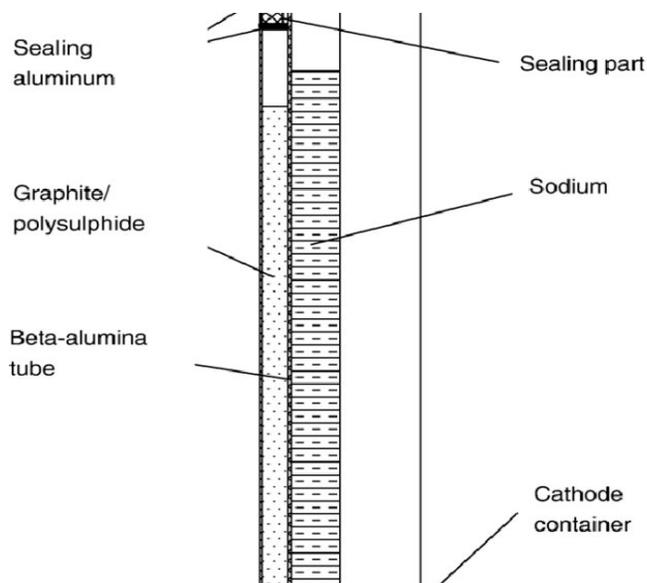


Fig 13. The Discharge Reaction Scheme of Sodium Nickel Chloride Battery.

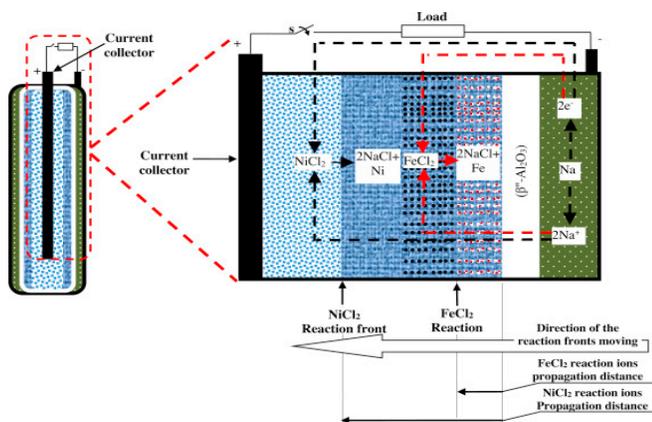


Fig 14. The Discharge Reaction Scheme of Sodium Nickel Chloride Battery.

5) *Flow Battery*: Flow Battery provides a well-balanced solution for a large-scale energy storage system for renewable energy due to its flexible modular design and the ability to store a large amount of energy up to Megawatts (MW) [49-51]. There are three major types of the flow battery can be found on the market where they are Vanadium Redox, Polysulphide Bromide and Zinc Bromine. The operation principle of a flow battery is different from other conventional batteries for example Lead Acid and Lithium-Ion. Its rechargeability is the result of two chemical components that dissolved in the electrolytes of the system which is normally separated by a membrane [45]. The reversible reaction between the two electrolytes allows the potential chemical energy to be stored in the electrolyte solution. The power and energy capacity is determined by the size of the cell while the energy capacity is defined by the volume of the electrolyte. A schematic diagram of a flow battery has been shown in Fig. 15 [52].

6) *Supercapacitor*: A supercapacitor is designed based on the same physical principles as other conventional capacitors but having a higher area and thinner electrodes. Supercapacitor uses polarized liquid layers between the conducting electrolyte and electrode to increase the overall capacitance [45]. The electrodes of supercapacitors are made of highly effective surface materials, for example, porous carbon and carbon aerogel [53]. The two electrodes of the supercapacitor allow a potential to be applied across its cell which presents two double layers in it. There will be an ion-permeable separator being installed between the electrodes to allow the flows of an ion from the electrolyte but blocks all electrical contact. This type of storage can suppress fast wind power fluctuations within a short time frame. However, it is having a lower energy density in comparison with other conventional batteries but comes with a higher initial cost. Thus, it can be integrated with the battery system as a support for the wind turbines instead of stand-alone as shown in Fig. 16 [54].

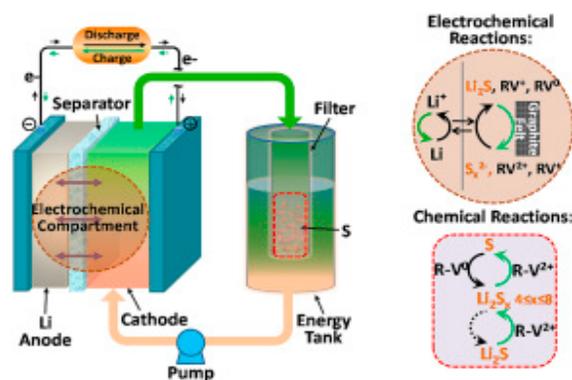


Fig 15. Schematic Diagram of the Flow of Discharge for a Flow Battery.

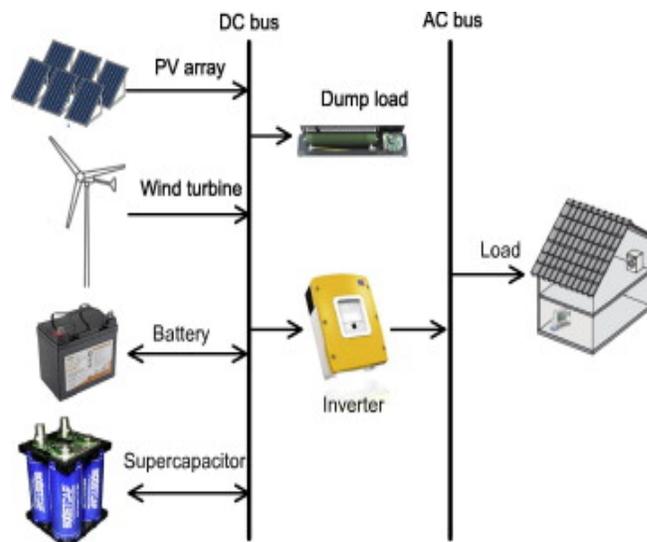


Fig 16. Sample Implementation of Supercapacitor with Battery Storage for Renewable Energy System.

## V. CONCEPT OF MICROGRID

Microgrid (MG) is an adjacent section of the main grid which can be formed in either single or multiple distributed generation (DG) units to provide sufficient power for the heavy load consumption within the grid [55]. Aziz et al., [56] stated that micro-grid is the integration of several distributed energy resources (DERs) that interconnected to each other to produce the maximum amount of electric power to a group of consumers when it operated as an island and switching the power with existing grid under grid-connected mode. Fig. 17 shows the overall structure of the micro-grid [56]. A micro-grid can be operated in the grid-connected mode or autonomous mode when it is connected to or disconnect from the main grid. It usually located at any site in the grid system to enhance system reliability and efficiency. Hatziaargyriou et al., [57] agreed that a micro-grid can be operated during grid-connected mode or autonomous mode when the system is disconnected from the main grid due to the faults occur in the system.

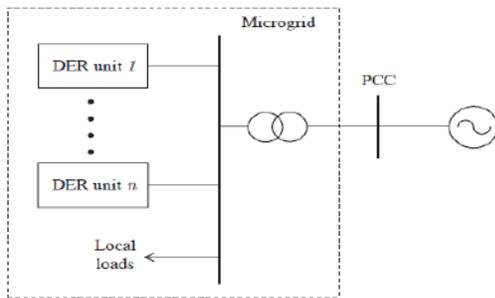


Fig 17. Micro-Grid Structure Formed by a Group of DER Units and Local Loads.

#### A. Power Control Strategies

Due to the intermittent characteristic of wind energy, the demand response (DR) is implemented into the wind energy generation system to enhance the grid system by controlling the energy consumption in the grid, maintaining its voltage profile and providing reliable energy for micro-grid [55]. This article had mentioned that the micro-grid can be used to improve power quality by regulating voltage profile, reducing harmful emissions and supplying consumers with low-cost energy supply [57]. Demand response devices with a storage system can be used to regulate the frequency caused by the intermittency of the wind [58]. With this demand response, the energy consumption is reduced during peak hours and the remained energy is shifted for off-peak hours used [55]. Besides, the consumers can incorporate with the system's producers through shifting load profiles to get incentives concerning the electricity price. This obviously would promote economic benefits for both consumers and utilities by reducing the capital investment cost and mitigate the rate of grid maintenance.

Consortium for Electric Reliability Technology Solutions (CERTS) micro-grid is one of the alternative ways to improve system efficiency by integrating more distributed energy resources (DERs) into the grid system [59]. In CERTS micro-grid, there are two components which are the static switch and micro source. The static switch is used to autonomously isolate the micro-grid from any disturbances whereas the micro source is used to power on the disconnected/isolated micro-grid using a power- frequency droop controller [59]. This is to ensure that new micro sources can be added to the existing system without causing any interruption to the system and each of the controllers can respond to the system changes. Fig. 18 shows the use of a static switch in the micro-grid system [59]. Under isolated grid mode, each micro source in the system must be able to autonomously respond to the system changes by adopting the local droop controllers for voltage and frequency regulation [58]. The author in [57] stated that the micro source controller controlled the voltage and frequency of the microgrid using the local information where it will execute the demands from a micro-grid central controller (MGCC) when the system is connected to the main grid whereas it will perform grid's optimization and load tracking autonomously when the system is being disconnected or isolated from the main grid [57]. The local load controller (LC) is installed at the controllable load and used for load management by providing the capability for load control with the instruction command

from MGCC. Furthermore, the Micro-grid Central Controller (MGCC) is used to maximize the grid value and optimize its operations [57, 60]. In this controller, the amount of energy supply will be drawn from the distribution system based on the electricity's prices in the market.

Besides that, this paper has stated two types of control strategies which are centralized control and decentralized control [58]. The decentralized control is the control of the system autonomously based on voltage and frequency measurement in a power system. In this case, a controller is used to generate maximum output power for the demand-supply [57]. The frequency restoration can be carried out by the application of decentralized proportional-integral frequency control. However, the errors in measurement may happen due to the application of control integral part to multiple resources. And thus, it is recommended to apply an integral part to the only single resource when there is no load sharing in the system [58]. Centralized control can be performed by sending control signals to different distributed energy resources through a series of communication channels that can be controlled from the control center. It is usually adopted with demand response to compensate for wind uncertainty. In this case, a low-pass filter is implemented in this control system to drive the slow changes of the power signal. The energy gained from the wind system will supply to distributed energy resources for the compensation due to the wind variability. The synchronous frequency control at the generator site is used to balance the uncompensated variable load and the wind by the centralized control system. During the contingencies condition when the generator is down, the frequency control will be carried out by the decentralized droop controllers [58]. A centralized micro-grid control system is typically built up with hierarchical control system architecture which consists of three operation layers including local micro source controller (MC) and load controller (LC), microgrid system central controller (MGCC) and distribution management system (DMS) [60]. The operation of MC, LC and MGCC had been mentioned previously. The distribution management system (DMS), also known as the distribution network operator (DNO) is responsible for the management of medium channel and low voltage area in micro-grid [57]. The DMS can be used to maintain the power quality within a stable and allowable limit in which it is controllable by the system operator [61]. Fig. 19 shows the hierarchical control system architecture of centralized micro-grid [62].

Moreover, in terms of cost-effective, the doubly-fed induction generators (DFIGs) can be used to provide sufficient energy supply. For the DFIGs based wind power system, the stator is directly connected to the grid while the rotor winding is connected through the slip rings to a back-to-back converter [56]. However, there are significant issues such as static and dynamic issues that need to be considered in this case. The static issues will be the profitability of distribution system operators (DSOs) based on the installation of DER. Hence, the location and sizing of DERs would be the main concern for the placement of DFIGs wind turbine in the micro-grid as it would have a direct impact on the DSOs' profit. On the other hand, for the dynamic issues, the IEEE standard 1547 was implemented to ensure the safety of the distribution systems. In

this rule, it protects the systems by tripping and disconnecting the DERs from the grid for a certain period whenever the voltage or frequency is running out of the range [56]. Fig. 20 shows the voltage disconnection requirement by IEEE Standard 1547 [56].

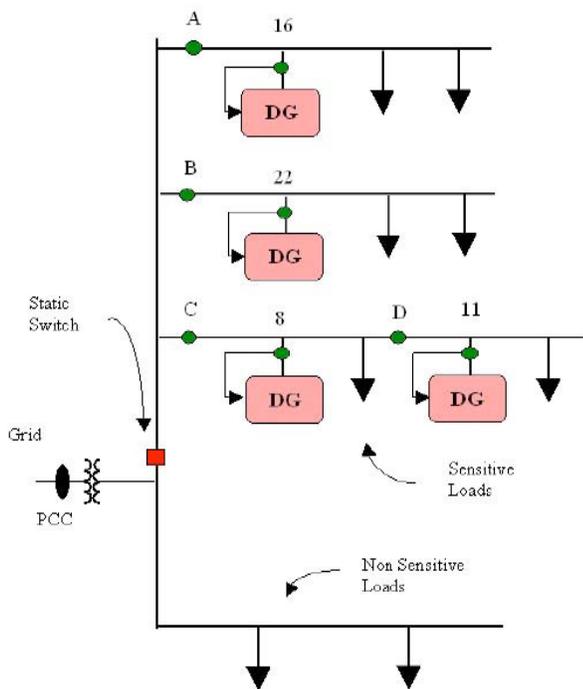


Fig 18. Use of Static Switch in Micro-Grid.

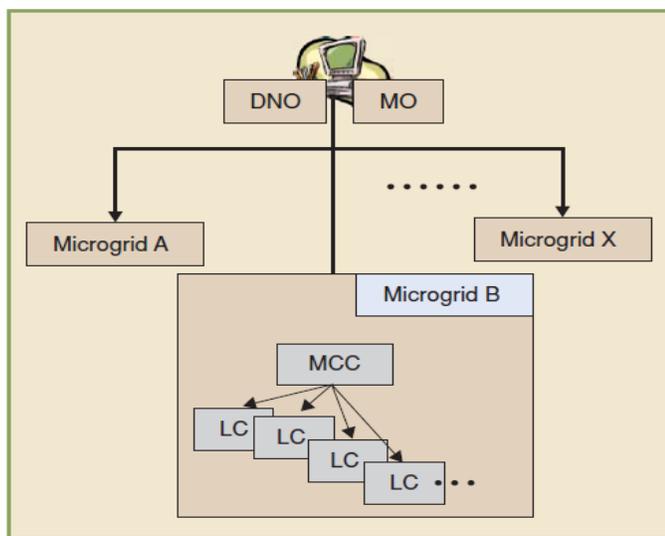


Fig 19. Hierarchical Control System Architectures.

Voltage Range (p.u.)	Clearing time (sec)
$V < 0.5$	0.16
$0.5 \leq V < 0.88$	2.00
$1.1 < V < 1.2$	1.00
$V \geq 1.2$	0.16

Fig 20. Voltage Disconnection Requirements by IEEE Standard 1547.

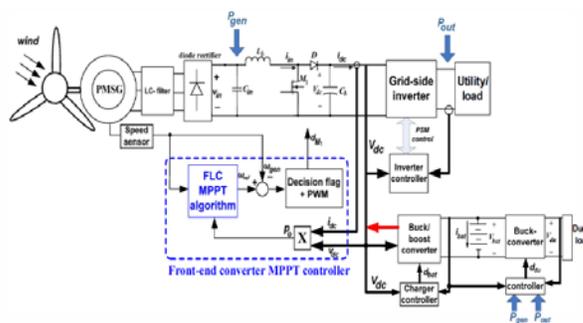


Fig 21. Front-End Converter MPPT Controller in the Wind Power System.

In [63], the authors proposed a control strategy based on the fuzzy logic controller to control the maximum power point tracking (MPPT) in the wind power system that connected to the low voltage grid. In this control system, there are two inputs and one output where the inputs are delta voltage and wind speed and the output is the duty cycle. The duty cycle is the amount of the pulse modulation that used to trip the boost converter to achieve the optimum voltage rating. As for the simulation results, it shows that the wind turbine system with MPPT controller can supply the power at its maximum ratings at a constant voltage and frequency [63]. Hui, J. and P. K. Jain also agreed that the generator load can be varied by monitoring the duty cycle of the front- end boost converter to achieve the MPPT and Fig. 21 shows the wind power system using the front- end converter MPPT controller [64].

### B. Energy Management Strategies

As the cost and efficiency of the grid operation are the main concerns, the energy management system plays an important role in the control system of the micro-grid [65]. Katiraei et al., [62], stated that a fast response to energy management strategy (EMS) is essential for a micro-grid that is operated under autonomous mode with more than two distributed energy resource units. It is because of the multiple DER units with different power capacities and lack of infinite bus in the system. Thus, the purpose of EMS is to distribute the real and reactive power equally among the DER units and restore the frequency by determining the power setpoints.

The use of energy management and power control (EMPC) systems in a wind-generated micro-grid especially at the low wind speed area has been investigated [66]. The energy management and power control (EMPC) system are used to stabilize the start-up- shutdown processes in each turbine in order to provide the maximum output to the grid that located in low wind speed area and also to protect the mechanical parts of the equipment from breakdown due to the frequent start-up and shutdown. The system is becoming less cost-effective as there are several parameters involved. To achieve cost-effective, a more adaptable control strategy called fuzzy- logic control is used [66]. These FLCs can provide maximum output power to the grid and ensure less mechanical stress on the mechanism parts yet to expand its lifespan. In [67], the authors proposed the use of the Economic model Predictive Control (EMPC) scheme to perform economic optimization in the operational costs of micro-grid. Fig. 22 and 23 shows the proposed EMPC system in wind power system at low wind speed region and the flowchart of FLCs in the system [66].

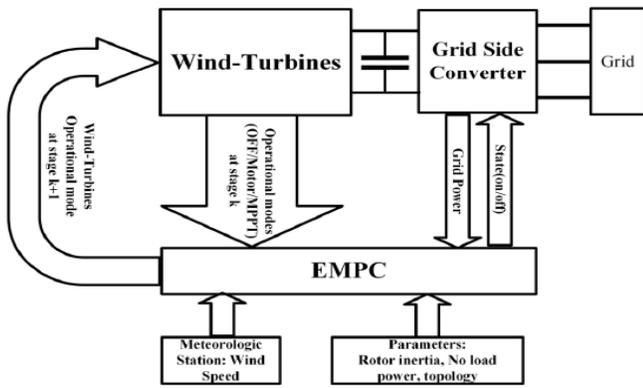


Fig 22. Proposed EMPC System for Low Wind Speed Area.

Other than that, the authors had proposed an energy management optimization model of wind-turbine based DC microgrid with batteries as the energy storage system [68]. This had overcome the intermittent characteristic of wind power in DC microgrid. The optimization of the energy resources can be achieved through the load and generation forecasting that related to electricity's price. Apart from that, the battery energy storage system (BESS) is proposed to be installed at the generation side to stabilize the uncertainty of wind speed [69]. This system can regulate the output power by incorporating other generators in the grid system to provide a balanced power supply. The energy produced during the strong wind will be stored in this BESS system and the stored energy will be supplied for grid used whenever the area is to be detected in the low wind speed situation.

VI. GRID INTEGRATION OF WIND TURBINE GENERATOR

A. Grid Connection of Wind Turbine Generator

Wind Turbine Generator (WTG) can be classified into a few types of connections with a power grid base on their different speed controls. They are Type 1, Type 2, Type 3 and Type 4 WTG. Camm et al. stated that Type 1 WTG has its rotating speed fixed to the power grid frequency and its connection to the power grid is shown in Fig. 24(a) [70]. It usually implemented on squirrel-cage induction generator (SCIG) with soft starter and parallel capacitor bank across the step-up transformer connected to the grid side. The soft starter and parallel capacitor bank eliminate the effects of consuming reactive power and drawing large currents because of its excitation field produced by the machine. The rotating speed of Type 2 WTG can be varied but it is limited. Fig. 24(b) shows the connection of Type 2 WTG to the main supply grid which is usually used in the wound-rotor induction generator [71]. The machine has a parallel connection with a variable resistor to maintain the power when gust and a capacitor bank to supply reactive power too.

The implementations of Type 3 and Type 4 WTG connections are getting more popular in recent years. Both Type 3 and Type 4 WTG are variable speed controlled. Type 3 WTG is usually implemented with a doubly-fed induction generator (DFIG) whereas Type 4 WTG is usually implemented with a permanent magnet synchronous generator (PMSG). The difference between Type 3 and Type 4 WTG is the connection and control of the power converter. Fig. 25(a) and (b) show the connection of Type 2 and Type 3 WTG with the power grid [70, 72]. Type 3 WTG is partially controlled whereas Type 4 WTG is fully controlled. The control method and capabilities of Type 3 WTG will be further discussed in Section 3. For Type 4 WTG, it is series connected with two back-to-back inverters with a Direct Current (DC) link in between them, to the power grid. The inverters are responsible for the control of power supply and regulate its voltage so that power grid synchronization and stability could be achieved. Zou and He agree with the statement and proposed a control model for it in [72]. The controller at the machine side controls the inverter to generate power from the wind turbine whereas the controller at the grid side controls the inverter to regulate the voltage supplied to the grid to remain synchronized. This type of WTG and Type 3 WTG is widely implemented as it can control the speed of the wind turbine to generate optimum power. Optimum power supply to the grid is important to reduce the case of power fluctuation happens on the power grid. Putri et al. emphasize on optimum power generation for PMSG and the proposed control method in [73].

B. Capabilities of Wind Turbine Generators

The voltage control of WTG is usually done by the centralized wind farm controller. It controls the amount of voltage supplied to the power grid. Type 1 and Type 2 WTG do not have the capabilities of voltage control whereas Type 3 and Type 4 do have it [70]. This is because Type 1 and Type 2 WTG require the capacitor bank to supply reactive power and correct its power factor. Type 3 and Type 4 can supply reactive power by itself. On the other hand, it means that the voltage can be controlled at a fixed terminal voltage and active power

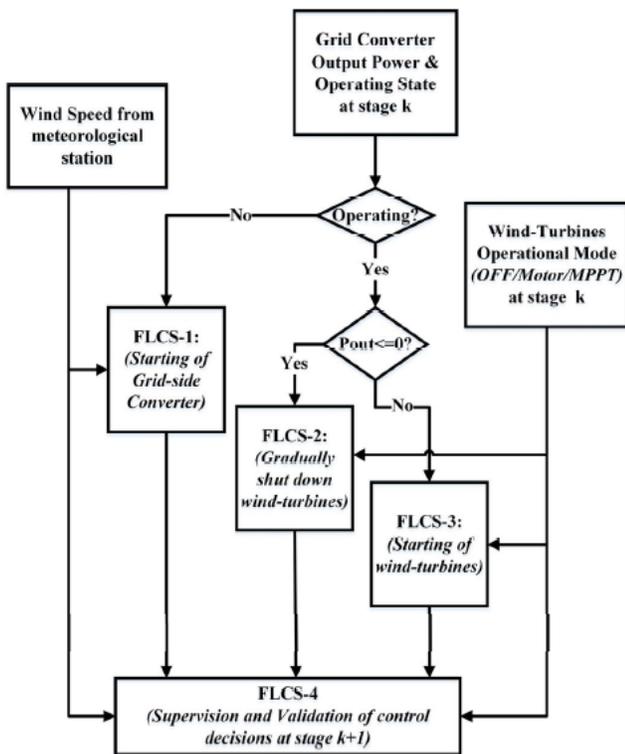


Fig 23. Flowchart of FLCs.

with varied reactive power. Besides, WTG needs to have reactive power capabilities so that the desired connection could be achieved in Wind Power Plant (WPP) at a specific range of power factors. The reactive power capabilities of Type 1 and Type 2 WTG is done by the capacitor bank installed together with the WTG [70-78]. The capacitor bank can set to 0.98 leading of the rated power to compensate for the no-load condition or 0.98 lagging of the rated power to compensate for the full-load condition. Type 3 WTG normally has 0.90 leading to 0.95 lagging power factor. Type 4 WTG has a wide range of power factor control since it has the control of current through the grid side converter [70, 79-84].

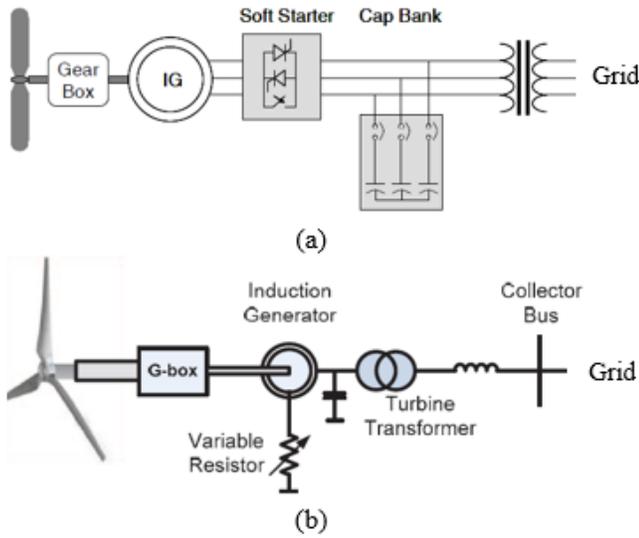


Fig 24. Connection of WTG to the Grid in Early-Stage (a) Type 1 WTG (b) Type 2 WTG.

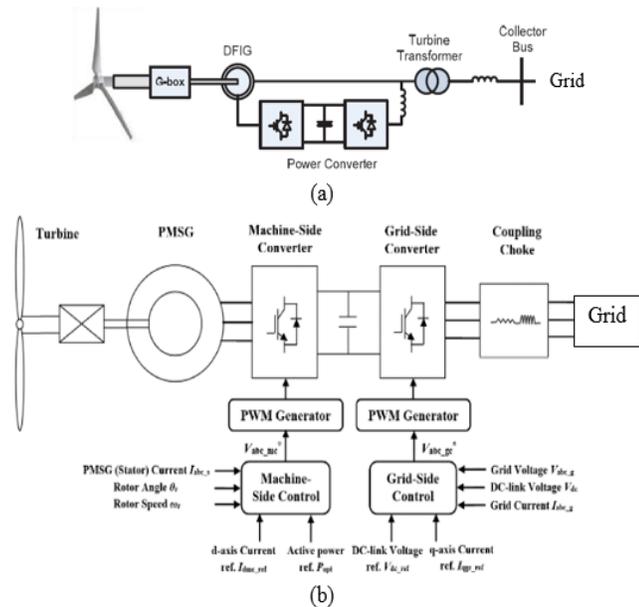


Fig 25. Modern Connection used between WTG and Power Grid (a) Type 3 WTG (b) Type 4 WTG.

There are a few problems regarding the stability of the grid met when connecting the WTG to the main grid. They include overvoltage in the network between WTG and grid, fluctuation of WTG output power, fluctuation of voltage in the weak distribution grid and more [74]. These problems are related to each other. Overvoltage usually occurs when the wind power generated is at its maximum and the power requirement by the load is at its minimum. Because of the variable wind speed and wind shear, the power fluctuation of WTG creates an aerodynamic power fluctuation. Power fluctuation can cause voltage fluctuation in a weak distribution grid and result in power quality issues. Thus, the control of voltage and power for WTG is important to maintain the stability of the power grid. This is also the reason why the popularities of implementing Type 3 and Type 4 WTG increase rapidly in the past decades. Besides, active voltage management by restricting the active power and compensate reactive power is one of the solutions to solve the overvoltage problem [74, 85-91]. Restriction of active power reduces the voltage supplied to the grid and reactive power compensation limits the voltage at the terminal of WTG.

### C. Doubly Fed Induction Generator

1) Proposed Model: Doubly Fed Induction Generator (DFIG) used Type 3 WTG connection to the power grid. As mentioned previously, it has a parallel connection of power converter or it can be referred to as two back-to-back inverters with a DC link connected in between. Fig. 26 shows the detailed schematic for Type 3 DFIG connection to the power grid proposed by Ledesma and Usaola in [75, 92-101]. Grid side inverter current is used to control active power and reactive power supplied to the grid; rotor side inverter current is used to control electromagnetic torque and excitation of the wind turbine. The voltage level of the DC link connected between two inverters is controlled by the active power of the grid side inverter.

Lei, Mullane, Lightbody and Yacamini also agree with Ledesma and Usaola as they suggested the same connection for DFIG to the power grid. Fig. 27(a) below shows the connection between the DFIG and power grid which is approximately the same compared to Fig. 26 [76]. The model proposed by Lei, Mullane, Lightbody and Yacamini assumed that both inverters to be ideal in their simulations. The researchers implemented two back-to-back voltage fed Pulse Width Modulation (PWM) inverter, with a DC link connected in between connecting the WTG to the supply grid. The inverters have the same functions as before which are controlling the power output to the grid. This can be done by controlling the switching frequency of the Insulated Gate Bipolar Transistors (IGBT) in the inverters. It can be as effective as connecting to a voltage source controller as shown in Fig. 27(b). The researcher replaced the use of inverters to a voltage source controller in their simulation [76, 102-103].

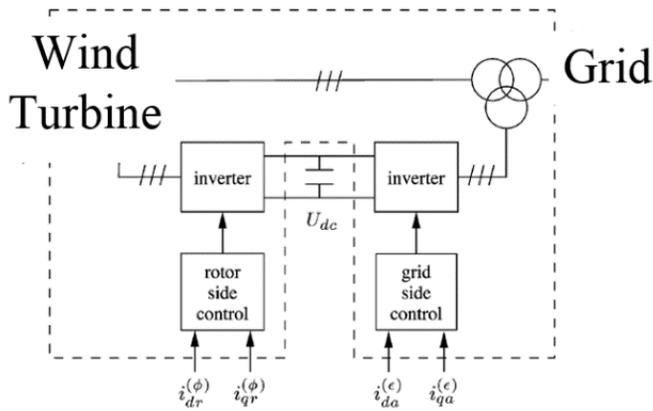


Fig 26. Connection of DFIG with the Power Grid.

Xu, Yao and Sasse proposed the use of voltage source converter (VSC) for DFIG in High Voltage Direct Current (HVDC) transmission in [77]. The large size DFIG wind farms are implemented and the power generated is transmitted to the power grid via two DC cross-linked polyethylene (XLPE) cables. A generator side VSC and a grid-side VSC is connected at both ends of the DC XLPE cable connecting the wind farm and power grid. A high pass filter is installed at both VSC to filter away high-frequency harmonics. Fig. 28 shows a clearer vision of the connection of the DFIG with the power grid proposed [77]. The voltage level and frequency generated by the wind farm are controlled by the generator side VSC whereas the voltage level supplied from the generator side is controlled by the grid side VSC at the power grid. To ensure the system operates normally, the voltage across the DC XLPE cable must be maintained so that it has a balanced power flow [77].

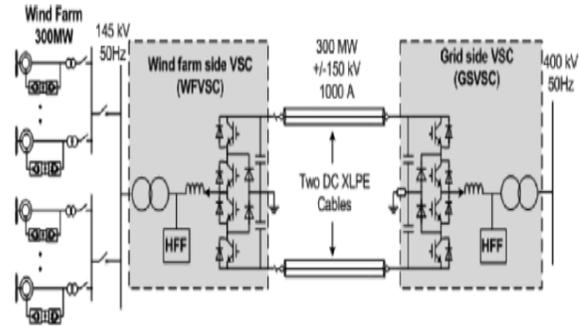


Fig 28. DC XLPE Cables with VSC at both Ends Connecting Large Size DFIG Wind Farm and Power Grid Together.

The first proposed model only consists of a variable-speed wind turbine control system, wind torque and mechanical system for completion in wind turbine modeling and assumed instantaneous control of current in inverters. The researchers also reduce the complexity in the simulation of the proposed model by using an integration step of 10 ms [75, 104]. The second proposed model allows the simulation of torque and reactive power controlled independently in a simple way. This model is suitable to analyze large size DFIG power systems and is efficient but not suitable for power converter study as both the PWM inverters were replaced by a voltage source controller to ease the simulation. The third proposed model provides advantages of fully controlled power flow of the system, can have long transmission distance regardless of the charging currents caused by the cable and high efficiency in the DC XLPE transmission cables. This proposed model also gives satisfactory results in operation and performance when no-fault occurred as shown in [77].

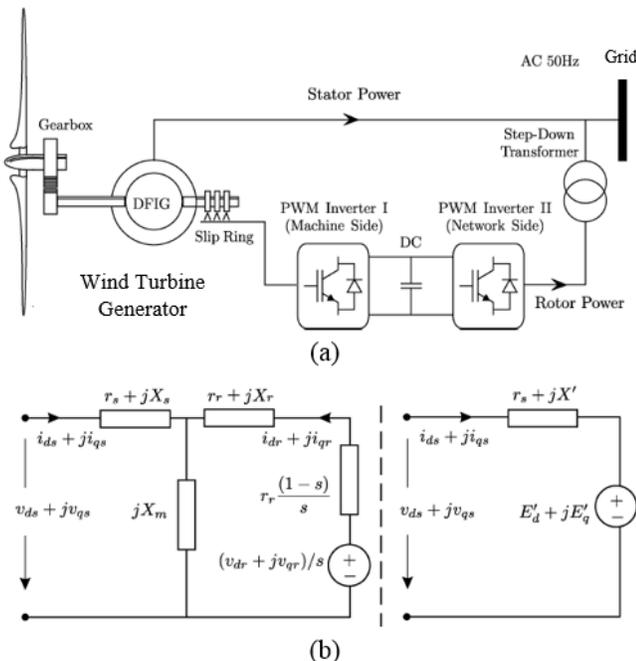


Fig 27. (a) Connection of DFIG to grid (b) Voltage Source Controller used to Replace two PWM Inverters in Simulation.

2) *Fault Ride-Through Capability*: Since DFIG is getting more popular to be implemented among other WTGs, it is important to study its behavior when the fault occurred. The sensitivity of DFIG to disturb the power grid when the fault occurred is very high. The control scheme and protection scheme is designed to allow a low voltage or fault ride-through of DFIG and protect its components. Crowbar protection is one of the protection schemes that is usually used for DFIG during fault ride through. This protection scheme is suggested by Sava et al. and Lopez, Gubia, Olea, Ruiz and Marroyo for protecting DFIG when the fault occurred and both agree that this protection scheme lost control in the reactive power when the protection is on in [78, 79]. It lost control of reactive power because the rotor side converter is blocked and thus causing the WTG is disconnected from the grid when the protection is on. The rotor side converter will only be gained control after the protection of the crowbar is off. The connection of crowbar protection to DFIG as mentioned in [79] is shown in Fig. 30(a) below. The crowbar protection mainly consists of a resistor and active switch device connected across the rotor side. Activation of crowbar protection depends on the DC link voltage level of the back-to-back inverters. When there is a fault occurred, the overvoltage or overcurrent will not destroy the rotor circuits but only activate and pass through the crowbar.

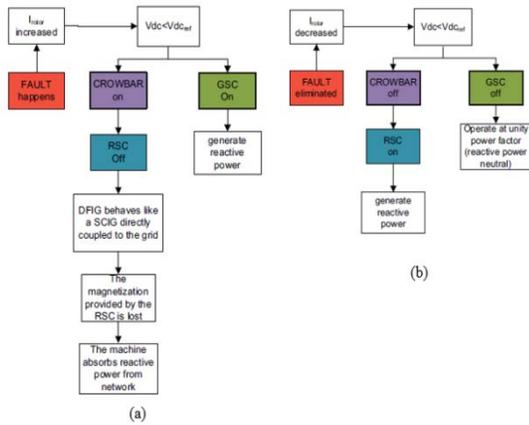


Fig 29. (a) Control Scheme when a Fault Occurs (b) Control Scheme when a Fault is Removed.

However, consideration of reactive power compensation is important in the system. Sava et al. have developed a control method for reactive power generation by grid side converter when the crowbar protection is on as shown in Fig. 29 [78]. The grid side converter will be controlled to supply reactive power when the crowbar protection is on. Furthermore, there are alternatives like an injection of current that opposes the natural flux which can protect the components in the rotor side converter without disconnecting it. Xiang, Ran, Tavner and Yang proposed the injection of opposed stator flux to decrease the rotor current [80]. The decreased amount of rotor current will protect the rotor side converter from any thermal destruction. But, the current required to oppose it usually is very large which happens only severe voltage sag as described in [79]. Thus, Lopez, Gubia, Olea, Ruiz and Marroyo suggest a strategy of combining the crowbar protection and oppose current injection for fault ride-through of DFIG [79]. Crowbar protection activates instantly when a fault occurs and demagnetization of the machine happens at the same time. After the machine is partially demagnetized, crowbar protection is off and the inverter injects the opposing current to generate reactive power. This method only reduces the activation period of the crowbar protection which means that there will still be a very short period without generating reactive power. This method accelerates the injection of reactive power when the fault occurred as shown in Fig. 30(b) [79].

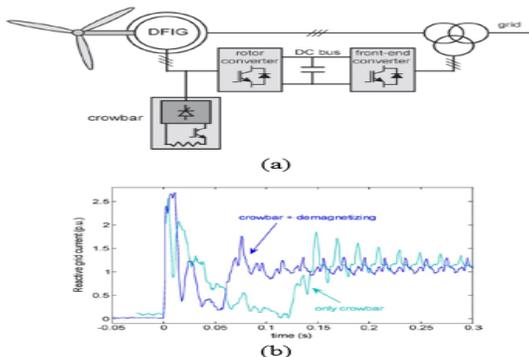


Fig 30. (a) Crowbar Protection Added to DFIG (b) Injection of Reactive Current into the Power Grid.

Yao, Li, Liao and Chen agree that crowbar can protect the rotor side converter by disconnecting the rotor circuit and short it with the resistor in crowbar [81]. They stated that the DC link voltage between two back-to-back inverters will be fluctuated and affect the control of the rotor current when the rotor circuit is connected back to the system (fault has been removed) in [81]. It can be said that the DC link voltage is important to maintain the stability of grid-supplied voltage. Its stability is maintained through balance active and reactive power supply to the power grid. The researchers proposed the improved grid side converter control strategies when the grid voltage is constant and when there is voltage sag happens in [81]. When the grid voltage is constant, the controller will keep on checking current at the grid side converter to ensure the stability of DC-link voltage. When there is a voltage sag happens, grid voltage will decrease rapidly and lead to a huge difference in power supplied. This will cause the fluctuation of DC-link voltage. Hence, the researchers introduced a control loop for a single inner current to limit the fluctuation effectively and it is proven in [81] that both control strategies are validated. The proposed control loop will regulate the grid current rapidly from the semiconductor converter to reduce the difference in output power. This helps in reducing DC link voltage fluctuation and limit the DC link voltage. But, this control strategy when voltage sag happens is not applicable if severe voltage sag happens at the power grid.

3) *Reactive Power Support:* As mentioned previously, the voltage control capability of WTG is important to maintain power grid stability. DFIG has the same voltage control capability as PMSG which is by controlling the generation of reactive power. As stated in [82], Santos-Martin, Arnaltes and Rodriguez Amenedo emphasized the importance of having reactive power capability for WTG required by the power grid when connecting WTG to the power grid. The reactive power usually generated by the stator and grid side inverter in the WTG. This reactive power generated is then fed into the power grid. Nevertheless, the grid side inverter has a unity power factor and this makes the only stator contributes to the reactive power supply in a commercial system. The researchers also stated that the voltage supplied to the power grid can be expressed in terms of reactive power in [82]. The stability of the power grid could be maintained if effective control of reactive power generation is established in WTG.

Control of both rotor side inverter and grid side inverter to generate reactive power is possible for voltage control purpose and this is suggested in [83]. Under the uncoordinated condition, both inverters are working either to supply or absorb reactive power. One of the inverters should absorb the reactive power generated by the other to prevent any extra current circulates in the circuit that will cause thermal destruction to the inverters. Either rotor side inverter or grid side inverter is used for reactive power generation to regulate the voltage supplied to the power grid and another of them will not work under coordinated condition. The rotor side inverter will not supply reactive power when the crowbar protection is on [83]. Since the rotor side inverter will be disconnected when crowbar protection is on, the grid side inverter is responsible for the supply of reactive power to the power grid.

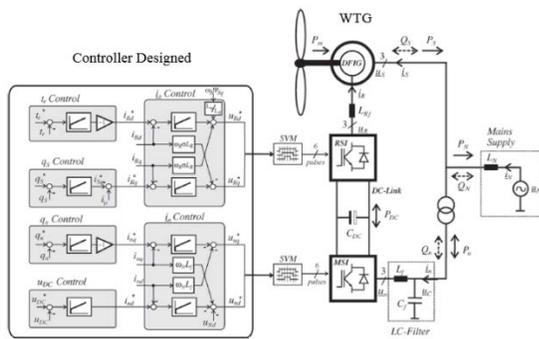


Fig 31. Rotor Side Inverter and Grid Side Inverter Control Scheme to Regulate Reactive Power

The maintenance of power grid stability is important when it has more developments and implementations in WTG. This can be achieved through voltage regulation by generating reactive power supply to the power grid. Rabelo, Hofmann, da Silva, de Oliveira and Silva also agree with the statements above and they proposed a controller designed to regulate voltage by supplying reactive power adequately using the rotor side inverter and grid side inverter [84]. Adequate reactive power generation increases system efficiency. The controller designed used linear control techniques and some improvements were made according to its limitations. The controller implemented is shown in Fig. 31 above and the detailed control method can be referred to in [84]. The control scheme allows the control of reactive power to be generated by both inverters to compensate it for the system. This reduces the power loss in the system and hence increases its efficiency.

### VII. PROSPECTS AND LIMITATIONS

Implementation of the wind power system in Malaysia does not give a lot of benefits. This is because wind speed in Malaysia is generally low and the direction of wind flow is unstable when compared with other countries. Wind power generated is insufficient to support the power grid. Ineffective control of wind turbine generator will lead to fluctuation of voltage and power supplied. But, these problems could be overcome by combining another source of energy in power generation. One of them is combining solar power with wind power in grid integration. Wind power can be used as a backup generator for solar power or vice versa to maintain a stable power generation. There are a few locations in Malaysia where the wind speed is higher and stable. These locations are suitable for this solar wind hybrid system. These power generations are also environment-friendly and sustainable when compared with the non-renewable energy used in Malaysia earlier like fossil fuels and natural gas.

### VIII. CONCLUSIONS

This research report reviewed different parts of the wind power system which include wind turbines, energy storage, microgrid and grid integration. It has been found that the implementation of HAWT is not suitable for Malaysia due to unstable wind flow direction. VAWT can be considered as it has omnidirectional characteristics. The use of an energy storage system in the wind power system allows energy to be used in a “smarter” way. The power generated can be stored or

supplied depends on load demand. The reliability of the wind turbine system in microgrid integration mainly depends on power control and energy management. Both strategies are important to maximize the efficiency of the microgrid. Type 3 and Type 4 WTG would be suggested for grid integration as they have control in wind turbine rotation. They allow the control of voltage and reactive power supplied to the grid for maintaining grid stability. Effective control of reactive power generation needs to be found in the future work to maintain the stability of the power grid in WTG.

### REFERENCES

- [1] G. Ahmad and U. Amin, "Design, construction and study of small scale vertical axis wind turbine based on a magnetically levitated axial flux permanent magnet generator," *Renewable Energy*, vol. 101, pp. 286-292, 2// 2017.
- [2] Z. Cheng, H. A. Madsen, W. Chai, Z. Gao and T. Moan, "A comparison of extreme structural responses and fatigue damage of semi-submersible type floating horizontal and vertical axis wind turbines," *Renewable Energy*, vol. 108, pp. 207-219, 8// 2017.
- [3] Z. Cheng, H. A. Madsen, Z. Gao and T. Moan, "A fully coupled method for numerical modeling and dynamic analysis of floating vertical axis wind turbines," *Renewable Energy*, vol. 107, pp. 604-619, 7// 2017.
- [4] A. O. Onol and S. Yesilyurt, "Effects of wind gusts on a vertical axis wind turbine with high solidity," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 162, pp. 1-11, 3// 2017.
- [5] K. H. Wong, W. T. Chong, N. L. Sukiman, S. C. Poh, Y.-C. Shiah and C.-T. Wang, "Performance enhancements on vertical axis wind turbines using flow augmentation systems: A review," *Renewable and Sustainable Energy Reviews*, vol. 73, pp. 904-921, 6// 2017.
- [6] A. Rezaeiha, I. Kalkman and B. Blocken, "Effect of pitch angle on power performance and aerodynamics of a vertical axis wind turbine," *Applied Energy*, vol. 197, pp. 132-150, 7// 2017.
- [7] X. Jin, G. Zhao, K. Gao and W. Ju, "Darrieus vertical axis wind turbine: Basic research methods," *Renewable and Sustainable Energy Reviews*, vol. 42, pp. 212-225, 2// 2015.
- [8] F. Toja-Silva, A. Colmenar-Santos and M. Castro-Gil, "Urban wind energy exploitation systems: Behaviour under multidirectional flow conditions—Opportunities and challenges," *Renewable and Sustainable Energy Reviews*, vol. 24, pp. 364-378, 8// 2013.
- [9] "HAWT versus VAWT: Small VAWTs find a clear niche," *Refocus*, vol. 4, no. 4, pp. 44-46, 7// 2003.
- [10] G. M. H. Shahariar and M. R. Hasan, "Design & construction of a Vertical Axis Wind Turbine," in *2014 9th International Forum on Strategic Technology (IFOST)*, 2014, pp. 326-329.
- [11] S. M. Camporeale and V. Magi, "Streamtube model for analysis of vertical axis variable pitch turbine for marine currents energy conversion," *Energy Conversion and Management*, vol. 41, no. 16, pp. 1811-1827, 11// 2000.
- [12] M. A. Biadgo, A. Simonović, D. Komarov and S. Stupar, "Numerical and analytical investigation of vertical axis wind turbine," *FME Transactions*, vol. 41, no. 1, pp. 49-58, 2013.
- [13] A. T. Zhuga, B. Munyaradzi and C. Shonhiwa, "Design of alternative energy systems: A self-starting Vertical Axis Wind Turbine for stand-alone applications (charging batteries)," *SPONSORS*, p. 72, 2006.
- [14] M. Ragheb, "Vertical axis wind turbines," *University of Illinois at Urbana-Champaign*, vol. 1, 2011.
- [15] Q. a. Li, T. Maeda, Y. Kamada, J. Murata, K. Furukawa and M. Yamamoto, "The influence of flow field and aerodynamic forces on a straight-bladed vertical axis wind turbine," *Energy*, vol. 111, pp. 260-271, 9/15/ 2016.
- [16] Q. a. Li et al., "Wind tunnel and numerical study of a straight-bladed vertical axis wind turbine in three-dimensional analysis (Part I: For predicting aerodynamic loads and performance)," *Energy*, vol. 106, pp. 443-452, 7// 2016.

- [17] M. F. Ismail and K. Vijayaraghavan, "The effects of aerofoil profile modification on a vertical axis wind turbine performance," *Energy*, vol. 80, pp. 20-31, 2/1/ 2015.
- [18] Z. Cheng, H. A. Madsen, Z. Gao and T. Moan, "Effect of the number of blades on the dynamics of floating straight-bladed vertical axis wind turbines," *Renewable Energy*, vol. 101, pp. 1285-1298, 2// 2017.
- [19] I. Ross and A. Altman, "Wind tunnel blockage corrections: Review and application to Savonius vertical-axis wind turbines," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 99, no. 5, pp. 523-538, 5// 2011.
- [20] S. Eriksson, H. Bernhoff and M. Leijon, "Evaluation of different turbine concepts for wind power," *Renewable and Sustainable Energy Reviews*, vol. 12, no. 5, pp. 1419-1434, 2008.
- [21] C.-J. Bai and W.-C. Wang, "Review of computational and experimental approaches to analysis of aerodynamic performance in horizontal-axis wind turbines (HAWTs)," *Renewable and Sustainable Energy Reviews*, vol. 63, pp. 506-519, 9// 2016.
- [22] B. Gaunt and D. A. Johnson, "Wind turbine performance in controlled conditions: Experimental results," *International Journal of Green Energy*, no. just-accepted, 2012.
- [23] Shikha, T. S. Bhatti and D. P. Kothari, "Early Development of Modern Vertical and Horizontal Axis Wind Turbines: A Review," *Wind Engineering*, vol. 29, no. 3, pp. 287-299, 2005.
- [24] F. Kanyako and I. Janajreh, "Implementation and economical study of HAWT under different wind scenarios," *Sustainable Cities and Society*, vol. 15, pp. 153-160, 7// 2015.
- [25] M. Ó. Óskarsdóttir, "A General Description and Comparison of Horizontal Axis Wind Turbines and Vertical Axis Wind Turbines," 2014.
- [26] J. Manwell, J. McGowan and A. Rogers, "Aerodynamics of wind turbines," *Wind Energy Explained: Theory, Design and Application*, Second Edition, pp. 91-155, 2009.
- [27] S. McFadden and B. Basu, "7 - Wind turbine gearbox design with drivetrain dynamic analysis," in *Offshore Wind Farms*: Woodhead Publishing, 2016, pp. 137-158.
- [28] C. Eisenhut, F. Krug, C. Schram and B. Klockl, "Wind-turbine model for system simulations near cut-in wind speed," *IEEE Transactions on Energy Conversion*, vol. 22, no. 2, pp. 414-420, 2007.
- [29] R. Bi, C. Zhou and D. M. Hepburn, "Detection and classification of faults in pitch-regulated wind turbine generators using normal behaviour models based on performance curves," *Renewable Energy*, vol. 105, pp. 674-688, 5// 2017.
- [30] R. Tiwari and N. R. Babu, "Recent developments of control strategies for wind energy conversion system," *Renewable and Sustainable Energy Reviews*, vol. 66, pp. 268-285, 12// 2016.
- [31] A. S. Yilmaz and Z. Özer, "Pitch angle control in wind turbines above the rated wind speed by multi-layer perceptron and radial basis function neural networks," *Expert Systems with Applications*, vol. 36, no. 6, pp. 9767-9775, 8// 2009.
- [32] B. Bavanish and K. Thyagarajan, "Optimization of power coefficient on a horizontal axis wind turbine using bem theory," *Renewable and Sustainable Energy Reviews*, vol. 26, pp. 169-182, 10// 2013.
- [33] K. Johnson, L. J. Fingersh, M. Balas and L. Pao, "Methods for increasing region 2 power capture on a variable speed HAWT," in 42nd AIAA Aerospace Sciences Meeting and Exhibit, 2004, p. 350.
- [34] P. J. Schubel and R. J. Crossley, "Wind turbine blade design," *Energies*, vol. 5, no. 9, pp. 3425-3449, 2012.
- [35] A. Sedaghat and M. Mirhosseini, "Aerodynamic design of a 300 kW horizontal axis wind turbine for province of Semnan," *Energy Conversion and Management*, vol. 63, pp. 87-94, 11// 2012.
- [36] M. Ragheb, "Optimal rotor tip speed ratio," Available from NetFiles at the University of Illinois at Urbana-Champaign, Last modified, 2009.
- [37] G. C. Venkatesh and S. V. Kulkarni, "Energy yield of passive stall regulated fixed speed wind turbine with optimal rotor speed," *Electric Power Systems Research*, vol. 76, no. 12, pp. 1019-1026, 8// 2006.
- [38] M.-S. Jeong, S.-W. Kim, I. Lee, S.-J. Yoo and K. C. Park, "The impact of yaw error on aeroelastic characteristics of a horizontal axis wind turbine blade," *Renewable Energy*, vol. 60, pp. 256-268, 12// 2013.
- [39] H. Hui, M. Chengxiong, L. Jiming and W. Dan, "Electronic Power Transformer Control Strategy in Wind Energy Conversion Systems for Low Voltage Ride-through Capability Enhancement of Directly Driven Wind Turbines with Permanent Magnet Synchronous Generators (D-PMSGs)," *Energies* (19961073), Article vol. 7, no. 11, pp. 7330-7347, 2014.
- [40] W. Shepherd and L. Zhang, "Electricity Generation Using Wind Power," ed: World Scientific, 2011.
- [41] H. Li and Z. Chen, "Overview of different wind generator systems and their comparisons," *IET Renewable Power Generation*, vol. 2, no. 2, pp. 123-138, 2008.
- [42] R. Cisneros, F. Mancilla-David and R. Ortega, "Passivity-Based Control of a Grid-Connected Small-Scale Windmill With Limited Control Authority," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 1, no. 4, pp. 247-259, 2013.
- [43] D. Kumar and K. Chatterjee, "A review of conventional and advanced MPPT algorithms for wind energy systems," *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 957-970, 3// 2016.
- [44] A. Nagliero, R. Mastromauro, V. Monopoli, M. Liserre and A. Dell'Aquila, "Analysis of a universal inverter working in grid-connected, stand-alone and micro-grid," in *Industrial Electronics (ISIE)*, 2010 IEEE International Symposium on, 2010, pp. 650-657: IEEE.
- [45] M. Świerczyński, R. Teodorescu, C. N. Rasmussen, P. Rodriguez and H. Vikelgaard, "Overview of the energy storage systems for wind power integration enhancement," in *Industrial Electronics (ISIE)*, 2010 IEEE International Symposium on, 2010, pp. 3749-3756: IEEE.
- [46] Z. Huang and G. Du, "Chapter 4 - Nickel-based batteries for medium- and large-scale energy storage A2 - Menictas, Chris," in *Advances in Batteries for Medium and Large-Scale Energy Storage*, M. Skyllas-Kazacos and T. M. Lim, Eds.: Woodhead Publishing, 2015, pp. 73-90.
- [47] T. Horiba, "Lithium-Ion Battery Systems," *Proceedings of the IEEE*, vol. 102, no. 6, pp. 939-950, 2014.
- [48] Z. Wen, J. Cao, Z. Gu, X. Xu, F. Zhang and Z. Lin, "Research on sodium sulfur battery for energy storage," *Solid State Ionics*, vol. 179, no. 27-32, pp. 1697-1701, 9/30/ 2008.
- [49] S. Dambone Sessa, G. Crugnola, M. Todeschini, S. Zin and R. Benato, "Sodium nickel chloride battery steady-state regime model for stationary electrical energy storage," *Journal of Energy Storage*, vol. 6, pp. 105-115, 5// 2016.
- [50] J. L. Sudworth, "The sodium/nickel chloride (ZEBRA) battery," *Journal of Power Sources*, vol. 100, no. 1-2, pp. 149-163, 11/30/ 2001.
- [51] W. Wang, Q. Luo, B. Li, X. Wei, L. Li and Z. Yang, "Recent Progress in Redox Flow Battery Research and Development," *Advanced Functional Materials*, vol. 23, no. 8, pp. 970-986, 2013.
- [52] J. Li, L. Yang, B. Yuan, G. Li and J. Y. Lee, "Combined mediator and electrochemical charging and discharging of redox targeting lithium-sulfur flow batteries," *Materials Today Energy*, vol. 5, pp. 15-21, 9// 2017.
- [53] M. A. Guerrero, E. Romero, F. Barrero, M. I. Milanés and E. González, "Supercapacitors: alternative energy storage systems," *Przegląd Elektrotechniczny*, vol. 85, no. 10, pp. 188-195, 2009.
- [54] T. Ma, H. Yang and L. Lu, "Development of hybrid battery-supercapacitor energy storage for remote area renewable energy systems," *Applied Energy*, vol. 153, pp. 56-62, 9/1/ 2015.
- [55] J. O. Petinrin and M. Shaaban, "A hybrid solar PV/wind energy system for voltage regulation in a microgrid," in 2013 IEEE Student Conference on Research and Development, 2013, pp. 545-549.
- [56] T. Aziz, S. I. M. Salman, M. S. Islam, A. Razaq, R. A. Chowdhury and M. I. H. Mitun, "Integration of wind energy system in microgrid considering static and dynamic issues," in 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2015, pp. 1-5.
- [57] N. D. Hatzigaryriou, A. Dimeas, A. G. Tsikalakis, J. A. P. Lopes, G. Karniotakis and J. Oyarzabal, "Management of microgrids in market environment," in 2005 International Conference on Future Power Systems, 2005, pp. 7 pp.-7.

- [58] S. Lu et al., "Control strategies for distributed energy resources to maximize the use of wind power in rural microgrids," in 2011 IEEE Power and Energy Society General Meeting, 2011, pp. 1-8.
- [59] P. Piagi and R. H. Lasseter, "Autonomous control of microgrids," in 2006 IEEE Power Engineering Society General Meeting, 2006, p. 8 pp.
- [60] A. G. Tsikalakis and N. D. Hatziargyriou, "Centralized control for optimizing microgrids operation," in 2011 IEEE Power and Energy Society General Meeting, 2011, pp. 1-8.
- [61] F. Chiou, J. P. Gentle and T. R. McJunkin, "Dispatchable renewable energy model for microgrid power system," in 2016 IEEE Conference on Technologies for Sustainability (SusTech), 2016, pp. 195-199.
- [62] F. Katiraei, R. Iravani, N. Hatziargyriou and A. Dimeas, "Microgrids management," IEEE Power and Energy Magazine, vol. 6, no. 3, pp. 54-65, 2008.
- [63] F. A. Pamuji and H. Miyauchi, "A new control design of maximum power point tracking by fuzzy logic controller for wind turbine connected to low voltage grid," in 2016 International Seminar on Intelligent Technology and Its Applications (ISITIA), 2016, pp. 369-372.
- [64] J. Hui and P. K. Jain, "Power management and control of a wind energy conversion system (WECS) with a fuzzy logic based maximum power point tracking (MPPT)," in IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, 2012, pp. 5966-5971.
- [65] M. Motevasel and A. R. Seifi, "Expert energy management of a microgrid considering wind energy uncertainty," Energy Conversion and Management, vol. 83, pp. 58-72, 7// 2014.
- [66] E. Tsioumas, N. Karakasis, N. Jabbour and C. Mademlis, "Energy management and power control strategy at the low wind speed region of a wind generation microgrid," in IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, 2016, pp. 4097-4102.
- [67] J. Patiño, A. Márquez and J. Espinosa, "An economic MPC approach for a microgrid energy management system," in 2014 IEEE PES Transmission & Distribution Conference and Exposition - Latin America (PES T&D-LA), 2014, pp. 1-6.
- [68] M. Yu, Y. Wang and Y. G. Li, "Energy management of wind turbine-based DC microgrid utilizing modified differential evolution algorithm," in International Conference on Renewable Power Generation (RPG 2015), 2015, pp. 1-6.
- [69] H. Babazadeh, W. Gao, Z. Wu and Y. Li, "Optimal energy management of wind power generation system in islanded microgrid system," in 2013 North American Power Symposium (NAPS), 2013, pp. 1-5.
- [70] E. H. Camm et al., "Characteristics of wind turbine generators for wind power plants," in 2009 IEEE Power & Energy Society General Meeting, 2009, pp. 1-5.
- [71] E. Muljadi, Y. C. Zhang, V. Gevorgian and D. Kosterev, "Understanding dynamic model validation of a wind turbine generator and a wind power plant," in 2016 IEEE Energy Conversion Congress and Exposition (ECCE), 2016, pp. 1-5.
- [72] Y. Zou and J. He, "Comprehensive modeling, simulation and experimental validation of Permanent Magnet Synchronous generator wind power system," in 2016 IEEE/IAS 52nd Industrial and Commercial Power Systems Technical Conference (I&CPS), 2016, pp. 1-9.
- [73] R. I. Putri, M. Rifa'i, L. Jasa, A. Priyadi, P. Margo and H. P. Mauridhi, "Modeling and control of permanent magnet synchronous generator variable speed wind turbine," in 2016 International Conference on Smart Green Technology in Electrical and Information Systems (ICSGTEIS), 2016, pp. 16-20.
- [74] M. Rahimi and H. Assari, "Addressing and assessing the issues related to connection of the wind turbine generators to the distribution grid," International Journal of Electrical Power & Energy Systems, vol. 86, pp. 138-153, 3// 2017.
- [75] P. Ledesma and J. Usaola, "Doubly fed induction generator model for transient stability analysis," IEEE Transactions on Energy Conversion, vol. 20, no. 2, pp. 388-397, 2005.
- [76] L. Yazhou, A. Mullane, G. Lightbody and R. Yacamini, "Modeling of the wind turbine with a doubly fed induction generator for grid integration studies," IEEE Transactions on Energy Conversion, vol. 21, no. 1, pp. 257-264, 2006.
- [77] L. Xu, L. Yao and C. Sasse, "Grid Integration of Large DFIG-Based Wind Farms Using VSC Transmission," IEEE Transactions on Power Systems, vol. 22, no. 3, pp. 976-984, 2007.
- [78] G. N. Sava, M. Q. Duong, S. Leva, M. Mussetta, S. Costinas and N. Golovanov, "Coordination control of active crowbar for doubly fed induction generators," in 2014 International Symposium on Fundamentals of Electrical Engineering (ISFEE), 2014, pp. 1-5.
- [79] J. Lopez, E. Gubia, E. Olea, J. Ruiz and L. Marroyo, "Ride Through of Wind Turbines With Doubly Fed Induction Generator Under Symmetrical Voltage Dips," IEEE Transactions on Industrial Electronics, vol. 56, no. 10, pp. 4246-4254, 2009.
- [80] R. Li, X. Dawei, P. J. Tavner and Y. Shunchang, "Control of a doubly fed induction generator in a wind turbine during grid fault ride-through," in 2006 IEEE Power Engineering Society General Meeting, 2006, p. 1 pp.
- [81] J. Yao, H. Li, Y. Liao and Z. Chen, "An Improved Control Strategy of Limiting the DC-Link Voltage Fluctuation for a Doubly Fed Induction Wind Generator," IEEE Transactions on Power Electronics, vol. 23, no. 3, pp. 1205-1213, 2008.
- [82] D. Santos-Martin, S. Arnaltes and J. L. Rodriguez Amenedo, "Reactive power capability of doubly fed asynchronous generators," Electric Power Systems Research, vol. 78, no. 11, pp. 1837-1840, 11// 2008.
- [83] M. Kayikci and J. V. Milanovic, "Reactive Power Control Strategies for DFIG-Based Plants," IEEE Transactions on Energy Conversion, vol. 22, no. 2, pp. 389-396, 2007.
- [84] B. C. Rabelo, W. Hofmann, J. L. d. Silva, R. G. d. Oliveira and S. R. Silva, "Reactive Power Control Design in Doubly Fed Induction Generators for Wind Turbines," IEEE Transactions on Industrial Electronics, vol. 56, no. 10, pp. 4154-4162, 2009.
- [85] Tabassum, M., Haldar, M.K. and Khan, D.F.S., 2016. Implementation and performance evaluation of advance metering infrastructure for Borneo-Wide Power Grid. *Frontiers in Energy*, pp.1-20.
- [86] Nabipour-Afrouzi, H., Yii, S.H.W., Ahmad, J. and Tabassum, M., 2018, October. Comprehensive Review on Appropriate Sizing and Optimization Technique of Hybrid PV-Wind System. In 2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC) (pp. 364-369). IEEE.
- [87] Maaji, S.S., Khan, D.F.S., Tabassum, M. and Haldar, M.K., PRELIMINARY STUDIES ON A Proposed BORNEO-WIDE SMART GRID.
- [88] Maaji, S.S., Khan, D.F.S., Haldar, M.K., Tabassum, M. and Karunakaran, P., 2013. Integration of solar and wind power to a borneo-wide power grid. *International Journal of Environmental Science and Development*, 4(6), p.681.
- [89] L. T. Hong, J. Ahmed, H. Nabipour-Afrouzi and S. Kashem, "Designing a PSCAD based PV simulator for partial shading to validate future PV application planning," 2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Kota Kinabalu, 2018, pp. 526-531, doi: 10.1109/APPEEC.2018.8566639.
- [90] J. Ahmed, Z. Salam, Y. L. Then and S. Kashem, "A fast MPPT technique based on I-V curve characteristics under partial shading," TENCON 2017 - 2017 IEEE Region 10 Conference, Penang, 2017, pp. 1696-1701, doi: 10.1109/TENCON.2017.8228132.
- [91] C. T. K. Kho, J. Ahmed, S. Kashem and Y. L. Then, "A comprehensive review on PV configurations to maximize power under partial shading," TENCON 2017 - 2017 IEEE Region 10 Conference, Penang, 2017, pp. 763-768, doi: 10.1109/TENCON.2017.8227962.
- [92] Chowdhury, M.A. and Kashem, S.B.A., 2018. H $\infty$  loop-shaping controller design for a grid-connected single-phase photovoltaic system. *International Journal of Sustainable Engineering*, pp.1-9.
- [93] Kashem, S.B.A., Sheikh, M.I.B., Ahmed, J. and Tabassum, M., 2018, April. Gravity and buoyancy powered clean water pipe generator. In 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018) (pp. 1-5). IEEE.
- [94] Mubarak, H. and Kashem, S.B.A., 2016. Comparison of different energy saving lights using solar panel. *Frontiers in Energy*, 10(4), pp.466-472.

- [95] M. B. M. Siddique, S. B. A. Kashem, K. Mathew "Home and Water Heating Using Biofuels" Proceedings of International Conference on Recent Innovations in Engineering and Technology, 2017.
- [96] F. Tay, S. B. A. Kashem, "Automated Miniature Greenhouse", *Advanced Science Letters* 23.6 (2017): 5309-5313.
- [97] Safe, A.A., Kashem, S., Moniruzzaman, M. and Islam, M.T., 2014, October. Design, fabrication & analysis of twisted blade vertical axis wind turbine (VAWT) and a simple alternator for VAWT. In *Strategic Technology (IFOST), 2014 9th International Forum on* (pp. 304-308). IEEE.
- [98] Mujahid Tabassum, S. B. A. Kashem, Kuruvilla Mathew, "Distributed energy generation – is it the way of the future?", Proceedings of the 1st Springer International Conference on Emerging Trends and Advances in Electrical Engineering and Renewable Energy, 2016.
- [99] Sheikh, M. Ismail Bilal, S. B. A. Kashem and Tanveer Choudhury. "Enhancing solar power generation using gravity and fresh water pipe." Proceedings of IEEE Xplore 2017, IEEE International Conference on Mechatronics, pp. 266-271, 2017.
- [100] Tabassum, M., Kashem, S. B. A. and Siddique, M.B.M., Feasibility of using Photovoltaic (PV) technology to generate solar energy in Sarawak. In *Computer and Drone Applications (IconDA), 2017 International Conference on* (pp. 11-16). IEEE 2017, November.
- [101] Siddique, M.B.M., Kashem, S.B.A. and Iqbal, A., Biofuels in Malaysian perspective: Debates and benefits. In *Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), 2018 IEEE 12th International Conference on* (pp. 1-6). IEEE. April, 2018.
- [102] Nushrat Shabrin, S. B. A. Kashem, et al., "Investment and Construction Cost Analysis on Net-Zero Energy Building Technology", *International Journal of Mechanical and Production Engineering*, ISSN: 2320-2092, Volume- 5, Issue-4, 2017
- [103] Nushrat Shabrin, S. B. A. Kashem, "A Comprehensive Cost Benefit Analysis of Green Building", *International Journal of Advances in Mechanical and Civil Engineering (IJAMCE)*, Volume 4 Issue 2, June 2017, ISSN: 2394-2827.
- [104] Saad Bin Abul Kashem, Muhammad E. H. Chowdhury, Mujahid Tabassum, Majid E. Molla, Azad Ashraf, Amith Khandakar, "A Comprehensive Study on Biomass Power Plant and Comparison Between Sugarcane and Palm Oil Waste", *International Journal of Innovation in Computational Science and Engineering*, Volume 1 Issue 1, page 26-32, May 2020, ISSN: 2708-3128.
- [105] Saad Bin Abul Kashem, Muhammad E. H. Chowdhury, Mujahid Tabassum, Majid E. Molla, Azad Ashraf, Jubaer Ahmed, "Feasibility Study of Solar Power System in Residential Area", *International Journal of Innovation in Computational Science and Engineering*, Volume 1 Issue 1, page 10-17, May 2020, ISSN: 2708-3128.
- [106] Kashem, S.B.A., Sheikh, M.I.B., Ahmed, J. and Tabassum, M., 2018, April. Gravity and buoyancy powered clean water pipe generator. In *2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018)* (pp. 1-5). IEEE.

# Voice Scrambling Algorithm based on 3D Chaotic Map System (VSA3DCS) to Encrypt Audio Files

Osama M. Abu Zaid

Lecturer in Computer Science Dept., Faculty of Computers &  
Information, Kafrelsheikh Univ., Egypt

**Abstract**—Here, a proposed voice scrambling algorithm established on one of two 3D chaotic maps systems (VSA3DCS) will be presented, discussed, and applied on audio signals file. The two 3D chaotic map systems in which any one of them is used to build VSA3DCS are Chen's chaotic map system and Lorenz chaotic map system. Also Arnold cat map-based scrambling algorithm will be applied on the same sample of audio signals. These Scrambling algorithms are used to encrypt the audio files by shuffling the positions of signals at different conditions with the audio file as one block or two blocks. Amplitude values of audio signals with signals' time are registered and plotted for original file versus encrypted files which are produced from applying VSA3DCS using Chen's, VSA3DCS using Lorenz, and Arnold-based algorithm. The spectrogram frequencies of audio signals with signals' time are plotted for original file versus encrypted files for all algorithms. Also, the histogram of the original file and encrypted audio signals are registered and plotted. The comparative analysis is presented by using some measuring factors for both of encryption and decryption processes, such as; the time of encryption and decryption, Correlation Coefficient of original and encrypted signals between the samples, the Spectral Distortion (SD) measure, Log-Likelihood Ratio (LLR) measure, and key sensitivity measuring factor. The results of several experimental and comparative analyses will show that the VSA3DCS algorithm using Chen's or Lorenz is a good algorithm to provide an effective and safe solution to voice signal encryption, and also VSA3DCS algorithm better than Arnold-based algorithm in all results with all cases.

**Keywords**—Lorenz chaotic map; Chen's chaotic map; Arnold cat map; scrambling algorithms; audio encryption

## I. INTRODUCTION

Now we are living amid a digital revolution that needs safe multimedia transmission. Visual encryption is essential when transmitting audio over communication networks to protect them from reading, altering their content, inserting false information, or deleting a portion of their content [1,2].

Multimedia encryption has recently become one of the key problems of great concern. It offers greater protections for the content, which may involve some private issues or save copyrights from being changed or violated [3]. Any cryptography process requires a simple algorithm with tiny processing time and high performance to protect the information. Besides that, it has a strong immune system against any external issue including noise and interference that can be faced in the channels of communication [3].

A chaotic map is a suitable solution for both issues (tiny processing time and high performance). As with other methods of encryption such as AES and DES, which have large processing then a long time, the chaotic map has a fair time to fit these tasks [3,4].

As it is possible any unauthorized person can receive the transmitted data with the simplest receivers, the security of audio conversations has recently become a crucial issue because of the successful development of crypt-analysis activities [1,5]. Chaos-based encryption mechanisms are considered to be ideal for practical use because they provide an honest combination of speed, high security, complexity [4,6,7].

In this work, it tries to solve these two challenges by producing a proposed voice scrambling algorithm (VSA3DCS) based on a 3D chaotic map system (Lorenz map system or Chen's map system). The VSA3DCS algorithm is compared with one of the 2D chaotic maps (Arnold Cat map) which used to permute the elements in the multimedia file (image or audio). Also, in this work, several metrics are evaluated to accomplish comparative analysis.

This research paper is arranged as follows: Section II will present the related work, motivation, and contribution. Section III will present the chaotic maps which are used in our work. Section IV will present the steps of the proposed algorithm VSA3DCS. Section V will present applying all algorithms on the same audio signals file. Section VI will discuss experiential results and comparative analysis. Section VII will discuss the conclusion. In the final, there are references which are being used.

## II. RELATED WORK, MOTIVATION AND CONTRIBUTION

Most of the research papers in the cryptography field use many chaotic maps systems of various dimensions or any other techniques in image encryption.

Many of the research papers apply only one-dimensional or two-dimensional chaotic maps systems in audio encryption, whereas most of these papers are is to produce an algorithm for changing in the values of signals (substitution encryption). In [3], E. Mosa et al. implemented a voice encryption method based on permutation of voice segments using a 2D chaotic map (Baker map) and substitution using masks in time and transform domains. In [8], Arnold cat map was applied by Mahmoud F. Abd Elzaher and others to permute voice samples, then either Henon or modified Henon or Unified or Lorenz chaotic systems were applied to produce the mask key and thus replace the permuted samples. In previous research

for me, it is now under review for publication in another valued journal, I applied both systems of 2D chaotic maps (Arnold Cat map and Baker map) which used in the permutation of locations for the elements of the audio signals file. Comparative analyses were made for the results showed that Arnold's application was the least in time of encryption/decryption and the best in performance in most cases.

So, here my research paper introduces a proposed multi-step voice scrambling algorithm that is developed using any one of two well-known types of 3D chaotic maps systems which are strength and sophistication in their use of cryptography operations; they are (Lorenz and Chen's). the proposed algorithm for encoding audio signals in a way that alters and confuses the locations of signals only (transposition encryption) without changing their values, and it is based on either of the two chaotic systems (Lorenz or Chen's) as will be evident from the part that explains the steps of that algorithm and as will be clear from A flow diagram that shows the details of the algorithm. The proposed algorithm is compared with the Arnold-based algorithm.

### III. THE CHAOTIC MAPS SYSTEMS

In this section, a concise description is provided about the two 3D chaotic maps systems used to construct the VSA3DCS algorithm. Also, Arnold cat map will be discussed here.

#### A. Lorenz Chaotic Map System

The Lorenz Chaotic map system is a three-equation scheme. The Lorenz system equations are defined as in Formula (1) [6,7,8,9,10].

$$\begin{cases} x = \sigma(y_0 - x_0) \\ y = rx_0 - y_0 - x_0z_0 \\ z = x_0y_0 - bz_0 \end{cases} \quad (1)$$

Where  $\sigma, r, b$  are the parameters of this chaotic system. The system displays unpredictable behavior when  $\sigma = 10, r > 24.74$  and  $b = 8/3$ . The initial state values  $x_0, y_0,$  and  $z_0$  act as the keys to the diffusion. A very good result for Lorenz chaotic map with the parameters  $\sigma = 10, r = 28, b = 8/3,$  and  $h = 0.1,$  the initial values  $x_0 = 10, y_0 = 20, z_0 = 30,$  where,  $h$  is the sequence step. The Lorenz system attractor is illustrated in Fig. 1.

#### B. Chen's Chaotic Map System

As one of the 3-D chaotic map systems defined by formula (2), Chen's chaotic map system is essential as a collection of the three differential equations of Chen's chaotic map system [7,10,11,12,13].

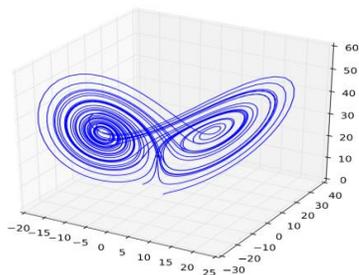


Fig 1. Chaotic Behavior of the Lorenz System.

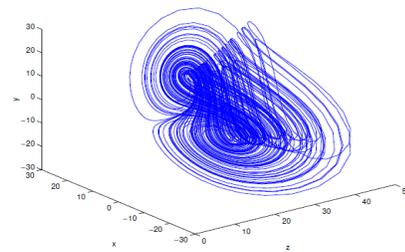


Fig 2. Chaotic Behavior of Chen's System.

$$\begin{cases} x = a(y_0 - x_0) \\ y = (c - a)x_0 - x_0z_0 + cy_0 \\ z = x_0y_0 - bz_0 \end{cases} \quad (2)$$

Where,  $a > 0, b > 0,$  and  $c$  so  $(2c > a)$  are system parameters. Chen's chaotic map system is chaotic with the parameters' values;  $a = 35, b = 3,$  and  $c \in [20, 28.4]$ .

There is also another parameter ( $h$ ), such that  $h$  is the increasing step value of  $x_0, y_0,$  and  $z_0$  for each round, that is,  $x_0 = x_0 + h, y_0 = y_0 + h,$  and  $z_0 = z_0 + h.$  If  $a = 35, b = 3,$  and  $c = 28;$  as shown in Fig. 2, it has a chaotic attractor. A very good result for this chaotic map with the parameters  $a = 35, b = 3, c = 28,$  and  $h = 0.05555,$  the initial values  $x_0 = 0, y_0 = 1, z_0 = 0,$  where  $h$  is the sequence step.

#### C. Arnold Cat Map

The Arnold Cat map is a chaotic map which is invertible in 2-D. For shuffling the pixel positions of the plain image or positions of signals in an audio file, we choose Arnold cat map method [6,14,15].

Without lack of generality, we assess the dimension of the multimedia file as  $N \times N$  (it may be  $N \times M$ ). Arnold map method as shown in Formula (3) [6,14,16,17]:

$$\begin{aligned} \begin{bmatrix} x_{m+1} \\ y_{m+1} \end{bmatrix} &= A \begin{bmatrix} x_m \\ y_m \end{bmatrix} \pmod{N} \\ &= \begin{bmatrix} 1 & p \\ q & pq + 1 \end{bmatrix} \begin{bmatrix} x_m \\ y_m \end{bmatrix} \pmod{N} \\ &= \begin{bmatrix} x_m + py_m \\ qx_m + (pq + 1)y_m \end{bmatrix} \pmod{N} \end{aligned} \quad (3)$$

Where  $p$  and  $q$  bear positive quantities because  $\det(A)=1$  from  $(pq+1)-pq.$  The  $(x_{m+1}, y_{m+1}),$  when Arnold map is applied once, is the novel position of the original pixel position  $(x_m, y_m).$  Where that  $m=0,1,2,3, \dots$  There are positive integer  $T$  at repeated  $R$  times, where at  $T, (x_{m+1}, y_{m+1}) = (x_m, y_m).$

### IV. VSA3DCS ALGORITHM

In this part of the paper, the proposed voice scrambling Algorithm (VSA3DCS) based on one of two 3D Chaotic Maps Systems (Chen's or Lorenz) is presented. VSA3DCS consists of a scrambling procedure to produce a shuffled audio file and return-scrambling procedure to reproduce the original audio file. The scrambling algorithm VSA3DCS is designed to shuffle the positions of signals of an audio file. VSA3DCS consists of seven steps of operations as following, and its Data-Flow diagram will be illustrated in Fig. 3:

**Step 1:** Obtain the  $au$  vector (1D matrix) of the audio signals file  $m \times 1.$  The length of  $au$  is  $L$  which is equal to  $m.$

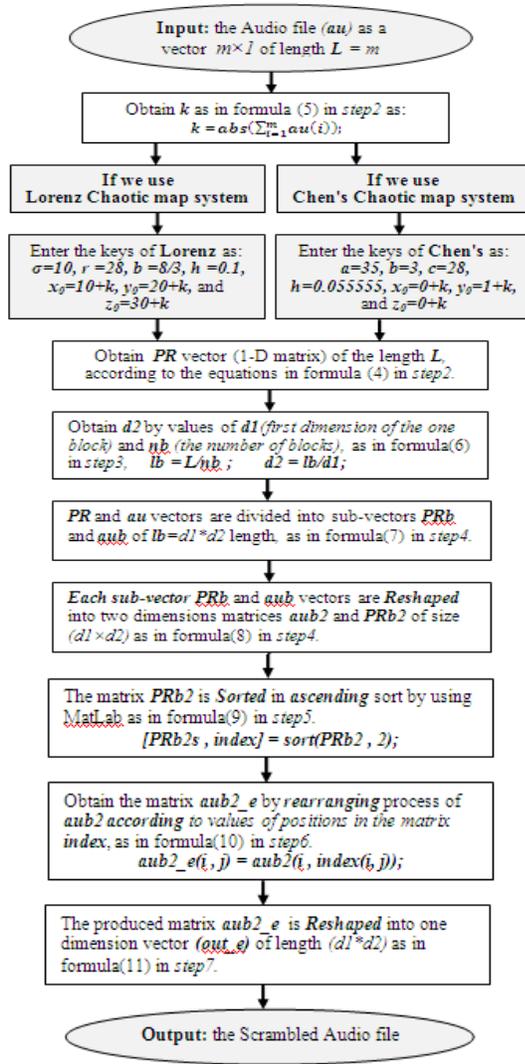


Fig 3. The Data-Flow Diagram for the VSA3DCS Algorithm.

**Step 2:** Obtain the  $PR$  vector as in formula (4), which is vector of numbers with the same length ( $L$ ) of  $au$ , and generated either by Chen's system at  $a=35, b=3, c=28$ , the initial values  $x_0=0+k, y_0=1+k, z_0=0+k$ , and  $h=0.055555$ , or by Lorenz system at  $\sigma=10, r=28, b=8/3$ , the initial values  $x_0=10+k, y_0=20+k, z_0=30+k$ , and  $h=0.1$ .

$$PR(i) = \text{mod}(\text{floor}(x), 256);$$

$$PR(i+1) = \text{mod}(\text{floor}(y), 256);$$

$$PR(i+2) = \text{mod}(\text{floor}(z), 256);$$

where  $i$  is from 1 to  $L$  with increasing step equal 3 for every round in the loop. The  $x, y$ , and  $z$  values are derived either from the three Lorenz system equations in formula(1) or from the three Chen's system equations in formula(2).  $k$  is obtained by formula(5), in which the keys in the proposed algorithm are modified.

$$k = \text{abs}(\sum_{i=1}^m au(i));$$

**Step 3:** Obtain  $d2$  the length of the second dimension of the one block from audio signals' file by using the first dimension

( $d1$ ) for the one block, and the number of blocks ( $nb$ ) which are needed to divide the  $au$  vector using its length ( $L$ ); this as in formula (6).

$$lb = L/nb; \quad d2 = lb/d1;$$

**Step 4:** By the two dimensions  $d1$ , and  $d2$ , both  $PR$  and  $au$  vectors are divided into sub-vectors of  $lb=d1*d2$  length, as in formula (7) to produce  $PRb$  and  $aub$  sub-vectors over the loop.

$$PRb = PR(1, i:i+lb-1); \quad aub = au(1, i:i+lb-1);$$

where  $i$  is from 1 to  $L$  with increasing step equal  $lb$  for every round in the loop. Also, each of both vectors  $PRb$  and  $aub$  is reshaped by MatLab into two dimensions matrices  $aub2$  and  $PRb2$  of size ( $d1 \times d2$ ), as in formula (8).

$$aub2 = \text{reshape}(aub, d1, d2);$$

$$PRb2 = \text{reshape}(PRb, d1, d2);$$

**Step 5:** Inside the previous loop, the matrix  $PRb2$  is sorted in ascending sort by using MatLab. The Matrix  $PRb2s$  is produced from this sorting process, as in formula (9).

$$[PRb2s, index] = \text{sort}(PRb2, 2);$$

Also, it returns the matrix of indices  $index$ , where  $\text{size}(index)=\text{size}(PRb2)$ . For example, if  $PRb2 = \begin{bmatrix} 3 & 7 & 0 & 5 \\ 0 & 4 & 5 & 2 \end{bmatrix}$ , then  $[PRb2s, index] = \text{sort}(A, 2)$  produces the following:

$$PRb2s = \begin{bmatrix} 0 & 3 & 5 & 7 \\ 0 & 2 & 4 & 5 \end{bmatrix}, \text{ and } index = \begin{bmatrix} 3 & 1 & 4 & 2 \\ 1 & 4 & 2 & 3 \end{bmatrix}$$

**Step 6:** The reshaped matrix  $aub2$  are rearranged according to the position of  $PRb2$  in  $PRb2s$ , i.e., according to values of positions in the matrix  $index$ , as in formula (10).

$$aub2\_e(i, j) = aub2(i, index(i, j));$$

$$\text{For example, let's suppose } aub2 = \begin{bmatrix} 115 & 30 & 50 & 110 \\ 30 & 45 & 65 & 120 \end{bmatrix},$$

Then formula (10) produces the following:

$$aub2\_e = \begin{bmatrix} 50 & 115 & 110 & 30 \\ 30 & 120 & 45 & 65 \end{bmatrix},$$

*Note that,* at decryption process, we obtain decrypted matrix ( $aub2\_d$ ) by using backward process of formula (10) as;  $aub2\_d(i, index(i, j)) = aub2\_e(i, j)$ ;

**Step 7:** At the end of every round of the loop, the matrix  $aub2\_e$  is reshaped by MatLab into one dimensions matrix (vector) of length ( $d1*d2$ ), as in formula (11).

$$\text{out}_e = \text{reshape}(aub2\_e, 1, d1*d2);$$

## V. APPLYING VSA3DCS AND ARNOLD ON AUDIO FILE

In this section, the debate and results of applying the VSA3DCS algorithm and Arnold-based algorithm on audio signals are presented. Original audio signals' file of a conversation between two persons in the time domain ( $TD$ ) which its patterns illustrated in Fig. 4. The length ( $L$ ) of this vector of audio signals for this file is equal to 60416.

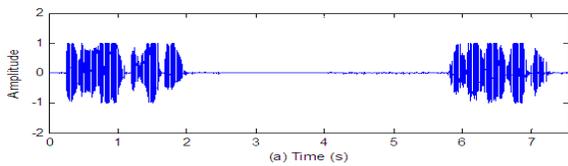


Fig 4. Original Audio Signals' Patterns in the Time Domain.

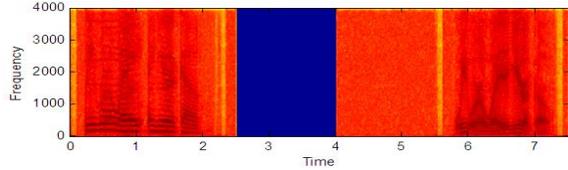


Fig 5. Spectrograms of the Audio Signals for the Original Audio.

The VSA3DCS algorithm is applied on the original audio file either based on Chen's system at  $a=35$ ,  $b=3$ ,  $c=28$ , the initial values  $x_0=0+k$ ,  $y_0=1+k$ ,  $z_0=0+k$ , and  $h=0.055555$ , or based on Lorenz system at  $\sigma=10$ ,  $r=28$ ,  $b=8/3$ , the initial value  $x_0=10+k$ ,  $y_0=20+k$ ,  $z_0=30+k$ , and  $h=0.1$ , whereas  $k$  is obtained by formula (5). The Arnold-based algorithm is applied to the same original audio file with choice  $p=1$ ,  $q=1$ , and  $R=1$ .

All algorithms are applied on the audio file with the first dimension for each one block  $d1 = 4, 8, 16, \text{ or } 32$ , and the number of blocks  $nb = 1\text{block, or } 2\text{blocks}$ . And the second dimension of each one block  $d2$  is obtained by formula (6).

Fig. 5 illustrates the spectrograms of the signals of the original audio file, which illustrated in Fig. 4.

#### A. Scrambled Audio Signals' Patterns

Fig. 6 illustrates scrambled audio signals' patterns for the Arnold-based algorithm and the VSA3DCS algorithm based on both systems (Lorenz and Chen's) at the case of  $d1=4$ ,  $nb=1$ , which led to  $d2=15104$ , whereas, Fig. 6(a) shows the result for Arnold, Fig. 6(b) shows the result of VSA3DCS with Lorenz, and Fig. 6(c) shows the result of VSA3DCS with Chen's.

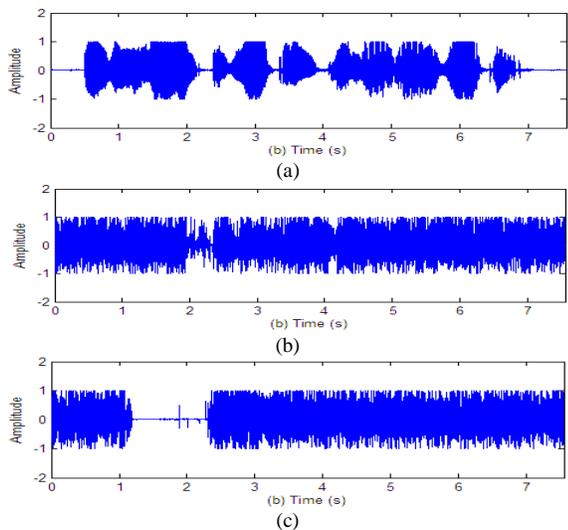


Fig 6. Scrambled Audio Signals' Patterns in TD at  $d1=4$ ,  $nb=1$ . (a) By Arnold-based Algorithm. (b) By VSA3DCS Algorithm with Lorenz. (c) By VSA3DCS Algorithm with Chen's.

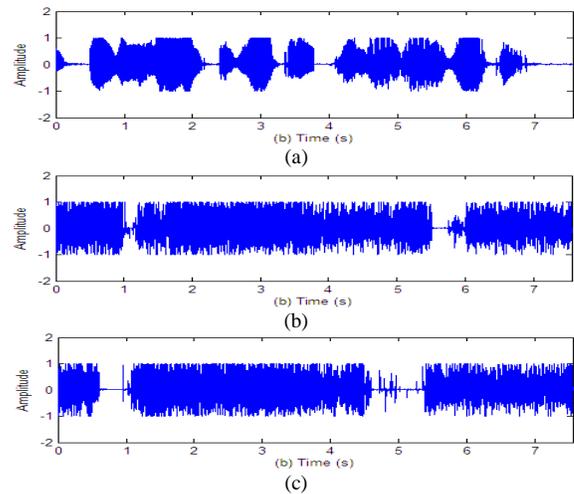


Fig 7. Scrambled Audio Signals' Patterns in TD at  $d1=4$ ,  $nb=2$ . (a) By Arnold-based Algorithm. (b) By VSA3DCS Algorithm with Lorenz. (c) By VSA3DCS Algorithm with Chen's.

Fig. 7 illustrates scrambled audio signals' patterns for the Arnold-based algorithm and the VSA3DCS algorithm with both chaotic systems at the case of  $d1=4$ ,  $nb=2$ , which led to  $d2=7552$ , whereas, Fig. 7(a) shows the result for Arnold, Fig. 7(b) shows the result of VSA3DCS with Lorenz, and Fig. 7(c) shows the result of VSA3DCS with Chen's.

Fig. 8 illustrates scrambled audio signals' patterns for the Arnold-based algorithm and the VSA3DCS algorithm with both chaotic systems at the case of  $d1=8$ ,  $nb=1$ , which led to  $d2=7552$ , whereas, Fig. 8(a) shows the result for Arnold, Fig. 8(b) shows the result of VSA3DCS with Lorenz, and Fig. 8(c) shows the result of VSA3DCS with Chen's.

Fig. 9 illustrates scrambled audio signals' patterns for the Arnold-based algorithm and the VSA3DCS algorithm with both chaotic systems at the case of  $d1=8$ ,  $nb=2$ , which led to  $d2=3776$ , whereas, Fig. 9(a) shows the result for Arnold, Fig. 9(b) shows the result of VSA3DCS with Lorenz, and Fig. 9(c) shows the result of VSA3DCS with Chen's.

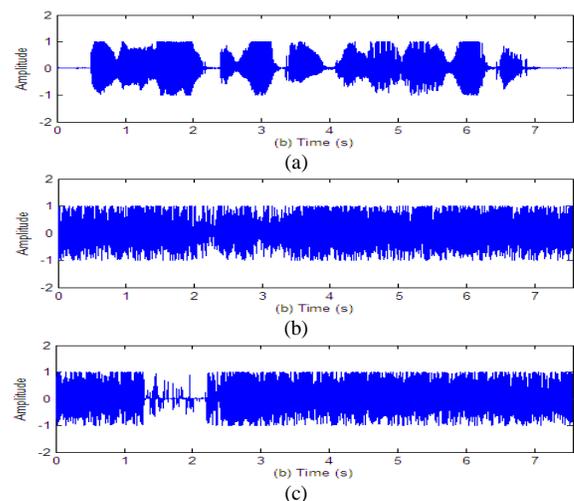


Fig 8. Scrambled Audio Signals' Patterns in TD at  $d1=8$ ,  $nb=1$ . (a) By Arnold-based Algorithm. (b) By VSA3DCS Algorithm with Lorenz. (c) By VSA3DCS Algorithm with Chen's.

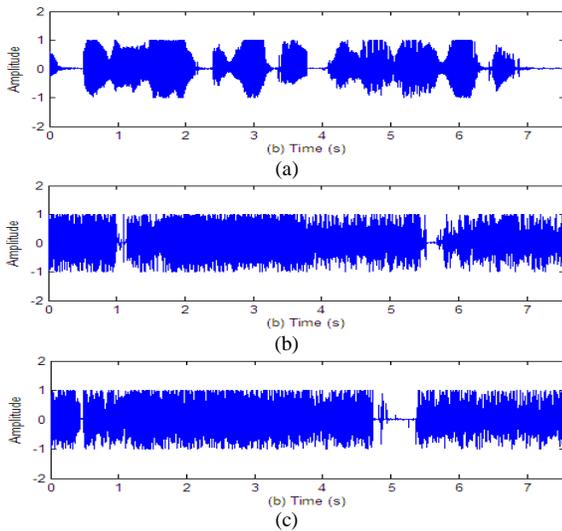


Fig 9. Scrambled Audio Signals' Patterns in TD at  $d1=8, nb=2$ . (a) By Arnold-based Algorithm. (b) By VSA3DCS Algorithm with Lorenz. (c) By VSA3DCS Algorithm with Chen's.

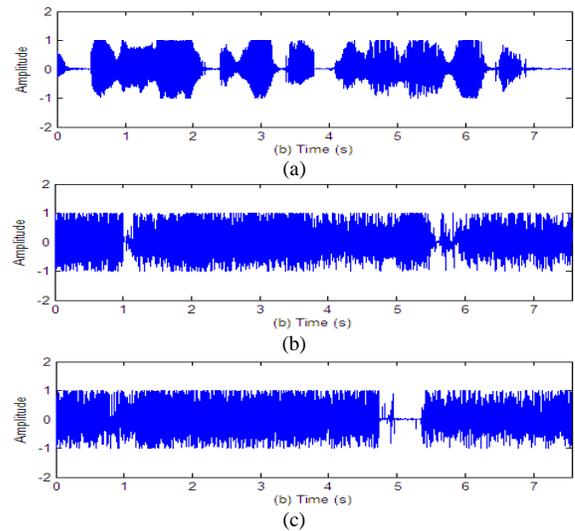


Fig 11. Scrambled Audio Signals' Patterns in TD at  $d1=16, nb=2$ . (a) By Arnold-based Algorithm. (b) By VSA3DCS Algorithm with Lorenz. (c) By VSA3DCS Algorithm with Chen's.

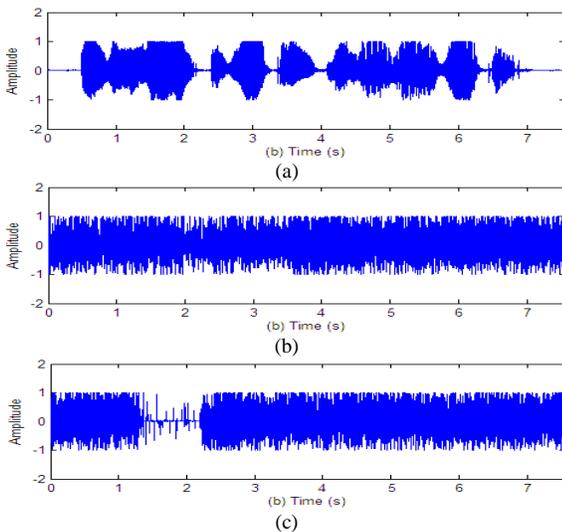


Fig 10. Scrambled Audio Signals' Patterns in TD at  $d1=16, nb=1$ . (a) By Arnold-based Algorithm. (b) By VSA3DCS Algorithm with Lorenz. (c) By VSA3DCS Algorithm with Chen's.

Fig. 10 illustrates scrambled audio signals' patterns for the Arnold-based algorithm and the VSA3DCS algorithm with both chaotic systems at the case of  $d1=16, nb=1$ , which led to  $d2=3776$ , whereas, Fig. 10(a) shows the result for Arnold, Fig. 10(b) shows the result of VSA3DCS with Lorenz, and Fig. 10(c) shows the result of VSA3DCS with Chen's.

Fig. 11 illustrates scrambled audio signals' patterns for the Arnold-based algorithm and the VSA3DCS algorithm with both chaotic systems at the case of  $d1=16, nb=2$ , which led to  $d2=1888$ , whereas, Fig. 11(a) shows the result for Arnold, Fig. 11(b) shows the result of VSA3DCS with Lorenz, and Fig. 11(c) shows the result of VSA3DCS with Chen's.

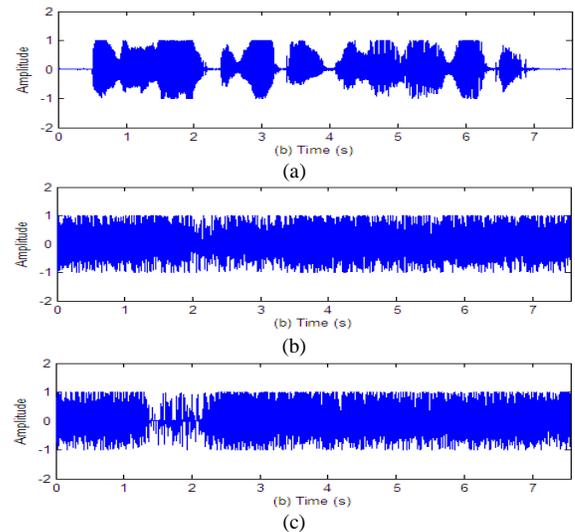


Fig 12. Scrambled Audio Signals' Patterns in TD at  $d1=32, nb=1$ . (a) By Arnold-based Algorithm. (b) By VSA3DCS Algorithm with Lorenz. (c) By VSA3DCS Algorithm with Chen's.

Fig. 12 illustrates scrambled audio signals' patterns for the Arnold-based algorithm and the VSA3DCS algorithm with both chaotic systems at the case of  $d1=32, nb=1$ , which led to  $d2=1888$ , whereas, Fig. 12(a) shows the result for Arnold, Fig. 12(b) shows the result of VSA3DCS with Lorenz, and Fig. 12(c) shows the result of VSA3DCS with Chen's.

Fig. 13 illustrates scrambled audio signals' patterns for the Arnold-based algorithm and the VSA3DCS algorithm with both chaotic systems at the case of  $d1=32, nb=2$ , which led to  $d2=944$ , whereas, Fig. 13(a) shows the result for Arnold, Fig. 13(b) shows the result of VSA3DCS with Lorenz, and Fig. 13(c) shows the result of VSA3DCS with Chen's.

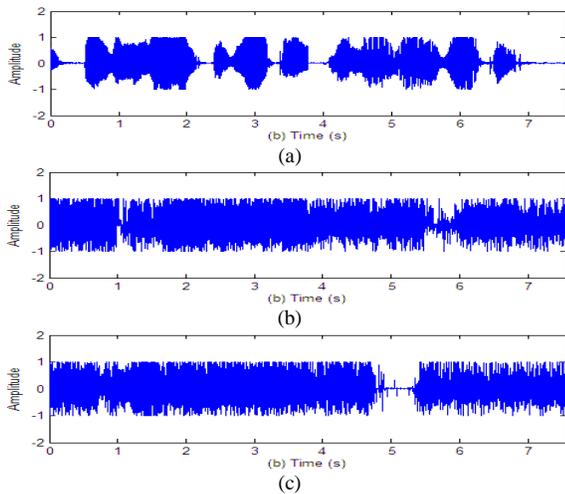


Fig 13. Scrambled Audio Signals' Patterns in TD at  $dI=32, nb=2$ . (a) By Arnold-based Algorithm. (b) By VSA3DCS Algorithm with Lorenz. (c) By VSA3DCS Algorithm with Chen's.

All Fig. 6 to 13, illustrate the results of scrambled audio by VSA3DCS algorithm with both Lorenz and Chen's are completely different than the original audio, and these results of applying the VSA3DCS algorithm are better than the results of applying Arnold-based algorithm at all cases of  $dI$  and  $nb$ .

### B. Spectrogram

A spectrogram reflects a visual representation of the frequency spectrum of a signal, as it varies over time. Generally a spectrogram is represented as an image with the intensity indicated by varying color or brightness.

Some of the results of applying all algorithms are presented, whereas, Fig. 14 illustrates Spectrogram of scrambled audio Signals for the Arnold-based algorithm and the VSA3DCS algorithm based on both systems (Lorenz and Chen's) at the case of  $dI=8, nb=1$ , Fig. 14(a) shows the result for Arnold, Fig. 14(b) shows the result of VSA3DCS with Lorenz, and Fig. 14(c) shows the result of VSA3DCS with Chen's.

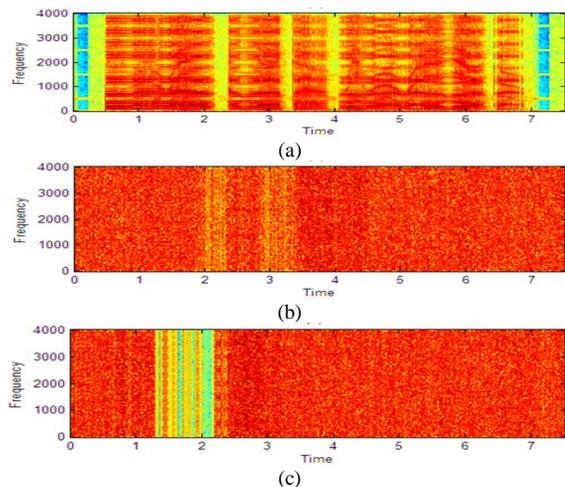


Fig 14. Spectrogram for Scrambled Audio in TD at  $dI=8, nb=1$ . (a) By Arnold-based Algorithm. (b) By VSA3DCS Algorithm with Lorenz. (c) By VSA3DCS Algorithm with Chen's.

Fig. 15 illustrates Spectrogram of scrambled audio Signals for the Arnold-based algorithm and the VSA3DCS algorithm based on both systems (Lorenz and Chen's) at the case of  $dI=16, nb=2$ , Fig. 15(a) shows the result for Arnold, Fig. 15(b) shows the result of VSA3DCS with Lorenz, and Fig. 15(c) shows the result of VSA3DCS with Chen's.

Fig. 16 illustrates Spectrogram of scrambled audio Signals for the Arnold-based algorithm and the VSA3DCS algorithm based on both systems (Lorenz and Chen's) at the case of  $dI=32, nb=1$ , Fig. 16(a) shows the result for Arnold, Fig. 16(b) shows the result of VSA3DCS with Lorenz, and Fig. 16(c) shows the result of VSA3DCS with Chen's.

Fig. 14 to 16 illustrate Spectrogram of scrambled audio with all algorithms is completely different from than spectrogram of the original audio at all cases, and the result of applying the VSA3DCS algorithm with both Lorenz and Chen's is better than the result of applying Arnold-based algorithm at most cases of  $dI$  and  $nb$ .

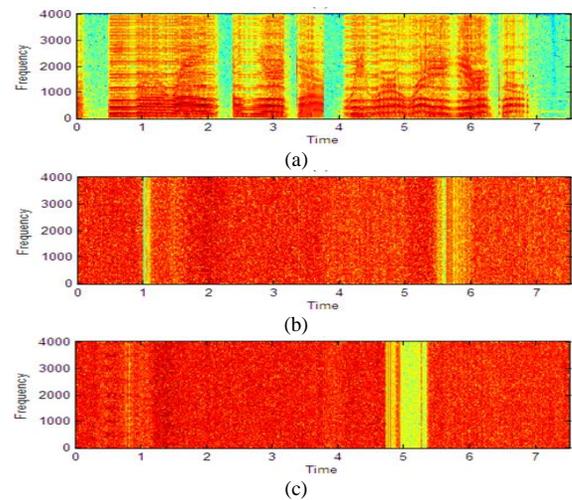


Fig 15. Spectrogram for Scrambled Audio in TD at  $dI=16, nb=2$ . (a) By Arnold-based algorithm. (b) By VSA3DCS algorithm with Lorenz. (c) By VSA3DCS algorithm with Chen's.

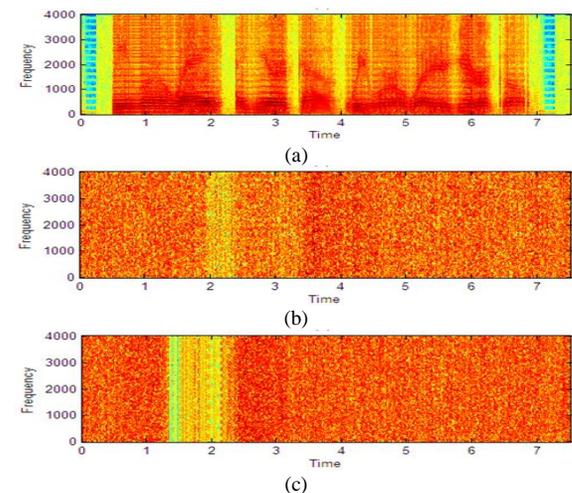


Fig 16. Spectrogram for Scrambled audio in TD at  $dI=32, nb=1$ . (a) By Arnold-based Algorithm. (b) By VSA3DCS Algorithm with Lorenz. (c) By VSA3DCS Algorithm with Chen's.

C. Histogram

For continuous data a histogram is used where the bins reflect data ranges. Also, a histogram is an approximate representation of the numerical or categorical data distribution.

Since all algorithms are used for the process of encryption by Scrambling (shuffling of signals' locations) for audio signals, so the histogram of the scrambled audio signals for all cases completely matched to the histogram of the original audio signals which illustrated in Fig. 17.

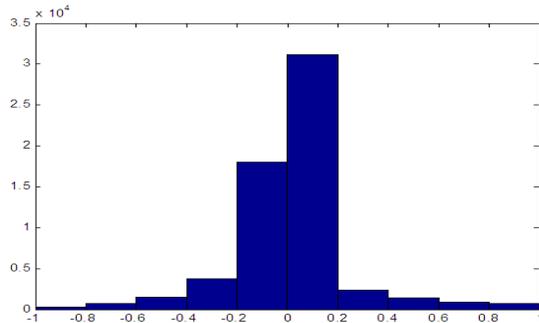


Fig 17. Histogram of the Audio Signals for Original Audio and Scrambled Audio Signals by applying all Algorithms at all Cases.

In the decryption process with all algorithms, all results and plots of audio signals' patterns, spectrogram, and histogram for decrypted audio signals are matched to all plots of the original audio signals illustrated in Fig. 4, 5, and 17, respectively. This indicates that the decryption process is equally successful and efficient with applying all algorithms in all cases.

VI. EXPERIENTIAL RESULTS AND COMPARITIVE ANALYSIS

Here we present experiential findings and comparative analysis using some of several experiential and statistical analyzes for both encryption and decryption procedures, such as encryption and decryption time, correlation coefficient (CC) of evident and encrypted signals between samples, measurement of spectral distortion (SD), measurement of log-likelihood ratio (LLR), and measurement of key sensitivity.

A. Encryption and Decryption Time

In this analysis, for applying all algorithms on the original audio signals file at all cases for both  $dI(4, 8, 16, \text{ and } 32)$  and  $nb(1, \text{ and } 2)$ , the execution time of encryption and decryption has been calculated by seconds.

TABLE I. ENCRYPTION TIME IN SEC. FOR ALL ALGORITHMS

Algorithms	nb	Results of Encryption Time in Sec. for all algorithms at all cases of dI and nb			
		dI=4	dI=8	dI=16	dI=32
Arnold-based Algorithm	nb=1	4.3000	1.1070	0.3120	0.1090
	nb=2	2.1840	0.5770	0.1870	0.0690
VSA3DCS with Lorenz	nb=1	2.1990	0.5930	0.1870	0.0930
	nb=2	1.1390	0.3280	0.1190	0.0670
VSA3DCS with Chen's	nb=1	2.1840	0.5930	0.1870	0.0940
	nb=2	1.1230	0.3280	0.1240	0.0670

Table I, shows the execution time of encryption procedure for applying all algorithms in all cases of  $dI$  and  $nb$ . Also, Table II, shows the execution time of decryption procedure for applying all algorithms in all cases of  $dI$  and  $nb$ .

Fig. 18 shows the plot for the results of encryption time of applying all algorithms in all cases of  $dI$  and  $nb$ . Also, Fig. 19 shows the plot for the results of decryption time of applying all algorithms in all cases of  $dI$  and  $nb$ .

TABLE II. DECRYPTION TIME IN SEC. FOR ALL ALGORITHMS

Algorithms	nb	Results of Decryption Time in Sec. for all algorithms at all cases of dI and nb			
		dI=4	dI=8	dI=16	dI=32
Arnold-based Algorithm	nb=1	5.3200	1.3570	0.3590	0.1090
	nb=2	2.6990	0.7020	0.2030	0.0690
VSA3DCS with Lorenz	nb=1	0.1090	0.0780	0.0780	0.0870
	nb=2	0.0820	0.0780	0.0780	0.0680
VSA3DCS with Chen's	nb=1	0.0860	0.0830	0.0830	0.0780
	nb=2	0.0780	0.0860	0.0780	0.0680

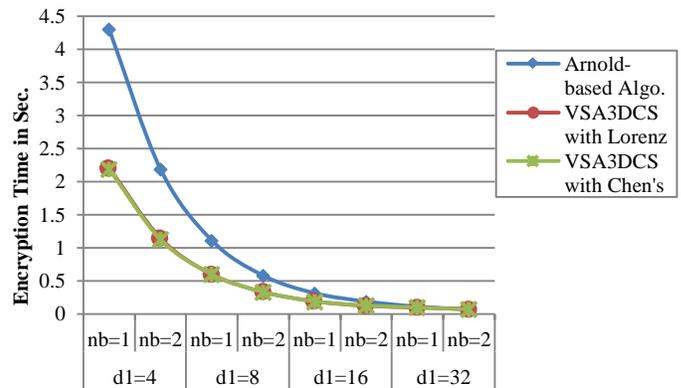


Fig 18. Encryption Time in Sec. of applying all Algorithms at all Cases.

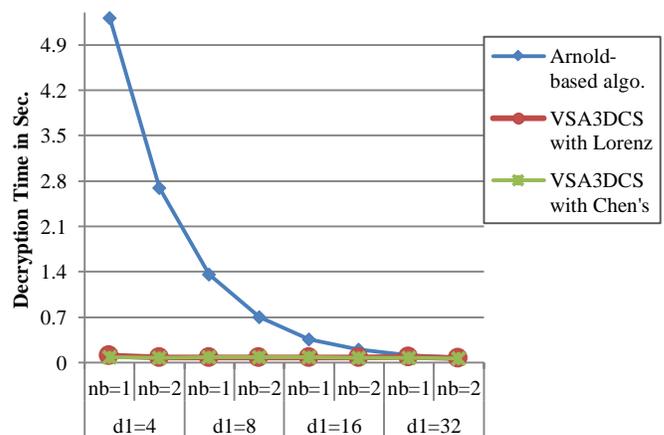


Fig 19. Decryption Time in Sec. of applying all Algorithms at all Cases.

Tables I and II, with Fig. 18 and 19, illustrate the execution time of encryption and decryption of the VSA3DCS algorithm with both Lorenz and Chen's is less than the time encryption and decryption of the Arnold-based algorithm at all cases of  $dI$  and  $nb$ . So, the VSA3DCS algorithm with both Lorenz and Chen's is better than the Arnold-based algorithm in all cases of  $dI$  and  $nb$ .

**B. Correlation Coefficient Measure**

If encrypted and original files are highly correlated, the coefficient of correlation equals one, i.e. the encryption method is ineffective in hiding the original signal information. If the coefficient of correlation is equal to zero then the initial voice signals and its encryption are entirely different. Progress of the encryption method thus implies lower CC values. The CC is computed using formula (12) [1,2,6,7]:

$$CC = \frac{\sum_{i=1}^N (x_i - E(x))(y_i - E(y))}{\sqrt{\sum_{i=1}^N (x_i - E(x))^2} \sqrt{\sum_{i=1}^N (y_i - E(y))^2}} \quad (12)$$

Table III shows the results of CC analysis for encrypting by applying all algorithms; Arnold-based algorithm, and VSA3DCS with both Lorenz and Chen's on original audio signals at all cases of  $dI$  and  $nb$ . Fig. 20 illustrates the plot for the results of CC for scrambled audio signals produced by all algorithms at all cases of  $dI$  and  $nb$ .

TABLE III. RESULTS OF CC ANALYSIS FOR ENCRYPTING WITH ALL ALGORITHMS IN ALL CASES

Algorithms	nb	Results CC for all Algorithms at all cases of $dI$ and $nb$			
		$dI=4$	$dI=8$	$dI=16$	$dI=32$
Arnold-based Algorithm	$nb=1$	0.0046	-0.0116	-0.0086	-0.0051
	$nb=2$	0.0045	-0.0117	-0.0085	-0.0050
VSA3DCS with Lorenz	$nb=1$	-0.0026	0.0042	0.0031	0.0011
	$nb=2$	-0.0018	-0.00086	0.000406	-0.0035
VSA3DCS with Chen's	$nb=1$	0.000081	-0.0019	0.0033	-0.0034
	$nb=2$	0.0026	-0.0012	0.0021	0.0050

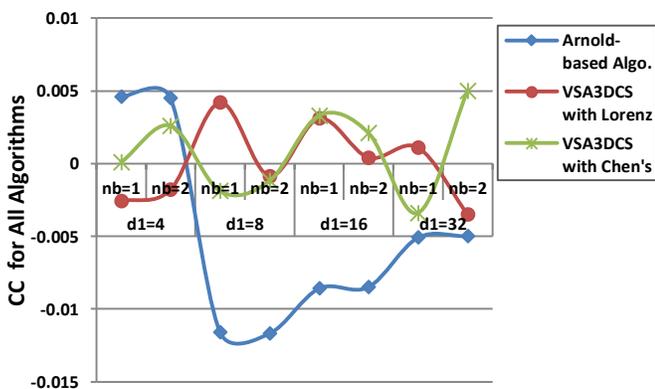


Fig 20. The CC of applying all Algorithms in all Cases.

Table III and Fig. 20 illustrate that the VSA3DCS algorithm with both Lorenz and Chen's achieves very small values (near to zero) of CC compared with the results of the Arnold-based algorithm, i.e., the VSA3DCS algorithm with both Lorenz and Chen's better than the Arnold-based algorithm. So, the proposed algorithm VSA3DCS is complex and strong for the encryption of the audio signal.

Results of the CC of decrypted audio signals equal to 1 for all algorithms in all cases of  $dI$  and  $nb$ , because decryption by all algorithms returns the decrypted audio signals file completely matched to the original audio signals file.

**C. Spectral Distortion (SD) Measure**

The SD is a type of measurements implemented in the frequency spectra of original and encrypted audio signals within the frequency domain. In dB it is calculated to demonstrate how far from that of the original audio signals the encrypted signal range is. The SD is calculable as in formula (13) [5,18]:

$$SD = \frac{1}{M} \sum_{m=0}^{M-1} \sum_{n=L,m}^{L,m+L-1} |V_s(k) - V_y(k)| \quad (13)$$

Where  $V_s(k)$  is the spectrum of the primary audio signal in dB for a given portion,  $V_y(k)$  is the spectrum of the encoded/decoded audio signal in dB for the same portion,  $M$  is the number of portions and  $L$  is the duration of the portion. The bigger the SD between the original and encrypted signals, the greater the encryption efficiency. On the other hand, between the primary audio signals and the decrypted signals, The SD must be as small as possible.

Table IV shows the results of the SD measure for encrypting by applying all algorithms in all cases of  $dI$  and  $nb$ . And, Fig. 21 displays the results of SD for encrypted audio signals produced by applying all algorithms at all values of  $dI$  and  $nb$ .

Table IV and Fig. 21 illustrate that all algorithms (VSA3DCS with both chaotic systems and Arnold) achieve good values for SD at all cases of  $dI$  and  $nb$ , whereas all results bigger than 13.91 (far from zero), so all of them are complex and strong algorithms for audio signals encryption. But in the most cases, the results of VSA3DCS with both chaotic systems is greater and better than the results of the Arnold-based algorithm.

TABLE IV. RESULTS OF SD MEASURE FOR ENCRYPTING WITH ALL ALGORITHMS AT ALL CASES

Algorithms	nb	Results SD for all algorithms at all cases of $dI$ and $nb$			
		$dI=4$	$dI=8$	$dI=16$	$dI=32$
Arnold-based Algorithm	$nb=1$	14.0178	13.9143	13.9700	13.9587
	$nb=2$	14.0046	13.9322	13.9867	13.9371
VSA3DCS with Lorenz	$nb=1$	14.0000	14.0400	14.0100	14.0000
	$nb=2$	14.0000	14.0376	13.9886	13.9500
VSA3DCS with Chen's	$nb=1$	14.0100	14.0220	13.9800	14.0400
	$nb=2$	14.0223	13.9700	14.1000	14.1176

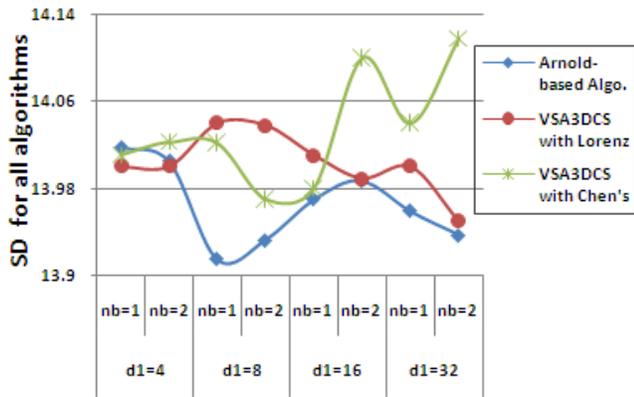


Fig 21. The SD of applying all Algorithms at all Cases.

Results of SD for decrypted audio signals equal to 0 with all algorithms at all cases, because decrypted audio signals file completely matched to the original audio signals file.

D. Log-Likelihood Ratio (LLR) Measure

The Audio signal LLR metric is based on the assumption that each component can be interpreted through a predictive linear all-pole model of the formula (14) [5,18]:

$$S(n) = \sum_{m=1}^{m_p} a_m s(n - m) + G_s u(n) \tag{14}$$

where  $a_m$  (for  $m=1, 2, \dots, m_p$ ) are all-polar filter coefficients,  $G_s$  is the filter gain and  $u(n)$  is a good source of excitation for the filter. The audio signal is fenced to form frames have lengths of 15 to 30ms. LLR metric is then determined as in [5]:

$$LLR = \left| \log \left( \frac{\bar{a}_s \bar{R}_y \bar{a}_s^T}{\bar{a}_y \bar{R}_y \bar{a}_y^T} \right) \right| \tag{15}$$

where,  $\bar{a}_s$  is the coefficient vector for LPCs;  $[1, a_s(1), a_s(2), \dots, a_s(m_p)]$  for the premier clear audio signal,  $\bar{a}_y$  is the coefficient vector for LPCs;  $[1, a_y(1), a_y(2), \dots, a_y(m_p)]$  for the encryption/decrypted audio signals, and  $\bar{R}_y$  is the autocorrelation matrix of the encryption/decrypted audio signals. The higher the LLR between the original and the encrypted signals, the greater the encryption efficiency. In comparison, the lower the LLR is to zero, the greater the decryption efficiency.

TABLE V. RESULTS OF LLR MEASURE FOR ENCRYPTING WITH ALL ALGORITHMS IN ALL CASES

Algorithms	nb	Results LLR for all algorithms in all cases of dl and nb			
		dl=4	dl=8	dl=16	dl=32
Arnold-based Algorithm	nb=1	0.7028	0.5524	0.6776	0.5927
	nb=2	0.4741	0.4674	0.4272	0.3375
VSA3DCS with Lorenz	nb=1	0.7746	0.7763	0.7671	0.8371
	nb=2	0.7133	0.7207	0.8081	0.6738
VSA3DCS with Chen's	nb=1	0.7238	0.7000	0.7799	0.7694
	nb=2	0.7463	0.7133	0.7415	0.7144

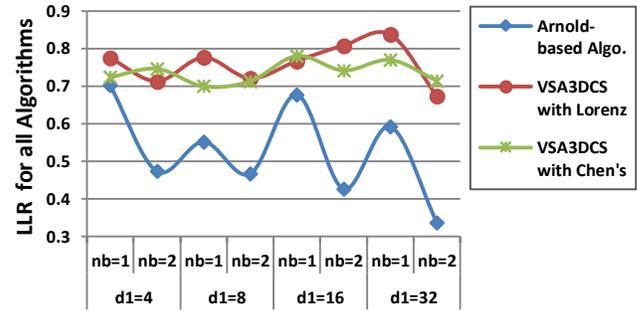


Fig 22. The LLR of applying all Algorithms in all Cases.

Table V shows the results of LLR measure for encrypting by applying all algorithms in all cases of  $dl$  and  $nb$ . Fig. 22 displays results of LLR of encrypted audio signals generated by all algorithms in all cases.

Table V and Fig. 22 illustrate that the VSA3DCS algorithm with both Lorenz and Chen's achieves very good results for LLR in all cases of  $dl$  and  $nb$ , i.e., the LLR results with VSA3DCS algorithm are better than the results with Arnold-based algorithm at all cases of  $dl$  and  $nb$ . So, the proposed algorithm VSA3DCS are complex and strong algorithm for audio signal encryption.

Results of LLR of decrypted audio signals equal to 0 for all algorithms at all cases of  $dl$  and  $nb$ , because the decrypted audio signals file completely matched to the original audio signals file.

E. Key Sensitivity Measure

The experimental results indicate that both the Arnold-based algorithm and the VSA3DCS algorithm with both Lorenz and Chen's are extremely sensitive to the mismatching of hidden keys. Table VI displays keys sensitivity results for all algorithms.

From Table VI, we can see that the VSA3DCS algorithm with both Lorenz and Chen's has greater space for the keys than the Arnold-based algorithm. Also, any of the keys with little movement (e.g.,  $10^{-17}$  is modified to  $h$ ) will generate an incorrect decrypted image. VSA3DCS algorithm is therefore very sensitive to the keys, and they can also withstand various sensitivity dependent attacks.

TABLE VI. RESULTS OF KEY SENSITIVITY MEASURING FACTOR FOR ALL ALGORITHMS

Name and precision For Arnold-based Algorithm							
q				p			
10 <sup>-15</sup>				10 <sup>-15</sup>			
Name and precision For VSA3DCS with Chen's							
a	b	c	h	k	x <sub>0</sub>	y <sub>0</sub>	z <sub>0</sub>
10 <sup>-14</sup>	10 <sup>-15</sup>	10 <sup>-14</sup>	10 <sup>-17</sup>	10 <sup>-10</sup>	10 <sup>-16</sup>	10 <sup>-15</sup>	10 <sup>-14</sup>
Name and precision For VSA3DCS with Lorenz							
σ	r	b	h	k	x <sub>0</sub>	y <sub>0</sub>	z <sub>0</sub>
10 <sup>-15</sup>	10 <sup>-14</sup>	10 <sup>-15</sup>	10 <sup>-17</sup>	10 <sup>-10</sup>	10 <sup>-16</sup>	10 <sup>-15</sup>	10 <sup>-14</sup>

Table VI illustrates the results of the precision of the keys for the VSA3DCS algorithm are better than the results of the Arnold-based algorithm. Therefore, VSA3DCS satisfies high quality of security better than the other.

## VII. CONCLUSION

In this paper, a proposed voice scrambling algorithm (VSA3DCS) based on one of 3D chaotic maps systems (Lorenz or Chen's) is presented and compared with the Arnold-based algorithm. VSA3DCS algorithm and Arnold chaotic algorithm are applied on audio signals file to encrypt it by scrambling process for its signals' positions. The encrypted audio signals which produced from applying all algorithms are compared and discussed by using some experiential measures and comparative analysis, such as; the encryption/decryption time, the Correlation Coefficient (CC) of the evident and encrypted signals between samples, the Spectral Distortion (SD) measure, Log-Likelihood Ratio (LLR) measure, and key sensitivity measure. The encryption/decryption time for all algorithms is very good, but the VSA3DCS algorithm with both Lorenz and Chen's achieves encryption/decryption time very close to zero and less and better than encryption/decryption time of Arnold-based algorithm in all cases of *dl* and *nb*. The results of CC are better with the VSA3DCS algorithm than the other with the Arnold-based algorithm in all cases of *dl* and *nb*. In the results of SD, in the most cases, the results of VSA3DCS with both chaotic systems are greater and better than the results of the Arnold-based algorithm. The results of LLR are better with the VSA3DCS algorithm than the other with the Arnold-based algorithm at all cases of *dl* and *nb*. Also, the VSA3DCS algorithm with both Lorenz and Chen's has greater space for the keys than the Arnold-based algorithm, also, the VSA3DCS algorithm is very sensitive to the keys. Also, the plots of scrambled audio signals' patterns and spectrogram illustrate the VSA3DCS algorithm with both Lorenz and Chen's is better than the Arnold-based algorithm. The final results show that the VSA3DCS algorithm is a strong algorithm to supply an efficient and stable approach for encrypting audio signals.

## REFERENCES

- [1] Osama M. Abu Zaid, Nawal F El Fishawy, Elsayed Nigm " Encryption Quality Measurement of a Proposed Cryptosystem Algorithm for the Colored Images Compared with Another Algorithm," Arab J. Inf. Technol. 13 (1), pp. 20-29, Jan. 2016.
- [2] El-Fishawy N., Osama M. Abu Zaid, "Quality of Encryption Measurement of Bitmap Images with RC6, MRC6, and Rijndael Block Cipher Algorithms," International Journal of Network Security, vol. 5, no. 3, pp. 241-51, 2007.
- [3] Roayat Ismail Abdelfatah, Mohamed E. Nasr, and Mohammed A. Alsharqawy, " Encryption for Multimedia Based on Chaotic Map: Several Scenarios," Multimedia Tools and Applications, Springer Nature, <https://doi.org/10.1007/s11042-020-08788-8>, 2020.
- [4] Osama M. Abu Zaid, El-fishawy N., Nigm E., Faragallah O., "A Proposed Encryption Scheme Based on Henon Chaotic System (PESH) for Image Security," International Journal of Computer Applications, vol. 61, no. 5, pp.29-39, 2013.
- [5] E. Mosa, N.W. Messiha, O. Zahran, F.E. Abd El-Samie. Dec.2011. "Chaotic encryption of speech signals", International Journal of Speech Technology,14(4), pp. 285-296, 2011.
- [6] Ankita Bisht, Mohit Dua, Shelza Dua and Priyanka Jaroli, "A Color Image Encryption Technique Based on Bit-Level Permutation and Alternate Logistic Maps," Journal of Intelligent Systems , vol. 29, no. 1, pp.1246-1260, 2020.
- [7] Chao-Feng Zhao, and Hai-Peng Ren, " Image Encryption Based on Hyper-Chaotic Multi-Attractors," Nonlinear Dynamics, Springer, vol. 100, pp.679-698, 2020.
- [8] Mahmoud F. Abd Elzاهر, Mohamed Shalaby, and Salwa H. El Ramly, "An Arnold Cat Map-Based Chaotic Approach for Securing Voice Communication," INFOS '16: Proceedings of the 10th International Conference on Informatics and Systems, pp. 329-331, May 2016.
- [9] Chong Fu, Wen-jing Li, Zhao-yu Meng, Tao Wang, and Pei-xuan Li, " A Symmetric Image Encryption Scheme Using Chaotic Baker map and Lorenz System," Ninth International Conference on Computational Intelligence and Security 2013, IEEE Xplore, pp. 724-728, 2013.
- [10] Huibin Lu, Xia Xiao., "A Novel Color Image Encryption Algorithm Based on Chaotic Maps," Advances in Information Sciences and Service Sciences (AISS), Vol. 3, No. 11, pp. 28-35, December 2011.
- [11] Osama M. Abu Zaid, Nawal F El Fishawy, Elsayed Nigm "Cryptosystem Algorithm Based on Chaotic System for Encrypting Colored Image," International Journal of Computer Science Issues 10 (4), pp. 215-224, 2013.
- [12] M Demba, and Osama M. Abu Zaid, "A Proposed Confusion Algorithm Based on Chen's Chaotic System For Securing Colored Images," International Journal of Signal Processing Systems Vol. 1, No. 2, pp. 296-301, December 2013.
- [13] Osama M. Abu Zaid, Nawal F El Fishawy, Elsayed Nigm "A Proposed Permutation Scheme Based on 3-D Chaotic System," International Journal of Computer Science Issues 10 (4), pp. 208-214, 2013.
- [14] Zhenwei Shang, Honge Ren, Jian Zhang., " A Block Location Scrambling Algorithm of Digital Image Based on Arnold Transformation ", the 9th International Conference for Young Computer Scientists (ICYCS2008), IEEE Xplore, pp. 2942-2947, 2008.
- [15] Anjana Savita, Phool Singh, A K Yadav, and Kehar Singh, "Asymmetric Audio Encryption System Based on Arnold Transform and Random Decomposition", Asian Journal of Physics, Vol. 27, Nos 9-12, pp. 711-719, 2018.
- [16] Zhi-Hong Guan, Fangjun Huang, and Wenjie Guan, "Chaos-Based Image Encryption Algorithm," Physics Letters A, Vol. 346, pp. 153-157, 2005.
- [17] Yuanzhi Wang, Guangyong Ren, Julang Jiang, Jian Zhang, and Lijuan Sun, "Image Encryption Method Based on Chaotic Map," 2007 Second IEEE Conference on Industrial Electronics and Applications, IEEE Xplore, pp. 2558-2560, 2007.
- [18] E. Mosa, N.W. Messiha, O. Zahran, F.E. Abd El-Samie, "Encryption of Speech Signal with Multiple Secret Keys in Time and Transform Domains", International Journal of Speech Technology, Springer, Vol. 13, No. 4, pp. 231-242, 2010.

# Air Quality Monitoring Device for Vehicular Ad Hoc Networks: EnvioDev

Josip Balen<sup>1</sup>

J. J. Strossmayer University of Osijek,  
Faculty of Electrical Engineering, Computer Science and  
Information Technology, Osijek, Croatia

Srdan Ljepic<sup>2</sup>

Institute RT-RK Osijek LLC  
Osijek, Croatia

Kristijan Lenac<sup>3</sup>

University of Rijeka  
Faculty of Engineering  
Rijeka, Croatia

Sadko Mandzuka<sup>4</sup>

University of Zagreb  
Department of Intelligent Transport Systems  
Zagreb, Croatia

**Abstract**—Urban air pollution has become a major concern for numerous densely populated cities globally since poor air quality may cause various health problems. The first crucial step towards solving this important problem is to identify the most critical areas with the highest air pollution over the allowed limit. Nowadays, air pollution is monitored by various stationary measurement systems that are expensive, large, consume a big amount of energy and gathered data has a low spatial resolution. This paper presents EnvioDev, a mobile air quality and traffic conditions measurement device for Vehicular Ad hoc Networks (VANETs) that can be used on any type of a vehicle. EnvioDev was tested in a real-world urban environment measuring *CO*, *CH<sub>4</sub>* and *LPG* concentrations, as well as air temperature and humidity in order to create a city pollution map and the results are presented in the paper. Moreover, in order to determine how many EnvioDev devices are required to obtain close to a real-time air quality map of an urban area, three experiments with Simulation of Urban Mobility (SUMO) simulator were conducted. In the experiments an urban city map was divided into five zones and data aggregation frequencies are varied during different traffic load periods in order to study the number of required vehicles with EnvioDev measurement device. The obtained results show that by increasing the data aggregation frequency the number of required vehicles with EnvioDev measurement device increases and it is depended on the size and topology of the testing area.

**Keywords**—Air pollution; air quality; Arduino; sensors; SUMO; VANET

## I. INTRODUCTION

Air pollution is one of the major global environmental problem and air quality is decreasing in numerous world's cities, especially in densely populated areas. Air pollution can be linked to many health issues, and therefore it is a main research subject for numerous scientists across the globe. One of the basic steps towards solving this important problem would be identifying the most critical areas with the highest air pollution over the allowed limit. It would be beneficial to have a real-time pollution map of the whole target area. However, stationary measurement systems are very expensive, need construction works, require special licenses and authority

permits, and have very expensive instruments and equipment for measuring atmospheric conditions. Furthermore, they are stationary and can conduct measurements only on one specific location. Mathematical models and appropriate software tools to extrapolate this "point" data to a specific area are with very limited possibilities.

By increasing the number of measuring devices and enabling them to become mobile it is possible to measure, monitor, track and process more accurately air quality parameters, create analytics, calculate air quality index (AQI) and create a real-time pollution map of a desired area. By providing real-time environmental and traffic data to relevant authorities, living conditions such as air quality, noise distribution and traffic load, especially in urban areas, can be dramatically improved. Moreover, such devices would enable scientists and epidemiologists to efficiently gather and process useful information, conduct research based on influence of ultrafine pollution particle effect on health conditions, fast react in a case of disasters and implement local and global pollution policies. In addition, appropriate traffic management strategies can be implemented based on this information.

The main research and development challenges and problems in this paper include building a reliable, affordable and as small as possible environmental condition measurement device that can be placed on various types of vehicles and used for real-time measurements of a target area. This paper describes an inexpensive and reliable air quality and traffic conditions measuring device EnvioDev - ENVIRONMENTAL Input/Output DEvice for Vehicular Ad hoc Networks (VANETs). EnvioDev is based on Arduino platform equipped with several sensors for measuring various atmospheric parameters and pollutants, as well as detecting device location, acceleration and angular velocity. EnvioDev can be placed on any type of a vehicle and is connected with a cloud system with GSM LTE module through cellular network. On a cloud system all data are shown on a map, and therefore are generating a real-time pollution map of a desired area. Generated data and pollution maps will be available to local authorities and will enable them to keep track of environmental

condition on health effect, detect emergency situations like gas and fuel leakage and react in a case of disasters.

EnvioDev was used in real world experiments to determine the air quality in the city of Osijek. By collecting spatially resolved measurements, five selected parameters were measured: temperature, humidity, carbon monoxide ( $CO$ ), methane ( $CH_4$ ) and liquefied petroleum gas (LPG, propane ( $C_3H_8$ ) and butane ( $C_4H_{10}$ )) concentration. Analysis of the obtained measurements results showed that on several locations LPG and  $CO$  concentrations are above expected average. Therefore, it is desirable to perform constant air quality monitoring in the future by using several EnvioDev devices in order to get more measurement results with a higher density, reliability and accuracy. In order to determine the optimal number of required EnvioDev devices that is necessary to efficiently cover a target area and provide high quality real-time measurement results an open source traffic simulator Simulation of Urban MObility (SUMO) was used [1]. Three experiments were conducted and an optimal number of devices is presented for selected scenarios of interest.

The paper is organized as follows: Section 2 describes state-of-the-art solutions to the highlighted problem in the paper. Section 3 describes EnvioDev full system architecture. Section 4 presents and discusses obtained real-world experimental results. Section 5 describes conducted simulations and discusses the results. Section 6 concludes the paper and presents the future work.

## II. RELATED WORK

Environmental pollution is an increasing problem in urban environments and has a significant impact on a climate change. Encouraged by this problem, the idea of enabling distributed and real-time air pollution monitoring has been intensively developed in the past few years mainly based on embedding mobile measurement devices in public transport vehicles since they cover most parts of urban areas and are in traffic most part of the day. Besides public transport vehicles, developed measurement devices can be mounted on other private and official vehicles such as taxi vehicles or delivery service vehicles (vans, motorbikes and even bicycles) in order to expand the area and frequency of the measurements. One important issue recognized in the related state-of-the-art work is calibration of mobile sensors that is crucial for obtaining accurate and reliable measurement results.

Solutions where public transport vehicles are used for providing mobility to mobile air quality monitoring devices are very common in the literature, as shown in [2], [3], [4], [5] and [6]. Development and deployment of mobile measurement devices mainly on city buses is described. Conducted experiments prove that there is a huge potential of such systems since they enable efficient data gathering and processing. Various pollutants were monitored across different cities and air pollution maps were created. The main issue in most of the papers was finding the optimal route of public transport vehicles in order to acquire the most reliable outcome and cover as much of the target area as possible.

Since most of the researches are using low cost sensors several papers raised the issues of sensors quality as described

in [7], [8], [9] and [10]. Therefore, measurements from low-costs sensors were compared with stationary monitoring stations equipped with certified instruments. It was concluded that the low-cost sensors accuracy was satisfied for many various applications, but the calibration is encouraged in the areas with the network of high-quality stationary measurement devices.

Researches in [11] and [12] are proposing models that rely on cloud computing and are using vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication for both, traffic and air pollution monitoring. In [13] a full ICT solution that besides using mobile measurement devices for collecting, processing and distribution of gathered data enables citizens to report any kind of pollution they detect by using their smartphones.

## III. ENVIODEV SYSTEM OVERVIEW

Major research and development challenges in this paper are related to design, development and implementation of efficient EnvioDev system architecture. Unlike the solution offered in [2] where public transport vehicles are used for providing mobility to heavy and big mobile air quality monitoring devices, one of the major goals of EnvioDev system architecture is to build small, low-cost, lightweight, low power and mobile measurement device that can be easily paced on any type of a vehicle. Furthermore, the major requirements include providing reliable real-time measurements and communication with the cloud. Moreover, EnvioDev has to be robust, resistant and weatherproof since it will be working in difficult and demanding working conditions.

In the related work several papers [2], [6] and [9] are dealing with the problem of efficient sensors calibration that heavily depends on the availability of high-quality stationary measurement stations in the target area. EnvioDev was first calibrated in the laboratory environment and afterwards in a real environment based on the obtained measurement from the one stationary measurement station that exists in the target area of city of Osijek [14]. This section introduces the architecture of developed mobile measurement device EnvioDev in terms of hardware architecture, network and communication capabilities, as well as software and cloud system design.

### A. Hardware Architecture

The core of the hardware architecture is Arduino Mega 2560 Rev3 microcontroller with 54 digital I/O pins that provide enough ports to connect multiple sensors. Values obtained from sensors provide analog information, and therefore microcontroller is also used for data conversion and processing. Fig. 1 shows EnvioDev hardware architecture, all included components and their interaction. Selected components took into account the size, design and purpose of the target EnvioDev device. Sensors with high sensitivity and fast response time were selected. The chosen sensors for air pollution measurement are the following: carbon monoxide sensor ( $CO$ ), methane sensor ( $CH_4$ ) and liquefied petroleum gas concentration sensor (propane ( $C_3H_8$ ) and butane ( $C_4H_{10}$ )). In addition, environmental parameters sensors are also included, such as temperature and humidity sensors. Moreover, geographical location and time are also collected by

using GPS sensor. Finally, all obtained data is transmitted to online cloud by using a GSM module.

### B. Network and Communication Structure

Communication and network structure are designed to provide efficient data collection, interpretation, processing and cloud storage. The sensors use I2C, a half-duplex communication protocol, and SPI, a full duplex communication protocol, to send data to the Arduino microcontroller. After receiving data from sensors, Arduino interprets and formats the data which is then sent to the GSM module. The GSM module supports bidirectional communication and is used to send data to the software system in the cloud where they will be processed, stored and used to create a pollution map of the target area. The goal is to provide proactive and reactive communication and data delivery with the lowest possible latency. In that regard the 4G mobile network used in the experiments showed satisfying performances. Fig. 2 shows EnvioDev communication and network structure. Communication capabilities of EnvioDev can be extended by using Dedicated Short Range Communication (DSRC) technology specifically designed for VANETs with V2V or V2I communication models [15] or Internet of Things (IoT) technology that is emerging in VANETs recently [16].

### C. Software System Design

In order to collect all measurements, derive air quality parameters, calculate indicators, create pollution map and present data to different types of users, software system was developed by using several different technologies and programming language at different programming layers, from a low level in Arduino to a high level in the cloud. The main requirement on the software system is to have a responsive backend system that manages all available resources efficiently and in timely manner [17]. EnvioDev software system design is shown in Fig. 3 and consisted of the following four modules and the cloud:

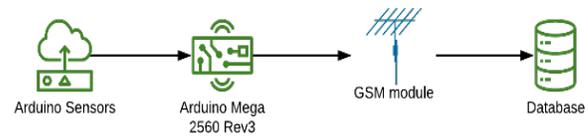


Fig 2. EnvioDev Communication and Network Structure.

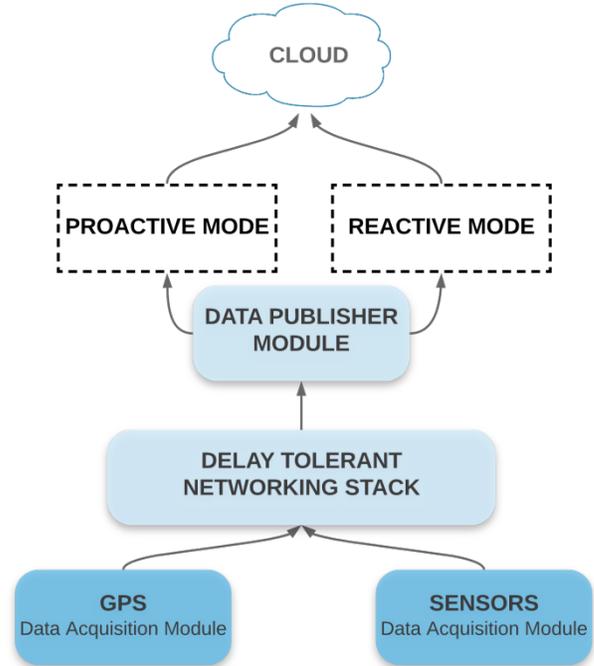


Fig 3. EnvioDev Software System Design.

- Sensor Data Acquisition Module - the module is used for receiving and converting data obtained from the sensors,
- GPS Data Acquisition Module - the module receives and converts EnvioDev location (latitude and longitude values), time and speed information,
- Delay tolerant networking (DTN) stack – is used as a temporary storage of measured data which increases EnvioDev reliability by preserving data on the device until the transmission to the Cloud is performed,
- Data Publisher Module - the module uses GSM component to periodically sends information to the Cloud and has two operating modes, *i*) proactive – data are sent based on predefined internal events and triggers, *ii*) reactive – data are sent based on the request from users,
- Cloud – consists of the database and application layer used for data presentation.

## IV. REAL-WORLD EXPERIMENTS AND RESULTS ANALYSIS

The real-world experiment was conducted in an urban area of city of Osijek in Croatia, as shown on a map in Fig. 4. On a map of tested area all locations where the measurements were conducted are marked with red dots. Most parts of the city

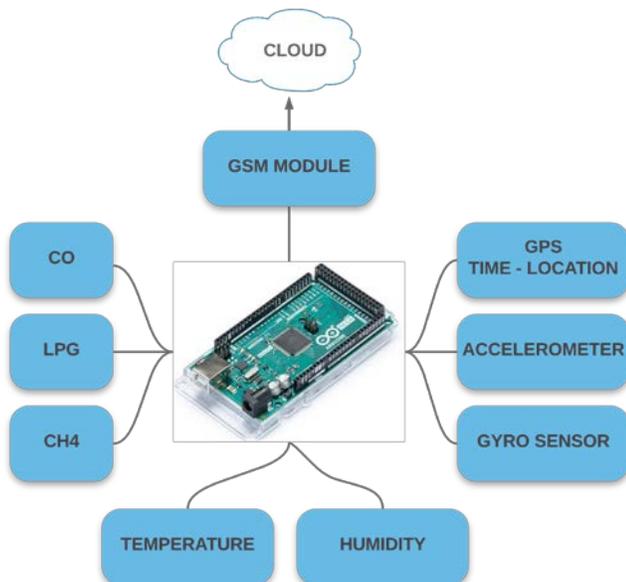


Fig 1. EnvioDev Hardware Architecture.

horizontal were covered but some locations could not be accessed due to the road works. Measurements were performed during several parts of the day but avoiding rush hours since a huge number of vehicle can influence measurement results. Due to sparse traffic during the experiment, measurement results were much more accurate since there was no influence of the gases produced by other vehicles that can dramatically influence the measurement results and provide a wrong picture of the city current pollution situation. In total five experiments were conducted by using one major route that goes through the main streets and major parts of the city of Osijek but each time the route was slightly changed. The duration of each experiment was 20 minutes and in average the route length was 25 kilometers.

Measured air quality parameters during the experiment are described in Table I. The chosen air quality metrics is based on three pollutants, namely  $CO$ ,  $CH_4$  and  $LPG$  concentration. The concentration is shown in parts-per-million (ppm,  $10^{-6}$ ) measurement unit. Furthermore, the air temperature  $T$  [ $^{\circ}C$ ] in degrees Celsius and the amount of relative humidity in the air  $H$  [%] expressed as a percentage which represents a measure of the amount of water vapor that air is holding compared the amount it can hold at a specific temperature, were measured.

EnvioDev was fixed on a roof of a personal vehicle and partially protected from the wind and air flow. However, in order to lower the influence of the air flow on the measurement results, a proactive mode in the Sensor Data Acquisition Module was used by implementing a trigger at a low level of EnvioDev software system, for taking measurements  $M_i$  only when the following conditions are satisfied: *i*) the testing

vehicle stops (GPS Data Acquisition Module: current vehicle speed  $v_{cur}$  equals to zero), *ii*) distance from the previous measurement is at least 0,1 km (GPS Data Acquisition Module: Euclidean distance between the current vehicle position  $V_{pos}(lat, lng)$  and the position of the previous measurements  $\sum_{i=0}^{N-1} M_i(lat, lng)$  recorded within the aggregation frequency interval  $f_{DA}$ ), and *iii*) previous measurements  $\sum_{i=0}^{N-1} M_{i-1}$  were not recorded within the 5% of aggregation frequency time interval (in order to reduce the number of redundant measurements), as shown in (1).

$$M_i \text{ if } \begin{cases} v_{cur} == 0 \frac{km}{h} \\ E_d(V_{pos}(lat, lng), \sum_{i=0}^{N-1} M_i(lat, lng)) > 0.1km \\ t_{cur} - f_{DA} \times 0.05 > \max \sum_{i=0}^{N-1} M_i(t) \end{cases} \quad (1)$$

Consequently, a vast majority of measurements were performed at intersections and during red traffic lights. During each measurement location coordinates (latitude and longitude values) and timestamps were retrieved from the GPS sensor (GPS Data Acquisition Module: `get_current_loaction()`; `get_timestamp()`). After each measurement all measured values were transferred to DTN stack and afterwards through GSM Publisher Module to the cloud. The cloud system gathers and processes all the measured values provided by EnvioDev since it is necessary to convert all sensor measures to the defined metrics.

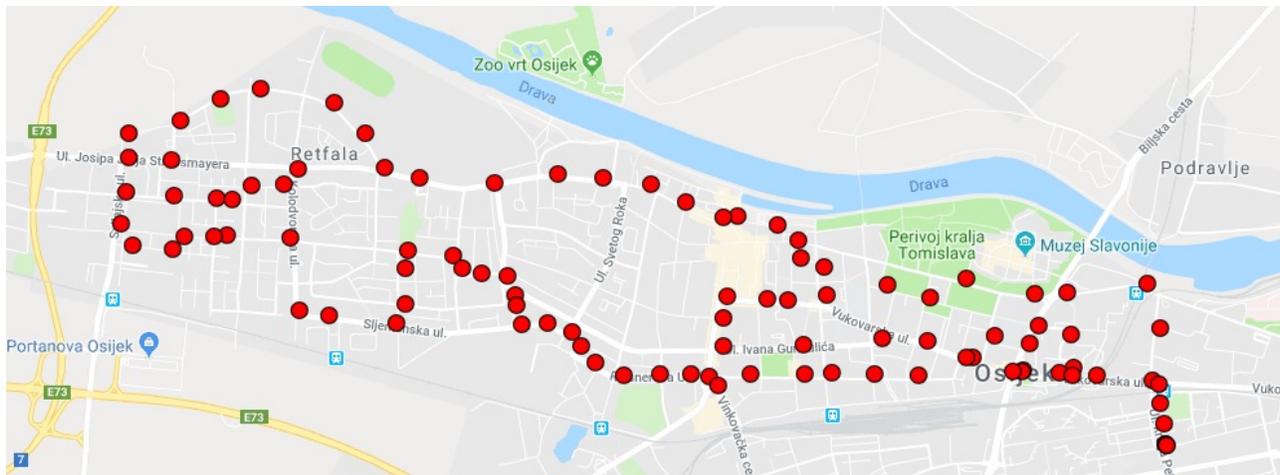


Fig 4. Map of the Experiment Environment with Locations of all Measurements taken during the Experiment.

TABLE I. AIR QUALITY PARAMETERS MEASURED DURING THE REAL-WORLD EXPERIMENT

Parameter	Measurement unit	Description
$CH_4$	ppm	Ppm concentration of methane in the air
$LPG$	ppm	Ppm concentration of propane ( $C_3H_8$ ) and butane ( $C_4H_{10}$ ) in the air
$CO$	ppm	Ppm concentration of carbon monoxide in the air
$T$	$^{\circ}C$	Air temperature in degrees Celsius
$H$	%	The amount of relative humidity in the air expressed as a percentage

Obtained measurement results are presented in Table II (only partial data is shown due to a high number of obtained results). According to World Health Organization (WHO) air quality guidelines for Europe [18] global CO concentrations range between 0.05 – 0.12 ppm (0.06 and 0.14 mg/m<sup>3</sup>). During the experiment measured CO values were ranging between 0.07 and 0.14 ppm. In most of the areas covered by the experiment, the measured values were in the recommended range, several were close to the top margin and three measurements showed a higher CO concentration than recommended that were measured at only one specific place. According to [19] ordinary edge level of LPG concentration is 400 ppm, while 600 ppm is marked as upper exposure limit. Measured results in the experiment show concernedly level of LPG concentration in most of the measurement results since majority of the tested area has a high level of LPG and several parts are above the upper exposure limit. The experiment was conducted during the period of the day with a sparse traffic when there is very little influence of the gases produced by other vehicles, so those measurement results are a warning to the local authorities since high levels of LPG can decrease the amount of the oxygen in the air and cause suffocation with

several serious health symptoms. Obtained measurement results of CH<sub>4</sub> level values in range from 550 ppm to 865 ppm. According to [20] maximum recommended safe methane concentration for a human during continuous exposure is 1.000 ppm. In comparison to average global CH<sub>4</sub> level [21] it can be concluded that CH<sub>4</sub> level in tested area is acceptable and bellow a global average.

AQI is differently calculated and used across the globe since local, regional, and national governments have its own air quality indices, corresponding to different national air quality standards. It is usually used to measure air pollution and apply measures in a case of unhealthy values since public health risks increase as the AQI rises. AQI is based on measurement of various factors such as particulate matters PM<sub>2.5</sub> and PM<sub>10</sub>, Ozone (O<sub>3</sub>), Nitrogen Dioxide (NO<sub>2</sub>), Sulfur Dioxide (SO<sub>2</sub>) and CO emissions. In our experiments CH<sub>4</sub>, LPG and CO are used for AQI calculations, as shown in (2) where M<sub>j</sub> represents a single measure from a specific sensor (M<sub>CH<sub>4</sub></sub>, M<sub>LPG</sub>, M<sub>CO</sub>), M<sub>min</sub> represents a breakpoint concentration that is ≤ M<sub>j</sub>, M<sub>max</sub> represents a breakpoint concentration that is ≥ M<sub>j</sub>, AQI<sub>max</sub> represents an index breakpoint corresponding to M<sub>max</sub>, AQI<sub>low</sub> represents an index breakpoint corresponding to M<sub>min</sub>.

TABLE II. EXCERPT OF THE MEASUREMENT RESULTS OBTAINED DURING THE EXPERIMENT

Location		T	H	CH <sub>4</sub>	LPG	CO
Latitude	Longitude	°C	%	ppm	ppm	ppm
45.55150604	18.70520591	16	52	864.06	650.78	0.1376
45.55150604	18.70520782	16	53	849.41	645.90	0.1372
45.55147171	18.70521354	16	53	739.55	643.46	0.1217
45.55142593	18.70529365	17	55	690.72	611.72	0.1123
45.55170822	18.70538520	17	56	688.28	594.63	0.1096
45.55292892	18.70497322	16	55	676.07	557.54	0.1064
45.55340195	18.70484352	16	57	646.78	567.77	0.1050
45.55550765	18.70469856	17	59	641.89	606.84	0.1041
45.55906677	18.60395469	16	56	627.25	579.98	0.1010
45.55876541	18.69502067	17	54	588.18	514.06	0.0974
45.55865097	18.68857002	18	53	563.77	526.27	0.0960
45.55622482	18.69618034	18	55	549.12	509.18	0.0936
45.55485916	18.69904136	17	51	554.00	499.41	0.0936
45.55637359	18.68939781	17	52	541.80	475.00	0.0912
45.55744552	18.67583274	16	49	556.21	445.70	0.0890
45.56023025	18.68218231	16	47	583.30	475.00	0.0896
45.56102752	18.68082809	16	49	585.74	475.00	0.0893
45.56163787	18.68025398	17	48	600.39	465.23	0.0898
...	...	...	...	...	...	...

$$AQI_j = \frac{(M_j - M_{min})x(AQI_{max} - AQI_{min})}{M_{max} - M_{min}} + AQI_{min} \quad (2)$$

For a specific city area or a time period the final  $AQI_f$  is calculated as the arithmetical mean of the  $N AQI_j$ , as shown in (3).

$$AQI_f = \frac{1}{N} \sum_{j=1}^N AQI_j \quad (3)$$

Based on the average measured results partially presented in Table II the calculated  $AQI_f$  for the city of Osijek during the experiments equals to 59.85 and according to [22] that uses US AQI metric scale it represents a moderate AQI. However, as mentioned before AQI has many various calculation principles so it is hard to compare this result to the others. Due to measurement results it can be concluded that further research and measurements regarding the air quality parameters in the tested area are required since at several locations  $LPG$  and  $CO$  concentrations are higher than the recommended level and especially since the measurements were conducted only during the part of the day with a sparse traffic. Furthermore, since  $CH_4$  level keeps increasing every year it is recommended to constantly monitor its concentration. In order to get more detailed data that can be used for a deep analysis of the air quality in the city of Osijek it is necessary to conduct more comprehensive measurements during the whole day and with a higher number of EnvioDev devices.

## V. SIMULATIONS AND DISCUSSIONS

### A. Simulation Setup

The experiment described in Section 4 was conducted by using only one EnvioDev measurement device on only one

vehicle that is not enough to cover the experiment area. In order to determine how many vehicles with implemented EnvioDev measurement devices are required to efficiently cover the whole area and satisfy the data aggregation frequency  $f_{DA}$  (how often EnvioDev should collect data) with satisfying data freshness index ( $D_{FI}$ ) defined in equation (4), a 24 hours of traffic in the experiment area was simulated by using SUMO. SUMO is an open source, microscopic, multi-modal traffic simulator that enables simulation of a given traffic demand through a real-world road network [1]. The simulation map size is 6.5 x 3.0 km since the simulated area covers the city of Osijek that is placed along a river and is much wider than it is higher. Therefore, the map is divided into five equally wide geographical zones. The full width of our map is 5500 pixels (corresponding to axis X), and therefore each geographical zone has a width of 1100 pixels, as shown in Fig. 5. The aggregation process maps each geo-located measure in the relative zone.

EnvioDev could be used on any type of a vehicle but the most efficient use would be in vehicles that are moving most of the day and are covering the whole target area. Therefore, the most convenient vehicles are public transport vehicles, taxi vehicles or even delivery service vehicles since they satisfy all the requirements. In this simulation it is assumed that EnvioDev will be implemented on taxi vehicles. Therefore, a SUMO tool called *randomTrips.py* that generates a set of random trips for a given network, in our case a map of city of Osijek, is used [23]. Taxi trips mostly have a random source and destination and are mostly driving 24/7, so the trips generated from *randomTrips.py* represent a realistic traffic sample.



Fig 5. Simulation Map of the Experiment Environment Divided into Five Geographical Zones.

During the day different traffic intensities occur and for the purpose of this simulation the experiments are divided into the following three categories:

- Rush hour (RH) period - from 7:00 - 9:00 h and from 14:00 - 17:00 (5 hours in total),
- Normal hour (NH) period - from 9:00 - 14:00 h and from 17:00-22:00 (10 hours in total),
- Late hour (LH) period - from 22:00 – 7:00 h (9 hours in total).

Accordingly, based on the data presented in [24] three types of traffic loads were generated, where RH had a total of 900 vehicles, NH had a total of 300 vehicles and LH a total of 50 vehicles. During the simulations taxi vehicles with implemented EnvioDev were randomly selected. Locations of those vehicles are monitored and mapped during simulation and measured data collection was conducted at three different frequencies: each 15, 30 or 60 minutes. Those frequencies were selected based on the experience since the concentration of the air pollutants parameter during the normal weather conditions do not change quickly but still they enable frequently data gathering with up-to-date information and usage for the research and study purposes. Since generated trips and selected vehicles with EnvioDev are randomly generated, occasionally simulated vehicles are on the same location for a longer period time which is also a realistic scenario since taxi sometimes have breaks or less workload and consequently are parked.

Since EnvioDev reads data from several sensors, input freshness values are combined and the average freshness of data at the moment of reading  $t_{Davg}$  is calculated, as shown in (4) where  $t_c$  is the current timestamp and  $t_{si}$  is the timestamp of each sensor last measurement.

$$t_{Davg} = \frac{1}{N} \sum_{i=1}^N (t_c - t_{si}) \quad (4)$$

Vehicles equipped with EnvioDev constantly move, and consequently the coverage of the city area dynamically changes over time and it is possible that some parts of the city

have better coverage than the others. Therefore, it is necessary to define  $D_{FI}$  that depends on the average freshness of data at the moment of reading  $t_{Davg}$ , the time needed for data processing and synchronization  $t_{PS}$  and  $f_{DA}$ . Three data aggregation frequencies are used in our experiments  $f_{DA} = \{15, 30, 60\}$  minutes. The equation (5) defines the  $D_{FI}$  value where the lower results gives higher freshness.

$$D_{FI} = \frac{t_{Davg} + t_{PS}}{f_{DA}} \quad (5)$$

Simulation results were exported as floating car data and the content is shown in the Table III [25]. From the floating car data, it is possible to retrieve all necessary information about any specific vehicle regarding driving parameters (*speed, position (x, y), lane, slope*) and vehicle (*id, type, angle*) as well as simulation timestamp. Simulation results are presented and discussed below in three separate experiments. Heat maps of each measurement location for each vehicle are shown in Fig. 6, 7 and 8, where the number of five vehicles containing EnvioDev is used as a reference value. The number of total measurements conducted  $n_i^{Mt}$  minus number of redundant measurement  $n_i^{Mr}$  per number of vehicles containing EnvioDev  $n_i^V$  per each geographic zone  $N_Z$  during each defined traffic load over the testing period  $T_p$  is shown in the corresponding Tables IV, V and VI. Based on the obtained results the required number of vehicles equipped with EnvioDev  $N_{vehs}$  required to satisfy the experiments' requirements is calculated in (6).

$$N_{vehs} = \frac{f_{DA} \times T_p}{\frac{1}{N} \sum_{i=1}^N \frac{n_i^{Mt} - n_i^{Mr}}{n_i^V}} \times N_Z \quad (6)$$

### B. First Experiment

In the first experiment the data aggregation frequency  $f_{DA} = 60$  minutes is used. In Table IV the number of measurements in first experiment during three various traffic loads RH, NH and LH, per each geographical zone for three various numbers of vehicles  $\{1, 5, 10\}$  containing EnvioDev, is shown. The locations of all measurements per each vehicle are shown separately in Fig. 6 where each vehicle is represented with a dot of a different color.

TABLE III. CONTENT OF FLOATING CAR DATA

Parameter	Unit	Description
<i>timestep</i>	Seconds	The time step described by the values within this timestep-element
<i>Id</i>	Id	The id of the vehicle
<i>Type</i>	Id	The name of the vehicle type
<i>speed</i>	m/s	The speed of the vehicle
<i>angle</i>	Degree	The angle of the vehicle in navigational standard
<i>X</i>	m	The absolute X coordinate of the vehicle
<i>Y</i>	m	The absolute Y coordinate of the vehicle
<i>Pos</i>	M	The running position of the vehicle measured from the start of the current lane.
<i>Lane</i>	Id	The id of the current lane.
<i>slope</i>	Degree	The slope of the vehicle in degrees (equals the slope of the road at the current position)

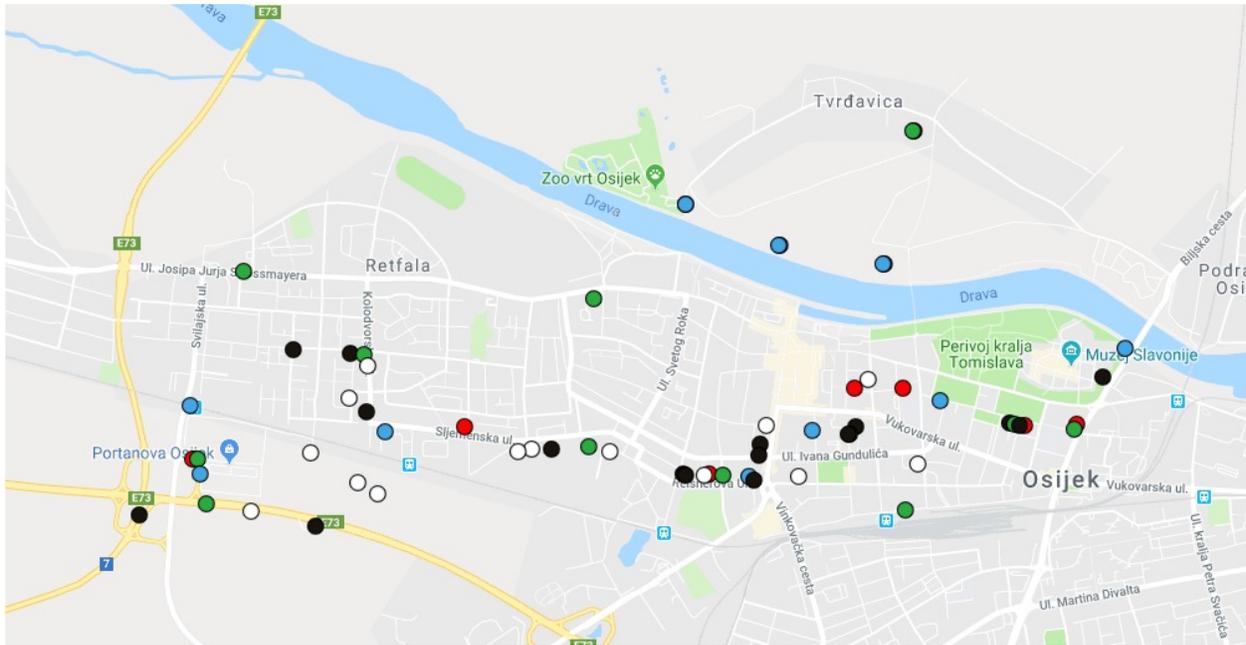


Fig 6. Locations of all Measurements taken during the First Experiment ( $f_{DA} = 60$  min;  $V_{NO} = 5$ ).

TABLE IV. NUMBER OF MEASUREMENTS IN FIRST EXPERIMENT

Vehicle No.	Traffic Load	Geographical Zone				
		Z1	Z2	Z3	Z4	Z5
1	RH	1	0	1	0	0
	NH	1	1	1	4	0
	LH	0	0	1	2	0
	$\Sigma$	2	1	3	6	0
5	RH	4	4	7	7	3
	NH	5	4	6	10	3
	LH	2	3	5	6	2
	$\Sigma$	11	11	18	23	8
10	RH	9	13	14	25	7
	NH	6	10	9	9	7
	LH	6	5	7	4	4
	$\Sigma$	21	28	30	38	18

Based on the measurement results and calculated  $D_{FI} = 0,15$  it is concluded that in order to achieve the data aggregation frequency of  $f_{DA} = 60$  minutes it is necessary to have nine vehicles on the tested area ( $N_{vehs} = 9.06$ ) to cover all five geographical zones efficiently. In Table IV it can be seen that some geographical zones are less represented in the measurement results than the others due to city structure since zones 3 and 4 are heavy urban areas where majority of people work and live mainly in buildings, therefore the traffic load is heavier than in other parts of the tested area.

### C. Second Experiment

Compared to the first experiment the data aggregation frequency is doubled and equals to  $f_{DA} = 30$  minutes. Three groups of vehicles {5, 10, 20} containing EnvioDev are used in this experiment. Table V shows the number of measurements

obtained during the second experiment for three traffic loads (RH, NH and LH) per each geographical zone. The locations of all measurements per each vehicle are shown separately in Fig. 7 where each vehicle is represented with a dot of a different color. Calculated  $D_{FI} = 0,21$  is slightly higher than in the first experiment but the data aggregation frequency is doubled, resulting in a double number of required vehicles equipped with EnvioDev ( $N_{vehs} = 18.34$ ) in the tested area to satisfy the requirements. However, this increase is not efficient since in to have almost double number of vehicles equipped with EnvioDev much higher resources are needed in terms of number of devices and vehicles as well as communication and maintenance expenses. Furthermore, from the Fig. 7, it can be seen a small coverage increase but the increased aggregation frequency results with redundant measurements and an uneven geographical zones coverage, as shown in Table V.

TABLE V. NUMBER OF MEASUREMENTS IN SECOND EXPERIMENT

Vehicle No.	Traffic Load	Geographical Zone				
		Z1	Z2	Z3	Z4	Z5
5	RH	8	8	9	13	2
	NH	10	8	6	19	11
	LH	2	7	3	10	7
	$\Sigma$	20	23	19	42	20
10	RH	17	21	24	43	14
	NH	9	12	13	27	14
	LH	5	5	6	12	7
	$\Sigma$	31	38	43	82	35
20	RH	39	44	58	85	34
	NH	22	21	18	44	33
	LH	9	7	18	21	15
	$\Sigma$	70	72	94	150	82



Fig 7. Locations of all Measurements taken during the Second Experiment ( $f_{DA} = 30$  min;  $V_{NO} = 5$ ).

#### D. Third Experiment

In the third experiment the highest data aggregation frequency is used, namely  $f_{DA} = 15$  minutes, and compared to the first and second experiment it is four and two times higher. The number of measurements obtained during the third experiment for three different traffic loads (RH, NH and LH) and three groups of vehicles {5, 15, 30} containing EnvioDev per each geographical zone is shown in Table VI. Fig. 8 shows locations of all measurements per each of the five vehicles containing EnvioDev, separately and each vehicle is marked as a dot with a different color.

Although  $D_{FI} = 0,41$  is much higher when compared to the first and second experiment the increase of the number of vehicles equipped with EnvioDev required to satisfy the experiment requirements is not dramatic and equals to 26.84. This is a result of the fact that data aggregation frequency is quadrupled when compared to the first experiment and doubled when compared to the second experiment and due to the fact that data freshness is dependent on the aggregation period. Therefore, if it is required to have a higher number of measurements in the tested area it is recommended to use the third experiment setup since the invested resources are not dramatically higher but will produce much higher number of measurement results. If it is needed to further increase  $D_{FI}$  it is necessary to increase the number of vehicles equipped with EnvioDev depended on the size and shape of the target area. The third experiment setup ensures almost real-time measurement results since the air quality during the normal weather conditions does not change quickly as well as the concentration of present air pollutants. The exception is in the

areas with heavy industry, highly populated urban areas or areas with very unstable weather conditions.

TABLE VI. NUMBER OF MEASUREMENTS IN THIRD EXPERIMENT

Vehicle No.	Traffic Load	Geographical Zone				
		Z1	Z2	Z3	Z4	Z5
5	RH	17	17	20	24	4
	NH	11	9	9	25	10
	LH	4	4	5	10	5
	$\Sigma$	<b>32</b>	<b>30</b>	<b>34</b>	<b>59</b>	<b>19</b>
15	RH	51	47	64	96	24
	NH	24	35	29	61	24
	LH	14	9	11	25	7
	$\Sigma$	<b>89</b>	<b>91</b>	<b>104</b>	<b>182</b>	<b>55</b>
30	RH	102	101	131	196	38
	NH	49	62	45	98	38
	LH	20	16	45	51	15
	$\Sigma$	<b>171</b>	<b>179</b>	<b>221</b>	<b>345</b>	<b>91</b>

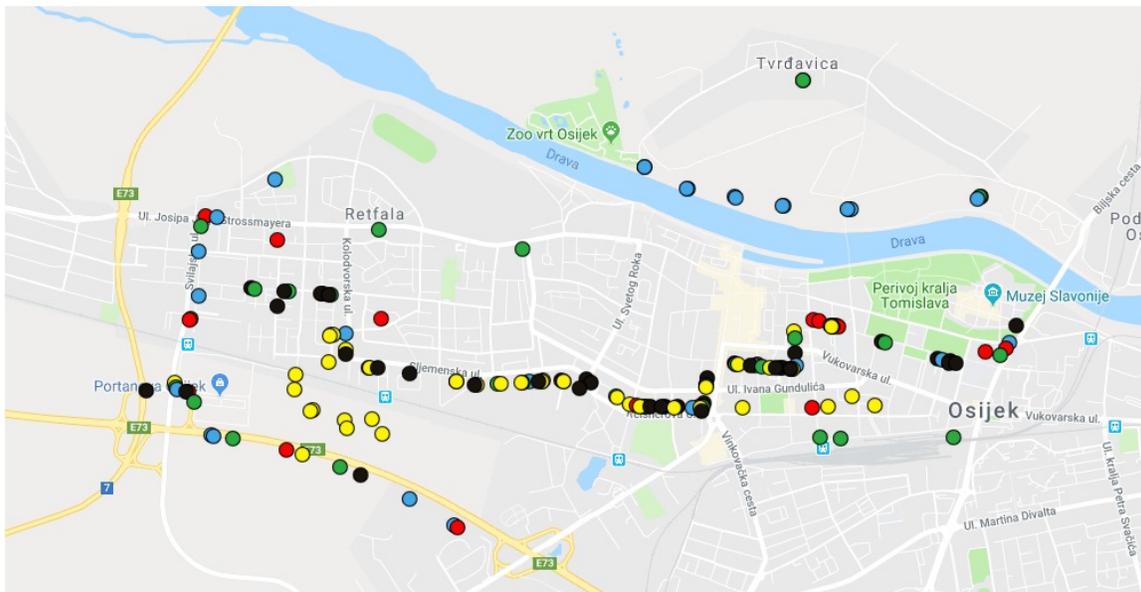


Fig 8. Locations of all Measurements taken during the Third Experiment ( $f_{DA} = 15$  min;  $V_{NO} = 5$ ).

## VI. CONCLUSION

Using mobile air quality measurement devices has many advantages when compared to the stationary ones in terms of cost-benefit trade-offs such as efficiency, overall deployment and maintenance costs, spatial coverage and implementation procedure. The presented EnvioDev measurement device can efficiently measure all air quality and traffic conditions parameters and provide useful real-time data. EnvioDev device was used in real-world experiments in the urban area of the city of Osijek. During the experiments high levels of *LPG* and *CO* concentration were recorded at several locations further stressing the need for continuous and detailed air quality monitoring. A fleet of vehicles equipped with the presented EnvioDev device can be used to conduct more comprehensive measurements for that purpose. Simulation results showed that for a lower data aggregation frequency of  $f_{DA} = 60$  minutes, nine vehicles equipped with EnvioDev devices are needed, whereas for a much higher data aggregation frequencies of  $f_{DA} = 30$  and 15 minutes, 18 and 27 vehicles equipped with EnvioDev devices are needed respectively to satisfy the requirements for efficient coverage of the city area and regular measurements in each part of the urban area.

## VII. FUTURE WORK

In the future work we plan to conduct more experiments in order to perform more detailed research on how the working environment conditions (winds, sun, rain, etc.) influence EnvioDev measurements. Furthermore, since we have additional sensors such as accelerometer, gyroscope, UV radiation and noise intensity sensor which measurements are not included in this work we plan to use them and based on measurements detect specific traffic conditions such as heavy traffic, traffic congestions, vehicle accidents [26] as well as area noise and UV radiation levels in a real-time [27]. Since EnvioDev is modular and expandable several additional sensors and camera can be added since from the picture is possible to extract very useful information for traffic purposes

[28] as well for recognizing environmental pollution. Development of traffic management strategies based on air pollution information is challenging. However, based on algorithms presented in [29] it is possible reroute urban traffic from critical parts of the street network (e.g. outside residential city areas). In addition, all generated data can be used with semantics, big data, and machine learning algorithms [30] in order to provide customized information, predict specific situations, improve security [31] and provide decision support.

## REFERENCES

- [1] SUMO, Simulation of Urban MObility, <http://sumo.sourceforge.net/> (accessed: 2020)
- [2] D. Hasenfratz, "Enabling Large-Scale Urban Air Quality Monitoring with Mobile Sensor Nodes", Swiss Federal Institute of Technology in Zurich, Zurich, Switzerland, Doctoral thesis, 2015.
- [3] A. Marjovi, A. Arfire and A. Martinoli, "High Resolution Air Pollution Maps in Urban Environments Using Mobile Sensor Networks," Proceedings of the International Conference on Distributed Computing in Sensor Systems, Fortaleza, Brazil, 10-12 June 2015, pp. 11-20.
- [4] Y. Gao, W. Dong, K. Guo, X. Liu, Y. Chen, X. Liu, J. Bu, C. Chen, "Mosaic: A low-cost mobile sensing system for urban air quality monitoring," Proceedings of the 35th Annual IEEE International Conference on Computer Communications, San Francisco, USA, 10-14 April 2016, pp. 1-9.
- [5] S. M. Biondi, V. Catania, S. Monteleone and C. Polito, "Bus as a sensor: A mobile sensor nodes network for the air quality monitoring", Proceedings of the IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Rome, Italy, 9-11 October 2017, pp. 272-277.
- [6] D. Hasenfratz, O. Saukh, C. Walsler, C. Hueglin, M. Fierz, T. Arn, J. Beutel, L. Thiele, "Deriving high-resolution urban air pollution maps using mobile sensor nodes", Pervasive and Mobile Computing, Vol. 16, 2014, pp. 268-285.
- [7] A.S. Mihaita, L. Dupont, O. Chery, M. Camargo, C. Cai, "Evaluating air quality by combining smart mobile pollution monitoring and data-driven modelling", Journal of Cleaner Production, Vol. 221, 2019, pp. 398-418.
- [8] S. Kaivonen, E. Ngai, "Real-time pollution monitoring with sensors on city bus", Digital Communications and Networks, Vol. 6, No. 1, 2019, pp. 23-30.
- [9] N. Castell, F.R. Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, D. Broday, A. Bartonova, "Can commercial low-cost sensor contribute

- to air quality monitoring and exposure estimates", *Environment International*, Vol. 99, 2016, pp. 293-302.
- [10] L. Morawska, et al., "Application of low-cost sensing technologies for air quality monitoring and exposure assessment", *Environment International*, Vol. 116, 2018, pp. 286-299.
- [11] S. A. Khandelwal and A. B. Abhale, "Monitoring vehicles and pollution on road using vehicular cloud environment," *Proceedings of the International Conference on Technologies for Sustainable Development (ICTSD)*, Mumbai, India, 4-6 February 2015, pp. 1-6.
- [12] R. Asorey-Cacheda, A. J. Garcia-Sanchez, C. Zúñiga-Cañón, P. Marco-Jornet, P. A. Moreno-Riquelme and J. Garcia-Haro, "Deployment of Air Quality Monitoring Sensors over a Delay Tolerant Mobile Ad-Hoc Network in Public Transportation Systems", *Proceedings of the 20th International Conference on Transparent Optical Networks*, Bucharest, Romania, 1-5 July 2018, pp. 1-4.
- [13] M. Bacco, F. Delmastro, E. Ferro and A. Gotta, "Environmental Monitoring for Smart Cities", *IEEE Sensors Journal*, Vol. 17, No. 23, 2017, pp. 7767-7774.
- [14] Air Quality in Republic of Croatia, Measurement Station Osijek-1, <http://iszz.azo.hr/iskzl/postaja.html?id=160> (accessed: 2020)
- [15] K. Paridel, J. Balen, Y. Berbers, G. Martinovic, "VVID: A delay tolerant data dissemination architecture for VANETs using V2V and V2I communication", *Proceedings of the MOBILITY, International Conference on Mobile Services, Resources, and Users*, Venice, Italy, 21-26 October, 2012, pp. 151-156.
- [16] A. Sharif, J.P. Li, M. Asim Saleem, T. Saba, R. Kumar, "Efficient Hybrid Clustering Scheme for Data Delivery Using Internet of Things Enabled Vehicular Ad Hoc Networks in Smart City Traffic Congestion," *Journal of Internet Technology*, Vol. 21, No. 1, 2020, pp. 149-157.
- [17] J. Balen, D. Vajak, K. Salah, "Comparative Performance Evaluation of Popular Virtual Private Servers," *Journal of Internet Technology*, Vol. 21, No. 2, 2020, pp. 343-356.
- [18] WHO Regional Office for Europe, Chapter 5.5 Carbon monoxide, Copenhagen, Denmark, 2000, [http://www.euro.who.int/data/assets/pdf\\_file/0020/123059/AQG2ndEd\\_5\\_carbonmonoxide.PDF](http://www.euro.who.int/data/assets/pdf_file/0020/123059/AQG2ndEd_5_carbonmonoxide.PDF) (accessed: 2020)
- [19] M. Abdul Hannan, A.S. Mohd Zain, F. Salehuddin, H. Hazura, S.K. Idris, A.R. Hanim, M. Yusoff, "Development of LPG Leakage Detector System using Arduino with Internet of Things (IoT)", *Journal of Telecommunication, Electronic and Computer Engineering*, Vol. 10, No. 2-7, 2018, pp. 91-95.
- [20] Methane (CH<sub>4</sub>) Safety, AGRI-FACTS, [http://www1.agric.gov.ab.ca/\\$department/deptdocs.nsf/all/agdex9038/\\$file/729-2.pdf?OpenElement](http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/agdex9038/$file/729-2.pdf?OpenElement) (accessed: 2020)
- [21] Global CH<sub>4</sub> levels, Atmospheric CH<sub>4</sub> Graph, <https://www.methanelevels.org> (accessed: 2020)
- [22] Air Quality Index (AQI) & Pollution Report, <https://air-quality.com/> (accessed: 2020)
- [23] SUMO Tools, <https://sumo.dlr.de/wiki/Tools/Trip> (accessed: 2020)
- [24] Master Plan for transport development of the City of Osijek and Osijek-Baranja County, [http://www.obz.hr/hr/pdf/strategija/2018/03\\_prilog\\_i\\_katalog\\_ulaznih\\_podataka.pdf](http://www.obz.hr/hr/pdf/strategija/2018/03_prilog_i_katalog_ulaznih_podataka.pdf) (accessed: 2020)
- [25] SUMO Simulation, FCD (floating car data) Output <https://sumo.dlr.de/wiki/Simulation/Output/FCDOuput> (accessed: 2020)
- [26] T. Addeb, W. Di, M. Ibrar, "Software-Defined Networking (SDN) based VANET Architecture: Mitigation of Traffic Congestion", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 11, No. 3, 2020, pp. 706-714.
- [27] S. Grubeša, A. Petošić, M. Suhanek, I. Đurek, "Mobile crowdsensing accuracy for noise mapping in smart cities", *Automatika*, Vol. 59, No.3-4, 2018 pp. 286-293.
- [28] J. Videković, J. Balen, "On-board smartphone application platform for real-time sharing of traffic information", *Proceedings of 37th Conference on Transportation Systems with International Participation - Automation in Transportation*, RIjeka, Croatia, 15-18. November 2017, pp. 27-31.
- [29] S. Mandzuka S., "Cooperative Systems in Traffic Technology and Transport", *New Technologies, Development and Application*, Vol. 42, 2019, pp. 299-308.
- [30] C.T. Yang, L.Y. Lin, Y.T. Tsan, P.Y. Liu, W.C. Chan, "The Implementation of a Real-time Monitoring and Prediction System of PM2.5 and Influenza-Like Illness Using Deep Learning", *Journal of Internet Technology*, Vol. 20, No. 7, 2019, pp. 2237-2245.
- [31] M. Dörterler, F. Bay Ömer, "Neural Network Based Vehicular Location Prediction Model for Cooperative Active Safety Systems", *PROMET*, Vol. 30, No. 2, 2018, pp. 205-215.

# Urbanization Change Analysis based on SVM and RF Machine Learning Algorithms

Farhad Hassan<sup>1</sup>, Tauqeer Safdar<sup>2</sup>, Aman Ullah Khan<sup>4</sup>  
Department of Computer Science & Engineering  
Air University Multan Campus, Multan, Pakistan

Syed Muhammad Husnain Kazmi<sup>5</sup>  
IAR Systems AB  
Strandbodgatan 1, 75323 – Uppsala, Sweden

Ghulam Irtaza<sup>3</sup>  
Department of Computer Science, University of Education  
Multan Campus, Multan, Pakistan

Farah Murtaza<sup>6</sup>  
School Education Department  
Multan, Pakistan

**Abstract**—To maintain sustainability in the development, measured the yearly change rate of the land through Land Cover classified maps that hold the data which is surveyed as an influential factor for environment management and urbanization. This paper measured the change rate, which is helpful for the management of the city to define the new policy and implement the best one to maintain the natural resources. Machine Learning algorithms are utilized to produce the most acknowledged Land Cover maps using the GEE cloud-based reliable platform using the LANDSAT8 satellite imagery. For the classification used the Random Forest (RF) and Support Vector Machine (SVM) Algorithm. This investigation also found that the Support Vector Machine (SVM) classifier accomplished better over-all accuracy and Kappa coefficient as compared to the Random Forest (RF) classifier while the training sample for both is the same.

**Keywords**—Random Forest (RF); Support Vector Machine (SVM); GEE; classification; machine learning classifier; multi-temporal change analysis; urban change analysis; LANDSAT8; Kappa co-efficient

## I. INTRODUCTION

Urbanization is a global phenomenon with various nations encountering different frequencies and precedents of rural-urban relocation. The whole world is changing into an urbanized center with a more significant part of the individuals moving to huge urban communities during the previous decades. Almost everyone is either way engaged with this procedure. Before mid-century, urban advancement was, for the most part, limited to developed nations yet has spread to developing nations now. Presently, practically all the developing nations on the planet are encountering urbanization. Advancing investigation suggests that the association between urbanization and advancement is not customized. Urbanization is ordinarily seen as solidly related to fiscal improvement, particularly in advanced countries. It is evaluated that urban territories make over 80% of the overall GDP. Enormous scale urbanization is changing the traditional structure prompting a significant upheaval in our general public. The development of urbanization in Pakistan mirrors the desires and ambitions of billions of new urbanites. It is a repetitive procedure a nation encounters as it advances from an undomesticated to modern culture. People migrate from

monetarily restrained territories to a spot where fitter possibilities are advertised [1].

Pakistan has an average level of urbanization inside the Asia-Pacific area, dependent on both the development of urbanization and urban evolution. Pakistan has an enormous shelter shortage of around ten million units and increasing. Urban masses' improvement in the country has not been facilitated by advancement in shelter units or unbiased access to land, causing shelter to lack and the preferment of ghettos. Given Pakistan's shortcoming to earthquakes and other natural hazards, the lawmaking organization should set up a practical framework for developing guidelines and their consistency to avoid the implied adverse effects of vertical shelter schemes. The modern approach to manage urban shelter has faced multiple challenges. Although, among the South Asian nations, Pakistan has the most raised portion of the population living in the urban zone. The urbanized percentage of Pakistan's population grew 4.1 percent in the five years between 2005 and 2010 and is expected to grow more in the near future years [2]. As per the UN, a fraction of the populace will be residing in urban areas by the year 2030. Altogether, urban networks in Pakistan contribute fifty-five percent of the GDP. Furthermore, Pakistan generates 95% of its government salary from 10 substantial urban zones [2].

A Multan district is in the southern area of the Punjab province, Pakistan. Multan is also famous with the name City of Saints and is found on the banks of the River of Chenab at 30.1575° N latitude and 71.5249° E longitude. Total covering area around 3,721. Sq.km and it is the seventh-largest city in Pakistan. Urbanization is the transformation of any land into urban land and globally, it is observing that the trend of urbanization is increasing rapidly. To measure this change needs the upper layer view of the Earth's surface like satellite images. Numerous sensors are catching tremendous volumes of high-resolution satellite images containing diverse information that is updated diurnally. Remote sensing is the process of receiving the information associated with the Earth by performing the scanning process with the assistance of satellites. Remote sensing is performed to achieve the precise and clear image of the Earth, and these high-resolution remote sensing images are free of any cost, and for the researcher and

analyst, it is an incredible possibility to achieve better analysis results in a specific region of interest (Lopez, R. D., & Frohn, R. C. (2017)). If these images utilized in an appropriate manner, then it gives a tremendous profit and investigates the surface of the Earth [1]. As per a near perspective, remote sensing imagery contrasts from the standard imagery, the significant explanation behind these matters is additional spectral information that the satellite imagery contains and which is not noticeable for the unaided eye [3]. On the high-resolution satellite images, Land Use Land Cover (LULC) change examination is one of the major because from this investigation report obtains the yearly trend of urbanization change that helps the policymaker of any establishment makes the correct choice and makers additionally understand the impacts of these progressions on people and their environments. Land Use and Land Cover (LULC) investigation has an exceptional influence on the future of the land, and this examination depends on the classification of the particular feature's class [4]. Land Cover Change Analysis screens the change either in loss or gain of any land type over the time range and measures the status of progress by the analysis for the scope of course of events. Yu et al. (2016) Land use have frequently affected land cover. The surface of the land has been changed in the part of appearance when seen on two distinctive time-series [5]. Furthermore, Huang et al. (2017) mentioned that Google Earth Engine (GEE) is the reliable and powerful platform created by Google, Carnegie Mellon University and the US Geological Survey, uninhibitedly accessible; a propelled distributed computing conditions for remote sensing data processing and geospatial analysis and no need of any fast CPUs or GPUs for high computational work aside from fast internet connection [6].

The paper is organized into sections like problem statement which describes the land loss due to urbanization. In the next section, a literature survey is done where the machine learning approaches are mentioned; later in that section, satellite image collection and analysis processes are mentioned by highlighting the work of different researchers. In the later section of methodology, the data collection methods and SVM and RF algorithms are used to obtain the results of the study which describes the transformation of land in 2017 and 2018 and its impacts. In the last section the paper is concluded by mentioned the land transformation changes.

## II. RESEARCH PROBLEM

The continuous urbanization may result in several concerns for the supervision of the city. In order to increase the sustainable progress of the city, there is a demand for measuring Land Cover Change analysis. From this investigation, obtain the influence of water-bodies and vegetation land on the advancement of urbanization with the help of the SVM and RF machine learning algorithms during the multi-temporal range of 2017 to 2018. The research is conducted for the lack of land cover change analysis in order to measure the change of urbanization. The motivation behind the research is to measure the changes in water bodies and vegetation using the classifier (SVM and RF) approach. Through the latest cloud-based powerful and the most reliable environment named, GEE and designing the new policies for the accommodation of the development.

## III. LITERATURE REVIEW

Agriculture is a practice of producing and harvesting of the crops in a methodical manner. Because of the ever-increasing demand for the food, there is a need to improve the yield of the crop and avoid the loss of crop in every possible manner. In order to do so, the scientific community is using the optimal resources for the cultivation of crops, and remote sensing is used for this purpose because of its enormous advantages and features. One of the most vital applications of remote sensing is the classification of the crops by differentiating between the varieties of the crop.

Varma et al. (2017) Satellite imaging are used for the viable investigation of the temporal changes that can affect the yield of the crop in the specific areas. By using satellite imaging methods, the growth of crops from sowing to the harvesting level can be monitored in a very efficient way. Geo-references and ortho-rectified satellite images are used to identify the land loss problems and the affected areas in a different region of the world, not just that the seasonal changes and crop variation can also be monitored by using these techniques. Moreover, on the information based on the activities like deciding the type of crop and its acreage, growth determination of different stages of the crop, and delineating the extent of the crop can be planned in advance. These methods are then used for the decision-making policies to increase the yield of the crop and avoid land loss.

Senthilnath et al. (2016) Another satellite image collection method called multi-spectral satellite imaging can aid the identification and classification of the crops because they take chances in the reflectance as a function for the specific crop type. The classification of the crop is used in the auditing of the land use, and it also facilitates the soil and water study of the particular area. Because of the variability in the cultivation of the crop in a geographical area the process of the classification is a very challenging factor. Moreover, the classification of the crop is done on the basis of spatial and spectral bands or also by combining both methods for the purpose of classification. The clustering method is used for the grouping of the explicit datasets in such fashion that the data point towards the same group is virtually analogous. The purpose of clustering is that it reduces the intra-cluster distances, and it maximizes the inter-cluster distance. The data extracted from the points are in the form of optimal cluster centres, and the clustering is performed on the satellite data in several ways, for example, the clustering methods are focused on the partitioning using the spatial patterns. Partition clustering is completed by dividing the data with the fixed number of clusters (priors) by using the similarity indexes. K-means is known as one of the most commonly used clustering methods.

Machine learning is one of the most unswerving classification methods. In machine learning, the classifier works to determine the category to which a new research sample belongs based on the instruction data of the marked kind. The machine learning technique makes its technique more suitable in classification, especially in the field of remote sensing images because it is impossible to have a detailed knowledge area of study. Satellite images broadly utilized for

land cover change investigation and more often than not the manager and policy designer need a reliable technique to get to the result of a land cover change by reviewing and recognizing changes of land types Karpatne et al. (2016).

The research yield portrays the diverse land includes that is available increases and lost on the surface of the Earth. GEE gives the most precise and dependable research data with a financially savvy parameter and takes lesser time contrasted with different gadgets that can be utilized for land cover change analysis. Mueller et al. (2017) reported that Landsat Satellite imagery is profitable for the various plans, for example, to measure the change of climate and disasters related to the environment and management of natural resources. Land cover change analysis is the under of natural resources either created by any machine learning classifier, threshold method or the hybrid approach. Landsat data are most commonly used in natural phenomena-oriented applications.

Borra et al. (2019) reported that the procedure of satellite image classification encompasses the gatherings of image's pixel esteem into the section of feature assortment, and for the satellite image classification, various strategies are accessible [7] to create the map. These maps can acquire various highlights or qualities to do the ideal work. A few different strategies incorporate the supervised and unsupervised plan to acquire the estimation of the precision of remote sensing information which is an essential requisite in classification.

Support Vector Machine Classifier is among one of the many mainstreams and the unanimously acknowledged classifiers in the domain of remote sensing, the essential purpose for this is profoundly precise classification results. It is the binary classifier and dependent on the concept that the training samples which are nearer concurrence to the boundaries of a class will differentiate the class superior to other samples of training [8]. So, it is evident that SVM classifier focuses on finding the optimal hyper-plane that separates the samples of training input into many classes and samples of training data is close to the boundaries of the class and at the lesser distance to hyper-plan are taken as support vectors, which is to be used for actual training [12]. Maxwell, A. E. (2018) the selection of kernels plays a significant role in the classification's results. RBF is also the type of kernel in which has a user-defined parameter that controls the impact of a sample of training on the boundary of decision and the value of the user-defined parameter is higher, then there is the chance of over-fitting. So that is why it is necessary to take the right balance for value [9]. Tehrany et al. (2015) conducted a study where the researchers implemented the Support Vector Machine (SVM) algorithms in different fields of study like flood susceptibility assessment and landslide susceptibility investigation. Genetic algorithms are acknowledged as the most advanced and pervasive developed heuristic search models in the field of artificial intelligence, and it has its application in the urban planning, ecological studies, climate modelling, and remote sensing. Kruber et al. (2019) also evaluated the performance of machine learning algorithms called RF and SVM and compared the results with each other and mentioned that RF and SVM are called a mapping

algorithm for the groundwater outbreaks, and these algorithms can provide very effective results.

#### IV. AREA OF STUDY

A Multan district is in the southern area of the Punjab province, Pakistan. Multan is also eminent with the name City of Saints and is found on the banks of the River of Chenab at 30.1575° N latitude and 71.5249° E longitude. Total covering the area around 3,721. Sq.km and it is the seventh-largest city in Pakistan. The variety of feature classes present in Multan makes it a good candidate for the study, and the selected area of study for this research is 756.52368 sq. kilometers. The coordinate's values of area of study are [71.34246, 30.29730], [71.34246, 30.08839], [71.68166, 0.088396] and [71.68166, 30.29730]. Fig. 1 indicates the location of the Multan District under the boundaries of Pakistan and also shows the selected area study of area for the research.

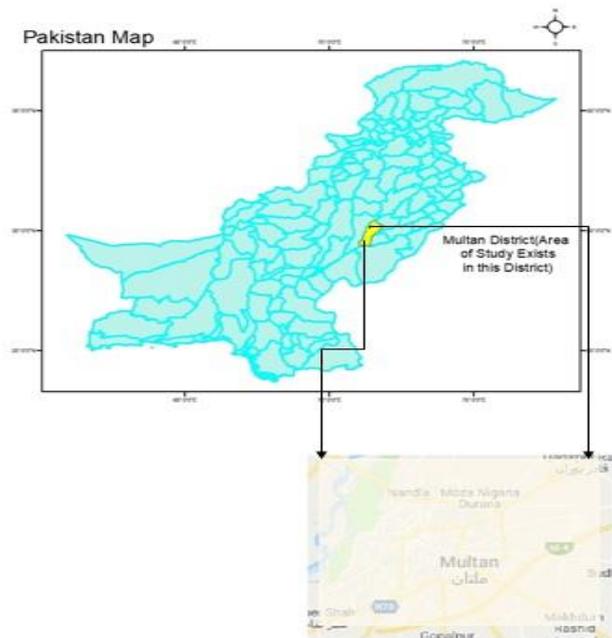


Fig 1. Area of Study.

#### V. DATA PREPROCESSING

The satellite imagery data use in the research is Landsat 8, Collection 1, Tier 1 and ID of this dataset is LANDSAT/LC08/C01/T1. The quality of this dataset is high, and it is considered a more suitable multi-temporal analysis. The total number of bands of this dataset is 12 and for this research just selected just 6 bands B2, B3, B4, B5, B6, and B7. The area bound was set according to the above-mentioned coordinate's value under the district of Multan. For the year 2017, the collection of images is 43, and for 2018 the numbers of an image are 45, and these 88 images are filtered as per area of study. After this select only two images from them one for 2017 and one for 2018. For this reduced selection, convert them into TOA reflectance, simple cloud score to be calculated and in the last apply the median on the least cloudy as shown in Fig. 2.

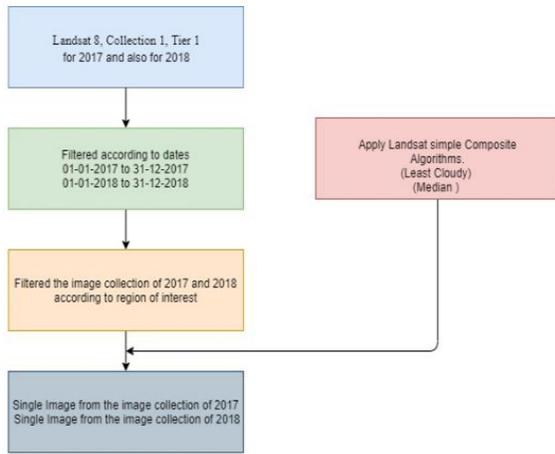


Fig 2. Data Preprocessing.

### VI. METHODOLOGY

To attain the urbanization of yearly drift the adopted methodology is shown in Fig. 3. From the LANDSAT8 take the image assortment and in 2017 assortment of satellite images are 43 however in 2018 the assortment which comprises of 45 images as indicated by the filtration of the study zone limit, then apply the composite algorithm which is exceptionally valuable to make an assortment of images into a solitary image. The composite calculation forms the median from the assortment of Landsat images and the exclusive image from the assortment ought to be the sans cloud image. The composite algorithm pertains to both image assortments of 2017 and 2018.

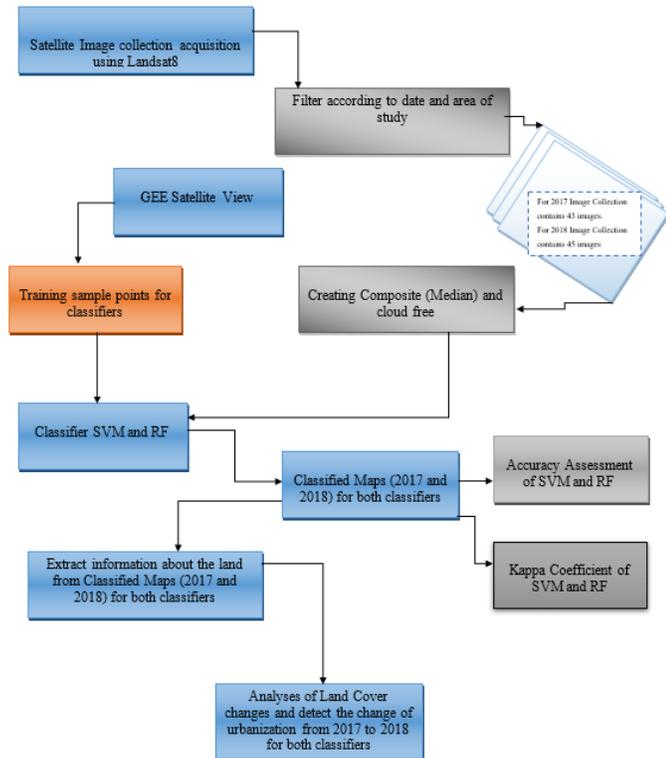


Fig 3. Research Methodology.

For the determination of the representations of training points, utilizing the geometry alternative such as polygon and point of the GEE frame for the classification and uses two unique classifiers SVM and RF to compose the classified maps as shown in the Fig. 3. Estimate the precision and kappa coefficient of both SVM and RF classifiers and additionally anticipates the varieties of land in square kilometer for 2017 and 2018. In the final prognosis, the variation in urbanization land from different types of land over time, operating the two classifiers.

### VII. RESULTS AND ANALYSIS

The characterized images for years 2017 and 2018 are created by applying machine learning classifiers (SVM and RF) for the apprehension of the variation in urbanization. We use SVM and random forest classifier on a comparable training data index. We analyze that the consequences produced by the SVM classifier are more exact than the random forest classifier. SVM classifier produces the effects by retaining the pixel's fallacious quality with low miss arrangement of urbanization.

The projected results for the year 2017 and 2018 utilize the SVM classifier for the classification with the subtleties of three sorts of land water, vegetation and urban under the selected area of study. Fig. 4 and 5 are representing the classified results using SVM classifier.

With the utilization of the RF classifier, created the detailed map for the year 2017 and 2018 are appeared in Fig. 6 and 7, these show the outcomes for 2017 and 2018. Both resultant images include three distinct kinds of land water bodies, vegetation and urban. Fig. 6 and 7 are representing the classified results using RF classifier.

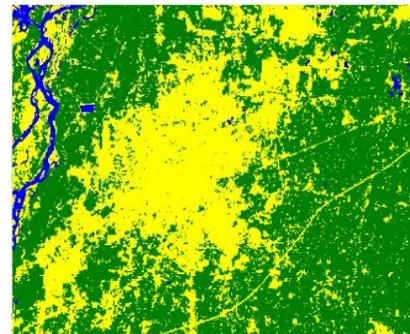


Fig 4. Classified Results for 2017 using SVM.

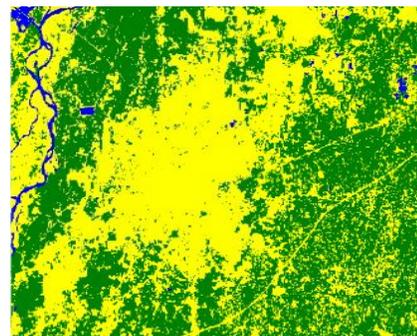


Fig 5. Classified Results for 2018 using SVM.

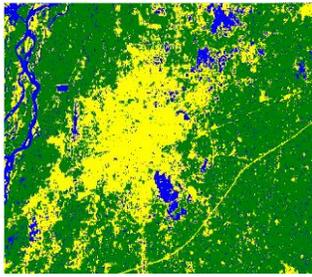


Fig 6. Classified Results for 2017 using RF.

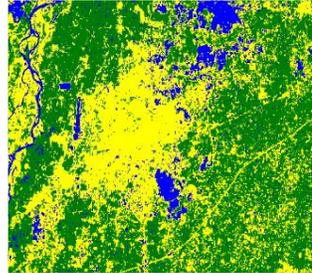


Fig 7. Classified Results for 2018 using RF.

#### a) Type of Land for 2017

Extracted each type of land from the classified results under the study area using the machine learning (SVM and RF) classifiers. The water area for the year 2017, which is obtained from the SVM and RF classifier is shown in Fig. 8 and 9, respectively. Blue color indicates the existence of water under the area of study while the black color indicates any other land type area.

The vegetation or greenish area as presented in the Fig. 10 and 11 identified using the Green color, indicates the existence of vegetation while black color indicates the non-vegetation area under the area of study.

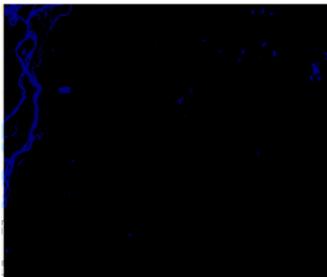


Fig 8. Water 2017 based on SVM.

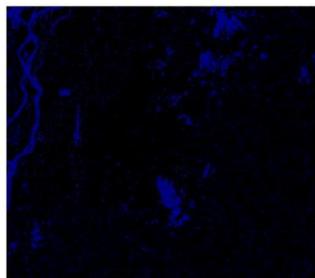


Fig 9. Water 2017 based on RF.

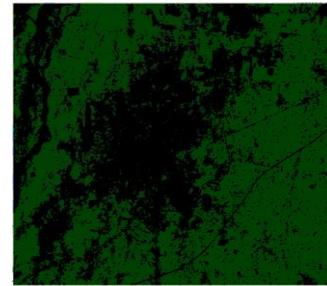


Fig 10. Vegetation Area in 2017 based on SVM.

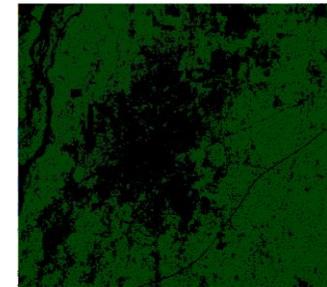


Fig 11. Vegetation Area in 2017 based on RF.

The urban area for the year 2017 which is obtained from the SVM classifier results is shown in Fig. 12 and urban area for the year 2017 which is obtained from the Random forest classifier is shown in Fig. 13. In both, Yellow color indicating the area of urban land presence while the black color indicates any other land type area.

#### b) Type of Land for 2018

For 2018 extract each type of land from the classified results under the study area using the machine learning (SVM and RF) classifiers. The water area for the year 2018 which is obtained from the SVM and RF classifier is shown in Fig. 14 and 15, respectively. Blue color indicates the existence of water bodies under the area of study, while the black color indicates any other land type area.

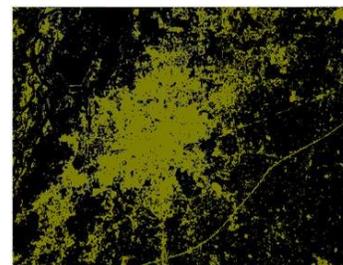


Fig 12. Urban Area in 2017 based on SVM.

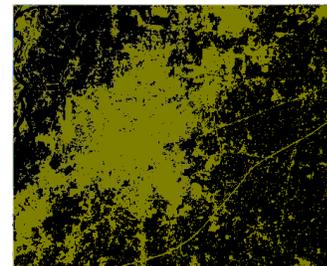


Fig 13. Urban Area in 2017 based on RF.

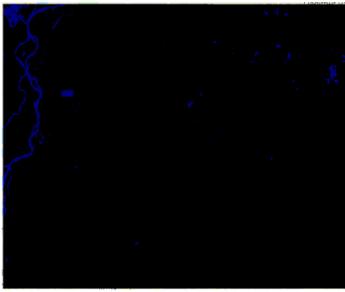


Fig 14. Water 2018 based on SVM.

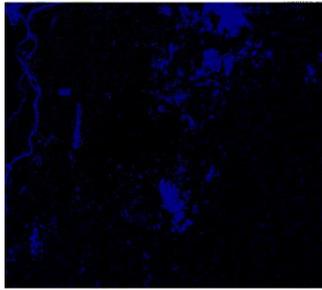


Fig 15. Water 2018 based on RF.

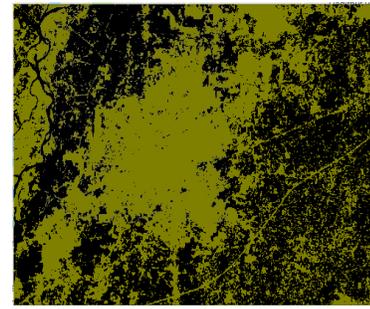


Fig 18. Urban Area in 2018 based on SVM.

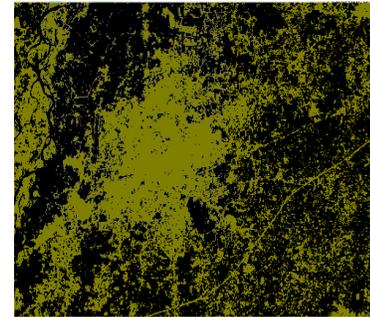


Fig 19. Urban Area in 2018 based on RF.

The urban area for the year 2018 which is obtained from the SVM classifier results are shown in Fig. 16 and urban area for the year 2018 which is obtained from the Random forest classifier is shown in Fig. 17. In both, Yellow color indicating the area of urban land presence while the black color indicates any other land type area.

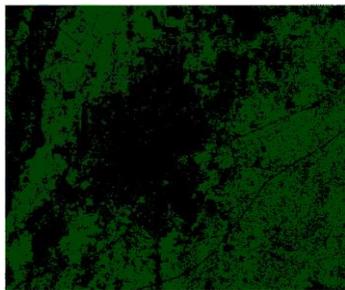


Fig 16. Vegetation Area in 2018 based on SVM.

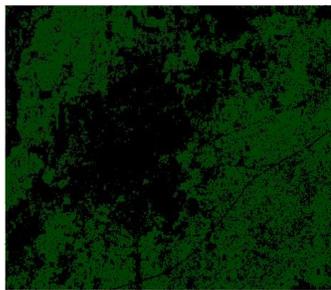


Fig 17. Vegetation Area in 2018 based on RF.

The urban area for the year 2018 which is obtained from the SVM classifier results are shown in Fig. 18 and urban area for the year 2018 which is obtained from the Random forest classifier is shown in Fig. 19.

1) *Measure Water Bodies Losses*: To calculate the water bodies' losses between the multi-temporal ranges from 2017 to 2018 firstly obtain the values of water bodies of both years separately by using both the classifier's results. In 2017, the area of water bodies according to the classified result of SVM is shown in Fig. 8 under the area of study while for the year 2018 is shown in Fig. 12. From both, extract the area of losses in waterbodies and it is shown in Fig. 20. While the loss area of water bodies based on RF classifier's results over the time series of 2017 to 2018 is shown in Fig. 21.

The area covered by water bodies that are obtained using the SVM classifier for the year 2017 is 17.733197689859043 sq. kilometer and in 2018 it is 10.974930358946 sq. kilometer under the area of study which is 756.5236827301217 sq. kilometer. The calculated difference between both extracted land type shows that the water bodies' area is decreased by 6.75826733091299 sq kilometers from the classified results of SVM and RF as shown in Fig. 20 and Fig. 21. Using RF classifier, we inspected that in 2017 water-bodies area is 71.40641692305495 sq. kilometers, in 2018 it is decreased by 8.188619262852404 sq. Kilometers and the remaining water area is 63.2177976602023 sq. Kilometers.

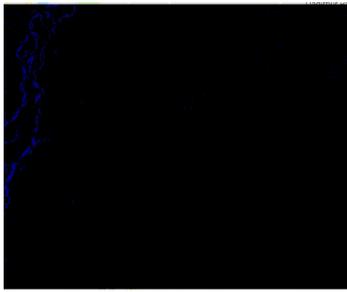


Fig 20. Waterbodies Losses Area 2017 to 2018 based on SVM.

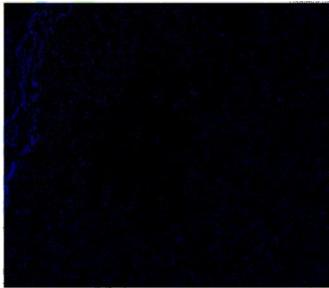


Fig 21. Water Bodies Losses Area 2017 to 2018 based on RF.

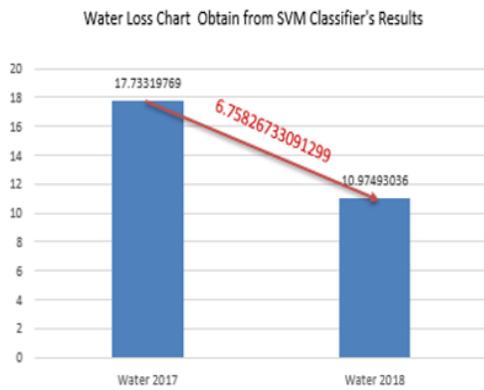


Fig 22. Water Loss Chart based on SVM.

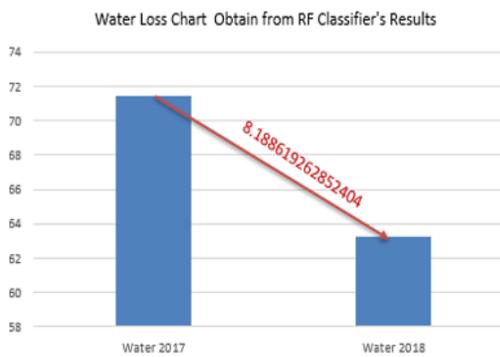


Fig 23. Water Loss Chart based on RF.

2) *Measure Vegetation Losses*: Extract the vegetation area from the classified resultant output of 2017 and 2018 using the SVM classifier and Random Forest, and then measure the difference from both years results to obtain the loss area of vegetation from 2017 to 2018 as shown the Fig. 22 and 23 based on SVM and RF.

The area covered by vegetation is obtained using the SVM classifier for the year 2017 is 435.1198872814169sq. Kilometers and for the year 2018, it remains 355.15992729514426 sq. kilometers under the area of study which are 756.5236827301217 sq. kilometers. The vegetation area decreased by 79.95995998627411sq. Kilometer, as shown in Fig. 24 and 25. Using RF classifier, in the 2017 vegetation area is 451.10853885372273 sq. kilometers but in 2018 it remains 362.966601021753 sq. kilometers. The loss of vegetation area from 2017 to 2018 is 88.14193783196974 sq. kilometer as shown in Fig. 26 and 27.

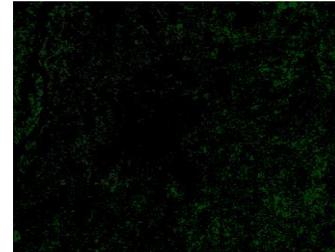


Fig 24. Vegetation Loss 2017 to 2018 based on SVM.

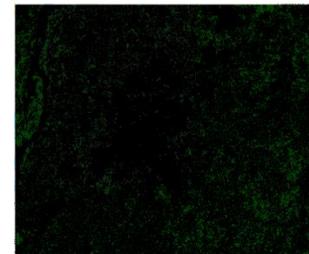


Fig 25. Vegetation Loss 2017 to 2018 based on RF.

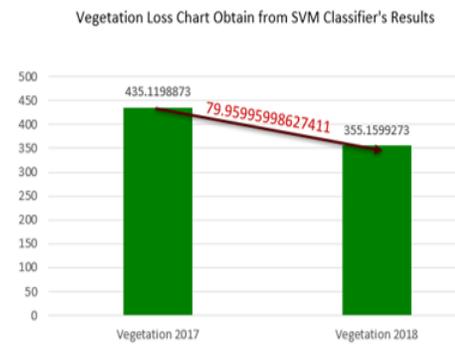


Fig 26. Vegetation Loss Chart based on SVM.

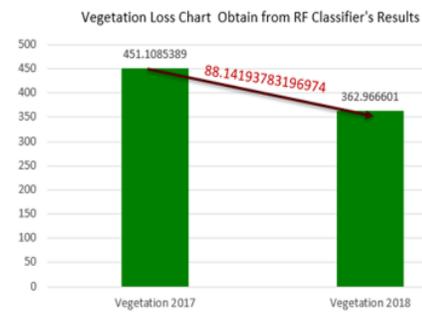


Fig 27. Vegetation Loss Chart based on RF.

3) *Urban Change*: From the classified results of SVM and RF measure the change in urban land over the multi-temporal range of 2017 to 2018. Fig. 28 shows the change of urban land-based on SVM, and Fig. 98 shows the changing area of land but for RF classifier.

From the classified resultants of SVM for the year 2017, the urban land area was 303.67059775885275 sq. Kilometers and in 2018 the area of urban land reaches 390.3888250760368 sq. kilometers. The urban land area increased by 86.71822731718704 sq. Kilometer using SVM, as shown in Fig. 28. Classified resultants of RF for the year 2017, the urban land area was 234.0087269533519 sq. Kilometers and in 2018 the area of urban land reaches 330.339284048179 sq. kilometers as shown in Fig. 29. The urban land area increased by 96.33055709482205 sq. Kilometers using RF and using SVM as shown in Fig. 30 and 31.

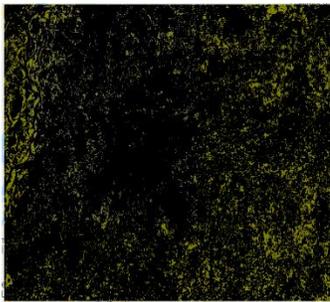


Fig 28. Urban Growth SVM.

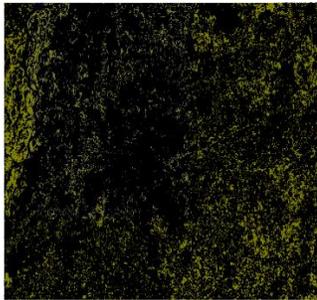


Fig 29. Urban Growth RF.

Urban Land Gain Chart Obtain from SVM Classifier's Results

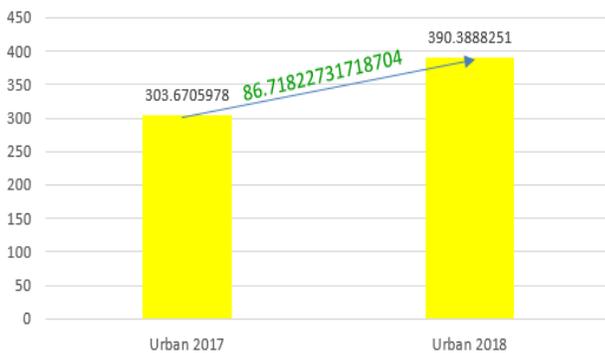


Fig 30. Urban Growth Chart based on SVM.

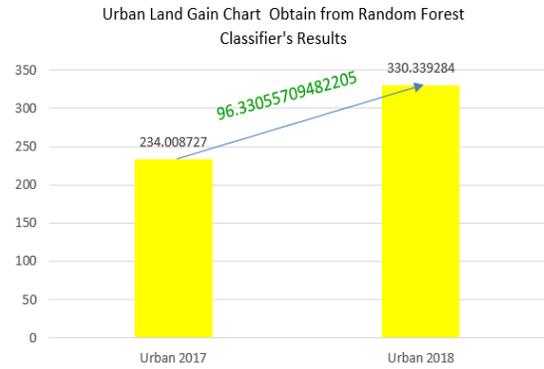


Fig 31. Urban Growth Chart based on RF.

a) *Vegetation Land into an Urban land*

From the classified results of both algorithms extract the vegetation land and obtained the transformation area of vegetation land into urban land. As shown in Fig. 32 for the SVM algorithm and Fig. 33 for the RF algorithm.

b) *Water Bodies into an Urban Land*

From the classified results of both algorithms extract the water-bodies and obtained the transformation area of water-bodies into urban land. As shown in Fig. 34 for the SVM algorithm and Fig. 35 for the RF algorithm.

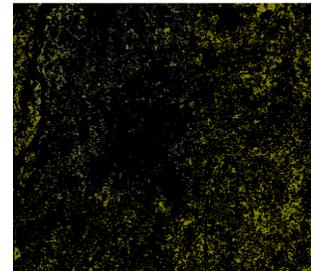


Fig 32. Urban in 2018 from Vegetation in 2017 based on SVM.

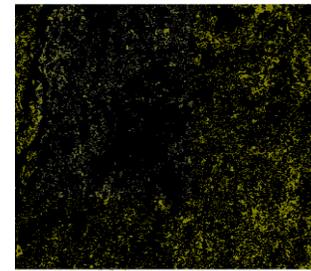


Fig 33. Urban in 2018 from Vegetation in 2017 based on RF.

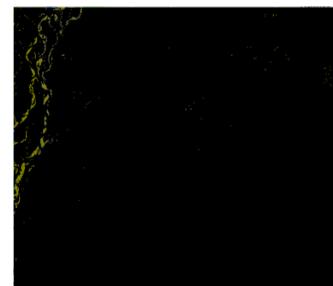


Fig 34. Urban in 2018 from Water in 2017 based on SVM.

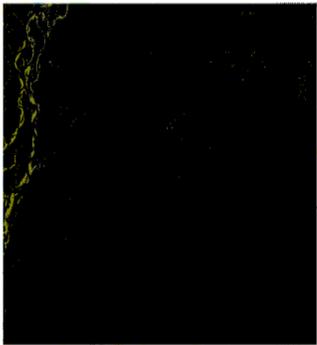


Fig 35. Urban in 2018 from Water in 2017 based on RF.

c) Total Change in Urban Land

The Total area of change from the vegetation loss and water bodies of 2017 in 2018 is extracted using the SVM, as shown in Fig. 36. While Fig. 37 indicates the total change based on RF.

From the classified resultants of SVM for the year 2017, the urban land territory was 303.67059775885275 sq. Kilometers and in 2018, the zone of urban land reaches 390.3888250760368 sq. Kilometers. The urban land territory extended by 86.71822731718704 sq. Kilometer utilizing SVM as shown in Fig. 38. Classified resultants of RF for the year 2017, the urban land zone was 234.0087269533519 sq. Kilometers and in 2018 the territory of urban land reaches 330.339284048179 sq. Kilometers. The urban land territory extended by 96.33055709482205 sq. Kilometers utilizing RF as shown in Fig. 39.

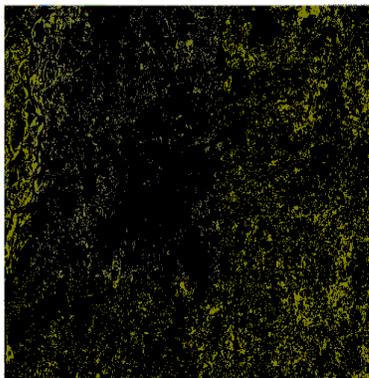


Fig 36. Total Urban Change based on SVM.

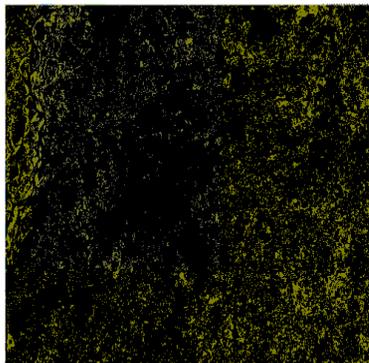


Fig 37. Total Urban Change based on RF.

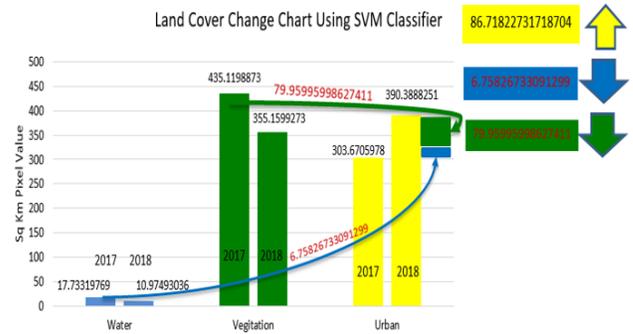


Fig 38. Land Transformation based on SVM Classifier.

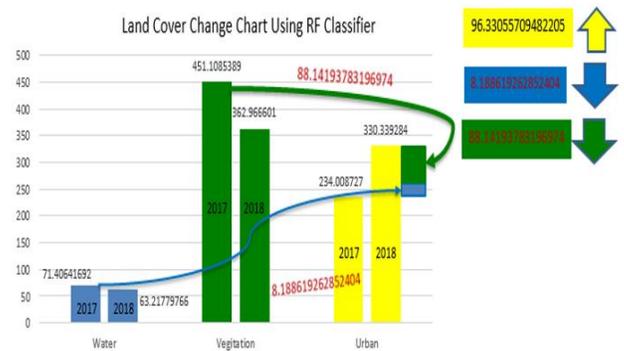


Fig 39. Land Transformation based on RF Classifier.

VIII. ACCURACY AND KAPPA COEFFICIENT ANALYSIS

1) Accuracy Assessment: For the assessment of accuracy uses the method of the Confusion matrix (error matrix). This method is the best approach for expressing accuracy of classification in supervised learning, and the error matrix (confusion matrix) provides the information of accuracy for every class in the classification [11].

$$\text{Classifier's accuracy} = (\text{TCP}/\text{TNP}) * 100$$

By using the above-mentioned equation, obtain the accuracy of the classifier and value of TCP and TNP can be calculated by a confusion matrix.

a) SVM Classifier Accuracy

To measure the accuracy of the model, use the values of the Confusion matrix (error matrix) as presented in Table I.

TCP is the sum of the diagonal values of the confusion matrix of the SVM Classifier.

$$\text{So, TCP} = 8+129+134=271$$

TNP is the sum of all values of the accuracy matrix.

$$\text{So, TNP} = 8+4+17+0+129+6+0+0+134=298$$

Put the calculated values of TCP and TNP into the above-mentioned equation.

TABLE I. CONFUSION MATRIX FOR SVM CLASSIFIER

	Water	Vegetation	Urban
Water	8	4	17
Vegetation	0	129	6
Urban	0	0	134

SVM Classifier's accuracy =  $(271/298) * 100$   
= **90.9395973154**

*b) RF Classifier Accuracy*

To measure the accuracy of the model, use the values of the Confusion matrix (error matrix). As given below here

*Confusion Matrix for RF classifier*

So, TCP =  $3+38+35=76$

TNP is the sum of all values of the accuracy matrix.

So, TNP =  $3+2+5+1+38+4+1+0+35=89$

Put values of TCP and TNP into the mentioned equation.

RF Classifier's Accuracy =  $(76/89) * 100$   
= **85.3932584269**

2) *Kappa Coefficient*: Kappa coefficient is used to control only those instances that may have been correctly classified by chance [10]. The value of this coefficient is between -1 to +1. The negative value of Kappa coefficient means low accuracy with between the classified image and the reference image, if the resultant value is 0 then it means no correlation at all between the classified image and the reference image and if the value is the higher then it indicates the higher accuracy of the classification [13].

Kappa Coefficient =  $((OLA) - (ELA)) / ((TNP) - (ELA))$

The OLA means the Observed Level of Agreement which is basically the diagonal's values sum of the confusion matrix and ELA indicates the Expected level of agreement which is calculated by multiplying the total row values with total column values and the resultant value further divides by TNP repeats this till ends. After that sum up all values. TNP is the Total Number of Pixels of the confusion matrix. The confusion matrix for SVM classifier using the Kappa Coefficient value is calculated and presented in Table II.

*a) Kappa Coefficient for SVM Classifier*

OLA =  $8+129+134 = 271$

ELA = row total \* col total / overall total

Use the above-mentioned equation for each feature and sum all the obtained values

$8*29/298 = 0.7785234899328859$

$135*133/298 = 60.25167785234899$

$134*157/298 = 70.59731543624162$

So,

The sum of all these =  $131.6275167785235$  and

Expected Accuracy =  $131.6275167785235$

Put these calculated values in the Kappa Coefficient Equation,

Kappa Coefficient =  $(271-131.6275167785235) / (298-131.6275167785235)$

Kappa Coefficient =  $0.8377135480747897$

The value of kappa Coefficient is in the range of 0.81–0.99, which shows almost perfect agreement.

*b) Kappa Coefficient for RF Classifier*

OLA =  $3+38+35=76$

ELA = row total \* col total / overall total

Use the above-mentioned equation for each feature and sum all the obtained values

$3*10/89 = 0.3370786516853933$

$40*43/89 = 19.32584269662921$

$44*36/89 = 17.79775280898876$

So,

The sum of all these =  $37.46067415730336$ .

Expected Accuracy =  $37.46067415730336$

Put these calculated values in the Kappa Coefficient Equation,

Kappa Coefficient =  $(76-37.46067415730336) / (89-37.46067415730336)$

Kappa Coefficient =  $0.7477654240244169$

The value of kappa Coefficient is in the range of 0.61–0.80 that shows substantial agreement.

SVM classifier accomplished 90.9% comprehensive classification precision with a kappa estimation of 0.83, while the Random Forest (RF) classifier acquired 85.3% in general classification precision with a kappa estimation of 0.74. RF. SVM classified maps are adjoining to the ground facts when contrasted with RF generated maps. The RF map contained miss-classified water pixel information. We expect to concentrate on urbanization development from 2017 to 2018. Using SVM, the yearly change is 86.71822731718704 sq. kilometers and using RF it is 96.33055709482205 sq. kilometer.

TABLE II. CONFUSION MATRIX FOR SVM CLASSIFIER BASED KAPPA COEFFICIENT

	Water	Vegetation	Urban
Water	3	2	5
Vegetation	1	38	4
Urban	1	0	35

## IX. CONCLUSION

The urbanization has affected the plans of settlement and use of the land around the sub-urban, increasing the population and areas concurrent. And by this mean, there have occurred the changes in the weather pattern raising the global warming and energy crisis. So, this study is conducted for the purpose to have a view about the land loss and how it is contributing to the energy crisis and greenhouse damaging. Analyzing the multi-temporal land cover change is very helpful for higher management and environmental departments to maintain sustainability in the development of urbanization. Realize the current change rate situation, then made the decisions and policy according to it. Also, for forecasting possible future changes with respect to growth to urban land/losses of water bodies and vegetation to avoid from any natural environmental problems that come in the near future. Uses the same size of training sample points and polygons for both classifiers and according to overall-accuracy and kappa coefficient comparison, observed that the SVM performed better as compared to RF, though both are superlative in performance when training sample data are smaller in size. Along with this, SVM classified maps are almost to the ground facts as compared to RF generated maps. The RF map contained miss-classified water pixel data. We utilize SVM and RF classifiers to assess the variation in urbanization. SVM and RF utilizing the prevailing platform GEE under an associated size of training sample points and polygons. Conferring to general precision and kappa-coefficient analysis, we observed that the SVM functioned better when contrasted with RF, while both are best in execution when training sample data are diminutive in size. This yearly change trend will be helpful for higher management and environmental departments to maintain sustainability in the development of urbanization by realizing the current situation, made better decisions and predicted future changes with respect to growth/loss land patterns including natural or unnatural environmental problems. The classification results of SVM algorithm are almost to the ground facts as compared to RF classification results and the majoris that the random forest classification results contained miss-classified water pixel data.

## X. FUTURE DIRECTIONS

- 1) For the future work we can do the comparison of the results we obtained from supervised classifiers with unsupervised classifier results.
- 2) Digital image classification of multi-spectral imagery for prominent surface features, particularly with reference to population growth, within the precincts of Multan district can be done in the future.
- 3) Using Satellite imagery measure the impact of water-bodies growth/losses on vegetation for the same area of study.

### REFERENCES

- [1] Liss, B., M.D. Howland, and T.E.J.J.o.A.S.R. Levy, Testing Google Earth Engine for the automatic identification and vectorization of archaeological features: A case study from Faynan, Jordan. 2017. 15: p. 299-304.
- [2] Blank, J., Clary, C., & Nichiporuk, B. (2014). Drivers of long-term insecurity and instability in Pakistan: Urbanization. Rand Corporation.
- [3] Ishii, T., et al. Detection by classification of buildings in multispectral satellite imagery. in 2016 23rd International Conference on Pattern Recognition (ICPR). 2016. IEEE.
- [4] Lam, N.S.-N., Methodologies for mapping land cover/land use and its change, in Advances in land remote sensing. 2008, Springer. p. 341-367.
- [5] Mountrakis, G., et al., Support vector machines in remote sensing: A review. 2011. 66(3): p. 247-259.
- [6] Thunig, H., et al. Land use/land cover classification for applied urban planning-the challenge of automation. in 2011 Joint Urban Remote Sensing Event. 2011. IEEE.
- [7] Borra, S., R. Thanki, and N. Dey, Satellite Image Clustering, in Satellite Image Analysis: Clustering and Classification. 2019, Springer. p. 31-52
- [8] Mantero, P., et al., Partially supervised classification of remote sensing images through SVM-based probability density estimation. 2005. 43(3): p. 559-570.
- [9] Maxwell, A.E., T.A. Warner, and F.J.J.J.o.R.S. Fang, Implementation of machine-learning classification in remote sensing: An applied review. 2018. 39(9): p. 2784-2817.
- [10] Karpatne, A., Jiang, Z., Vatsavai, R. R., Shekhar, S., & Kumar, V. (2016). Monitoring land-cover changes: A machine-learning perspective. IEEE Geoscience and Remote Sensing Magazine, 4(2), 8-21.
- [11] Thanh Noi, P. and M.J.S. Kappas, Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. 2018. 18(1): p. 18.
- [12] Yang, J., et al. Attribute weighted Naive Bayes for remote sensing image classification based on the cuckoo search algorithm. in the 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC). 2017. IEEE
- [13] Rwanga, S. S., & Ndambuki, J. M. (2017). Accuracy assessment of land use/land cover classification using remote sensing and GIS. International Journal of Geosciences, 8(04), 611.

# Development of a Recurrent Neural Network Model for English to Yorùbá Machine Translation

Adebimpe Esan<sup>1\*</sup>, John Oladosu<sup>2\*</sup>, Christopher Oyeleye<sup>3\*</sup>, Ibrahim Adeyanju<sup>4</sup>  
Olatayo Olaniyan<sup>5</sup>, Nnamdi Okomba<sup>6</sup>, Bolaji Omodunbi<sup>7</sup>, Opeyemi Adanigbo<sup>8</sup>

Department of Computer Engineering, Federal University Oye-Ekiti, Ekiti state, Nigeria<sup>1,4,5,6,7,8</sup>  
Ladoke Akintola University of Technology, Ogbomoso<sup>2,3</sup>

**Abstract**—This research developed a recurrent neural network model for English to Yoruba machine translation. Parallel corpus was obtained from the English and Yoruba bible corpus. The developed model was tested and evaluated using both manual and automatic evaluation techniques. Results from manual evaluation by ten human evaluators show that the system is adequate and fluent. Also, results from automatic evaluation shows that the developed model has decent and good translation as well as higher accuracy because it has better correlation with human judgment.

**Keywords**—Recurrent; tokenizer; corpus; translation; evaluation; correlation

## I. INTRODUCTION

The demand for translation and translation tools currently exceeds the capacity of available solution [1], hence, the need to intensify research in the field of machine translation [2]. Machine Translators (MT) accept characters of source language and map to the characters of the target language to generate the words with the help of various rules and other learning process techniques [3]. Previous researchers have employed various approaches to develop machine translators and the approaches were categorized into two by [4], namely; single and hybrid approaches. Single approaches include: rule-based, knowledge-based, statistical and direct approaches while Hybrid approaches are: word-based, phrase-based, syntax-based, forest-based and neural machine translation models.

Neural Machine Translation (NMT) is an improvement in the field of machine translation where a large neural network is built and trained to read a sentence and output a correct translation [5]. The approach consists of the encoder and the decoder for encoding a source sentence and decoding it to a target sentence [6]. Neural machine translators have shown promising results than previous MT approaches through the incorporation of some neural components to existing translation systems like phrase-based systems [7]. In addition, research revealed that NMT produces automatic translations that are significantly preferred by humans when compared to other machine translation approaches.

However, the most widely used model for NMT is the Recurrent Neural Network model which is a supervised machine learning model that is made of artificial neurons with one or more feedback loops. In order to train a RNN, a parallel corpus is trained so as to minimize the difference between the output and target pairs by optimizing the weights of the

network [8]. In addition, a portion of the corpus is used as the validation dataset [9] to watch the procedure during training and prevent the network from underfitting or overfitting. RNNs have distributed hidden states used for storing information about the past efficiently and non-linear dynamics for updating their hidden state [10]. Hence, this research developed a recurrent neural network model for English to Yoruba machine translation.

## II. RELATED WORKS

Neural Machine Translation (NMT) is an improvement in the field of machine translation and it is based purely on deep neural networks. The encoder–decoder architecture [5] which is a conventional approach to neural machine translation, encodes a whole input sentence into a fixed-length vector from which a translation was decoded. Research show that the use of a fixed-length context vector is a challenge for the translation of longer sentences, hence, the research was extended by developing a model that soft-search for a set of input words, or their annotations computed by an encoder, when generating each target word [11]. The method prevents the model from encoding all the source sentences into a fixed-length vector but focuses only on relevant information that will help to generate target word. This approach outperformed the conventional encoder-decoder model significantly.

However, as training and decoding complexities increase proportionally to the number of target words in previous NMT systems, the size of the target vocabulary was extended by using an approach that enables training a model with much larger target vocabulary without substantial increase in computational complexity [12]. Decoding was efficiently done using a very large target vocabulary by selecting a small portion of the target vocabulary. Research show that the models trained outperformed the baseline models with a small vocabulary size. Though, it is unable to translate words which could not be found in the vocabulary. Therefore, alignment-based technique was used by [13] to mitigate this problem. The technique was carried out by training the model on data that is augmented by the output of a word alignment algorithm, allowing the NMT system to emit, for each Out of Vocabulary (OOV) word in the target sentence, the position of its corresponding word in the source sentence.

Moreover, [14] developed a multi-task learning model by training a unified neural machine translation model. In the research, an encoder is shared across different language pairs and each target language has a separate decoder. The

\*Corresponding Author

challenge with this model is the inability to address the data scarcity problem of some resource-poor language pairs. Thus, attention mechanism was incorporated in the models by [15] to overcome the problem. The attention mechanism was incorporated into the multi-task neural machine translation model and this method helps eliminate the data scarcity problem of the baseline model. Despite this achievement, the model still relies on word-level modeling.

Therefore, to reduce reliance of MT systems on word-level modeling, an attention-based encoder– decoder with a sub word-level encoder and a character-level decoder were developed for NMT [16]. The approach focused on the target side, in which a decoder generated one character at a time, while soft-aligning between a target character and a source sub-word. Research showed that the character-level decoder outperformed the sub-word-level decoder. Finally, [17] also addressed the data scarcity problem by developing a multi-task learning model by training a unified neural machine translation model with the decoder shared over all language pairs and each source language has a separate encoder. Research showed that given small parallel training data, the model was effective in learning the predictive structure of multiple targets.

### III. METHODOLOGY

#### A. Design of the Recurrent Neural Network Model

The RNN model was designed to include three layers: input layer, hidden layer and output layer. The input layer was designed to have  $N$  input units and the inputs to this layer is a sequence of vectors through time  $t$  such that  $\{x_{t-1}, x_t, \dots, x_{t+1}\}$ , where  $x_t = (x_{t1}, x_{t2}, \dots, x_{tN})$ . The encoder RNN reads the input sentence which is a sequence of vectors  $x = (x_1, \dots, x_{T_x})$  into a vector  $c$  as in equations 1 and 2:

$$h_t = f(x_t, h_{t-1}) \quad 1$$

and

$$c = g(h_1 \dots h_{T_x}) \quad 2$$

Where  $h_t \in \mathbb{R}^n$  is a hidden state at time  $t$ , and  $c$  is a vector generated from the sequence of the hidden states  $f$  and  $g$  are non-linear functions. The input units were connected to the hidden units in the hidden layer, where the connections are defined with a weight matrix  $W_c$ .

In addition, at the hidden layer, this research modified the recurrent neural network of [5] by estimating the distribution with an attention mechanism [13] to overcome the shortcoming of previous research [7]. The hidden layer has  $M$  hidden units  $h_i = (h_{i1}, h_{i2}, \dots, h_{iM})$ . The source encoder recurrent neural network (RNN) maps each source word from the input unit to a word vector and processes these to a sequence of hidden vectors  $h_1 \dots h_t$  as shown in equation 1. The source hidden vectors influence the distribution through an attention pooling layer  $h_t$  that weighs each source word relative to its expected contribution to the target prediction as shown in equation 3.

$$h_t^2 = \tan h(W_c [C_t ; h_t]) \quad 3$$

From equation 3,  $\tan h$  is the activation function,  $W_c$  is the weight matrix,  $C_t$  is the context vector and  $h_t$  are the hidden states. This research used  $\tan h$  activation function at the hidden layer to overcome the shortcoming of the sigmoid function in previous work [5].

Moreover, at the output stage, previous model [7](Cho et al. 2014) mapped the input vector to the target sequence with another RNN during sequence learning and this prevented the model from learning long term dependencies while training the RNNs. Therefore, this research computed the output layer by combining the RNN (GRU) hidden representation of previously generated words ( $w_1, \dots, w_{t-1}$ ) with source hidden vectors to predict scores for each possible next word as shown in equation 4. The activation in the GRU was modeled as equation 4:

$$y_t = W_t h_t \quad 4$$

where

$$h_t = (1 - Z_t)h_{t-1} + Z_t \tilde{h}_t \quad 5$$

From equation 5,  $Z_t$  is the update gate which controls the update value of the activation as shown in equation 6.

$$Z_t = (W_z X_t + U_z h_{t-1}) \quad 6$$

From equation 3.6,  $W$  and  $U$  are weight matrices to be learnt.

The candidate activation is shown in equation 7;

$$\tilde{h}_t = \tan h(W_h X_t + U_h (r_t \odot h_{t-1})) \quad 7$$

Where  $r_t$  is a set of reset gates defined as equation 8:

$$r_t = \sigma(W_r X_t + U_r h_{t-1}) \quad 8$$

The diagram of the designed RNN model is shown in Fig. 1.

#### B. Method and Size of Data Collection

Fourteen thousand two hundred and forty five sentences were extracted from the bible parallel corpus for training the system and five thousand sentences extracted from the bible parallel corpus were used in validating the system. The developed system was tested automatically using two different data sets from two literatures: five hundred and eighty eighth (588) sentences and one thousand (1,000) sentences respectively. Manual testing was also done using one two hundred (200) sentences from the third literature.

#### C. Text Preprocessing

Pre-processing of this corpus was carried out in three phases, namely; data loading, tokenization and vocabulary building.

1) *Data loading*: The parallel corpus was loaded as strings into memory. Every English sentence is placed on a line with its corresponding Yoruba translation and separated by a TAB. Cases were ignored and spaces were added between words and punctuation marks.

2) *Tokenization*: Morphology-based and frequency based tokenization approaches were used in tokenizing the corpus

used. Morphology based tokenization was employed to split off punctuation and numbers. Frequency-based tokenization was carried out using byte-pair encoding (BPE) [18].

3) *Vocabulary building*: Tokens that rarely appeared were mapped into a special unknown (“<unk>”) token. Special tokens like: “<pad>”, “<bos>” and “<eos>” were added for padding, beginning of sentence and end of sentence respectively.

**D. Training of the Model**

Training of data was done at every 1000 checkpoint. The model was validated on a dataset of 5000 sentences with an accuracy of 60.051. The sequence diagram of the designed model is shown in Fig. 2 while the class diagram of the designed model is shown in Fig. 3.

**E. The Developed RNN Model**

The developed model is shown in Fig. 4. From the diagram, source words were depicted by colour yellow while target words were depicted by blue colour. The source words were first mapped to word vectors and then fed into a recurrent neural network (RNN). At the end of sentence <eos> symbol was displayed and the final time step initializes a target RNN. At each target time step, attention was applied over the source RNN and combined with the current hidden state to produce a prediction of the next word  $p(w_t|w_{1:t-1}, X)$ . This prediction was then fed back into the target RNN. The developed system was tested and implemented for English to Yoruba translation. Fig. 5 shows a sample page of the developed system and Fig. 6 shows a sample page that confirms the ability of the system to translate long sentences.

**F. Evaluation of the Developed Model**

The Modified Recurrent Neural Network model was evaluated using Human judgment and Bilingual Evaluation Understudy (BLEU). Ten human evaluators evaluated the developed model using adequacy and fluency metrics [19] on a 5 point likert scale (over 0 to 4). The guidelines for evaluation required that the following score be given to a sentence by looking at each output sentence on a 5 point Likert scale (over 0- 4): 4=: All Meaning,3 =: Most meaning, 2 =: Much meaning, 1=: Little meaning and 0=: No meaning. The overall adequacy of the system was computed using the formula by [20] where the total number of sentences with scores 2, 3 and 4 were added and divided by the total number of sentences N as shown in equation 9.

$$Adequacy = \frac{Scores\ 2,3,4}{N} \tag{9}$$

The correct grammatical constructions present in the translated sentences were evaluated using fluency metric based on the research by [21]. The guidelines for evaluation required that the following scores be given to a sentence by looking at each output sentence on a 5 point Likert scale (over 0- 4): 4 := for Perfect., 3 := for Good, 2 := for Non-native, 1 := for Diffluent, 0 := for Incomprehensible. The overall fluency of the system was computed using the formula by [22] as shown in equation 10. Scores above 2 were considered. Scores 3 and 4 were penalized by multiplying their count by 0.8 and 0.6 respectively so as to make the estimated score better.

$$(F = 100 * ((S4 * 0.8 + S3 * 0.6 + S2) )/N) \tag{10}$$

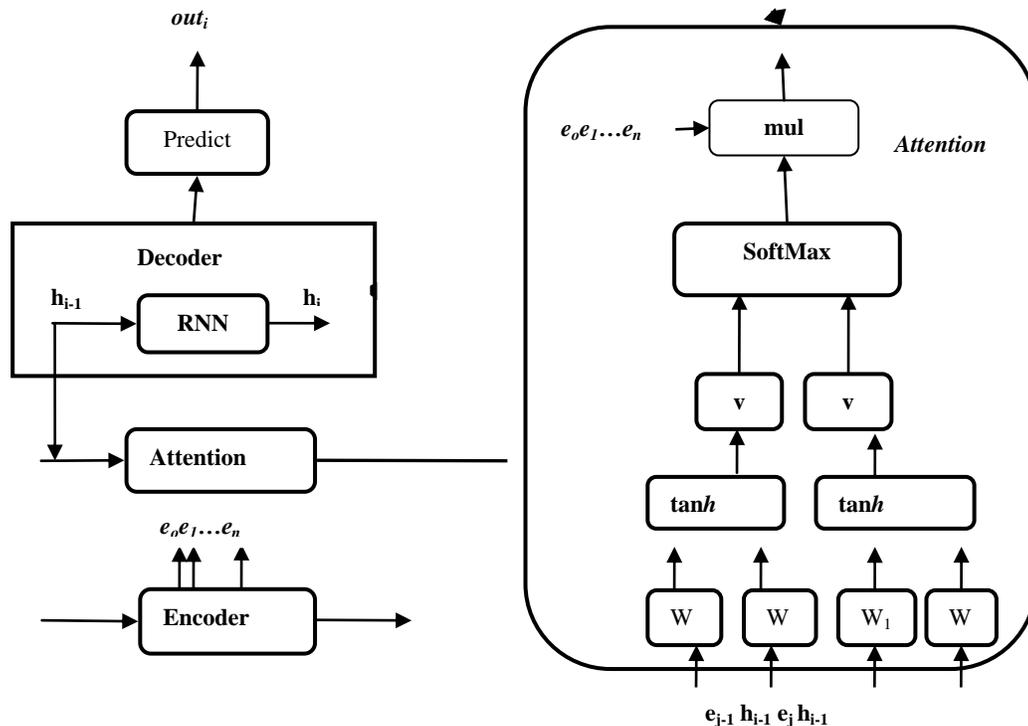


Fig. 1. Design of the RNN Model.

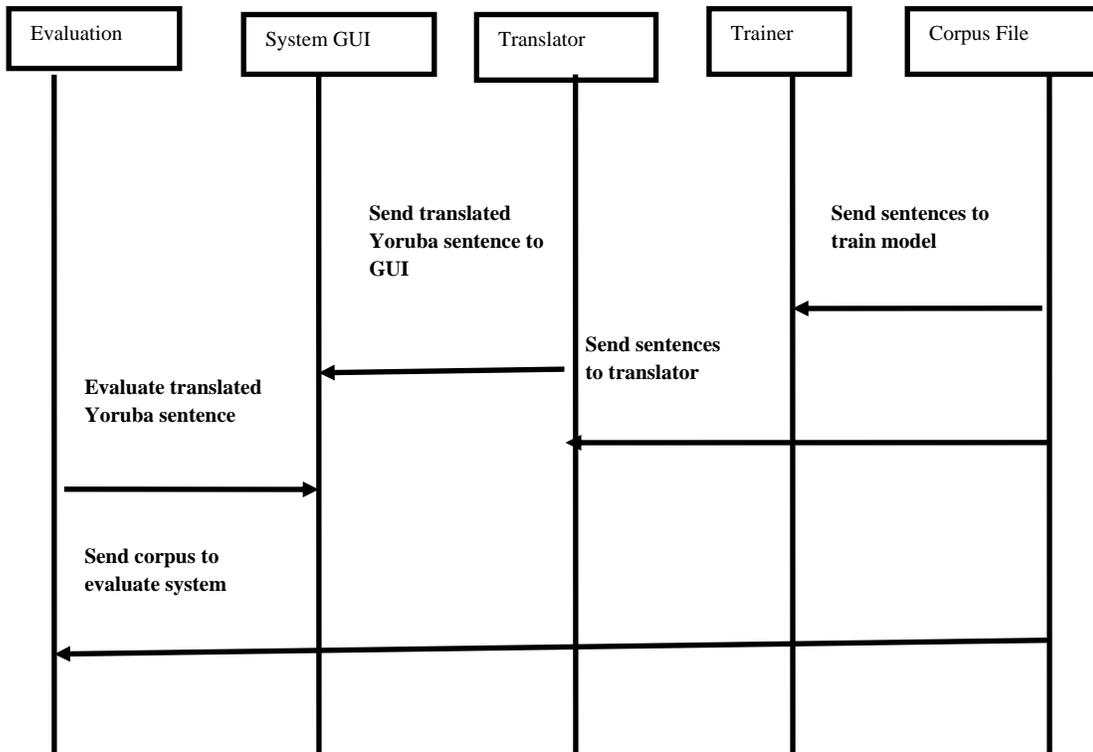


Fig. 2. Sequence Diagram of the Developed Model.

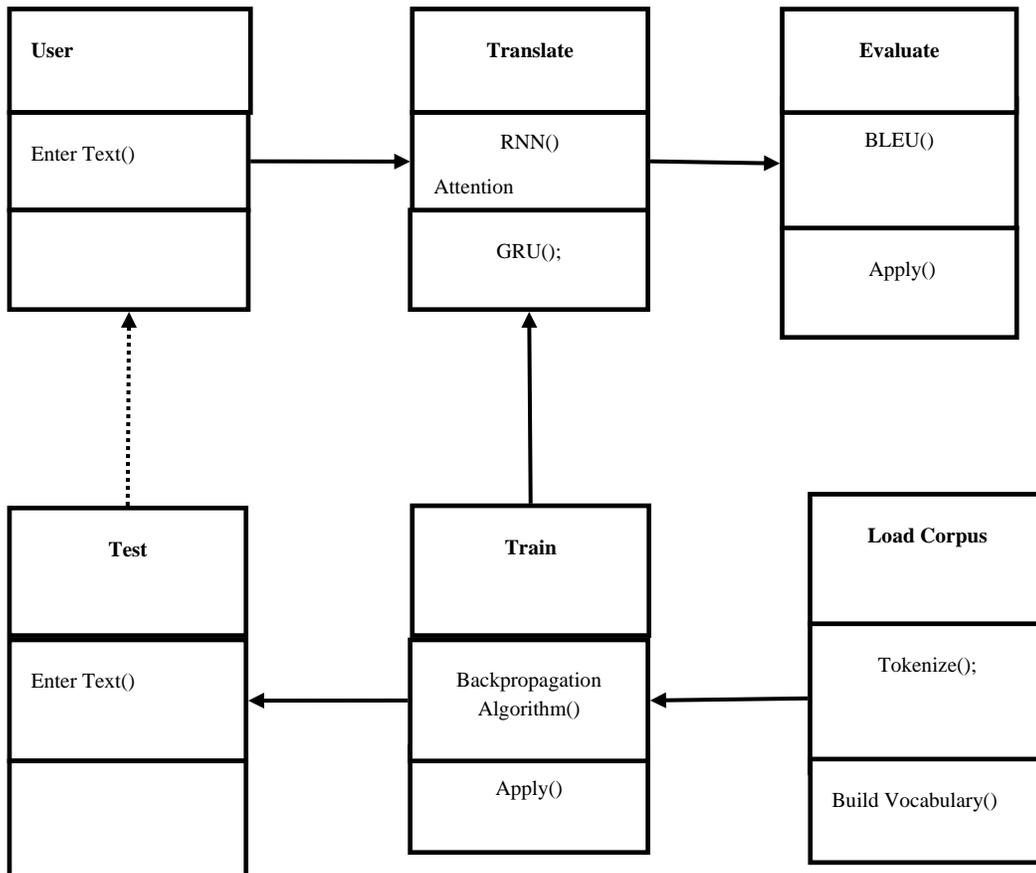


Fig. 3. Class Diagram of the Developed Model.

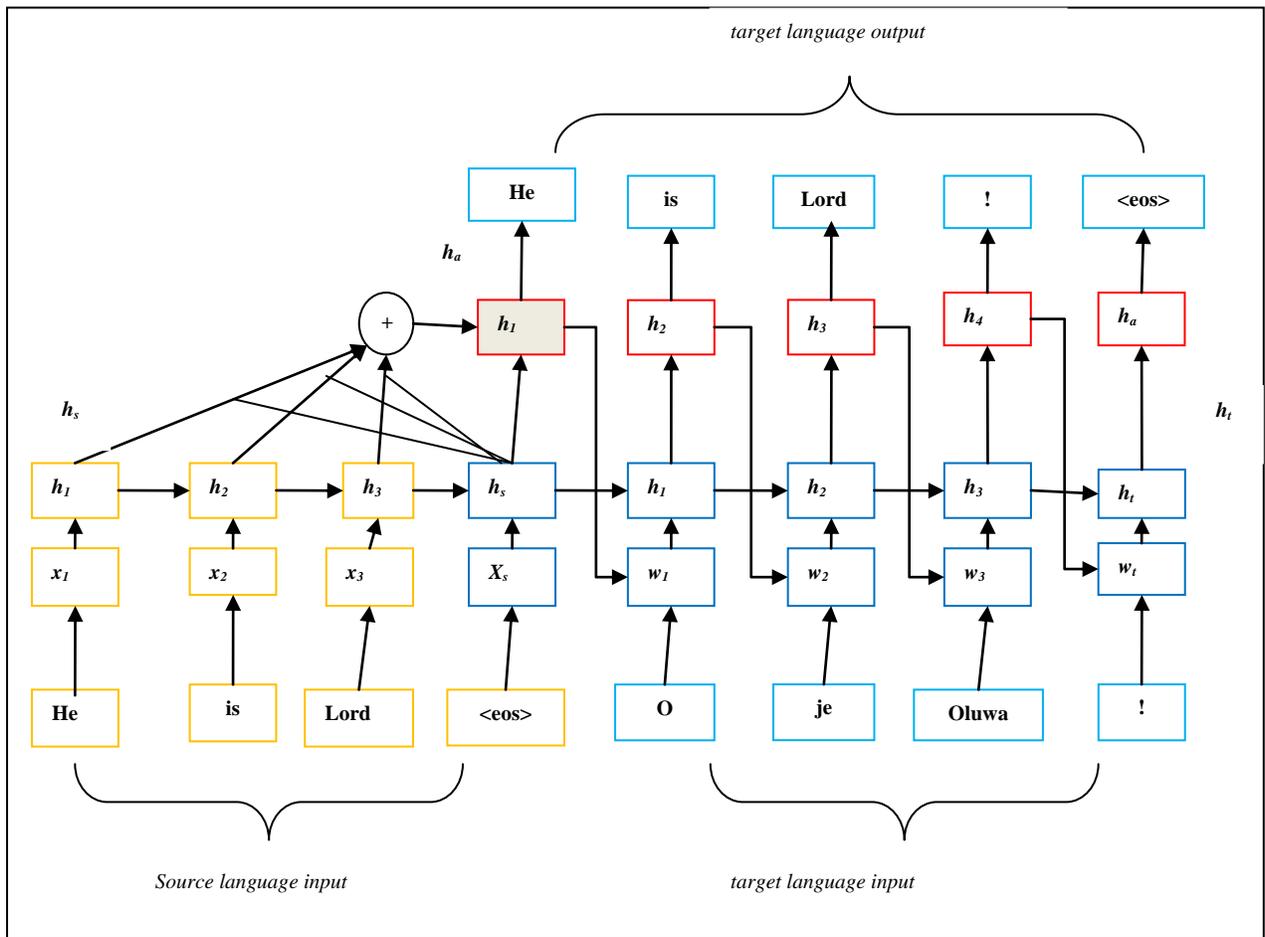


Fig. 4. The Developed RNN Model.

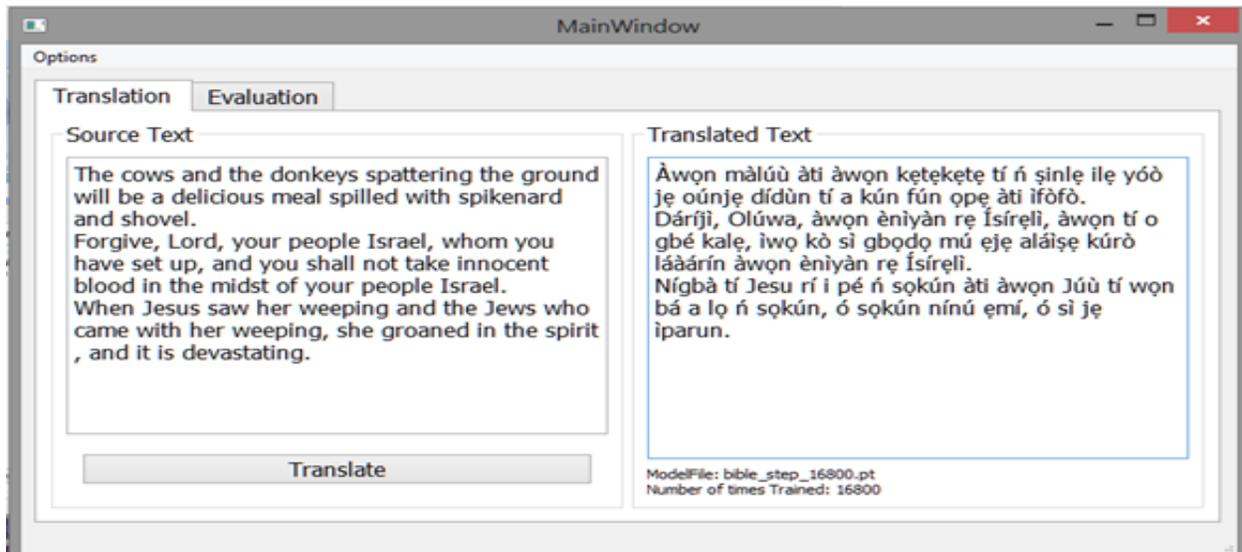


Fig. 5. Sample Page of the Developed NMT Model.



Fig. 6. Sample Page that Shows the Ability of the Model to Translate Long Sentences.

#### IV. RESULTS

Results from ten human evaluators were computed using the adequacy formula from [20] as shown in Table I and results from ten human evaluators were computed using the

fluency formula from [22] and the results obtained are shown in Table II. The average scores for adequacy and fluency metrics for the developed system were computed and recorded in Table III. The overall average for adequacy and fluency metrics are 86.65 and 70.72, respectively.

TABLE I. ADEQUACY METRIC SCORES FOR THE DEVELOPED SYSTEM

SCALE 0-4	RNMT		ADEQUACY		METRIC		SCORES			
	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10
0	9	8	10	7	6	4	5	7	5	6
1	25	21	23	18	19	20	17	20	19	18
2	42	38	40	35	37	32	30	31	36	34
3	50	52	49	54	53	53	50	51	54	51
4	74	81	78	86	85	91	98	91	86	91

TABLE II. FLUENCY METRIC SCORES FOR THE DEVELOPED SYSTEM

SCALE 0-4	RNMT	FLUENCY	METRIC	SCORES						
	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10
0	6	10	7	3	8	9	9	5	6	4
1	12	15	13	7	10	10	11	10	11	9
2	49	42	45	47	40	43	41	43	42	46
3	60	62	67	69	70	68	66	65	72	71
4	73	71	68	74	72	70	73	77	69	70

TABLE III. COMPUTED SCORES FOR ADEQUACY AND FLUENCY METRICS

METRIC	CALCULATED				METRIC	SCORES				
	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10
Adequacy	83.0	85.5	83.5	87.5	87.5	88.0	89.0	86.5	88.0	88.0
Fluency	71.7	68.4	69.8	73.8	69.8	69.9	69.5	71.8	70.2	72.3

## V. DISCUSSION

Results from adequacy and fluency metrics of the developed system were compared and it was discovered that the system's adequacy score is higher than the fluency score. This is in line with the research by [20] where Google translate' Comprehensibility (adequacy) score was found to be higher than its fluency score. Also, the overall adequacy score of the developed system show that neural machine translators are more adequate than other MT approaches. This is according to the research by [22] where English language was translated to Hindi using Rule based and Statistical approaches and percentage adequacy scores recorded was significantly lower than the developed system. Moreover, the developed system has an improved quality output when compared to previous systems (Google and Bing SMT systems) [20] because its adequacy and fluency scores are higher than the previous systems. The results obtained in this research also confirm that NMT systems give relatively high accuracy when trained on a larger dataset and can yield good predictions as well [23]. The results from manual evaluation also affirms that GRU based RNN provides a stronger and more robust translation model with high resourced languages [24].

Automatic evaluation of the developed system gives a percentage BLEU score of 54.8 when tested on a dataset of five hundred and eighty eight sentences and 57.3 when tested on a data set of one thousand sentences. The average BLEU score obtained from the two datasets is fifty-six percent (56%). The research by [25] reveals that BLEU scores up to 50 and above generally reflect good and fluent translations. Author in [26] also confirms that a BLEU score of 50 implies a decent translation. Hence, the result shows that the system has good and fluent translation.

In addition, research by [6] revealed that neural machine translators have higher quality output than phrase-based machine translators. Hence, the results obtained from automatic evaluation of the developed system were compared to the results from previous system [1] and the BLEU score is significantly higher than the previous system. The performance of neural machine translators was also confirmed to be higher than statistical machine translators by [27] when English was translated to Arabic language. The NMT model outperforms SMT model by 1.5 BLEU in the out-of-domain testing.

In other words, this research establishes that NMT systems have higher quality output than previous rule based and statistical MT systems. This was established by comparing the result obtained from the developed system to the research by [22] and a significant improvement in BLEU score was recorded. The result also revealed that Recurrent Neural network Models coupled with an attention mechanism produces higher quality output than previous neural machine translators. This was confirmed by [28] where different neural machine translators were compared and the recurrent neural MT outperformed other approaches considered.

## VI. CONCLUSION

This research developed a recurrent neural network model for English to Yoruba machine translation. The model was

tested and evaluated using both human and automatic evaluation techniques and the system was found adequate and fluent with decent and good translation. Hence, this research established that neural machine translators outperformed previous machine translation approaches and affirms that the addition of attention mechanism and gated recurrent units improves the quality of translation. It is recommended that future work extends the vocabulary size of this research and modify the model to handle multiple tasks for translation into different languages.

## REFERENCES

- [1] Ayogu, I.I., Adetunmbi, A.O. and Ojokoh, B.O. (2018). Developing Statistical Machine Translation System for English and Nigerian Languages. *Asian Journal of Research in Computer Science*; 1(4):1-8.
- [2] Esan,A.O, Omodunbi, B. Olaniyan, O. Odiase, P. and Olaleye, T. (2018). Development of Adjectival Phrase-Based English to Yorùbá Machine Translator. *International Digital Organization for Scientific Research Journal of Applied Sciences*, 3(1) 29-42.
- [3] Pankaj, K., and Er.Vinod, K. (2013). Statistical Machine Translation Based Punjabi to English Transliteration System for Proper Nouns. *International Journal of Application or Innovation in Engineering & Management (IIAEM)*, 318-321.
- [4] Oladosu J.B., Esan A.O., Adeyanju I. A., Adegoke B.O, Olaniyan O.M. and Omodunbi B.A. (2016). Approaches to Machine Translation: A Review. *FUOYE Journal of Engineering and Technology*,1 (1), 120-126.
- [5] Sutskever, I., Vinyals,O. and Le, Q. (2014). Sequence to sequence learning with neural networks. *Proceedings of International conference in Advances in neural information processing systems*, (3104–3112).
- [6] Bentivogli, L., Bisazza, A., Cettolo, M. and Federico, M. (2016). Neural versus Phrase-Based Machine Translation Quality: a Case Study.
- [7] Cho, K. Bahdanau, D. Bougares, F. Schwenk, H. and Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, pp. 1724–1734.
- [8] Agrawal, R. and Sharma, D. M. (2017). Building an Effective MT System for English -Hindi Using RNN 'S. *International Journal of Artificial Intelligence and Applications (IIAIA)*, 8(5).
- [9] Salehinejad, H., Sankar, S., Barfett, J., Colak, E. and Valaee, S. (2018). Recent Advances in Recurrent Neural Networks. <http://arxiv.org/abs/1801.01078v3> pp. 1-20.
- [10] Mahata and Das, (2018). Machine Translation Using Recurrent Neural Network on Statistical Machine Translation. *Journal of Intelligent Systems*, 34-43.
- [11] Bahdanau, D., Cho, K. and Bengio, Y. (2015). Neural Machine Translation By Jointly Learning To Align And Translate. In *proceedings of ICLR*.
- [12] Jean,S., Cho, K., Memisevic, R., and Bengio, Y. (2015). On using very large target vocabulary for neural machine translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China, 1–10.
- [13] Luong, M., Pham, H. and Manning, C.D. (2015a). Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*.
- [14] Dong, D., Wu, H., He, W.,Yu, D. and Wang, H. (2015). Multi-Task Learning for Multiple Language Translation. In *Proceedings of ACL*, Beijing, China.
- [15] Firat,O. Cho, K. and Bengio, Y. (2016). Multi-Way, Multilingual Neural Machine Translation with a Shared Attention Mechanism. *arXiv:1601.01073v1 [cs.CL]*.
- [16] Chung, J., Cho, K. and Bengio, Y., (2016). A Character-Level Decoder without Explicit Segmentation for Neural Machine Translation.

- [17] Baniata, L.H., Park, S. and Park S. (2018). A Neural Machine Translation Model for Arabic Dialects that Utilizes Multitask Learning (MTL). *Computational Intelligence and Neuro-Science*; (2018):1-10.
- [18] Sennrich, R., Haddow, B. and Birch. A., (2015). Improving neural machine translation models with monolingual data. arXiv preprint arXiv:1511.06709. [19] Papineni, K., Roukos, S., Ward, T. and Zhu, W. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Stroudsburg, PA, USA, pp. 311-318.
- [19] Papineni, K., Roukos, S., Ward, T. and Zhu, W. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Stroudsburg, PA, USA, pp. 311-318.
- [20] Ojha, A., Bansal, A., Hadke, S. and Jha, G. (2014). Evaluation of Hindi-EnglishMT Systems. *Proceedings of second workshop on Indian Language Data: Resources and Evaluation*, Harpa Conference Centre, Reykjavik, Iceland.
- [21] Pathak, S., Ahmad, R. and Ojha, A. (2012): A Language Engineering Approach to Enhance the Accuracy of Machine Translation Systems, 2(1):205-214.
- [22] Sreelekha, S., (2016). Statistical Vs Rule Based Machine Translation; A Case Study on Indian Language Perspective. *World Journal of Computer Application and Technology*. 4(4): 46-57.
- [23] Greenstein, E. and Penner, D. (2015). Japanese-to-English Machine Translation Using Recurrent Neural Networks. pp. 1-7.
- [24] Wang, R., Panju, M. and Gohari, M. (2017). Classification-Based RNN Machine Translation using GRUs. arXiv:1703.07841v1 [cs.NE].
- [25] Klein, G., Kim, Y., Deng, Y., Senellart, J. and Rush, A.M. (2017). OpenNMT: Open-Source Toolkit for Neural Machine Translation. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics-System Demonstrations*, Vancouver, Canada, 67-72.
- [26] Lavie, A. (2011). Evaluating the Output of Machine Translation Systems. In *Proceedings of 13th Machine Translation Summit Tutorial*, Xiamen, China.
- [27] Oudah, M., Almahairi, A. and Habash, N. (2019). The Impact of Pre-processing on Arabic-English Statistical and Neural Machine Translation, arXiv:1906.11751v1 [cs.CL].
- [28] Chen, M., Firat, O., Bapna, A., Johnson, A., Macherey, A., Foster, G., Jones, L., Parmar, N., Shazeer, N., Vaswani, A., Uszkoreit, J., Kaiser, L., Wu, Y., Hughes, M., and Zhifeng, M. (2018). The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers)* (pp. 76-86), Melbourne, Australia.

# Thinging-Oriented Modeling of Unmanned Aerial Vehicles

Sabah Al-Fedaghi<sup>1</sup>, Jassim Al-Fadhli<sup>2</sup>  
Computer Engineering Department  
Kuwait University  
Kuwait

**Abstract**—In recent years, there has been a dramatic increase in both practical and research applications of unmanned aerial vehicles (UAVs). According to the literature, there is a need in this area to develop a more refined model of UAV system architecture—in other words, a conceptual model that defines the system’s structure and behavior. The existing models mostly are fractional and do not account for the entire important dynamic attributes. Progress in this area could reduce ambiguity and increase reliability in the design of such systems. This paper aims to advance the modeling of UAV system architecture by adopting a conceptual model called a thinging (abstract) machine in which all of the UAV’s software and hardware components are viewed in terms of the flow of things and five generic operations. We apply this model to a real case study of a drone. The results—an integrated conceptual representation of the drone—support the viability of this approach.

**Keywords**—Unmanned Aerial Vehicle (UAV); drone; conceptual modeling; diagrammatic representation; system architecture

## I. INTRODUCTION

Individuals are becoming more reliant on automated systems for a wide range of applications. According to Culus Schellekens, and Smeets [1], “It looks like our 21st century will be the century of robots, with a lot of buzz concerning a fast-growing subfamily of these machines, namely drones.” The recent drone attacks on Saudi Arabia’s oil installations highlight the importance of giving this technology high research priority. The introduction of unmanned aerial vehicles (UAVs) has raised profound questions around the world with regard to issues such as accountability, transparency, privacy, legality, use of force, and safety [2]. In recent years, although UAVs have been an in-demand research topic, “there still exist many unanswered questions” [3].

From a technical standpoint, unmanned airplanes can be categorized as UAVs, remotely piloted vehicles, or drones. These types differ mainly in the type of the mission, the size, and (importantly) the level of autonomy in their operation [4]. The term “drone,” which is in general use among both the media and the public, refers to all types of UAVs.

Nowadays, the number of possible uses of UAVs is large and increasing [5]. One planned development in this direction is the use of a centimeter-scale quadcopter with a driving application over vast regions. An immense number of such vehicles are used for various purposes (e.g., providing climatic and meteorological data) [3]. One especially important

application of UAVs is carrying out so-called D missions [6]: those that are dangerous, dirty, or dull.

This paper focuses on the high-level modeling and control of UAVs. Its general objective is to offer a schematic language for UAVs so as to enhance the understanding of their functionalities. The understanding of technology is a constitutive part of human life and helps to address issues of survival and improve the use and practice of technology. According to Sellars [7], contemporary society hangs together largely through technology. This context requires understanding “both the practice of designing and creating artifacts (in a wide sense, including artificial processes and systems) and the nature of the things so created” [8]. Accordingly, there is a need to develop a language that ensures good technical specification. Such a specification is similar to a script for performing a task in that it allows stakeholders (engineers, team members, legislators, technocrats, managers, officials, etc.) to understand the roles they need to play and that helps them to avoid either stepping on one another’s toes or overlooking a critical piece of information [9].

This paper’s specific objective is thus to provide a modeling language that can be used to specify the system architecture as an integral phase of the UAV development process. The system architecture is the conceptual model that defines the structural, behavioral, and other views of a system [10]. Indeed, it is necessary to develop, for UAVs, both “architecture generation and assessment models. Architecture assessment models that presently exist tend to be fractional and do not account for all dynamic attributes that should be considered in the architecture assessment” [11]. Further development in this area can reduce systems’ ambiguity and increase their tangibility.

To accomplish these aims, we propose applying a diagrammatic modeling technique called a thinging machine (TM). This modeling apparatus is viable in the area of UAV systems architecture because it can provide a precise description of the total system. This claim is substantiated by contrasting TM models with the current UAV diagramming methods. Additionally, this paper shows the feasibility of the TM approach by describing a real case study.

Section II includes a partial survey of the works in the area of diagram-based modeling for UAVs. Section III provides a review of the TM modeling tool, and a detailed example is given in Section IV. Section V describes the case study, in which an actual drone is modeled. Section VI clarifies the TM

model itself and Section VII present the application of TM in the case study and describe the possible utilization of the resulting model in a simulation. Section VIII discuss the simulation aspects, and finally Section IX is the conclusion.

## II. SOME RELATED WORKS

Extensive research has been conducted on the conceptual modeling of UAVs at all levels [12]. Drones provide unprecedented levels of access to airspace, and such new access could fundamentally change business, shipment, and travel [13]. However, we focus here on a few architectural models that facilitate contrasts with the TM approach. A comprehensive survey of the field can be found in Renaul [11].

Pastor et al. [6] presented a hardware/software architecture for UAVs, using block diagrams to provide a general view of the architecture of a mission-control computer and ad hoc diagrams to describe operational scenarios. This is an example of such a scenario: The mission control decides to take a georeferenced video. For this task, it will need the services provided by storage, the flight-computer system, and the camera and sensing modules.

Diem, Hien, and Khanh [14] sought to analyze, design, and implement controllers of a standard UAV platform; they thus adopted a model-driven architecture with a real-time unified modeling language (UML). This architecture contains three models that are used to separate the specifications for a system’s operation: the computation-independent, platform-independent, and platform-specific models. To capture the general requirements based on the object-oriented paradigm, Diem, Hien, and Khanh [14] presented a model with abstract classes using UML stereotypes and a class diagram [15] in order to describe the main functional components of quadrotor UAVs (see Fig. 1-3). According to Diem, Hien, and Khanh [14], in Fig. 1, “The Guidance System’s class is responsible for giving the desired trajectory for a quadrotor UAV to follow. This responsibility is completed by taking the desired waypoints defining pre-mission with the possible inclusion of external environmental disturbances issued from the Air Environment’s class; then, it generates a path for this quadrotor UAV to follow.

This is the level of modeling and description that TM modeling targets. The style of diagramming in Fig. 1-3 is included to provide a contrast with the TM diagrams for the model UAV, as developed later in this paper.

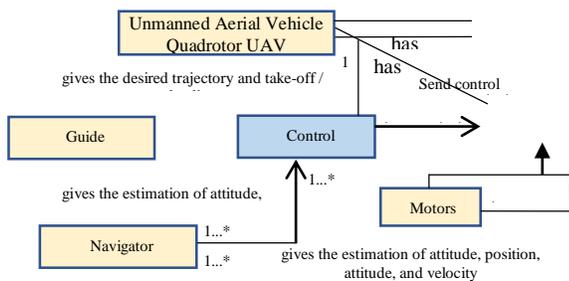


Fig. 1. A UML Class Diagram for Presenting the main Functional Components of Quadrotor UAVs (Redrawn and Adapted from [14]).

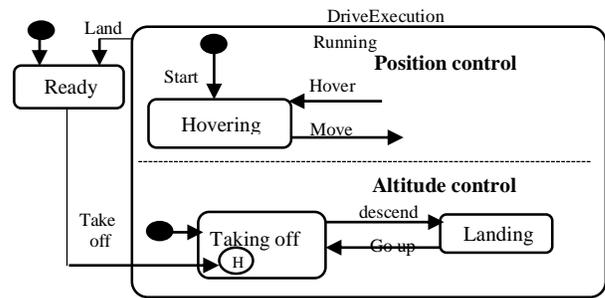


Fig. 2. Local State Machine for the Drive use case. (Redrawn and Adapted from [14]).

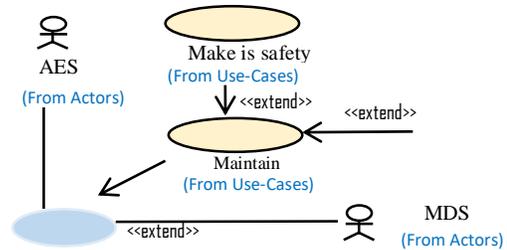


Fig. 3. Main use-case Model for a Quadrotor UAV. (Redrawn and Adapted from [14]).

Although nothing prevents a diagram from presenting multiple views of a system, these diagrams are heterogeneous and contain awkward symbols; thus, there is a need for more systematic depictions that, according to [16], help to meet the challenge of defining a single coherent architecture. TM both presents the totality of the system in a conceptual form and distinguishes between a model’s static and dynamic aspects.

## III. THINGING MACHINES

We adopt a conceptual model that is centered on a system’s things and (abstract) machines. The philosophical foundation of this approach is Heidegger’s notion of thinging [17]. According to Riemer, Johnston, Hovorka, and Indulska [18], Heidegger’s philosophy gives an alternative analysis of “(1) eliciting knowledge of routine activities, (2) capturing knowledge from domain experts and (3) representing organizational reality in authentic ways” [18]. More information about TM’s philosophical foundation can be found in Al-Fedaghi [19–21].

The simplest type of the thing/machine combination is a TM, which is a generalization of the known input-process-output model. In a TM, the flow of things is the exclusive conceptual movement among the five operations (stages), as shown in Fig. 4. A thing is created, transferred, processed, released, and/or received in a machine.

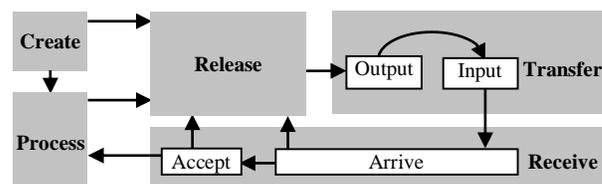


Fig. 4. A Thinging Machine.

Accordingly, the stages of a TM can be described as operations that transform, modify, or otherwise affect things abstractly or concretely. They are as follows.

- **Arrive:** A thing flows to a new machine (e.g., packets arrive at a router's port).
- **Accept:** A thing enters a TM after arrival; on the assumption that all arriving things are accepted, arrive and accept can be combined as the **Receiving** stage.
- **Release:** A thing is marked as ready for transfer outside the machine (e.g., in an airport, passengers wait to board after passport clearance).
- **Process:** A thing's descriptions are changed (rather than the thing itself).
- **Create:** A new thing is created (e.g., a forward packet is generated in a machine).
- **Transfer:** A thing is input to or output from a machine.

TM also includes *triggering* (denoted by a dashed arrow), or the initiation of a new flow (e.g., electricity triggers a flow of air). TM modeling is used in many applications (e.g., see Al-Fedaghi [22–25]).

#### IV. EXAMPLE

Transporting things is a main application area for UAVs [26–29]; this includes delivering medicines and immunizations. For the delivery, a transaction message is sent to the UAV with the GPS coordinates and the identifier of the order's package docking device [30]. At the delivery location, the control unit checks to ensure that the identifier matches the one in the transaction message and then performs the package transfer [28–30]. Without loss of generality, we model just the delivery system's pickup, as shown in Fig. 5.

In the figure, a packet-transfer request is created (circle 1) and sent to the UAV system (2), where it is processed (3) to extract the pickup-location address (4). This address—with proper processing (5)—flows (6) to the tracking device (7, the antenna and communication) and is sent (8) to the satellites (9). The pickup location's GPS coordinates then flow from the satellites (10), through the tracking device, to the UAV (11), which processes them (12). This process triggers the creation of control instructions (13). These instructions flow to the actuator (14), which is responsible for moving and controlling the UAV's mechanism. Accordingly, the UAV moves to the pickup location (15).

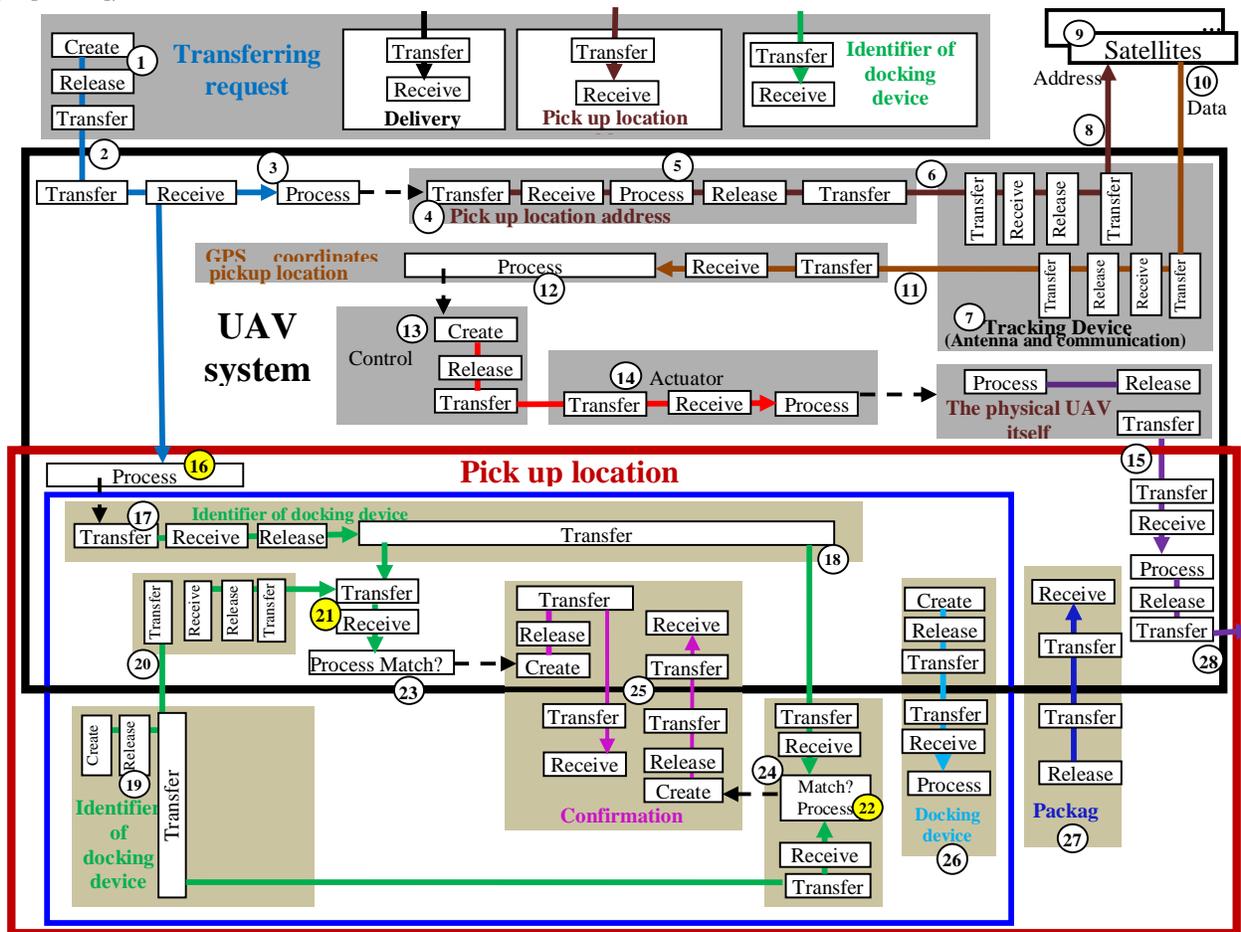


Fig. 5. The TM Model of the UAV Delivery System's Pickup.

Upon arrival at the pickup location, the transfer request is processed again (16) to extract the docking device's identifier (17) and send it to the local location (18). At the same time, the docking device's local-location identifier (19) flows (20) to the UAV, which compares the two identifiers (21, 22). If the identifiers' UAV (23) and local location (24) match, confirmations are exchanged (25).

Upon confirming its identifier from both sides, the docking device (26) moves to the local location and transfers the involved package to the UAV (27). Lastly, the UAV leaves for the delivery location (28).

In a TM model, an event is a machine with at least three submachines: the time, the region, and the event itself. Accordingly, Fig. 6 shows each event, represented by its region, and Fig. 7 shows the UAV system's behavior in terms of the chronology of events (listed below).

- Event<sub>1</sub> (E<sub>1</sub>): A package-transfer request is created.
- Event<sub>2</sub> (E<sub>2</sub>): The request arrives at the UAV, where the local-area address is extracted and sent to the tracking device, which sends the GPS coordinates.

- Event<sub>3</sub> (E<sub>3</sub>): The GPS coordinates are received.
- Event<sub>4</sub> (E<sub>4</sub>): The GPS coordinates are sent to the control, which issues instructions to the actuator.
- Event<sub>5</sub> (E<sub>5</sub>): The UAV is processed (moved) according to the incoming coordinates.
- Event<sub>6</sub> (E<sub>6</sub>): The UAV moves to the pickup location.
- Event<sub>7</sub> (E<sub>7</sub>): The docking-device identifier is extracted from the request and sent to the pickup location.
- Event<sub>8</sub> (E<sub>8</sub>): The docking-device identifier is received and checked at the pickup location.
- Event<sub>9</sub> (E<sub>9</sub>): Confirmations are exchanged between the UAV and the pickup location.
- Event<sub>10</sub> (E<sub>10</sub>): The package is picked up.
- Event<sub>11</sub> (E<sub>11</sub>): The UAV reaches the delivery location.

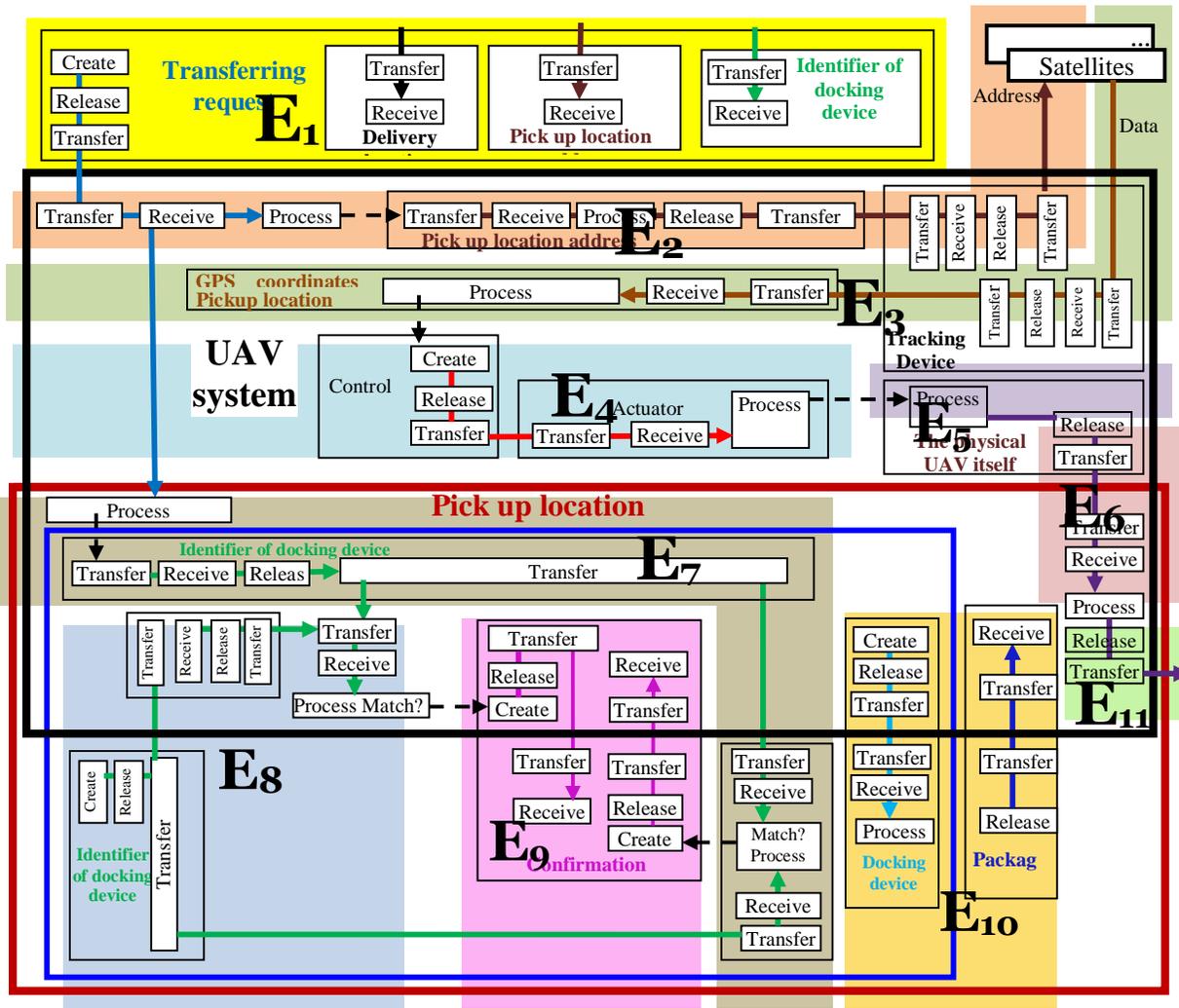


Fig. 6. The TM Model of Part of the UAV Delivery System.

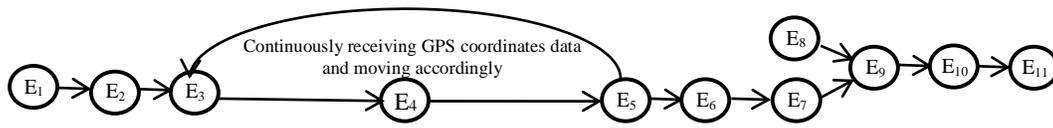


Fig. 7. The Chronology of Events in the TM Model of Part of the UAV Delivery System.

### V. CASE STUDY

This case study is from an actual project called RECON that involves a drone and its control unit, as initially reported in [25]. RECON was implemented to monitor, analyze, inspect, and intervene in data collection for traffic planners, safety managers, and commuters. RECON was originally built using non-TM notions. Fig. 8, 9, and 10 show some of the diagrams used for the original drone (see [25]). No further details about RECON are given because this paper is focused on elaboration of the use of TM and not on description of a UAV project.

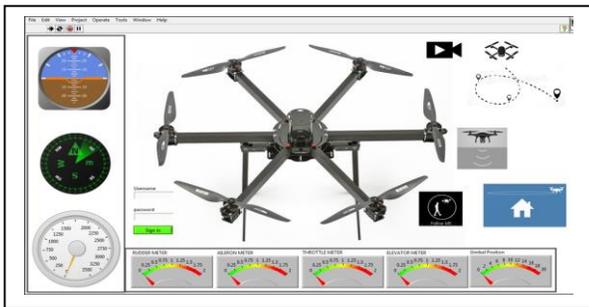


Fig. 8. The RECON user interface.

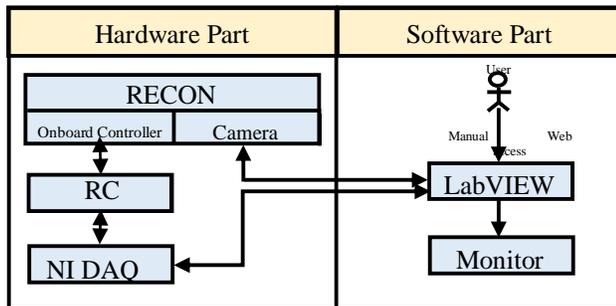


Fig. 9. An Overview of RECON.

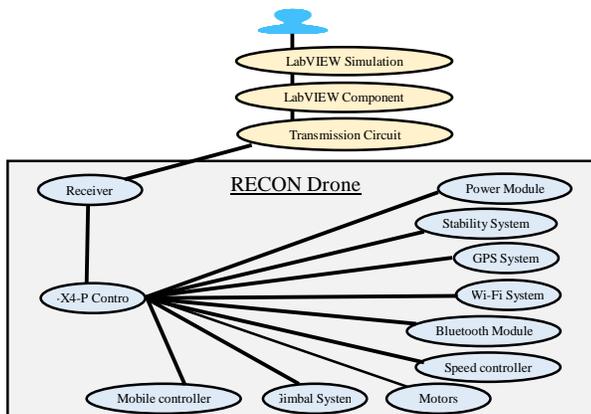


Fig. 10. The RECON user interface Modeled as a UML use Case.

### VI. MODELING UNMANNED AERIAL VEHICLES

Fig. 11 shows a TM-based static UAV model that consists of a user interface (UI) (circle 1), a server with a control panel (2), and the drone itself—comprising a controller (3) and a physical body (4). The UI contains several pointers (i.e., buttons; see Fig. 8), which are used to manage the flight.

If the drone is turned on, creating a click (5) in the UI causes one of the following signals to be sent (6): point-to-point, auto-landing, elevating signal, lateral balance, throttle signal, rudder signal, fail-safe, and follow-me.

#### A. Point-to-Point

This pointer causes the drone to move to a new point (position) from its current point in all directions. To accomplish this point-to-point movement, the UI process (6) creates a signal that is released and transferred (7) to the server, where it is stored (8) in the database. The signal also flows to the drone controller (9), where it is processed to trigger the physical drone's movement (10).

The point-to-point movement (11, in the lower right corner of the figure) creates digital data (12) that flow to the server (13), where they are stored. The data also flow to the UI (14), where they are displayed on the user's screen (15).

#### B. Auto-Landing

This pointer causes the drone to land automatically in a given location. To accomplish this, a signal is created, released, and transferred (16) to the server, where it is stored (17). The signal is also sent to the controller (18), where it is processed, which triggers the physical drone's auto-landing operation (19). In addition, the auto-landing operation (20) creates related data (21) that flow to the server (22), where they are stored. The data then flow to the UI (23) for display on the user's screen (24).

#### C. Elevating Signal

This pointer causes the drone to move up or down. To accomplish this, a signal is created, released, and transferred (25) to the server, where it is stored (26). The signal then flows to the drone controller (27). In the controller, it is processed to trigger the physical drone's movement (28). Furthermore, this movement (29) creates related data (30) that flow to the server (31), where they are stored. The data then flow to the UI (32), where they are displayed on the user's screen (33).

#### D. Lateral Balancing

This pointer measures the rate of rotation and helps keep the drone balanced. Flying with unbalanced props can harm the drone's motors, reduce its flight quality, and affect its stability. Stabilization technology provides navigational information to the controller to enhance flight safety. The drone's lateral balance needs to work almost instantly to act against gravity, wind, and so on. To accomplish this, a signal is created,

released, and transferred (34) to the server, where it is stored (35). The signal then flows to the drone controller (36). In the controller, it is processed to balance the physical drone (37).

This stabilizing motion (38) creates related data (39) that flow to the server (40), where they are stored. The data then flow to the UI (41), where they are shown on the user's screen (42).

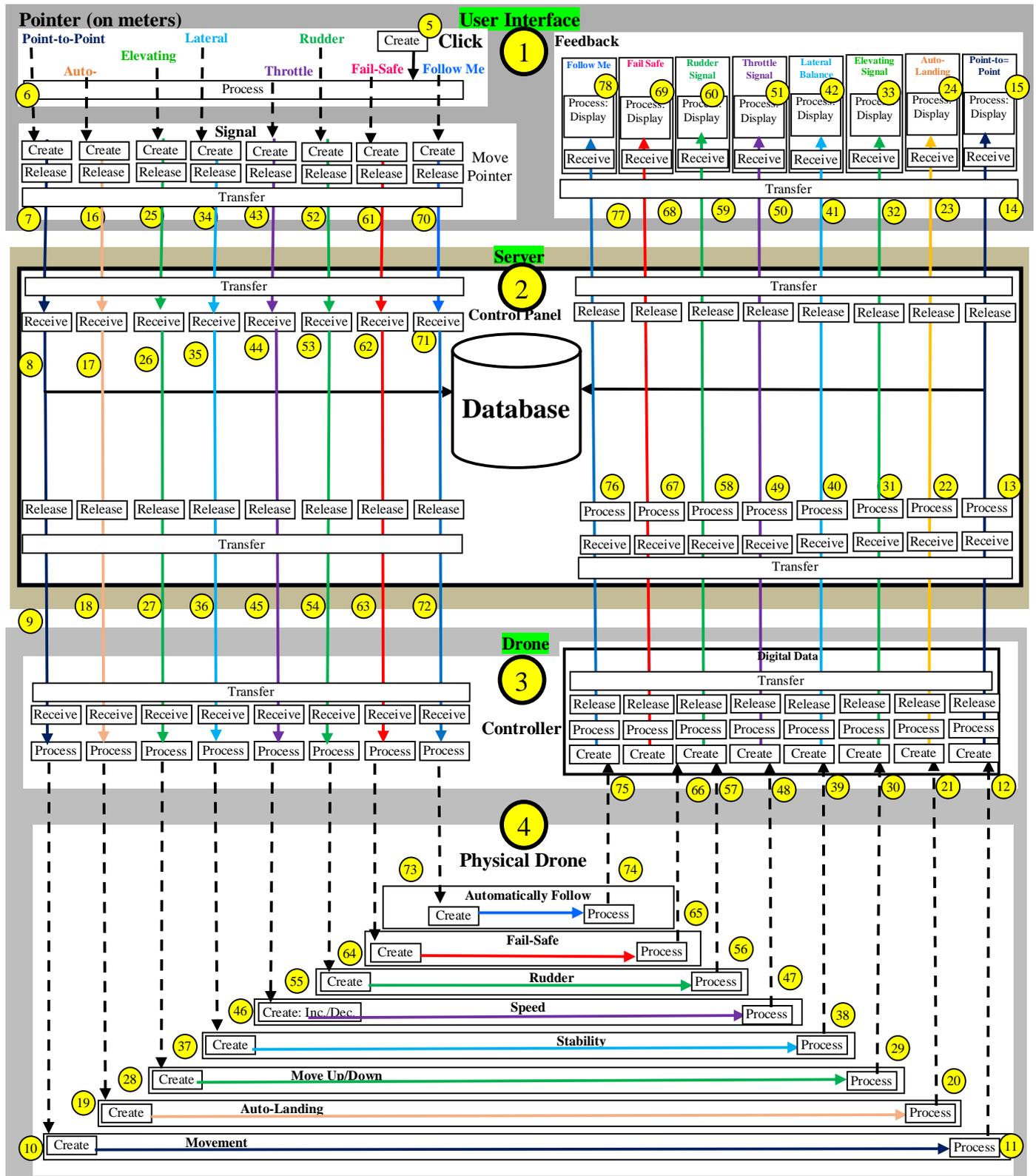


Fig. 11. The TM's Static UAV Model.

### E. Throttle Signal

This pointer affects the speed of the drone's electric motors (which increase and decrease its speed). Increasing the throttle generates more thrust. To accomplish this, a signal is created, released, and transferred (43) to the server, where it is stored (44). The signal then flows to the drone controller (45), where it is processed to increase or decrease the physical drone's speed (46). In addition, the change in movement speed (47) creates related data (48) that flow to the server (49), where they are stored. The data then flow to the UI (50), where they are displayed on the user's screen (51).

### F. Rudder Signal

This pointer involves altitude adjustments. To accomplish this, a signal is created, released, and transferred (52) to the server, where it is stored (53). The signal then flows to the drone controller (54), where it is processed to increase or decrease the physical drone's speed (55). In addition, the change in altitude (56) creates related data (57) that flow to the server (58), where they are stored. The data then flow to the UI (59), where they are presented on the user's screen (60).

### G. Fail-Safe

This procedure supports the drone in case of an error. This mode sets the conditions that the model's servos and motors revert to when it loses the transmitter's control signal. For instance, the fail-safe could automatically cause the drone to return home or to the nearest base station. To accomplish this task, a signal is created, released, and transferred (61) to the server, where it is stored (62). The signal then flows to the drone controller (63), where it is processed to move the physical drone (64). In addition, this movement (65) creates related data (66) that flow to the server (67), where they are stored. The data then flow to the UI (68), where they are displayed on the user's screen (69).

### H. Follow Me

This pointer causes the activation of follow-me mode, which gives the drone the ability to autonomously track a target without piloting. For example, a drone can be programmed to automatically follow its operator around. To accomplish this, a signal is created, released, and transferred (70) to the server, where it is stored (71). The signal then flows to the drone controller (72), where it is processed to move the physical drone in follow-me mode (73). In addition, this movement (74) creates related data (75) that flow to the server

(76), where they are stored. The data then flow to the UI (77), where they are shown on the user's screen (78).

## VII. DYNAMIC THINGING MACHINE MODEL

The space limitations and density of overlapping events do not permit us to diagram all the events for the processes in the UI. Accordingly, we show the events for only two of them: Point-to-Point and Follow Me, as shown in Fig. 12. Clicking on Point-to-Point causes the following events.

- Event 1 (E<sub>1</sub>): *Point-to-Point* is clicked on the pointer.
- Event 2 (E<sub>2</sub>): A flow signal to the server is created.
- Event 3 (E<sub>3</sub>): The signal is stored in the database.
- Event 4 (E<sub>4</sub>): The signal flows to the drone controller.
- Event 5 (E<sub>5</sub>): The physical drone begins to move.
- Event 6 (E<sub>6</sub>): The movement operation takes its course.
- Event 7 (E<sub>7</sub>): The signal is processed to trigger up or down movement in the physical drone.
- Event 8 (E<sub>8</sub>): The signal is processed to balance (stabilize) the physical drone.
- Event 9 (E<sub>9</sub>): The signal is processed to accelerate or slow the physical drone's movement.
- Event 10 (E<sub>10</sub>): The signal is processed to increase or decrease the physical drone's speed.
- Event 11 (E<sub>11</sub>): The signal is processed to trigger the physical drone's movement; instructions are sent to the flight controller to execute to this mode.
- Event 12 (E<sub>12</sub>): The signal is processed to move the physical drone and thus follow the selected target.
- Event 13 (E<sub>13</sub>): The signal is processed to trigger flows to the controller and shift from analog to digital.
- Event 14 (E<sub>14</sub>): The signal flows to the control panel.
- Event 15 (E<sub>15</sub>): The signal is stored in the database.
- Event 16 (E<sub>16</sub>): The signal flows to the server.
- Event 17 (E<sub>17</sub>): The signal flows to the UI.

Fig. 13 shows the chronology of these events.

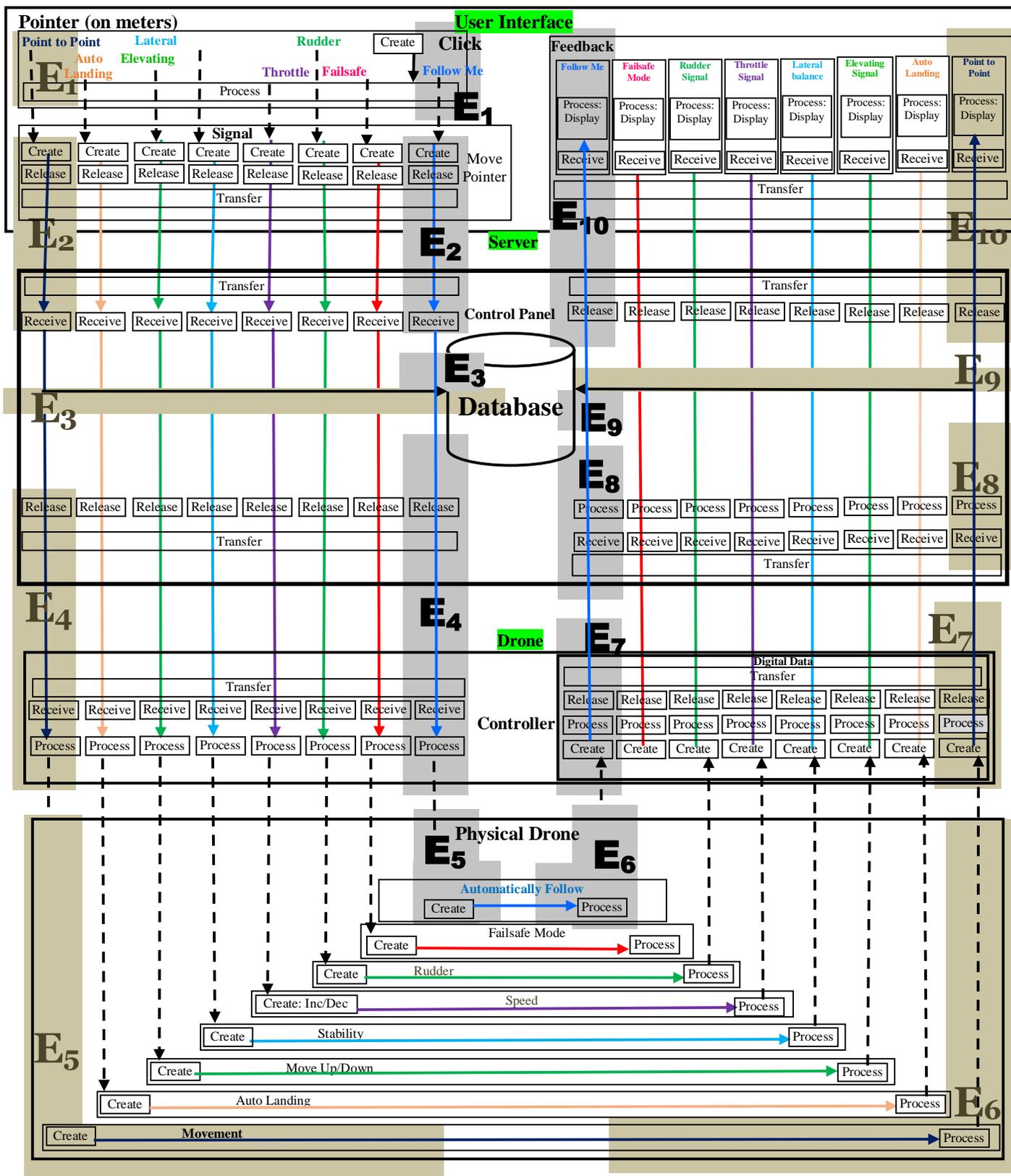


Fig. 12. Events in the TM Model of the UAV Delivery System.

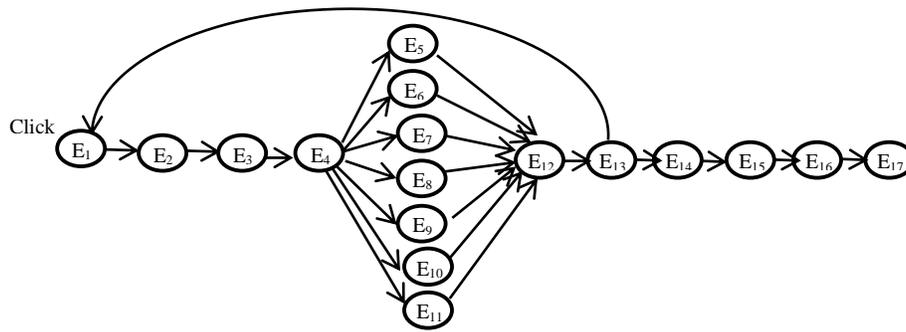


Fig. 13. The Chronology of Events in the TM Model of the UAV.

VIII. SIMULATION

The TM diagram can be used to simulate drone processes. TM events are fine-grained activities that result in the integration of a static description and a dynamic model of events. Without loss of generality, we focus on flowcharting using the simulation language Arena. In Arena, the flowchart plays an important role, and the success of a simulation depends on how well the flowchart projects the identification of events; the notion of an event is used informally. TM can assist with this process.

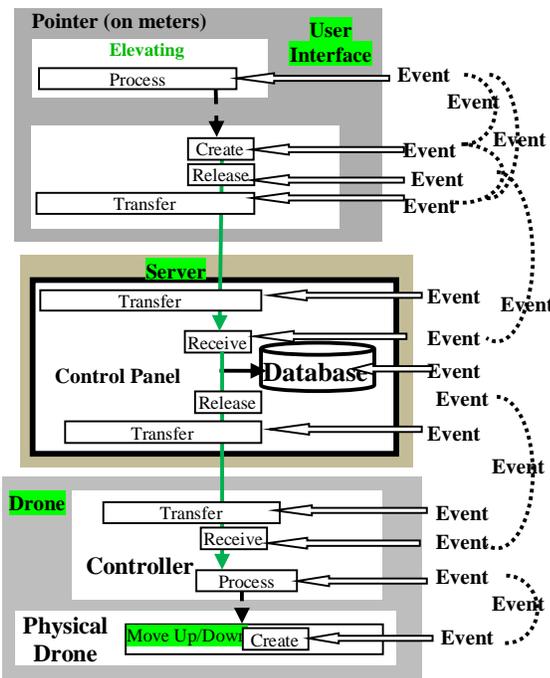


Fig. 14. Some Combinations in a Series of Elementary Elevating Events.

By contrast, a TM diagram specifies all elementary events (create, process, release, transfer, and receive). For example, Fig. 14 shows the series of elementary events involved in elevating the drone (see Fig. 11, circle 5), including the click and the movement of the drone, as well as some more complex events that can be formed from the elementary events. Many possible selections of events are possible. For example, release-transfer/transfer-receive can be considered one event in which a thing flows from one submachine to another. Alternatively, it can be considered two events: leaving (release-transfer) and arriving (transfer-receive). We start with a TM diagram and

identify event boundaries from the elementary events until we attain the required level of granularity. We are experimenting with Arena flowcharts produced by an ad hoc method as well as those developed using TM.

IX. CONCLUSION

This paper addresses the development of conceptual modeling for UAVs. We propose a flow-based specification called TM as a good vehicle in this area and demonstrate the TM methodology through a case study involving the construction of a drone.

A shortcoming of TM regards its (visual) diagramming complexity, which originated in the various machines' and submachines' completeness. The TM diagram can be simplified to whatever level of granularity is required for the original TM description. For example, Fig. 15 was produced from a static TM representation of a UAV.

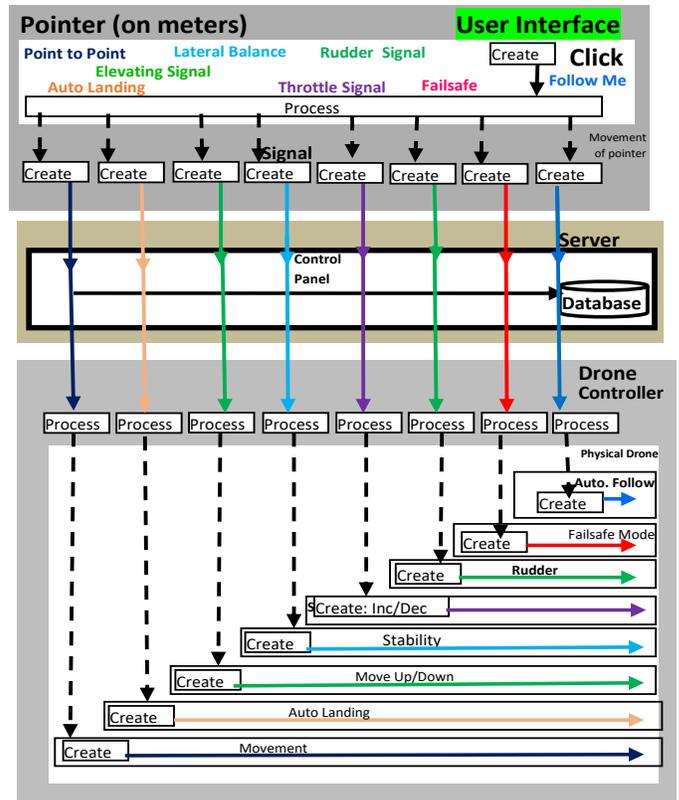


Fig. 15. Simplification of the TM's Static UAV Model (Partial).

Further research is needed to directly apply the TM methodology to more sophisticated UAV systems. Further investigation is also required to develop TM tools and supporting apparatus. Specifically, additional synchronization, constraints, and logical notation need to be superimposed on the base TM description. As mentioned previously, TM can be used as a basis for simulation.

#### REFERENCES

- [1] J. Culus, Y. Schellekens, and Y. Smeets, "A drone's eye view," PwC, Media centre, Brussels, Belgium, Report, May 2018. [Online]. Accessed August 12, 2019.
- [2] B. T. Smith, "Ethics on the fly: toward a drone-specific code of conduct for law enforcement," M.S. thesis, Dept. Abbrev., Naval Postgraduate School, Monterey, CA, USA, 2016.
- [3] S. N. Akhtar, "The use of modern tools for modelling and simulation of UAV with haptic," M.Sc.Res. thesis, Cranfield Univ., Shrivensham, U.K., 2017.
- [4] P. G. Fahlstrom and T. J. Gleason, Introduction to UAV Systems, West Sussex, U.K.: Wiley, 2012.
- [5] P. Kardasz, J. Doskocz, M. Hejduk, P. Wijekut, and H. Zarzycki, "Drones and possibilities of their using," J. Civil Environ. Eng., vol. 6, no. 3, pp. 1–7, January 2016.
- [6] E. Pastor, J. Lopez, and P. Royo, "A hardware/software architecture for UAV payload and mission control," in Proc. 25th Digit. Avionics Syst. Conf., Portland, Oregon/US: IEEE/AIAA, October 2006, pp. 5B4-1–5B4-8.
- [7] W. Sellars, "Philosophy and the scientific image of man," in Frontiers of Science and Philosophy, R. Colodny, Ed. Pittsburgh: Univ. of Pittsburgh Press, 1962, pp. 35–78.
- [8] M. Franssen, G.-J. Lokhorst, and I. van de Poel, "Philosophy of technology," in Stanford Encyclopedia of Philosophy, February 20, 2009. [Online]. Available: <https://plato.stanford.edu/entries/technology/>
- [9] Tetra Site, Technical Specification, 2019. Accessed October 5, 2019.
- [10] H. Jaakkola and B. Thalheim, "Architecture-driven modelling methodologies," in Proc. Conf. Inf. Model. Knowl. Bases XXII, A. Heimbürger et al., Eds. pp. 97–166, 2011.
- [11] A. Renaul, "A model for assessing UAV system architectures," Procedia Comput. Sci., vol. 61, pp. 160–167, 2015.
- [12] Li B., "Study on modeling of communication channel of UAV," in Proc. Int. Congr. Inf. Commun. Technol. 2017, in Procedia Comput. Sci., vol. 107, pp. 550–555, 2017.
- [13] Ministry of Transport and Ministry for Business, "Drones: Benefits study: High level findings," M.E. consulting report, Auckland, New Zealand, Jun. 2019.
- [14] P. G. Diem, N. V. Hien, and N. P. Khanh, "An object-oriented analysis and design model to implement controllers for quadrotor UAVs by specializing MDA's features with hybrid automata and real-time UML," WSEAS Transactions on Systems, vol. 12, no. 10, pp. 483–496, Oct. 2013.
- [15] OMG, Specifications of MDA, ver. 1.01, 2003. Accessed October 2, 2019.
- [16] T. Wolf et al., "Choice as a principle in network architecture," ACM Conf. Appl. Techn., Architectures, Protocols Comput. Commun., Helsinki, Finland, Aug. 13–17, pp. 105–106, 2012.
- [17] M. Heidegger, "The thing," in Poetry, Language, Thought, A. Hofstadter, Trans. New York: Harper & Row, pp. 161–184, 1975.
- [18] K. Riemer, R. B. Johnston, D. Hovorka, and M. Indulska, "Challenging the philosophical foundations of modeling organizational reality: The case of process modeling," Int. Conf. Inf. Syst., Milan, Italy, Dec. 15-18, 2013.
- [19] S. Al-Fedaghi, "Five generic processes for behaviour description in software engineering," Int. J. Comput. Sci. Inf. Secur., vol. 17, no. 7, pp. 120–131, July 2019.
- [20] S. Al-Fedaghi, "Thing/machine-s (thimacs) applied to structural description in software engineering," Int. J. Comput. Sci. Inf. Secur., vol. 17, no. 8, pp. 01–11, August 2019.
- [21] S. Al-Fedaghi, "Toward maximum grip process modeling in software engineering," Int. J. Comput. Sci. Inf. Secur., vol. 17, no. 6, pp. 8–18, June 2019.
- [22] S. Al-Fedaghi and G. Aldamkhi, "Conceptual modeling of an IP phone communication system: A case study," 18th Annu. Wireless Telecommun. Symp., New York, NY, USA, April 9–12, 2019.
- [23] S. Al-Fedaghi and O. Alsumait, "Toward a conceptual foundation for physical security: Case study of an IT department," Int. J. Saf. Secur. Eng., vol. 9, no. 2, pp. 137–156, 2019.
- [24] S. Al-Fedaghi and M. BehBehani, "Thing machine applied to information leakage," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 9, pp. 101–110, 2018d.
- [25] S. Al-Fedaghi and J. Al-Fadhli, "Modeling an unmanned aerial vehicle as a thinging machine," 5th Int. Conf. Control, Automat., and Robot., Beijing, China, April 19–22, 2019.
- [26] G. N. Fandetti, "Method of drone delivery using aircraft," 2015, uS Patent App. 14/817,356.
- [27] A. Reyna, C. Martín, J. Chen E. Solerand M. Díaz, On blockchain and its integration with IoT. Challenges and opportunities, Future Generation Computer Systems, Vol. 88, pp. 173-190, November 2018.
- [28] RedStag Fulfillment, "The future of distribution." <https://redstagfulfillment.com/the-future-of-distribution/> (accessed December 14, 2017).
- [29] RedStag Fulfillment Site, "The future of distribution Part II: Product distribution in emerging markets," Accessed December 21, 2017.
- [30] H. Shakhathreh et al., "Unmanned aerial vehicles: a survey on civil applications and key research challenges," IEEE Access 7, January 2019, Art. no. 8682048.

# Deep Learning Model for Identifying the Arabic Language Learners based on Gated Recurrent Unit Network

Seifeddine Mechti<sup>1</sup>  
MIRACL Laboratory  
Sfax University,  
Sfax, Tunisia

Roobaea Alroobaea<sup>2</sup>  
College of  
Computers & Information  
Technology  
Taif University, Taif,  
Saudi Arabia

Moez Krichen<sup>3</sup>  
Faculty of CSIT,  
AlBaha University,  
AlBaha, Saudi Arabia  
ReDCAD Laboratory,  
Sfax University,  
Sfax, Tunisia

Saeed Rubaiee<sup>4</sup>, Anas Ahmed<sup>5</sup>  
Department of  
Industrial & Systems  
Engineering,  
College of Engineering,  
University of Jeddah,  
Jeddah, Saudi Arabia

**Abstract**—This paper focuses on identifying the Arabic Language learners. The main contribution of the proposed method is to use a deep learning model based on the Gated Recurrent Unit Network (GRUN). The proposed model explores a multitude of stylistic features such as the syntax, the lexical and the n-grams ones. To the best of our awareness, the obtained results outperform those obtained by the best existing systems. Our accuracy is the best comparing with the pioneers (45% vs 41%), considering the limited data and the unavailability of accurate tools dedicated to the Arabic language.

**Keywords**—Arabic; Native Language Identification (NLI); deep learning; Gated Recurrent Unit Network (GRUN)

## I. INTRODUCTION

Technological progress and the unprecedented sharing of resources on the Internet has generated a huge number of documents on the web and especially on social networks. These documents and / or publications belong to different author profiles. Unfortunately, many Internet users do not reveal their real identity and give false information regarding their age, sex, nationality, level of education, mother tongue, etc. For this, several works have been interested in identifying the source of information.

In fact, in the commercial sector there is a need to know the age, gender, origin, and other details in order to offer potential buyers' products that are suitable to their profiles. Also, the products should be offered to them in perfect harmony with their preferences and moods. Furthermore, the origins of clients from their texts and their languages should be known. In this same framework, our work aims to detect the mother tongue of users.

Another application of mother tongue detection is the educational field. Indeed, for the learners of a given language one needs to know the level of mastery of the language in order to classify them into different learning groups corresponding to different levels of education. For example, for learners of the Arabic language, three levels of learning can be used, which are non-native learners, medium learners and native learners.

This article are interested in the detection of the mother tongue of the authors for learners of the Arabic language. This

task is part of computational linguistics. We have based on the series of experimentation on the Gated Recurrent Unit (GRU) model. Our model contributes to overcoming the limitations of RNN.

This paper is divided into five sections. Section II is given a short overview about related works. Section III discusses our Baseline approach of ANLI. Section IV presents our new deep learning approach based on GRU. Finally, concluding remarks are detailed and upcoming outlines of research are provided in Section VI.

## II. RELATED WORK

Nowadays, other languages apart from English language [1] paying attention to researchers in order to evaluate the applicability of Natural Language Interaction (NLI) methods to other languages [2].

To the best of our knowledge, Malmasi and Drass [3] and Lan and Hayato [4] focused on the Chinese language. The former research proposed a system that introduced the first expansion of Natural Language Interaction applied to non-English data using a set of features such as “n-grams”, “part-of-speech tags”, “context-free grammar production rules” and “function words”. The system found that the adoption of integrated features surpassed the employ of single features with 70.61% precision.

In [4], the authors resort to “skip-grams” as to solve “Natural Language Interaction” problem using lexical attributes built on JCLC (“Jinan Chinese Learner Corpus”). As the dimension of the “skip-gram” function increases tremendously, they decide to take as informative features “n-grams” with 10 occurrences. A simple example is proposed in Fig. 1.

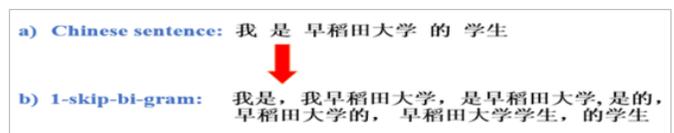


Fig. 1. An Example of a Chinese Sentence.

Unlike most of the Natural Language Interaction researches which used TF (“term frequency”) or TF-IDF (“term frequency–inverse document frequency”). The big advantage of this analysis is that careful consideration is given to assigning each function the appropriate weight. They followed the “BM25 term-weighting” process [5]. Using hierarchical “linear SVM” classifiers, their proposed method achieved a higher score with 75% per cent accuracy.

Additionally, other languages were considered such as the Finnish and the Norwegian languages in [6], [7], the portuguese [8] and the indien[9]. These works aimed to identify if the NLI methods earlier used in level two English can be effective to other languages. Their findings provided encouraging signs that the NLI strategies are applicable to other languages.

A. Arabic Native Language Identification

Arabic is generally viewed as a language that is vital and of strategic use. However, the work [10] by Malmasi and Dras is the first experience which deals with Arabic. Their objective was to examine the utility of syntactic characteristics, primarily “CFG development laws”, “Arabic function words”, and “n-grams part-of-speech”. They used a controlled approach to classification of multiple classes. As a result, these studies appeared to be effectively usable for “Arabic NLI”.

In addition, it is noteworthy that merging features resulted in a fair precision of approximately 41%. That was attributed, first, to the reason that Arabic’s morphological and syntactic diversity varies substantially from English and, on the other side, to the compact size of the dataset that is used in learning process.

B. NLI Shared Task

The growing interest in the NLI field reflected by a number of papers that have been published motivated research groups to organize shared Tasks [11] (to our knowledge this the first and the only shared task). The key goal of the mission was to further homogenise the group and support the field advance by creating a favorable framework for direct comparison of the systems.

In this task, 29 teams from different countries participated and 24 teams were elected to write papers describing their systems. These 24 teams competed across three different subtasks. The same test set of data was used for each task. Only the training data changed from a task to another. The teams developed systems trained on Data compiled from the TOEFIL11 corpus only, from External corpora and from both, respectively in the closed-task, open1-task and open2-task.

The teams were free to choose the convenient learner methods and features. Based on the report of [11], it is observed that “word”, “character” and “POS n-gram” features were the most common features (see Table I). Unsurprisingly, “Support Vector Machines” was the most used among other machine learning algorithms.

C. Gated Recurrent Unit (GRU)

This technique is similar to “Long Short-Term Memory” (LSTM) [12]. It was proposed by [13]. GRU was developed to solve the problem of short-term memory. It has two gates

TABLE I. COMMON FEATURES ADOPTED IN THE SHARED TASK.

Feature	Type	# of teams
Character N-Grams	N between 1 and 9	16
POS N-Grams	N between 1 and 5	15
Word N-grams	N between 1 and 5	18
Function N-grams		2
Syntactic Features	TSG, Dependencies, Adaptor Grammars, Productions	6

which are “update gate” and “reset gate” to control the memory flow as shown in Fig. 2. It uses the memory to store the value for a certain amount of time and at a critical point dragging that value out and utilizing it with the present state to update at a future date. To sum up, it has less tensor operations than LSTM. To some extent, it is a slight quicker to train than LSTM [14], [15].

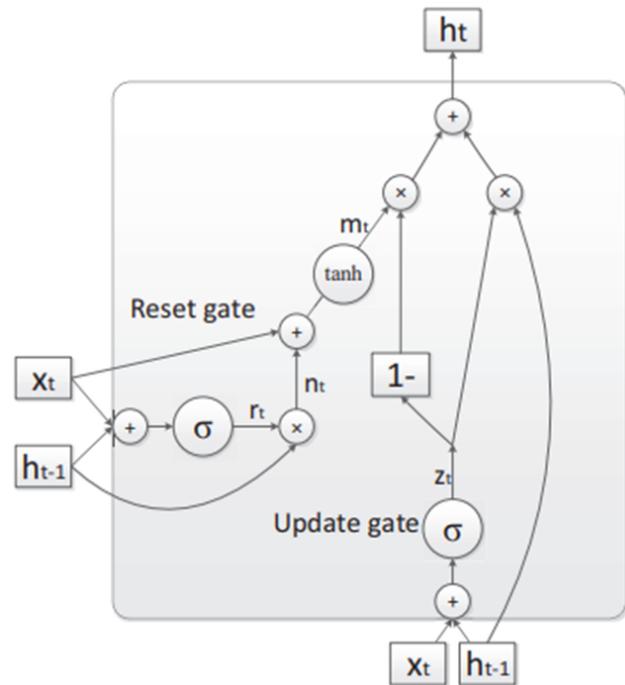


Fig. 2. Update Gate and Reset Gate in GRU [14].

III. METHODOLOGY

This paper aims to predict Arabic learners’ native language as encouraged by the work of [16]. To get the best classification model, the feature-selection step was used as it was not given great attention in most of the aforementioned studies. The best classification model refers to improve the performance and reduce the features.

Three stages were adopted in the proposed methods here to get the best classification model. The first stage is pre-processing. The text that will be used in the next stage was well-prepared. The second stage is feature-extraction. The set of features that seem to be useful will be extracted for level one learners' background discrimination. The final stage is a classification algorithm that will be applied to build the classification model. Noticeably, the last two stages are achieved by a sub-stage of feature selection.

#### A. Text Pre-Processing

This stage aims to prepare for the deep Learning (DL) stage. Indeed, the texts are written by non-Arab individuals from all over the world, who studies in the Kingdom Saudi Arabia the Arabic language. Analysis of the texts has shown that there are several inconsistencies and many errors in the corpus. Words, characters and URLs appear in the texts as shown in the example given in Fig. 3. However, notes are inserted in the texts during the transcription.

Note	Meaning
كلمة غير موجودة (indefinite word)	Indicates that the considered word does not exist in the Standard Arabic language.
معلومة شخصية محذوفة (Personal information Deleted)	Indicates that some personal data concerning the text's author was deleted (e.g. learner's name, contacts, etc.).

Fig. 3. Notes and Corresponding Meanings.

In this case, deleting these annotations can change the structure of the sentences. This explains why these notes were treated case by case by replacing them with suitable words to keep a good syntactic structure of the sentence.

In the extreme case where we do not find an appropriate word, so it was deleted. Once the text pre-processing phase is achieved (i.e. the corpus data are transformed into usable data), the texts are ready for the next phase where features will be extracted from them (as shown in Fig. 4).

#### B. Extraction and Selection of Features

Three syntactic feature categories were discovered: POS n-grams, function words and production rules. Thus, three collections of features were generated for each text. For every individual feature, frequency (TF) was calculated.

a) *Function words*: Namely empty words, In this study, 411 common Arabic function words were adopted and classified into 17 types. Fig. 5 shows examples of the Arabic function words listed by types.

b) *POS n-grams*: These features highpoint the words' linguistic class. The tagger was applied to assign the grammatical category for each word.

c) *Production Rules*: This terminology define both the syntactic class of the "words" and "sentences" structures. Fig. 6 shows some production rules extracted from the corresponding parse tree of a given sentence.

The first production rule "S  $\rightarrow$  VP | VB NP" indicates that the sentence (S) is constituted by a verb phrase (VP) or by verb phrase (VP) followed by noun phrase (NP). The second rule indicate that Verb phrase (VP) is constituted by verb followed by prepositional phrase introduced by a subordinating (SBAR) conjunction and so on. The Arabic syntactic tag set is slightly different from the tagset used for English given the major differences between the two syntactic systems. The full list of syntactic tags used in this study is detailed [17].

Since the rules extracted are often errored and not acceptable by Arabic syntax, we think of how to decide if a rule is valid or not. The solution is to compare it with an existed list of rules that we know that it is correct in advance.

Thus, we use a base of rules extracted from the Penn Arabic Treebank (ATB) by [18] which is a collection of text gathered from the Lebanese newspaper An-Nahar. While ATB texts are written by Journalists specialized in editing, we assume that they respect the Arabic syntax rules and so on we accept only rule appear in that base and throw away the rest.

For the step of feature selection, the idea was to use standard deviation to select features that contribute most to the classification. We calculated the standard deviation for each feature and sorted them in ascending order as described in the Algorithm shown in Fig. 7.  $x_{ij}$  is the weight of feature  $f_i$  in document  $doc_j$ . The idea of the Algorithm is to use standard deviation to select features that contribute most to the classification.

We then muted features which have the lesser standard deviations to pick only the most important features. This was achieved since a lower standard deviation implies that the values of the function are placed in close proximity to the mean, that is not appropriate for class discrimination. Then we were training our model utilizing the latest features sub-set. We repeated the method until we omitted no features without compromising precision. We trained the ultimate model later, with the features picked. The process is described in the following the informal algorithm shown in Fig. 8.

The algorithm starts with the full feature set and, for each step, the "p" worst features (in terms of Standard deviation) are excluded from the set. The number of removed features p is determined dynamically at the beginning of the algorithm ( $P_iM$  where M is the size of the feature set). Then, the new feature set is evaluated by applying a given classification algorithm in order to make a comparison of the performance of the new set with respect to the precedent set. The process is run repeatedly making sure that no loss in prediction performance occurred (the stop criterion is not verified).

#### C. Classification Model

Once the attributes are defined and selected, We train the final model by executing a learning algorithm. The output would be a classification model that is able to predict the native language for the response to new data.

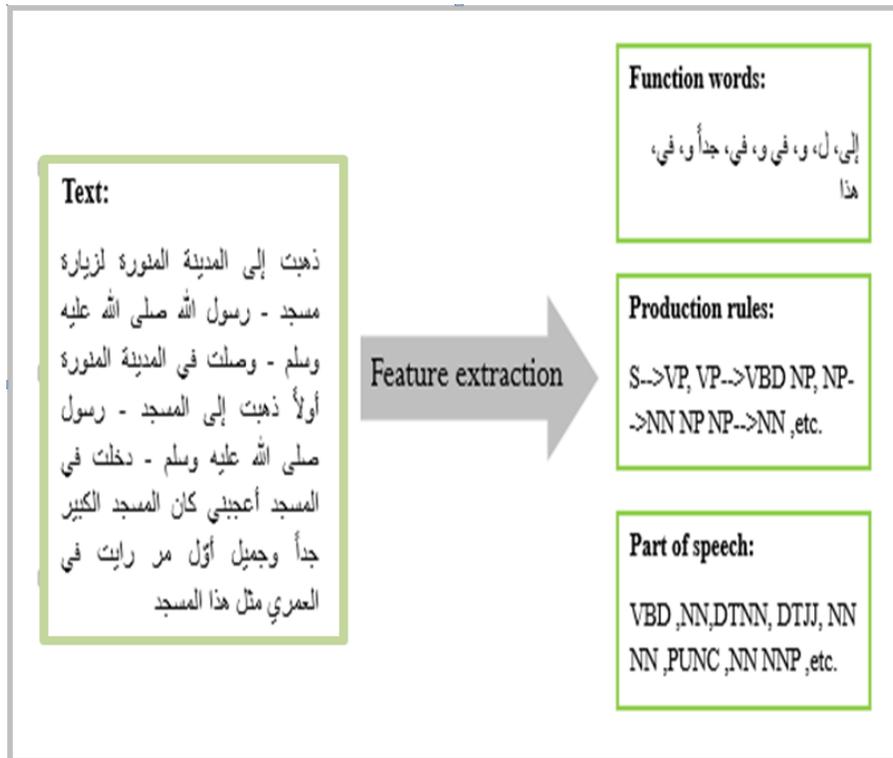


Fig. 4. Feature Sets Extraction.

Type	Examples
Linking words	<ul style="list-style-type: none"> <li>furthermore = علاوة على</li> <li>despite = بالرغم</li> <li>whereas = حيث أن</li> <li>etc.</li> </ul>
Conjunctions	<ul style="list-style-type: none"> <li>or = أو</li> <li>but/ rather = بل</li> <li>and = و</li> <li>etc.</li> </ul>
Prepositions	<ul style="list-style-type: none"> <li>from = من</li> <li>to = إلى</li> <li>in = في</li> <li>on = على</li> <li>etc.</li> </ul>

Fig. 5. Function Words

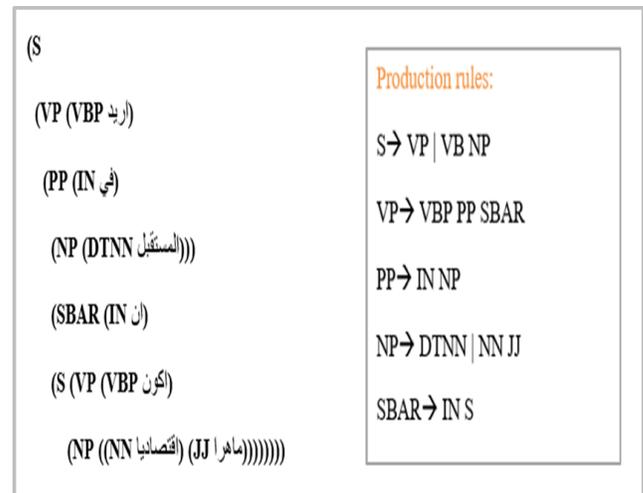


Fig. 6. Constituent Parse Tree and Grammar Production Rules.

#### IV. EXPERIMENTS

Several series of experiments were carried out on the test corpus. These experiences have been validated and evaluated by the following techniques.

One of the most popular cross-validation techniques is K-fold. It consists in dividing the data into k subsets; one subset serves as validation data and the others act as training data. The validation process is then repeated k times. This technique becomes the de facto standard for communicating the results of the NLI; therefore, we reported our experimental results under cross validation K, with k = 10.

Because our training data set is imbalanced, to test the classification model, the adoption of different performance indicators can be a useful approach to tackle this issue. Thus, we have been using three variables frequently adopted in data mining assessment to estimate the efficiency of our strategy: recall, accuracy and precision.

##### A. Data Description

Our model was trained to [19], the section of the second edition of the “Arabic learner corpus” (ALC). The above

---

**Algorithm:** Calculation of the Standard deviation of the features  
**Input:** feature set, term-document matrix weight  
**Output:** set of features sorted by standard deviation

**Begin**

- (1) Calculating the mean for each feature
- (2) Calculating the Standard deviation for each feature
- (3) Sorting the feature's sets of values in ascending order
- (4) Return the Sorted set

**End**

---

Fig. 7. Calculation of Standard Deviation.

---

**Algorithm:** Feature selection Using Standard Deviation  
**Input:** sorted feature set, a given classification algorithm (classifier) and desired number of features to exclude in each step (p)  
**Output:** subset of most confident features

**Begin**

- (1) Apply classifier using the full set
- (2) Update feature set by removing the p first features
- (3) Evaluate the new set by applying classifier using the new subset
- (4) If (stop criterion not verified) return to (2)
- (5) Return the new set

**End**

---

Fig. 8. Feature Selection.

includes texts produced by Native and non-native persons speaking 66 distinct Native language. In this trial we have included the seven top L1s with text length of 166 words in terms of text numbers. Table II shows the distribution of L1 broken down by word number and text number.

ALC texts are available in two computerized formats, TXT and XML formats. Those texts are annotated by author's native language within other metadata such as age, gender, etc. Each text has a title and content. The title specifies the topic of the text. Fig. 9 shows an example of XML text used in this study.

### B. Architecture Layers

The different layers of the GRU model are:

- **Input Layer:** In this layer, Each unit directly transfers its allocated value to the Embedding layer.

TABLE II. L1 DISTRIBUTION BY NUMBER OF TEXTS AND WORDS.

Native Language	# of Texts	# of Words
Chin	76	~ 11000
Urd	64	~ 12300
Mal	46	~ 6700
Fren	44	~ 6000
Ful	36	~ 5800
Eng	35	~ 5800
Yor	28	~ 5800
TOTAL	329	~ 52200

- **Embedding layer:** To initialize the GRU's embedding layer, we used a bag of words that were strained via a shallow neural network. This bag defines words for determining the resemblance between words by a vector. In reality, the similitude search is based on "word2vec" techniques. In reality word2vec is a two-way combination.
- **BI-GRU Layer:** There are two gates in the GRU cell: an "upgrade gate", and a "reset gate". It diminishes the 3 gates that are specified in LSTM.
- **Activation Layer:** For the hidden layers, the most recent deep learning networks used rectified linear units (ReLU). Many frameworks, such as "TF Learn" and "Tensor Flow" and allow the use of ReLUs on hidden layers simpler.
- **Drop Out Layer:** Since the size of our model is fairly large and we have a bent to implement dropout to regulate the network size and to adjust the number of hidden choices among the recurrent layers to prevent overfitting drawback.
- **Dense Layer:** Sigmoid was adopted as an activation function to complete the flow of information within the two gates created by the Bi-GRU sheet.

### C. Results using the GRU

For the GRU, we disseminate a batch size of 1000. We use an unfold dimension of 20-time steps. We apply dropout, with a 0.8 probability for the item. To clip enormous gradients that may otherwise cause drop minima, we tend to apply a gradient cap of 5. For the training, we apply 10 iterations. We run 5 algorithms: Adam, RMSprop, Adagrad, Adadelta, and SGD. Our model is trained best based on Adam optimizer with a 0.001 learning rate. For the evaluation, we based accuracy, precision and recall. The confusion matrix shown in Fig. 10

```

▼<doc ID="S024_T1_M_Pre_NNAS_W_C">
  ▼<header>
    ▼<learner_profile>
      <age>25</age>
      <gender>ذكر</gender>
      <nationality>صيني</nationality>
      <mothertongue>الصينية</mothertongue>
      <nativeness>ناطق بغير العربية</nativeness>
      <No_languages_spoken>3</No_languages_spoken>
      ...
    </learner_profile>
    ▼<text_profile>
      <genre>سردى</genre>
      <where>فى الصف</where>
      <year>1434</year>
      ...
    </text_profile>
  </header>
  ▼<text>
    <title>الرحلة إلى دمام.</title>

  ▼<text_body>
    كان لنا فرصة السفر إلى دمام ، وذلك بدأ بتسجيل أسماء الراغبين للسفر، عرفنا تاريخ السفر أثناء التسجيل، واستعدنا بعض الأشياء له قبل تاريخه، وجاء وقت السفر فجأة بدون إدراك! لأن الدراسة كانت متوترة ومليئة بتنظيم الأوقات. لذا، لم نشعر ببطء مرور الأيام، بل بلعكس، شعرنا بسرعتها! على كل حال. السفر لا شك أنه سار وممتع، وكنا مسرورين بركوب الحافلة الجامعية إلى المقصد، وهو دمام . من المعروف أن دمام مدينة جميلة سياحية تستحق الزيارة لها . لذا ، كنا نشيطين داخل الحافلة، تعجبنا بالمناظر الطبيعية الجميلة، غنينا غناء ذا ميزات محلية ، ونزلنا من الحافلة بقرب البحر، سبحنا مباشرة فى البحر بسبب درجة حرارة الجو لعالية . وتجولنا فى شوارع تلك المنطقة . ! وتترهنا فى استراحة ما هناك . وما شاء الله لهذه الرحلة

  </text_body>
</text>
</doc>

```

Fig. 9. XML Files Containing Metadata and Text Content.

displays the number of samples that were classified correctly and falsely.

## V. RESULTS AND INTERPRETATIONS

We run two sets of experiments in order to evaluate the performance of our suggested method. The first sets of experiments aim to evaluate the performance of learning algorithms (classifiers) and consequently we choose the most efficient one to be used in the next set of experiments. The second sets of experiment is dedicated to evaluate the contribution of our features set in different configurations: individually, together, with and without passing by the selection process.

We performed multiple 10-fold cross-validation experiments to test our features both separately and in combination. Table III summarizes the full classification accuracies of the different set of features both with and without using our proposed feature selection step. Malmasi and Drass [10]

Confusion matrix		Predicted class	
		Class A	Class b
True class	Class A	True Positives (TP)	False Negatives (FN)
	Class b	False Positives (FP)	True Negatives (TN)

Fig. 10. Confusion Matrix.

developed the first and the only NLI method addressed Arabic language. They report that the best accuracy is obtained using combination of syntactic features similar of that used in the current study. Our results outperform those reported by them, around 5% up in accuracy.

TABLE III. NUMBER OF FEATURES AND ACCURACY.

Features	Without feature selection		With feature selection	
	# of features	Accuracy	# of features	Accuracy
Production rules	1124	30.5 %	106	36.5 %
Function Words	17	31.0 %	11	31.0 %
POS unigrams	33	30.0 %	16	34.9 %
POS bigrams	594	35.4 %	145	38.0 %
POS trigrams	580	29.0 %	347	29.0 %
Combined	2348	41.9 %	278	45.0 %

Based on the experimental results described in the Table III, we found that removing lowest deviation features in term of standard enhanced the prediction capability of our solution. Indeed, applying our selection algorithm enable our system to obtain a gain in accuracy ranging from 3.9% (case of Production rules) to 5% (case of combined features), as well as a gain in terms of memory space: we managed to reduce the size of feature vector 10 times less than the size of the initial vector from 2348 to 278.

Fig. 11 shows the last four iterations of our feature selection algorithm (iteration #205 to iteration #208) with p=10 (i.e. the ten lowest values in term of standard deviation are removed each iteration). We can see that we reach a higher accuracy of 45% at the iteration #207 with set contains 287 features. After the 207th iteration the classification performance is dramatically decreased even when we force the algorithm to continue running (iteration #209 and iteration #210).



Fig. 11. Variation of Accuracies in the Last Iterations of Features Selection Procedure.

In the following, we will detail the result of individual and combined features after have been selected. For individual

features, function words and production rules have demonstrated their capacity to distinguish L1 learners with and 31% accuracy for function words and 36.5% for production rules. Unsparingly, the POS n-grams consistently beaten the other syntactic features with those provided in past studies. The highest precision was obtained at 38%, with n=2. We notice that while 3-grams gave 29% of fair precision, it seems that coupling POS 3-grams with other attributes do not yield good results. The general result was quite underperformed. It may be attributed to the fact that, opposed to the other feature sets, these trigrams represent redundant information. And when we used features together, we ruled it out. We integrated 278 features divided as following: 145 bigrams, 16 unigrams, 106 production rules and 11 classes of function words. This set permitted to achieve a higher Arabic NLI classification result (45%), which prove the efficacy of the use of standard deviation.

TABLE IV. L1 DISTRIBUTION BY NUMBER OF TEXTS AND WORDS.

L1	Classified AS						
	Chin	Urd	Mal	Fren	Ful	Eng	Yor
Chin	56	14	3	3	-	-	-
Urd	13	40	6	5	-	-	-
Mal	9	11	21	2	2	1	-
Fren	8	16	6	13	1	-	-
Ful	3	12	7	7	7	-	-
Eng	10	12	3	3	-	7	-
Yor	8	10	3	4	-	-	4

The confusion matrix illustrated in Table IV presents the distribution of misclassified and correctly classified samples for the different native languages. A combination of production rules, POS and function words were adopted as classification features. The performance of the different native languages is slightly spaced. In fact, the experimental results reveal that it was possible to identify Asiatic Arabic learners better than European Arabic learners. For Chinese and Urdu, we obtained an precision rate of approximately 80% while this rate was 30% for English writers and 36% for French authors. In addition, we find out that most mispredicted samples are labelled as Chinese or Urdu sample. It is probably because Chinese and Urdu, compared to the other class, are over represented in term of samples number in training set, which is attributed to the idea of unbalanced training data and its impact in the effectiveness of the classification model.

Consequent to the tow above point, it was proven that Asian languages are effectively distinguished in the context of Arabic NLI. On the other side, the two closely related European often misclassified as Asian. African languages are the hardest to distinguish and represent the higher error rate. Especially for Yoruba, only one of seven texts is correctly classified. This may be because the deficiency of training data allocated for it.

## VI. CONCLUSION

In this research work, we investigated the efficacy of language transfer to identify the first language of non-native Arabic speakers based on their text written in Arabic. In particular, we focused on the transfer related to the syntax. For this purpose, we presented a supervised method for Arabic NLI task based on syntactic features extracted automatically from text written by non-Arabic learners. Essentially, our method consisted of three steps where the input is a set of text and the output is a classification model able for predicting the class of unseen text: we started by pre-processing the text, in this step, we dealt with the inappropriate characters, words and marks by removing or replacing them depending on the case. Then texts passed to the next step where syntactic feature types were extracted. Therefore, the initial set transformed into space vector representation at final, the new text representation used as input for a deep learning algorithm that served to build the classification model. We found out that the features space is higher compared with number of simples. Indeed, it exceeded two thousands when we use all the features together. We assumed that many of them were redundant and non-informative. Based on this hypothesis we proposed an algorithm using a statistical metric (standard deviation) that enabled us to select the non-useful features. To accomplish the task we used the second version of ALC corpus. We included the seven top native languages: three Asian languages (Chinese, Urdu and Malay), two European (French and English) and two African (Fulani and Yoruba). In all we experimented using 329 texts of average 160 words per text.

It is worth pointing out that our results are promising, we outperform the state-of-art accuracy (45% vs 41%), given the issues that we faced in this study concerning the limited data and the unavailability of accurate tools dedicated to the Arabic language. Our methodology currently uses a static-learning model which adopts ALC as a dataset for training and testing. Therefore, we intend in future works to address this problem by developing new Arabic learner corpus which may be adopted to evaluate the generalisability of our method and more broadly to serve linguistic and computational research areas. Furthermore, the analysis of ALC texts showed that learners committed several errors of different types (orthography, morphology, syntax, semantics, etc.) when they express their ideas [20]. Exploitation of errors presents a perspective. Indeed, these errors reflect one of the main aspect of language transfer resulting from the difference between the learner's native language and that of Arabic.

## REFERENCES

- [1] A. K. Hultgren, "English as the language for academic publication: on equity, disadvantage and 'non-nativeness' as a red herring," *Publications*, vol. 7, no. 2, p. 31, 2019. [Online]. Available: <https://doi.org/10.3390/publications7020031>
- [2] C. S. C. Dalim, M. S. Sunar, A. Dey, and M. Billinghurst, "Using augmented reality with speech input for non-native children's language learning," *Int. J. Hum. Comput. Stud.*, vol. 134, pp. 44–64, 2020. [Online]. Available: <https://doi.org/10.1016/j.ijhcs.2019.10.002>

- [3] S. Malmasi and M. Dras, "Chinese native language identification," in *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, 2014, pp. 95–99.
- [4] W. Lan and Y. Hayato, "Robust chinese native language identification with skip-gram," in *DEIM Forum*, 2016.
- [5] L. Wang, M. Tanaka, and H. Yamana, "What is your mother tongue?: Improving chinese native language identification by cleaning noisy data and adopting bm25," in *2016 IEEE International Conference on Big Data Analysis (ICBDA)*. IEEE, 2016, pp. 1–6.
- [6] S. Malmasi and M. Dras, "Finnish native language identification," in *Proceedings of the Australasian Language Technology Association Workshop 2014*, 2014, pp. 139–144.
- [7] S. Malmasi, M. Dras, and I. Temnikova, "Norwegian native language identification," in *Proceedings of the International Conference Recent Advances in Natural Language Processing*, 2015, pp. 404–412.
- [8] I. del Río, "Native language identification on L2 portuguese," in *Computational Processing of the Portuguese Language - 14th International Conference, PROPOR 2020, Evora, Portugal, March 2-4, 2020, Proceedings*, ser. Lecture Notes in Computer Science, P. Quresma, R. Vieira, S. M. Aluísio, H. Moniz, F. Batista, and T. Gonçalves, Eds., vol. 12037. Springer, 2020, pp. 87–97. [Online]. Available: [https://doi.org/10.1007/978-3-030-41505-1\\_9](https://doi.org/10.1007/978-3-030-41505-1_9)
- [9] A. A. Chowdhury, V. S. Borkar, and G. K. Birajdar, "Indian language identification using time-frequency image textural descriptors and gwo-based feature selection," *J. Exp. Theor. Artif. Intell.*, vol. 32, no. 1, pp. 111–132, 2020. [Online]. Available: <https://doi.org/10.1080/0952813X.2019.1631392>
- [10] S. Malmasi and M. Dras, "Arabic native language identification," in *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*, 2014, pp. 180–186.
- [11] J. Tetreault, D. Blanchard, and A. Cahill, "A report on the first native language identification shared task," in *Proceedings of the eighth workshop on innovative use of NLP for building educational applications*, 2013, pp. 48–57.
- [12] F. Adeeba and S. Hussain, "Native language identification in very short utterances using bidirectional long short-term memory network," *IEEE Access*, vol. 7, pp. 17 098–17 110, 2019. [Online]. Available: <https://doi.org/10.1109/ACCESS.2019.2896453>
- [13] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.
- [14] J. Kang, W.-Q. Zhang, and J. Liu, "Gated recurrent units based hybrid acoustic models for robust speech recognition," in *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*. IEEE, 2016, pp. 1–5.
- [15] L. Jing, C. Gulcehre, J. Peurifoy, Y. Shen, M. Tegmark, M. Soljagic, and Y. Bengio, "Gated orthogonal recurrent units: On learning to forget," *Neural computation*, vol. 31, no. 4, pp. 765–783, 2019.
- [16] S.-M. J. Wong and M. Dras, "Exploiting parse structures for native language identification," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011, pp. 1600–1610.
- [17] A. Bies and M. Maamouri, "Penn arabic treebank guidelines," in *Proceedings of the Conference on Arabic language resources and tools*, 2003, pp. 466–467.
- [18] M. Maamouri, A. Bies, T. Buckwalter, and W. Mekki, "The penn arabic treebank: Building a large-scale annotated arabic corpus," in *NEMLAR conference on Arabic language resources and tools*, vol. 27. Cairo, 2004, pp. 466–467.
- [19] A. Alfaifi, E. Atwell, and I. Hedaya, "Arabic learner corpus (alc) v2: a new written and spoken corpus of arabic learners," in *Proceedings of Learner Corpus Studies in Asia and the World 2014*, vol. 2. Kobe International Communication Center, 2014, pp. 77–89.
- [20] C. C. Aggarwal and C. Zhai, "A survey of text classification algorithms," in *Mining text data*. Springer, 2012, pp. 163–222.

# Quantifying Feature Importance for Detecting Depression using Random Forest

Hatoon AlSagri<sup>1</sup>

Information Systems Department<sup>1</sup>  
College of Computer and Information Sciences  
Al Imam Mohammad Ibn Saud Islamic  
University Riyadh, Saudi Arabia

Mourad Ykhlef<sup>2</sup>

Information Systems Department<sup>1,2</sup>  
College of Computer and Information Sciences  
King Saud University  
Riyadh, Saudi Arabia

**Abstract**—Feature selection based on importance is a fundamental step in machine learning models because it serves as a vital technique to orient the use of variables to what is most efficient and effective for a given machine learning model. In this study, an explainable machine learning model based on Random forest, is built to address the problem of identification of depression level for Twitter users. This model reflects its transparency through calculating its feature importance. There are several techniques to quantify the importance of features. However, in this study, random forest is used as both a classifier, which has over-performing aspects over many classifiers such as decision trees, and a method for weighting the input features as their importance imply. In this study, the importance of features is measured using different techniques including random forest, and the results of these techniques are compared. Furthermore, feature importance uses the concept of weighting the input variables inside a complete system for recommending a solution for depressed persons. The experimental results confirm the superiority of random forest over other classifiers using three different methods for measuring the features importance. The accuracy of random forest classification reached 84.7%, and the importance of features increased the classifier accuracy to 84.9%.

**Keywords**—Machine learning; random forest; feature selection; feature importance; depression; emotions; twitter

## I. INTRODUCTION

Depression is a leading cause of disability worldwide and a common mental illness. Globally, more than 300 million people are estimated to suffer from depression every year. Face-to-face clinical diagnose is need to diagnose depression but 70% of the patients would not consult a doctor when they are at early stages of depression. This might cause patients to reach advance stages in their condition [1].

Several studies have reported that the diagnosis of mental illnesses has increased because of the use of social media platforms [2] [2], and these mental illnesses are one of the leading causes of disability and among the most of the devastating diseases that individuals suffer from worldwide according to the World Health Organization [3], [4], [5]. Therefore, to detect the users at risk for early referral to psychological assistance and treatment, machine learning algorithms have been employed.

Now-a-days, the data collected from millions of Internet of Things (IoT) devices, sensors, social media, etc. enables extremely enriched datasets. Although this is beneficial for

machine learning researchers, this makes the data high dimensional it is quite common for datasets to have hundreds of features or more in most of the cases. Therefore, feature selection is an extremely vital process in the machine learning project lifecycle. Feature selections methods help reduce the dimensions without much loss of the total information. In addition, they help in understanding the features and their importance

Previously published papers have demonstrated that exploiting irrelevant features, along with the redundant ones, can impact the accuracy of classification significantly [2], [3], [4]. Considering feature selection as a major step in any machine learning algorithm, it contains a step for measuring feature importance. Therefore, an effective feature selection technique that relies on computing the importance of features and remove irrelevant features those that may cause no impact or negative impact [5].

Prominent perspective to feature selection besides enhancing the accuracy, is weighting the features or in other words “feature importance.” These weights could be exploited as weighting factors in further steps of recommending a remedy via recommendation techniques. Features’ importance represents the statistical significance of each feature and to what extent it contributes to the model.

Random Forest (RF), among other machine learning algorithms, has been an excellent tool to learn feature representations [6], [7] because of its robust classification power and easily interpretable learning mechanism [8]. Features’ importance can be estimated using different measures after being computed using RF. In this study, we apply RF as a classifier to detect depressed Twitter users with respect to features extracted from the users’ Twitter content and activity. RF has proved to have an accuracy higher than those of the other classifiers namely decision tree (DT), Naïve Bayes (NB), and support vector machine (SVM) (kernel and linear) where they were implemented and tested and gained results of 82% for the SVM linear as the highest accuracy among the others. SVM, DT, and NB presented new features that increased the accuracy of identifying depressed users. By applying RF to the same data, we could find features’ importance using three feature importance measures: overall, permutation, and tree interpreter feature importance measures. We were able to conclude that Tree interpreter feature importance measure proved to have the highest accuracy results when RF was recomputed after removing the least important features. In

addition, when the highest important features were removed, the accuracy of the classifier decreased significantly, proving the importance of these features.

Main Contributions of this paper can be summarized as:

- Applying RF to the RRACF model to classify depressed Twitter users more accurately
- Tree interpreter feature importance measure concludes best results of feature importance that has higher effect on classification accuracy.

The remainder of the paper is organized as follows. A literature review is provided in Section 2. Section 3 presents the background of RF and feature importance. Section 4 details the methodology used in this study. Section 5 describes the experiments and results. Finally, Section 6 outlines the conclusions of the study.

## II. LITERATURE REVIEW

Efforts to detect mental illness and more specifically depression have increased gradually with the increase in social media usage [9], [10]. Guntuku et al. [11] indicated that tweets containing negative emotional sentiments are posted by depressed Twitter users more than by healthy users

Various studies have used different classifiers to detect depression and other mental illnesses. For clinical outcome prediction using gene expression data, Kong and Yu [8] presented a new classifier, where RF is integrated with deep neural network, and demonstrated that the accuracy is higher compared to those of the other classification models using simulation experiments.

Jotheeswaran and Koteeswaran [12] proved the efficiency of RF on a system developed for emotion detection, knowledge transformation, and predictive analysis using a Twitter dataset. From the experimental results, they concluded that the decision forest-based feature extraction increases the precision of classifier in contrast to decision tree-based feature selection [12].

Reece et al. [9] found that the computational analysis of Twitter data can be used to detect major changes in individual psychology. They extracted predictive features from users' tweets and built models with supervised learning classifiers using these features. The classifiers were trained to distinguish between depression and post-traumatic stress disorder (PTSD) in affected and healthy users.

The 1200tree RF outperformed other classifiers by exhibiting accuracy results higher than those of the classifiers used by Mitchell et al. [13] and Choudhury et al. [14] in depression classification reaching (0.866) and by Taubman-Ben-Ari et al. [15] and Nadeem [16] in PTSD classification reaching (0.934) [9].

Sau et al. [17] conducted a study to predict anxiety and depression among geriatric population and concluded that the RF algorithm delivers the best results with a predictive accuracy of 90% compared to the other machine learning classifiers.

TABLE I. TAXONOMY OF RANDOM FOREST APPROACHES FOR FEATURE SELECTION AND DEPRESSION

Author	Technique	Data	Assessment
Mowery et al.,[18]	RF + DT	Twitter	Depression symptoms
Kong and Yu,[8]	RF + neural network	gene expression data	Feature representation for ranking gene importance
Reece et al.,[9]	RF	Twitter	Rank predictive features Depression and PTSD
Sau et al. ,[17]	RF	geriatric patients evaluated for depression and anxiety	Predicting depression and anxiety

Mowery et al. [18] demonstrated that the machine learning algorithms used with Twitter data improved precision in detecting symptoms of depression compared to the use of keywords alone. Decision trees and RF resulted in a higher precision than that achieved by other machine learning algorithms.

Table I indicates approaches of RF for feature selection and depression. Mowery et al. [18] showed that feature representation increased classification of depressed people. Also, Kong & Yu [8] indicated that using RF to represent features fed to deep neural network increased the accuracy of the system.

Similarly, our study uses RF to find important features but using feature importance measures that up to our knowledge, has not been introduced for detection of depression.

## III. RANDOM FOREST AND FEATURE IMPORTANCE BACKGROUND

### A. Random Forest (RF)

RF is an ensemble learning classification algorithm developed from multiple sub-decision trees [19]. The sub-decision trees are built using bagging and feature randomness to create an uncorrelated forest of trees that have a higher accuracy in prediction than that of any individual tree [20]. Bagging is a common ensemble method that uses bootstrap sampling. Using the bagging procedure, a number of decision trees are generated from the original sample set through bootstrap sampling, and the features that are randomly selected from the original set are used for partitioning at each node [6], [21]. Node is an elementary unit in any tree based algorithm. RF reduces the likelihood of over fitting generated during the use of single decision tree model [5]. In addition, the use of bootstrap sampling helps produce an optimal generalization ability and a higher accuracy classification model [19].

### B. Features' Importance

Along with improving the accuracy that has been shown in a majority of RF studies, RF provides feature importance measures as one of its useful derivatives that has contributed to its popularity [22], [23], [24], [25], [26]. Feature importance measures are the Overall, Permutation, and Tree Interpreter feature importance [25], [26].

**Overall feature importance** is calculated as the decrease in node impurity weighted by the probability of reaching that node. The node impurity should be decreased since we are going deeper into tree levels, so the impact of node can be

objectively quantified by the drop of impurity through the node. Gini impurity is calculated for each node where it is possible to calculate the node probability based on the number of samples that reach the node divided by the total number. In this case, higher values correspond to more important features. Overall feature importance starts by:

- 1) Calculating nodes importance  $n_j$  of node  $j$  for every decision tree, using the following equation:

$$n_j = W_j C_j - W_{left(j)} C_{left(j)} - W_{right(j)} C_{right(j)} \quad (1)$$

where  $W_j$  is the node  $j$  reachability probability and  $C_j$  is Gini impurity of the node. The same is for the *right* node and *left* node children of node  $j$ .

- 2) The importance of each feature (F) in the tree is calculated using Eq. 2, where  $m$  is total number of nodes:

$$F(j) = \frac{n_j}{\sum_{i=1}^m n_i} \quad (2)$$

- 3) The importance for each feature in RF (collection of  $k$  Trees) is calculated using the following equation:

$$\text{Feature Importance } (i) = \frac{\sum_{j=1}^m F(j)}{k} \quad (3)$$

**Permutation Features Importance:** starts by training the baseline model and recording the score by evaluating the validation set or training set.

For all features in features set, do:

- 1) Re-shuffle one feature values in the evaluated dataset.
- 2) Re-pass the dataset to the model and re-calculate the metric for the modified dataset.
- 3) The feature importance is computed as the difference between the benchmark score and the score from the permuted dataset.

**Tree Interpreter Features Importance:** begins with training the baseline model and recording the score by evaluating the validation set or training set. For all features in features set do:

- 1) Drop the node (feature)
- 2) Re-pass the dataset to the model and re-calculate the metric for the modified dataset.
- 3) The feature importance is computed as the difference between the benchmark score and the score from the modified dataset.

#### IV. METHODOLOGY

In this study, we focus on identifying the importance of features that help detect depression from users' accounts including both tweets' content and activity. The model contains different hyper parameters such as number of trees, depth, validation set, etc. The optimal combination among these hyper parameters has been found through executing exhaustive grid search. The system depression detection using activity and content features-random forest (DDACF-RF) proposed is in Fig. 1. This system uses RF for classifying users' mental conditions and identifying the importance of features. Data preparation, feature

extraction, and classification tasks are performed using various R packages, and in R version 3.3 [27], they are performed using Rstudio IDE [28]. The RF classifier is trained using 10-fold cross validation, each contains both training and testing set, to avoid over fitting, and it is then tested on a held-out test set. Initially, all tweets from the accounts of depressed and non-depressed users are retrieved along with their information and activities such as number of followers, number of following, and total number of posts. Next, text preprocessing is applied to all the documents through tokenization, normalization, and stemming which is done through splitting words, removing punctuation and returning word to its stem. Then, a document term matrix (DTM), which designates the frequency of words in each tweet, is created for each account. The weights of words are measured using Term Frequency-Inverse Document Frequency (TF-IDF). The features applied on the DTM are then merged with the account measures extracted from the social network and user activities as illustrated in Fig. 3. Finally, the results of the merge are treated as independent variables in the RF classification algorithm to predict whether a user is depressed or not. Fig. 1 illustrates the DDACF-RF classification model.

Three different feature importance measures are applied to find the best importance measure to weigh the features. Diaz-Uriarte and Alvarez de Andres strategy is later applied to conclude the best feature importance measure among the three [29]. Diaz-Uriarte and Alvarez de Andres strategy depended on computing RF, then removing 20% of the most important features, and then computing RF again [29], [30]. In this study, we removed the most and least important features and then recomputed RF in both cases. Using the three different feature importance measures discussed previously, the importance of features was calculated; then, the most and least important features were removed independently and RF was recomputed.

##### A. Data Collection

This study dataset concentrates on Twitter users who suffer from depression. Using a regular expression ("diagnosed with depression"), self-reported tweets are collected from Twitter. All Tweets are chosen to be in English and gathered in May-July 2018. Candidate users are filtered manually. These tweets are then processed by a human annotator to certify that the users are revealing their own depression and not talking about someone else. The manual labeling is done unanimously between two different psychologists. If any case has conflicts between them, it has been eliminated from final dataset.

Later, all their recent tweets are continuously crawled using the Twitter Search API. Total number of 500 users were collected with more than 1M tweets, 334 users were classified as depressed. For each user, up to 3000 of their most recent public tweets are included in the dataset, and each user is isolated from the others. Note that this 3000-tweet limit is derived from Twitter's archival policies [31]. Non-depressed users are collected randomly and checked manually to ensure that they have never posted any tweet containing the character string "depress". In an effort to minimize the noisy and unreliable data, users with fewer than five Twitter posts are excluded.

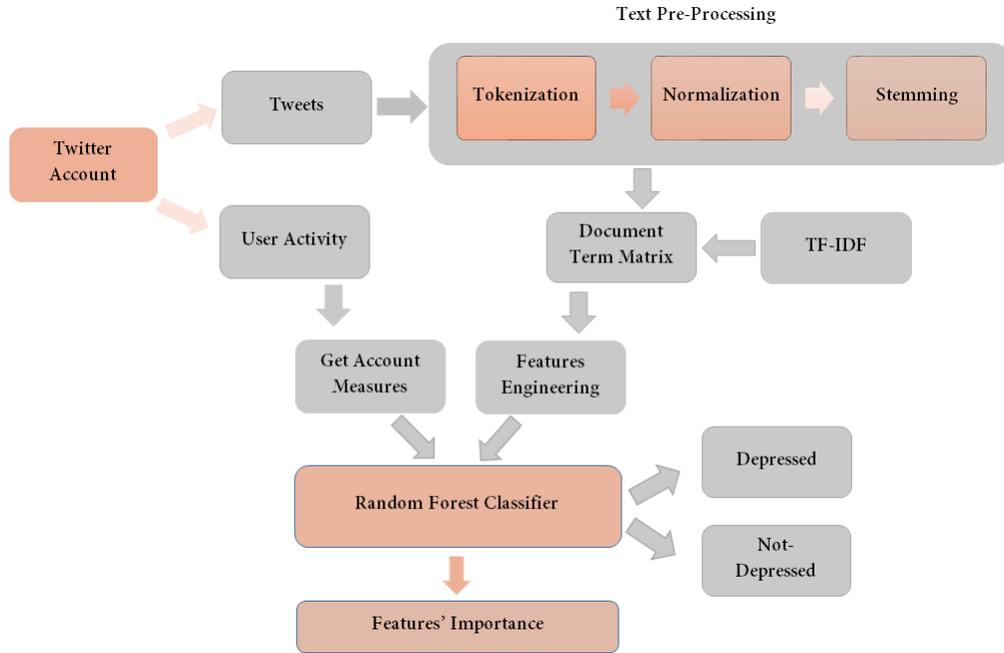


Fig. 1. DDACF-RF Classification Model

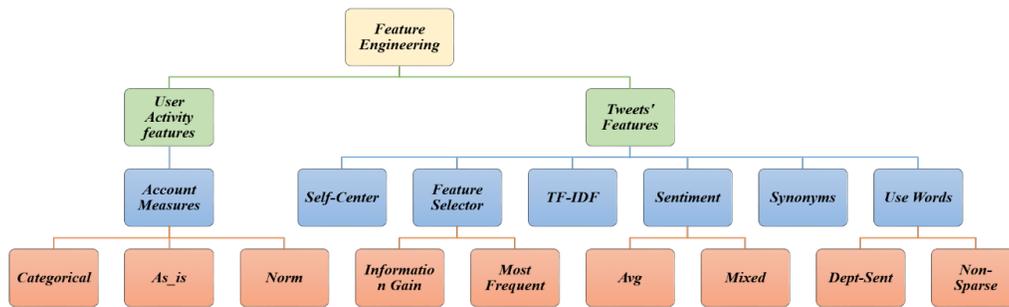


Fig. 2. Visualization of Features used in the Study

### B. Feature Engineering

In machine learning, feature engineering is referred to as “the process of using domain knowledge of the data to create features that can be used by machine learning algorithms to find patterns” [10]. The information that are recognized by machine learning and might be beneficial for prediction are extracted by generating the features [10].

The activity histories and tweets of Twitter users are used to extract various features reaching more than 150 thousands input features (raw features) containing words from tweets and account activities Fig. 3. This information undergoes preprocessing Fig. 1 before the engineered features are obtained, and once the engineered features are obtained Fig. 4, they are computed for both the training and test sets. Fig. 4 shows the features obtained from the tweets and activities of user accounts. These features are used as the variables for the classification model. Table II lists the features and their

possible values used for the classification model, Where T (true) and F (false) for possible values indicate the use of this feature or not. For example, when the possible value for TF-IDF is T meaning TF-IDF is used for the experiment and if it’s F that means word frequency is used instead.

### C. Self-Center

Previous studies have shown that first-person pronouns are useful predictors of depression. De Choudhury and Jamil [32], [10] indicated that the use of singular pronouns in comparison to second- and third-person pronouns is also an indicator of depression. Thus, we skip removing the first-person pronouns with other stop words in the normalization step in the proposed classification model to increase the efficiency of the classification algorithm.



Fig. 3. User Activity Features Extracted from User Account

TABLE II. DESCRIPTION OF FEATURES AND THEIR POSSIBLE VALUES

Features	Possible Values	Description
Self-Center	T	Use first-person pronouns.
	F	Remove all stop words associated with the first-person pronouns.
TF-IDF	T	Determine the relative frequency of words in a specific document compared to the inverse proportion of that word over the entire document corpus.
	F	Use word frequency.
Feature Selector	Information gain (IG)	· Measures the number of bits of information obtained for category prediction by determining the presence or absence of a term in a document. · Words are selected according to the higher IG.
	Most Frequent	Selects the most frequent words according to the words' higher frequency.
Sentiment	Avg	For each user, sentiment is calculated for each tweet using sentence sentiment, and then, the average of all tweets is calculated.
	Mixed	Selects a higher sentiment for sentences that are negative or positive with a hidden negative indication.
Use-Words	Dept-Sent	Sentiment words—positive and negative words—extracted from depressed user's tweets.
	Non-Sparse	Words with zeros more than 95% are removed.
Account Measures	As-is	User activities are taken as they are (number of posts, average number of posts a day, time of posts, number of replies, number of mentions, etc.).
	Norm	Activities are normalized, and average is calculated according to the number of user posts.
	Categorical	Activities are categorized according to 4 quartiles (low, below average, average, and high).
Synonyms	T	Words in the matrix are grouped, and the frequency is added based on their synonyms.
	F	Words are used as they are without reducing them based on their synonyms.

D. Feature Selector

For selecting features there are two possible values, either Most-frequent which select the most frequent words according to the words' higher frequency or Information gain. Inspired by Prieto et al. [21], information gain (IG) is added as a feature selector for the model. Prieto et al. [21] used IG to reduce features that improve the classification of depressed users, and reduced the time needed to generate the model. IG is used in machine learning as a term for goodness criterion.

E. Sentiment

Sentence sentiment is used for each tweet in the user's account, then the average of all tweets' sentiment is calculated and this is the Avg feature. Mixed feature calculates sentence sentiment for sentences that are either negative or positive but have hidden negative indication.

F. Use Words

This feature has two possible values, either non-sparse meaning non-sparse words are used and sparse words having more than 95% zeros are removed, or Dept\_Sent. Depression Sentiment (Dept\_Sent) is a feature, inspired by De Choudhury et al. [32], concentrates on depressed users' sentiment words. From tweets crawled for this study, sentiment words, positive and negative, are extracted from depressed users' tweets and put into files and all other words are removed for all users. The exploited feature in this study, Dept\_Sent, is distinguished by the fact that it does not use static lexicon words for representing depression. Dept\_Sent generalizes the depression lexicon and can be extended easily.

G. Account Measures

Tsugawa et al. [31] showed that features obtained from user activities can be used to predict user depression with 69% accuracy. In addition, De Choudhury [32] used features obtained from the records of individual user activities on Twitter to identify depressed users. Tsugawa et al. and Del Vicario et al. [31], [33] indicated that the more a user is active, the higher is his/her tendency to express negative emotion when commenting, which will help indicate whether the user is depressed.

As a result, aggregated features are used in this paper to help detect depressed users on Twitter. Activities extracted from each user account such as retweets, mentions, etc. used in this study are shown in Fig. 2.

Three different possible values for this feature (As\_is, Norm, Categorical). As\_is uses user's activities as it is while Norm uses the activities after calculating the average according to the number of user's posts. Categorical is a new feature that has been introduced uniquely in this study. It relies on categorizing activities of each user into four types (low, below average, average, and high), whose delimiters are defined using percentile values from quartile distribution (Q1, Q2, and Q3).

H. Synonyms

Tsugawa et al. [31] used the bag-of-words approach to reduce the number of words and found that it helped increase

TABLE III. DISTRIBUTION OF POSSIBLE VALUES OF EACH FEATURE

Feature	Possible Value (v)	Distribution p(v)
Self-Center	T	0.54
	F	0.45
TF-IDF	T	0.53
	F	0.46
Feature Selector	Information gain (IG)	0.33
	Most Frequent	0.33
	None	0.34
Sentiment	Avg	0.33
	Mixed	0.33
	None	0.34
Use-Words	Dept-Sent	0.45
	Non-Sparse	0.54
Account Measures	As-is	0.34
	Norm	0.33
	Categorical	0.33
Synonyms	T	0.09
	F	0.91

the accuracy. This feature reduces the number of words in the matrix by finding similar words and adding frequencies of synonyms, using Word Net.

This will make the word stronger for detecting depression and reduce the number of words in the corpus, thus decreasing the computation time.

Tree based methods have been picked for this study due to the categorical nature of the features Table III.

## V. EXPERIMENTAL RESULTS

### A. Results

This study was conducted on all possible combinations of feature values, using RF classifier. The expected labels for any training/testing sample are depressed/not depressed. Feature importance, used to find the features that mostly help increase the classification accuracy and determine the user’s mental condition, was an important result of the study.

Feature engineering used with NB, DT, and SVM used for detecting depressed users proved that utilizing a rich, diverse, and discriminating feature set that contains both tweet text and behavioral trends of different users helped increase the classification accuracy.

For that this study follows the same experimental steps and proves that the conclusion evaluation metrics increased when new features were added.

TABLE IV. RESULTS OF RANDOM FOREST CLASSIFICATION MODEL EXPERIMENTS

Features	Accuracy %	Precision	Recall	f-measure	RF tree
Initial features	67.8	0.36	0.615	0.457	2000
Dept-Sent	69.5	0.38	0.615	0.470	2000
Dept-Sent +Categ	72.9	0.45	0.642	0.529	2000
Dept-Sent +Categ +Synonyms	84.7	0.52	0.9	0.667	2000

TABLE V. COMPARISON OF RF RESULTS WITH THOSE OF OTHER CLASSIFIERS

Classifier	Accuracy %	Precision	Recall	f-measure
DT	77.5	0.65	0.59	0.619
NB	80	0.65	0.81	0.723
SVM-L	82.5	0.74	0.85	0.791
SVM-R	77.5	0.71	0.63	0.667
RF	84.7	0.53	0.9	0.667

### Metrics

### Formula

$$\text{Accuracy } Acc = \frac{\text{truepositives} + \text{truenegatives}}{\text{truepositives} + \text{truenegatives} + \text{falsepositives} + \text{falsenegatives}}$$

$$\text{Precision } P = \frac{\text{truepositives}}{\text{truepositives} + \text{falsepositives}}$$

$$\text{Recall } R = \frac{\text{truepositives}}{\text{truepositives} + \text{falsenegatives}}$$

$$\text{F-measure } F1 = \frac{2 * P * R}{P + R}$$

The experiments are compared with respect to four metrics, namely accuracy, precision, recall, and f-measure. All the experiments used “first-person pronouns” and “TF-IDF” that have already been proven to discriminant for depression identification [32]. In addition, considering “InfoGain” as the feature selector and “mixed” as the sentiment feature, the results were obtained for the first experiment. Further, in the second experiment “Dept-Sent” was added as feature along with features from first experiment. “Categorical” as the account measures feature was added in the third experiment, and “synonyms” in the last experiment. As a result, we observed an increase in all evaluation measures, where the accuracy reached 84.7% and recall was 0.9 as shown in Table IV.

Table V reveals the increase in accuracy and recall obtained using RF when compared to the other classifiers used in our previous work[34].

After training the RF, the importance of features was found to have a significant impact on the outcome values. Feature importance measures help find each feature’s importance as a measure by which the accuracy is decreased when that feature is removed and vice versa—by which the accuracy is increased when that feature is included. Fig. 4 shows the most important

TABLE VI. EFFECT OF THE REMOVAL OF MOST AND LEAST IMPORTANT FEATURES ON THE CLASSIFIER RESULTS

Features Importance Measure	Removing most important features	Removing least important features
Overall	78.1	83.81
Permutation	77.85	83.29
Tree Interpreter	72.9	84.908

features developed from the model.

We can conclude that the account measures (retweets, hash tags, ...) and the words extracted from users' contents have great significance on the detection of depressed users as their importance indicates. It's noticeable that the number of retweets appears with high importance in identification of the depression level. However, it can be explained through that as much time the user is online on Twitter or having more interactions reflects how much he is disconnected in reality. The three different feature importance measures resulted in different outcomes. These measures were then validated to obtain the best way of calculating feature's importance using Diaz-Uriarte and Alvarez de Andres strategy [29], [30].

The five most and least important features were removed independently and RF was recomputed. For example, the overall feature importance measure was used to compute RF. From the outcomes of overall feature importance, the five most important features were removed and RF was recomputed, and then, the five least important features were removed and RF was recomputed. Same strategy was repeated for the permutation and tree interpreter feature importance measures. On applying the strategy to the sample with an accuracy of (84.7), we found that the tree interpreter feature importance exhibited the highest accuracy (84.908) when the five least important features were removed and exhibited the least accuracy (72.9) when the five most important features were removed. Results of all feature importance measures are summarized in Table VI. shows the increase in accuracy of the classifier results after removing the least important features, which are less significant to the model. In addition, it shows the decrease in accuracy of the classifier results when the five most important features, demonstrating the importance of these features to the model efficiency, are removed. From the results, we can observe that the tree interpreter importance measure exhibits the highest results.

Table VI results show that the higher the decrease in accuracy reveals that the features removed were more important. It shows that tree interpreter found the most important features which caused more decrease in the accuracy than the overall and permutation importance measures. Also, removing the least important features increase the accuracy showing that tree importance measure was able to find the least important features that needed to be removed to get better classification accuracy.

The increase in accuracy was very small which was sufficient for our study to find and validate the best and most representative feature importance measure aiming to find a quantitative method to weight features. This method is needed in future work and will be employed to find a remedy for depressed Twitter users.

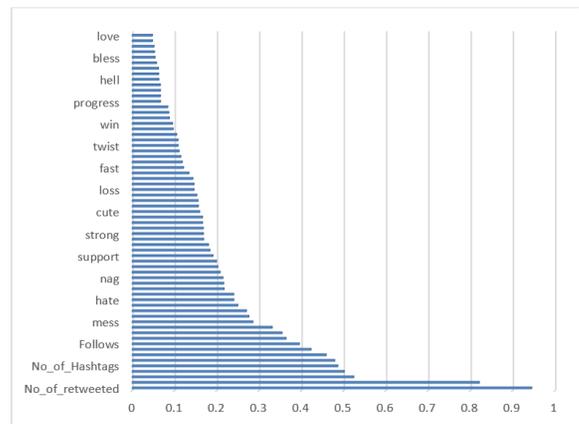


Fig. 4. Importance of Features

## VI. CONCLUSION

RF has proven to be an efficient classifier with respect to DT, NB, SVM-L, and SVM-R. In addition, it offers feature importance as an average gain achieved during forest construction. The feature importance revealed the features that do not add value to the classifier's performance and helped increase the accuracy. The uniqueness of this study was indicated in the different importance measures used, where the tree interpreter importance measure outperformed the other importance measures. The application of importance measures to the features extracted from both tweets and activities of user accounts helped classify the depressed users in the dataset more accurately.

Results of this study prove the benefit of feature importance in obtaining the best solution for depressed people and for mentally ill people in general. In future study, feature importance can be used to obtain the values of features that increase the efficiency of any model. In addition, the least important features that decrease the productivity and increase the time elapsed to obtain desired results can be eliminated in early stages of the study.

## ACKNOWLEDGMENTS

The authors would like to thank the Deanship of scientific research for funding and supporting this research through initiative of DSR Graduate Students Research Support (GSR) at King Saud University. They would also like to thank Dr. Fathy Mostafa from Ain Shams University Specialized Hospital and Dr. Afnan Alwabili from King Faisal Specialist Hospital for their grateful assistance in certifying that Twitter users in the dataset are practicing.

## REFERENCES

- [1] G. Shen, J. Jia, L. Nie, F. Feng, C. Zhang, T. Hu, T.-S. Chua, and W. Zhu, "Depression detection via harvesting social media: A multimodal dictionary learning solution." in *IJCAI*, 2017, pp. 3838–3844.
- [2] D. Chutia, D. K. Bhattacharyya, J. Sarma, and P. N. L. Raju, "An effective ensemble classification framework using random forests and a correlation based feature selection technique," *Transactions in GIS*, vol. 21, no. 6, pp. 1165–1178, 2017.

- [3] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *2014 Science and Information Conference*. IEEE, 2014, pp. 372–378.
- [4] A. Chinnaswamy and R. Srinivasan, "Hybrid feature selection using correlation coefficient and particle swarm optimization on microarray gene expression data," in *Innovations in bio-inspired computing and applications*. Springer, 2016, pp. 229–239.
- [5] S. Nakariyakul, "High-dimensional hybrid feature selection using interaction information-guided search," *Knowledge-Based Systems*, vol. 145, pp. 59–66, 2018.
- [6] A. Tang and J. T. Foong, "A qualitative evaluation of random forest feature learning," in *Recent Advances on Soft Computing and Data Mining*. Springer, 2014, pp. 359–368.
- [7] C. Vens and F. Costa, "Random forest based feature induction," in *2011 IEEE 11th International Conference on Data Mining*. IEEE, 2011, pp. 744–753.
- [8] Y. Kong and T. Yu, "A deep neural network model using random forest to extract feature representation for gene expression data classification," *Scientific reports*, vol. 8, no. 1, pp. 1–9, 2018.
- [9] A. G. Reece, A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, "Forecasting the onset and course of mental illness with twitter data," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.
- [10] Z. Jamil, D. Inkpen, P. Buddhitha, and K. White, "Monitoring tweets for depression to detect at-risk users," in *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*, 2017, pp. 32–40.
- [11] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "Detecting depression and mental illness on social media: an integrative review," *Current Opinion in Behavioral Sciences*, vol. 18, pp. 43–49, 2017.
- [12] J. Jotheeswaran and S. Koteeswaran, "Feature selection using random forest method for sentiment analysis," *Indian Journal of Science and Technology*, vol. 9, no. 3, pp. 1–7, 2016.
- [13] A. J. Mitchell, A. Vaze, and S. Rao, "Clinical diagnosis of depression in primary care: a meta-analysis," *The Lancet*, vol. 374, no. 9690, pp. 609–619, 2009.
- [14] N. Choudhury and S. A. Begum, "A survey on case-based reasoning in medicine," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 8, pp. 136–144, 2016.
- [15] O. Taubman-Ben-Ari, J. Rabinowitz, D. Feldman, and R. Vaturi, "Post-traumatic stress disorder in primary-care settings: prevalence and physicians' detection," *Psychological medicine*, vol. 31, no. 3, pp. 555–560, 2001.
- [16] M. Nadeem, "Identifying depression on twitter," *arXiv preprint arXiv:1607.07384*, 2016.
- [17] A. Sau and I. Bhakta, "Predicting anxiety and depression in elderly patients using machine learning technology," *Healthcare Technology Letters*, vol. 4, no. 6, pp. 238–243, 2017.
- [18] D. L. Mowery, Y. A. Park, C. Bryan, and M. Conway, "Towards automatically classifying depressive symptoms from twitter data for population health," in *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES)*, 2016, pp. 182–191.
- [19] H. Cai, Y. Chen, J. Han, X. Zhang, and B. Hu, "Study on feature selection methods for depression detection using three-electrode eeg data," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 10, no. 3, pp. 558–565, 2018.
- [20] T. Yiu, "Understanding random forest," *Medium*, 2012.
- [21] V. M. Prieto, S. Matos, M. Alvarez, F. Cacheda, and J. L. Oliveira, "Twitter: a good place to detect health conditions," *PLoS one*, vol. 9, no. 1, 2014.
- [22] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [23] A. Liaw, M. Wiener *et al.*, "Classification and regression by random-forest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.
- [24] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and qsar modeling," *Journal of chemical information and computer sciences*, vol. 43, no. 6, pp. 1947–1958, 2003.
- [25] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures," in *Workshop on Statistical Modelling of Complex Systems*. Citeseer, 2006.
- [26] K. J. Archer and R. V. Kimes, "Empirical characterization of random forest variable importance measures," *Computational Statistics & Data Analysis*, vol. 52, no. 4, pp. 2249–2260, 2008.
- [27] R. C. Team *et al.*, "R: A language and environment for statistical computing," 2013.
- [28] R. Team *et al.*, "Rstudio: integrated development for r. rstudio," *Inc., Boston, MA*, vol. 639, p. 640, 2015.
- [29] R. Díaz-Uriarte and S. A. De Andres, "Gene selection and classification of microarray data using random forest," *BMC bioinformatics*, vol. 7, no. 1, p. 3, 2006.
- [30] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern recognition letters*, vol. 31, no. 14, pp. 2225–2236, 2010.
- [31] S. Tsugawa, Y. Mogi, Y. Kikuchi, F. Kishino, K. Fujita, Y. Itoh, and H. Ohsaki, "On estimating depressive tendency of twitter users from their tweet data," in *IEEE Virtual Reality*, vol. 2, 2013, pp. 29–32.
- [32] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Seventh international AAAI conference on weblogs and social media*, 2013.
- [33] M. Del Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Echo chambers: Emotional contagion and group polarization on facebook," *Scientific reports*, vol. 6, p. 37825, 2016.
- [34] H. AlSagri and M. Ykhlef, "Machine learning-based approach for depression detection in twitter using content and activity features," *IEICE TRANSACTIONS on Information and Systems*, vol. E103-D, no. 07, 2020.

# Parallel QR Factorization using Givens Rotations in MPI-CUDA for Multi-GPU

Miguel Tapia-Romero<sup>1</sup>, Amilcar Meneses-Viveros<sup>2</sup>  
Departamento de Computación  
Cinvestav-IPN  
Mexico City, Mexico

Erika Hernández-Rubio<sup>3</sup>  
Instituto Politécnico Nacional  
SEPI-ESCOM  
Mexico City, Mexico

**Abstract**—Modern supercomputers incorporate the use of multi-core processors and graphics processing units. Applications running on these computers take advantage of these technologies with scalable programs that work with multicores and accelerator such as graphics processing unit. QR factorization is essential for several numerical tasks, such as linear equations solvers, compute inverse matrix or compute a diagonal matrix, to name a few. There are several factorization algorithm such as LU, Cholesky, Givens and Householder, among others. The efficient parallel implementation of each parallelization algorithm will depend on the structure of the data and the type of parallel architecture used. A common strategy in parallel programming is to break a problem into subproblems to solve them in different processing units. This is very useful when dealing with complex problems or when the data is too large to work with the available memory. However, it is not clear how data partitioning affects subtask performance when mapping to processing units, specifically to graphical processing units. This work explores the partitioning of large symmetric matrix data for QR factorization using Givens rotations and its parallel implementation using MPI and CUDA is presented.

**Keywords**—Givens factorization; CUDA; heterogeneous programming; scalable parallelism

## I. INTRODUCTION

Every time it is more common to work with large amounts of data. One of the most commonly used tasks in processing these large volumes of data is QR factorization for square matrices. QR factorization is used in processes such as solving linear equations, inverting matrices, and in the process of diagonalizing matrices, to name a few. There are various methods for factoring such as LU, Cholesky, Householder, or Givens. These numerical tasks are a tool of common use in areas such as physics, chemistry and engineering. Also, artificial intelligence papers using Givens rotations on large volumes of data have been reported [1], [2], [3], [4].

Modern supercomputers incorporate the use of multi-core processors and accelerators such as graphics processing units (GPUs). Applications running on these computers take advantage of these technologies with scalable programs that work with multicores and GPUs. Although it has been reported that applications with GPUs can speed up a lot, GPUs suffer from the amount of memory they have available for data management. There are already GPU cards with more than 12GB of memory, however many computers still have cards with 6GB of memory or less, so the use of various GPUs helps resolve this limitation. Modern applications running on

supercomputers must be able to take advantage of various architectures that help speed up computing and must have the ability to scale, that is, to work on different nodes.

Methodologies have been proposed to develop programs for the new supercomputers [5], [6]. These methodologies include phases such as partitioning, aggregation, and mapping phases, among others. Partitioning refers to break a problem into subproblems to solve them in different processing units. This is very useful when dealing with complex problems or when the data is too large to work with the available memory. Aggregation refers to grouping subtasks, which is useful when identifying processes that can work with shared memory. The mapping phase refers to the association of tasks with processing units.

These methodologies have assisted in the development of scalable parallel programs and ensure the use of the various types of processing units. However it is not clear how a partitioning strategy can affect performance when tasks are mapped to graphics processing units. This work explores the partitioning of large symmetric matrix data for QR factorization using Givens rotations and its parallel implementation using MPI and CUDA is presented. A single GPU card version is made for comparison and analysis purposes. Its means, the program can use different GPU cards that are on the same node or on different nodes of a computer cluster. For the purpose of studying the different partitioning of the matrix, this work focuses on large symmetric matrices. The results show that the row or column partitioning of the matrix play an important role in the performance of CUDA kernels, also the communication between the main memory and the memory of the GPUs is important. Also changes in nVidia GPU technologies can affect the performance of the application.

This paper is organized as follows. Section 2 present the related work. Section 3 describe the Givens rotation procedure. Section 4 presents the design of the CUDA parallel program of QR factorization using Givens rotations for dense matrices. Section 5 describes scalable parallel implementation prioritizing row partitioning of matrices. Section 6 describe the experiments that were performed. Section 7 discusses the consequences of row and column partitioning on the performance of tasks running on GPUs. Finally, Section 8 presents the conclusions of this work.

## II. RELATED WORK

Traditionally the Intel MKL Math Library is used for Givens factorization. This library is highly optimized for the use of multicore processors [7][8][9][10]. Other mathematical libraries for lineal algebra have also been developed of which use parallel implementations such as MAGMA, PLASMA, ViennaCL, Armadillo and dmath, to name a few [11][12][13][14][15]. This libraries are designed for multicore architectures and GPUs. In addition, this work has been ported to Xeon Phi [16] and other library was optimized for multicore [17].

Sameh, in 1978, present a parallel algorithm for solving a system of linear equations using Givens rotations [18]. This work considers a parallel computer with shared memory to solve a dense tridiagonal linear system and it shows that the complexity to solve the tridiagonal system is  $O(n)$  steps, compared to  $O(n \log n)$  steps reported from previous work through Gaussian elimination [19]. Later, other authors complement Sameh's work to construct the QR factorization with Givens rotations of a dense rectangular matrix [20] or to construct QR factorization using Givens rotations [21].

There are other works where different factorization methods (LU, Cholesky or Householder) are parallelized for multicore architectures or GPUs [22], [23], [24], [25], [26][25], [27], [28]. Some of these works are oriented to the study of communications between processes, optimizing implementations for multi-core architectures, optimizing partitioning for architectures with GPUs, among others. The parallelization strategies vary even with the same factorization method. This is because the strategies depend on the parallel architecture in which the factorization is implemented.

## III. GIVENS ROTATION

The main idea in Givens rotations is to rotate a vector to annihilate, or zero, one of its elements. Therefore a rotation matrix is used. Then if two row vectors,  $u^t$  and  $v^t \in \mathbb{R}^m$ , are rotated.

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} \mu_1 & \cdots & \mu_n \\ v_1 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} \mu'_1 & \mu'_2 & \cdots & \mu'_n \\ 0 & v'_2 & \cdots & v'_n \end{pmatrix} \quad (1)$$

The values for  $v'_i$  and  $\mu'_i$  are

$$\begin{aligned} \mu'_1 &= (\mu_1^2 + v_1^2)^{1/2}, \\ \mu'_i &= c\mu_i + sv_i, \quad 2 \leq i \leq m, \\ v'_i &= -s\mu_i + cv_i, \end{aligned} \quad (2)$$

and the values  $c$  and  $s$  of the matrix rotation are

$$\begin{aligned} c &= \mu_1/\mu'_1, \\ s &= v_1/v'_1. \end{aligned} \quad (3)$$

As discussed in [18], a sequential program need  $n(n-1)/2$  rotations to get an upper triangular matrix from  $n \times n$  square matrix .

## IV. ALGORITHM ADAPTATION FOR GPU

There are various strategies to harness the computing power of GPUs. The main one is that the programs must be SPMD, work with fine granularity, reduce data transfer between card memory and main memory, and avoid synchronization between

threads, among others. In the CUDA programming environment, the CPU and its memory are called a host, and a GPU card is called a device. The part of a program that runs on a GPU is called the kernel. To run a kernel, the GPU memory must have the input data and a memory space to store the results. The part of a program that runs on a GPU is called the kernel. To run a kernel, the GPU memory must have the input data and a memory space to store the results. This implies that information must be moved between the host and the device. When designing a CUDA program, you must be careful with memory management, otherwise you can generate a large overhead.

Matrix factorization is an operation that consumes a lot of memory and CPU time. Selecting a matrix factorization method depends on the type of matrix and the architecture where it will be implemented. In this work, the Givens rotations are used for the QR factorization, since a method that adapts to the shared memory architecture that the GPUs use, and that can also avoid the synchronization of threads with a good implementation.

As explained in [18], when applying Givens rotation to a matrix  $A$  to annihilate the  $a_{ij}$  element, the rotation matrix affects two rows of a matrix  $A$ , rows  $i-1$  and  $i$ . The result of the rotation makes the  $j$ th element of row  $i$  zero. It is possible to parallelize the computation of the columns of rows  $i-1$  and  $i$ , because they are computations that do not generate dependencies between the computations of the columns. So a CUDA thread can be assigned to the calculation of each column.

The implementation that was carried out involves communication between MPI processes and the synchronization of work between the CPU and the GPU, so the implementation is heterogeneous. MPI allows to distribute data between processes and control synchronization with GPUs; and with CUDA computations are performed on the GPU.

Algorithm 1 presents the QR factorization algorithm using Givens rotations in GPU card. Lines 5 and 6 of Algorithm 1 are executed in GPU. The rest of algorithm run in a CPU.

---

### Algorithm 1 QR factorization with Givens rotation

---

**Require:**  $A \in \mathbb{R}^{n \times n}$ , a symmetric square matrix;  $I \in \mathbb{R}^{n \times n}$ , an identity matrix.

**Ensure:**  $R \in \mathbb{R}^{n \times n}$ , an upper triangular matrix;  $Q \in \mathbb{R}^{n \times n}$ , an orthogonal Matrix.

- 1:  $R \leftarrow A$ .
  - 2:  $Q \leftarrow I$ .
  - 3: Copy  $R$  and  $Q$  to GPU memory.
  - 4: **for**  $i = 0, 1, \dots, n-1$ . **do**
  - 5:   To get  $i$  column  $i$  from  $R$ .
  - 6:   To apply Givens rotations to  $i$  column to  $R$  and  $Q$ .
  - 7: **end for**
- 

1) *To get  $i$  column from  $R$ :* As mentioned, applying a Given rotation affects two lines. Therefore, when the Givens rotation is applied to the  $i$  row, you must wait for the  $i-1$  row to change all its values in order to apply the rotations to this row. This must be done with a synchronization between the threads involved in the rotation.

In order to avoid synchronization, each thread compute the values  $c$  and  $s$ , according to equation 3, to carry out the rotations in the corresponding column. Therefore, a copy of the column to which the rotations are applied is stored, this copy is stored in read-only memory for all threads, so each thread can obtain the values of this column and then use them to calculate the  $c$  and  $s$ .

In order to get and store the  $i$  column from  $R$  matrix, line 5 in the Algorithm 1, each thread in the kernel gets a value from the  $i$  column of  $R$  matrix and stores it into array in global memory of the GPU card, this is achieved with the help of thread identifier. So in a single call to this kernel all the values of the required column are stored.

2) To apply Givens rotations to  $i$  column to  $R$  and  $Q$ : Algorithm 2 shows how to apply Givens rotations. From equation 2, it is possible notice that Givens rotation affects 2 rows. This algorithm is implemented in a CUDA kernel. Each thread executes the same process, the only thing that changes is that each thread works on a different column, so at the end of calling the algorithm once 2 we will modify to 0's the values of a column under the diagonal main and after to call  $n-1$  times the algorithm we will have the upper triangular matrix  $R$  and  $Q^T$ .

From

$$A = QR \Rightarrow Q^T A = R \Rightarrow Q^T = G_n \dots G_1 I,$$

results in  $Q^T$  matrix by applying the Givens rotations to the identity matrix. Because  $Q$  is orthogonal,  $Q^T = Q^{-1}$ .

**Algorithm 2** To Compute Givens Rotations Per Column

**Require:**  $R \in \mathbb{R}^{n \times n}$ , symmetric square matrix;  $Q \in \mathbb{R}^{n \times n}$ , identity matrix;  $L \in \mathbb{R}^n$ , column from  $R$ ;  $col$ , column identifier.

**Ensure:**  $R$  and  $Q$  matrices with Givens rotations applied to column  $col$ .

- 1:  $i =$  thread identifier.
- 2:  $j = n$ .
- 3:  $\mu'_i = \sqrt{l_{j-1}^2 + l_j^2}$ .
- 4:  $c = \frac{l_{j-1}}{\mu'_i}$  y  $s = \frac{l_j}{\mu'_i}$ .
- 5: **while**  $j > col$  **do**
- 6:  $\mu_i = r_{i,j-1}$  y  $\nu_i = r_{i,j}$ .
- 7:  $\alpha_i = q_{i,j-1}$  y  $\beta_i = q_{i,j}$ .
- 8:  $r_{i,j-1} = c\mu_i + s\nu_i$  y  $r_{i,j} = -s\mu_i + c\nu_i$ .
- 9:  $q_{i,j-1} = c\alpha_i + s\beta_i$  y  $q_{i,j} = -s\alpha_i + c\beta_i$ .
- 10:  $j = j - 1$ .
- 11:  $a = \mu'_i$ .
- 12:  $\mu'_i = \sqrt{l_{j-1}^2 + a^2}$ .
- 13:  $c = \frac{l_{j-1}}{\mu'_i}$  y  $s = \frac{a}{\mu'_i}$ .
- 14: **end while**

Fig. 1 shows how algorithm 2 affects a matrix. So, this algorithm needs to be applied to the first  $n-1$  columns of a matrix to get the matrices  $R$  and  $Q^T$ , Fig. 2.

From Fig. 1 and 2, it can be seen that, as the process progresses, there are some threads that are left unworked, and this happens for the last steps. But when the number of threads is less than the number of columns in the matrix, the threads

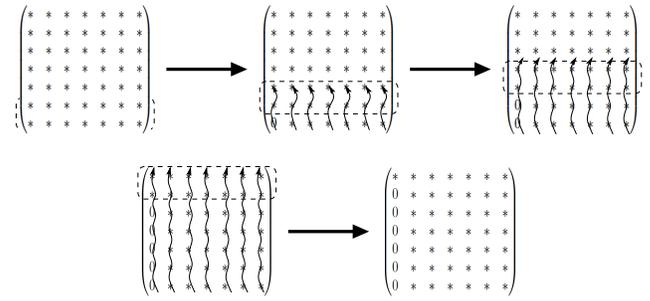


Fig. 1. Givens Rotations to the First Column.

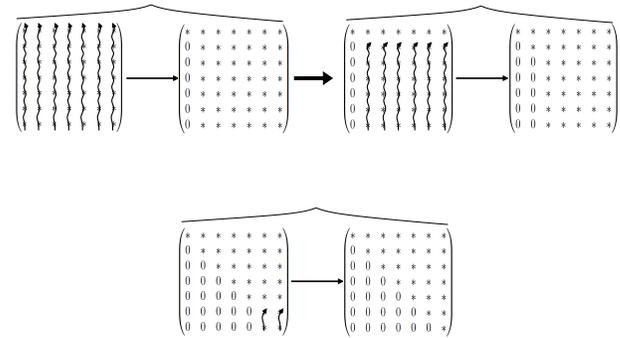


Fig. 2. Process to get  $R$ , an Upper Triangular Matrix with Givens Rotations.

are traversed. Although interesting results have been shown when the number of threads is  $n-1-col$ , and although there are several threads that do few computations, the results in time are better, since CUDA threads are kept in fine grain.

V. MULTI GPU STRATEGY

One of the main restrictions on the use of GPUs is memory. Many GPU cards have 8GB of memory or less. While nVidia has provided memory sharing strategies across multiple GPUs or directly communicate the main memory with the GPU memory, these solutions can be expensive for many users. One strategy could be the use of multiple streams to get a speedup and use short pieces of data to process, but this strategy is not always easy to implement in many problems.

To work with multiple GPUs, a strategy for partitioning data is designed. A strategy would be to divide the columns so that they are processed by different cards, as shown in Fig. 3. Because matrices are stored by row in C, this partitioning method generates an overhead. Since the input matrix  $A$  in Algorithm 1 is symmetric, then the partitioning of  $A$  by rows can be performed. This matrix partitioning makes the communication between the processes that handle these sub-matrices more efficient. And when applying Givens rotations the  $R^T$  matrix is obtained.

In the implementation of this work, it was decided to use MPI with CUDA, so each MPI process is in charge of communicating with the GPU card, and the partitioning of the data is done by lines, instead of columns. Thus, each MPI process has a GPU card assigned to execute CUDA processes. The main MPI process, or root process, is in charge

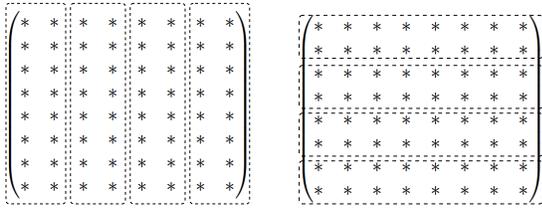


Fig. 3. Column and Row Partitioning of Data.

of doing the read operations and doing a scatter operation of the matrix  $A$  to the other processes. Subsequently, the root MPI process send to the other processes the row of  $A$ , to which Givens rotations will be applied. Each MPI process sends to the memory of the GPU card that has assigned the sub-matrix  $A$ , the row of  $A$ , and sends to execute the kernel of the Givens rotations. The MPI process must send the first  $n - 1$  lines among them, one line at a time, to each MPI process so that they apply the rotation to the sub-matrices it has. Every time an MPI process receives a row from the root process, it must upload it to the GPU memory and run the Givens rotation kernel. At the end each MPI process will have a partial result of the partial triangular matrix  $R^T$ . If each MPI process initializes the corresponding identity sub-matrix, a communication between the root process and the other MPI processes is avoided, and the corresponding Givens rotation is applied to it, at the end an  $Q$  submatrix is also obtained. Finally, the data of each process must be gathered in a single data set to the root MPI process, to be able to show the results, in this case the  $Q$  and  $R^T$  matrices.

In Algorithm 3 is shown the parallel  $QR$  factorization using MPI processes for multi GPU.

In Algorithm 3, there are some lines that deserve a review. In the line 8, select the GPU card that will be assigned to each process, the CUDA function to select the card is `cudaSetDevice()`. The assignment of the card will depend on the hardware configuration of each node. For example, if we have nodes with a multicore processor with two GPU cards, each node associates 0 and 1 with the identifiers of the cards. If MPI processes  $i, i + 1$ , with  $i = 0, 1, \dots, np$ , are assigned on each node. Then a module operation is used to assign the GPU card.

$$idGPU = (idProcess) \bmod 2.$$

Where 2 is the number of GPU cards.

The line 9, refers to the row partitioning of the matrices  $R$  and  $Q$ . Thus, each MPI process corresponds to sub-matrices  $R, Q \in \mathbb{R}^{(n/np) \times n}$  with the same number of lines. That is, each MPI process is sent  $(n * n)/np$  elements to be processed. For this task, `MPI_Scatter` function is used. The `MPI_Scatter` function is used to divide an array of data into equal parts between the different processes.

Line 21 of the Algorithm 3 gathers the partial results of each MPI process to form the resulting  $R^T$  and  $Q$  matrices. To do this, the `MPI_Gather` function is used.

In the following sections, other lines of Algorithm 3 will be described in more detail. These lines are: 12, 13 and 13.

---

### Algorithm 3 $QR$ Factorization using Givens Rotations for Multi GPUs

---

**Require:**  $A \in \mathbb{R}^{n \times n}$ , a symmetric square matrix ;  $np =$  total number of process.

**Ensure:**  $R^T \in \mathbb{R}^{n \times n}$  transposed upper triangular matrix;  $Q \in \mathbb{R}^{n \times n}$  orthogonal matrix.

- 1: Start MPI.
  - 2:  $np =$  total number of process.
  - 3:  $idProcess =$  process number.
  - 4: **if**  $idProcess == 0$  **then**
  - 5:      $R \leftarrow A$ .
  - 6:      $Q \leftarrow I$ .
  - 7: **end if**
  - 8: Associate a GPU card with the MPI process.
  - 9: MPI Scatter of  $R$  and  $Q$  to all processes.
  - 10: To copy the sub-matrices of  $R$  and  $Q$  to the memory of the GPUs.
  - 11: **for**  $i = 0, 1, \dots, n - 2$ . **do**
  - 12:     **if** Row  $i$  belongs to the  $R$  sub matrix of the process  $idProcess$  **then**
  - 13:         To get the  $i$  row from  $R$  sub matrix of the  $idProcess$  process.
  - 14:     **end if**
  - 15:      $idProcess$  process sends the row obtained to the other MPI processes.
  - 16:     Wait until all processes have received the row.
  - 17:     Copy the row to the memory of the GPUs card.
  - 18:     To apply Givens rotations to sub-matrices  $R$  and  $Q$ .
  - 19: **end for**
  - 20: Copy the  $R$  and  $Q$  sub-matrices from the GPUs card to the Host.
  - 21: Join submatrices to get  $R^T$  y  $Q$ .
  - 22: Finalize MPI.
- 

Lines 12 and 13 show how to obtain the necessary row of a specific process. And the line 18 Givens rotations are applied to the matrix partition of each process.

#### A. To Get the $i$ -th Row of the Process that has the Sub Matrix

As discussed in the previous sections, it is required to avoid synchronization between processes to achieve performance. To avoid this, a copy the row to which the rotations are applied to each GPU card is made. It is important to extract the line that is being processed. It must be inferred which process is the one that contains the line to be processed. The Algorithm 4 is responsible for obtaining the row and sending them to the MPI processes. This algorithm is used in lines 12 and 15 of Algorithm 3.

In Algorithm 4, the  $i$  is the row that is send to all processes,  $k$  is the partition that has the row,  $d$  is the number of rows that each process has,  $n$  is the number of rows that the original matrix has,  $np$  is the number of MPI processes, this value is equal to the number of partitions that were made from the matrix, and  $idProcess$  is the identifier of each process.

This process runs  $n$  times, as seen in the Algorithm 3, where  $j$  and  $k$  are initialized to 0 before starting the `for` cycle, it is possible yo obtain the row at the time it is required and assures us to go through all the rows.

**Algorithm 4** To Get the  $i$ -th Row of  $R$  Sub Matrix from MPI Process.

**Require:**  $i$ ,  $i$ -th row;  $k$  is the process with the  $i$  row;  $d = n/np$  number of rows;  $idProcess$ , process identifier.

- 1: **if**  $idProcess == k$  **then**
- 2:   To get  $i$  row of sub matrix  $R$  from  $idProcess$ .
- 3: **end if**
- 4: To send the row from  $idProcess$  to the other process.
- 5:  $i = i + 1$
- 6: **if**  $i == d$  **then**
- 7:    $i = 0$
- 8:    $k = k + 1$ .
- 9: **end if**

**B. To Apply Givens Rotations to Row  $i$  with Multi-GPUs**

Because rotations are applied to row  $i$  of  $A$  matrix, this operation affects rows  $i + 1$  to  $n - 1$  of this matrix. Since each process has this row, the sub-matrices of  $R$  and  $Q$  can be modified.

The algorithm is similar to Algorithm 2, although it has some changes to be able to apply the rotations to each partition. In Algorithm 5 shows the steps that must be applied to each sub matrix of  $R$  and  $Q$  to get the partition with the rotations applied.

**Algorithm 5** To Compute Givens Rotations Per Row on Each GPU Card

**Require:**  $R, Q \in \mathbb{R}^{(n/np) \times n}$ , sub matrices;  $L \in \mathbb{R}^n$ , is the  $i$ -row;  $fil$ , is the row identifier;  $n$ , matrix column size;  $np$ , number of process.

**Ensure:**  $R, Q \in \mathbb{R}^{(n/np) \times n}$ , sub matrices with Givens rotations applied;

- 1:  $i =$  thread identifier.
- 2:  $j = n$ .
- 3: **while**  $l_j$  sea 0 **do**
- 4:    $j = j - 1$
- 5: **end while**
- 6:  $tempj = j$
- 7: **while**  $i < n/np$  **do**
- 8:    $\mu'_i = \sqrt{l_{j-1}^2 + l_j^2}$ .
- 9:    $c = \frac{l_{j-1}}{\mu'_i}$  y  $s = \frac{l_j}{\mu'_i}$ .
- 10:   **while**  $j > fil$  **do**
- 11:      $\mu_i = r_{i,j-1}$  y  $\nu_i = r_{i,j}$ .
- 12:      $\alpha_i = q_{i,j-1}$  y  $\beta_i = q_{i,j}$ .
- 13:      $r_{i,j-1} = c\mu_i + s\nu_i$  y  $r_{i,j} = -s\mu_i + c\nu_i$ .
- 14:      $q_{i,j-1} = c\alpha_i + s\beta_i$  y  $q_{i,j} = -s\alpha_i + c\beta_i$ .
- 15:      $j = j - 1$ .
- 16:      $a = \mu'_i$ .
- 17:      $\mu'_i = \sqrt{l_{j-1}^2 + a^2}$ .
- 18:      $c = \frac{l_{j-1}}{\mu'_i}$  y  $s = \frac{a}{\mu'_i}$ .
- 19:    **end while**
- 20:     $j = tempj$
- 21:    increase  $i$  to block size
- 22: **end while**

Every CUDA thread work with a specific row. If the number of threads is less than the number of lines of the sub-

matrices, then the thread identifier is increased with the total of CUDA threads that have been requested per MPI process. Because CUDA threads traverse the sub matrix per row, using a moderate number of threads is not recommended to avoid the sparse cache problem. It was observed that an adequate number of CUDA threads is 128 threads per streaming multiprocessor in GPU cards.

A CUDA kernel runs the Algorithm 5. This kernel runs on every GPU card associated with an MPI process. The key for this algorithm to work is to have the row to which the rotations in memory will be applied, since without it the algorithm would not work, or it only work on the GPU card where the row data is kept. Fig. 4 shows how the algorithm works on the sub matrices.

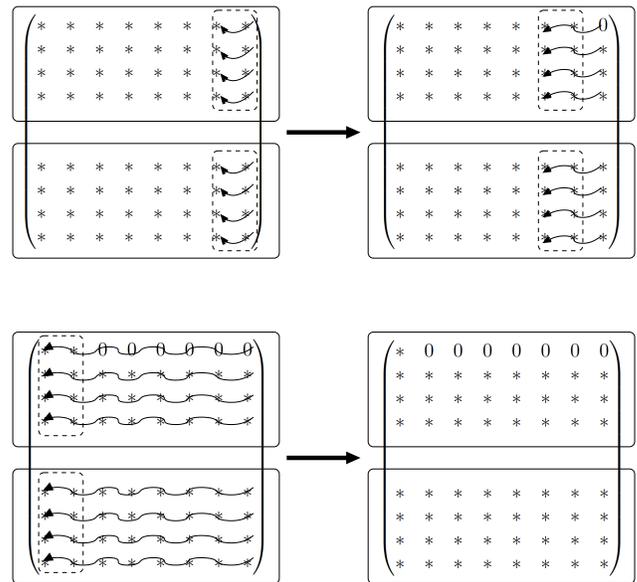


Fig. 4. Kernel Running on 2 GPU Cards.

It is known that although the cards run the same kernel, it does not imply that they end at the same time. So MPI processes must wait for all cards to finish executing the kernel that applies rotation to row  $i$  before moving on to row  $i + 1$ . This is accomplished by synchronizing the MPI processes with a call to the MPI\_Barrier function.

As in the method for a GPU card, in the method of Givens multi GPUs, each card must make  $n - 1$  calls to the kernel. Fig. 5 shows how the whole method works using two GPU cards.

Fig. 5 shows that after half the process, it seems that one of the cards, the one with the upper part of the matrix, no longer performs any operation; but this is not the case, although it does not work on the partition of  $R$ , on the partition of  $Q$  it does, since each call to the kernel changes the partition of  $Q$ . Fig. 6 shows how even after the middle of the process the two cards continue to work.

Like a parallel method that works on a single card, for the multi-GPU method, if the number of rows in each partition is greater than the number of threads, then the threads start traversing to completely cover the partitions. Thus, it is not

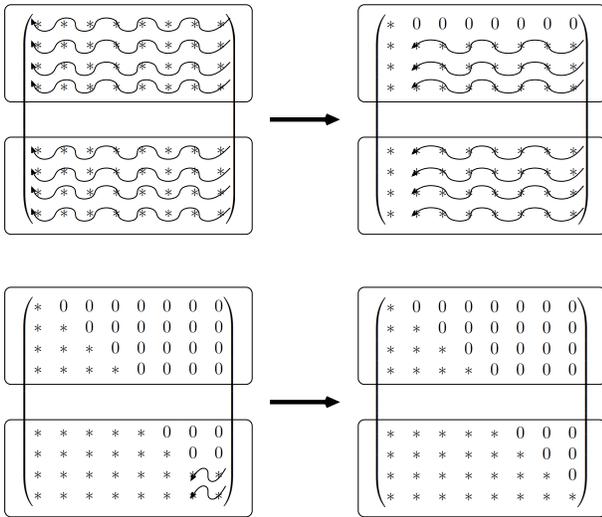


Fig. 5. Givens Method for 2 GPU Cards Working in the  $R$  Matrix.

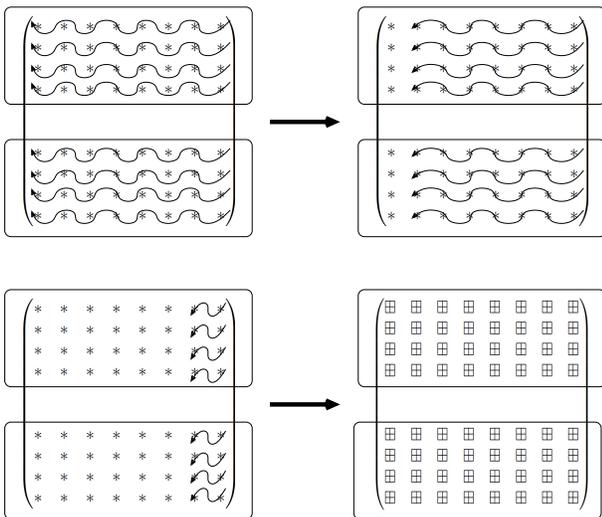


Fig. 6. Givens Method for 2 GPU Cards Working in the  $Q$  matrix.

necessary to have the same number of threads as rows per partition, this takes on importance since it can be the case that the two cards are different, so there is no need to change the number of threads in the program execution. In addition to avoid scattered cache problems, it is recommended to have 128 threads per block.

## VI. TESTS

To test the performance of the multi GPU parallelization of the QR factorization, it was compared against Intel Math Kernel Library (MKL) functions. These comparisons were made with single-GPU and multi-GPU experiments.

The hardware used for testing was as follows:

- Server with two **Intel Xeon X5675** [29] to run MKL. Each process has::

- Number of CPU core: 6. Clock Speed: 3.06 GHz. Memory bandwidth: 32 GB/s.
- Three different GPU cards were used for testing with CUDA:
  - 2 **NVIDIA Tesla K20X** card [30]:
    - Memory size: 6 GB. CUDA Cores: 2688. Clock speed per core: 732 MHz. Memory Bandwidth: 250 GB/s.
  - 1 **NVIDIA Tesla C2070** [31]:
    - Memory size: 6 GB. CUDA Cores: 448. Clock speed per core: 1.15 GHz. Memory Bandwidth: 144 GB/s.
  - 1 **NVIDIA GeForce GTX 460** [32]:
    - Memory size: 1 GB. CUDA Cores: 336. Clock speed per core: 1.53 GHz. Memory Bandwidth: 115.2 GB/s. Cuenta con *overclocking*

### A. Experiments with a Single GPU

The parallel implementation of the QR factorization for multi GPU was executed in each one of the cards. The 2 MKL functions were used: `?geqrf1`, which allows to get the matrix  $R^T$  and the function `?orgqr`, which is used after the function `?geqrf`, to get the matrix  $Q^T$ . In CUDA and MKL tests, single and double precision were used.

Three different types of test matrices were used: tridiagonal, pentadiagonal, and heptadiagonal. The results of the three types of matrices were very similar, since the algorithm considers symmetric dense matrices, so only the results of the tests with heptadiagonal matrices will be exposed. One of the objectives of this work is to be able to process large matrices. The experiments were carried out with single precision square matrices of size  $5000 \times 5000$  to  $20000 \times 20000$ , and  $5000 \times 5000$  to  $16000 \times 16000$  for double precision square matrices.

Fig. 7 shows the difference in execution times of the different tests that were carried out for simple precision. And the Table I shows the data in more detail for your study. These first experiments are for QR factorization with a single GPU card. In general, it is appreciated that the MKL functions for QR factorization are very competitive up to sizes between  $9000 \times 9000$  or  $10000 \times 10000$ .

Table I shows the execution times. Because the MKL experiments with 12 threads have better performances, these times will be taken as a reference for CUDA experiments. The K20X card, which is the one with the lowest performance among GPU cards. The test on the K20X has a maximum acceleration of 1.2x over MKL with 12 threads when working with an matrix size of  $20000 \times 20000$ . The GTX 460 card could only process matrices up to  $10000 \times 10000$  in single precision because it only has 1MB of memory, however, when processing this matrix size it achieves a 1.36x acceleration compared to MKL. The C2070 card begins to show better performance compared to MKL from the matrix size of  $10000 \times 10000$  with a 1.2x acceleration and achieves an acceleration of 2.8x at the size  $20000 \times 20000$ , which is the maximum it supports.

<sup>1</sup>Where the symbol ? changes for a s if it is a single precision or a d if it is a double precision.

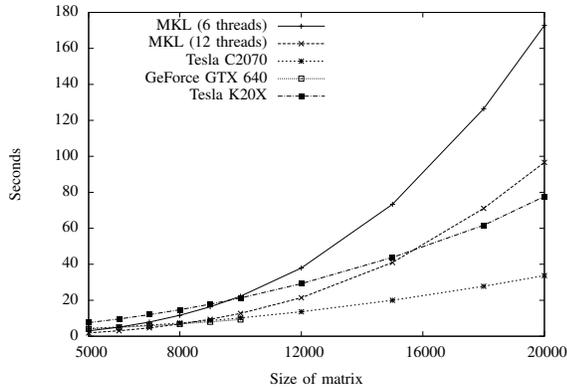


Fig. 7. Execution Times of QR Factorization for Single Precision Matrices.

TABLE I. EXECUTION TIMES OF QR FACTORIZATION FOR SINGLE PRECISION MATRICES.

Size	Execution times (seconds)				
	MKL 6 threads	MKL 12 threads	C2070	GTX 640	K20X
5000	3.005	1.854	4.315	4.127	7.567
6000	5.057	3.054	5.181	4.893	9.583
7000	7.887	4.664	6.175	5.775	11.967
8000	11.621	6.795	7.345	6.823	14.708
9000	16.361	9.428	8.689	7.991	17.812
10000	22.249	12.701	10.153	9.289	21.251
12000	37.928	21.379	13.598	-	29.217
15000	73.342	40.967	19.986	-	43.838
18000	126.449	71.037	27.771	-	61.680
20000	172.728	96.641	33.677	-	77.676

Fig. 8 shows the execution times of the different tests with double precision with MKL and CUDA with a single GPU card. Like the simple precision case, MKL experiments with 12 threads perform better than execution with 6 threads, so comparisons of the CUDA implementation will be made with the results of this experiment.

Table II shown the performance of CUDA programs with respect to the MKL version. The maximum size for GTX 640 card es  $8000 \times 8000$ , this card begins to show an acceleration of 1.1x in the size of  $6000 \times 6000$  and reaches 1.7x in the

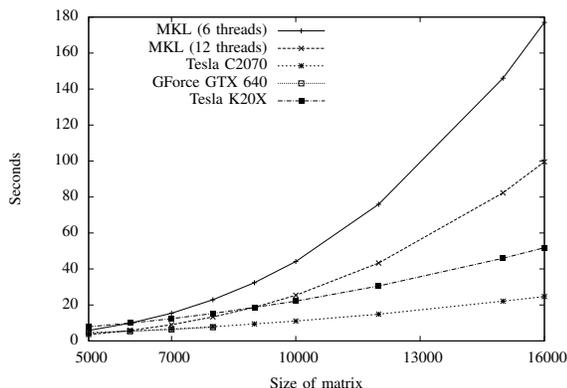


Fig. 8. Execution Times of QR Factorization for Double Precision Matrices.

TABLE II. EXECUTION TIMES OF QR FACTORIZATION FOR DOUBLE PRECISION MATRICES.

Size	Execution times (seconds)				
	MKL 6 threads	MKL 12 threads	C2070	GTX 640	K20X
5000	5.856	3.471	4.532	4.405	7.886
6000	9.863	5.783	5.507	5.333	9.893
7000	15.407	9.001	6.592	6.354	12.401
8000	22.894	13.338	7.882	7.599	15.271
9000	32.359	18.678	9.449	-	18.640
10000	44.139	25.375	11.043	-	22.124
12000	75.976	43.294	14.861	-	30.562
15000	146.037	82.342	22.118	-	45.949
16000	177.032	99.588	24.683	-	51.798

TABLE III. EXECUTION TIME OF QR FACTORIZATION WITH 2 K20X CARDS FOR SINGLE PRECISION SQUARE MATRICES.

Size	Execution times (seconds)		
	MKL 6 threads	MKL 12 threads	2 K20X
10000	22.24998125	12.70196825	38.5489915
15000	73.34226025	40.967238	84.951077
20000	172.7288738	96.6412415	150.7447375
22000	229.2751283	129.0128638	182.637326
24000	296.6546478	165.9274763	217.4389965
26000	376.5570043	210.0910878	254.3220125

maximum size of the card. The C2070 card presents better performance than MKL from  $7000 \times 7000$  matrix size, with 1.3x, and reaches 4.14x acceleration compared to MKL. The K20X starts its acceleration in the size of  $10000 \times 10000$  with 1.15x and reaches almost 2x speedup. Again, the C2070 performs better than the K20.

Finally, a very peculiar behavior is observed in Table I and II tables, since the execution times are maintained for matrix sizes from 5000 to 1000 for double and single precision cases.

### B. Experiments with Multi-GPU

Two experiments were carried out on the scalable implementation of GPU cards: One using 2 Tesla K20X cards and the other using a Tesla C2070 card and a GTX 640. The results of these experiments are compared against MKL of 6 and 12 threads. The first configuration of cards is homogeneous, since they are two Tesla K20 cards. Because K20 cards have 6GB of memory, its possible work with  $26000 \times 26000$  single precision square matrix for the experiments. The second configuration is heterogeneous, since they are cards with different characteristics. In the case of the experiments with the C2070 and the GTX 460x, it will only be possible to carry out experiments on stable matrices of sizes 6000 to 18000. This limitation in the size of the matrix is due to the memory restriction of the GTX 460 card, since it only has 1GB of memory.

It start with the experiments of the 2 K20X, in the Fig. 9 and Table III it can be see the times for MKL and the implementation proposed in this work to perform the QR factorization. In this the MKL experiments for 12 threads have better performance that the multi-GPU program. However, the multi GPU program is capable of processing a larger matrix than its single card version.

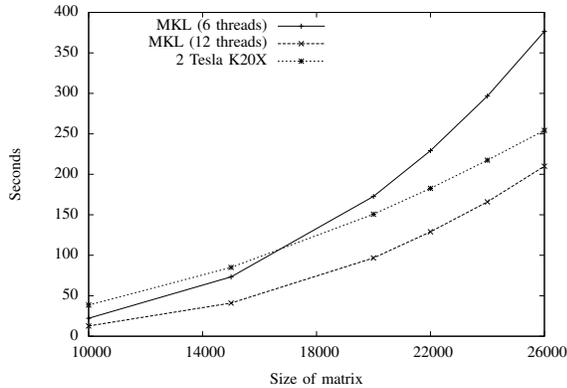


Fig. 9. Execution Time of QR Factorization with 2 K20X Gpus Cards for Single Precision Matrices.

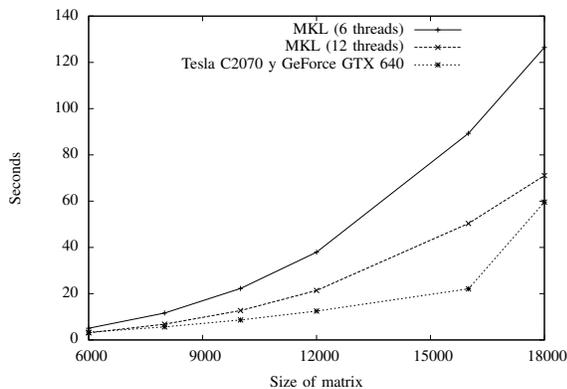


Fig. 10. Execution Times for QR Factorization with Multi GPU using C2070 and GTX 460 Cards for Single Precision Matrices.

Unfortunately for the configuration with the K20X cards the same behavior is maintained in single and double precision. In both cases the performance of the multi-GPU version is lower than that of MKL with 12 threads.

For multi-GPU experiments with the C2070 and GTX 640, we started showing the results for single precision matrices. In this case, the maximum size of the matrix used to process was  $18000 \times 18000$ , due to the memory limitation of the GTX 460 card. Fig. 10 presents the times of the MKL experiments and the multi GPU implementation. It is interesting to see that the same multi-GPU program has better performance than running for 2 K20X and better performance than MKL functions with 12 threads.

From the values in the Table IV, it is observed that the multi GPU program with the C2070 and GTX 640 cards, starts with similar times to the program with MKL and 12 threads, and accelerates until reaching 1.2x acceleration for the maximum size of the test matrix.

Fig. 11 shows the run times for experiments in the case of matrices with double precision. For these experiments the maximum size of the test matrices was  $12000 \times 12000$ . It is observed that the performance of the version of the QR multi GPU factorization is superior to the version of MKL

TABLE IV. EXECUTION TIMES FOR QR FACTORIZATION WITH MULTI GPU USING C2070 AND GTX 460 CARDS FOR SINGLE PRECISION MATRICES.

Size	Execution times (seconds)		
	MKL 6 threads	MKL 12 threads	C2070 y GTX 640
6000	5.05746	3.05409	3.343184
8000	11.62196075	6.79531825	5.625851
10000	22.24998125	12.70196825	8.625267
12000	37.928559	21.37907675	12.4785045
16000	89.31594025	50.3303555	22.0926865
18000	126.449971	71.03795625	59.6142455

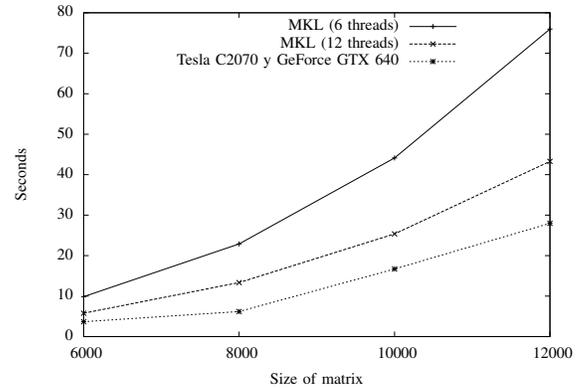


Fig. 11. Execution times for QR factorization with multi GPU using C2070 and GTX 460 cards for double precision matrices.

with 12 threads. Table V shows that the multi GPU program that uses the C2070 cards and the GTX 640, has better performance since the first experiments, with a performance of 1.6x compared to the MKL program with 12 threads. For the experiment with the maximum size of the matrices, the acceleration reached 1.6x over the version of MKL with 12 threads.

## VII. DISCUSSION

After carrying out the experiments and comparing the results, its possible analyze the behavior of the programs, including the technologies that were used in this work. This gives a set of lessons learned.

First it can be see that the performance of the Tesla K20X cards was lower than that of the Fermi Tesla C2070 card. When analyzing the execution of the CUDA program on these cards, it is appreciated that the communication time between the Host and Device (both to send data to the GPU and to receive data from the GPU) increased with the K20X cards.

TABLE V. EXECUTION TIMES FOR QR FACTORIZATION WITH MULTI GPU USING C2070 AND GTX 460 CARDS FOR DOUBLE PRECISION MATRICES.

Size	Execution times (seconds)		
	MKL 6 threads	MKL 12 threads	C2070 y GTX 640
6000	9.86363625	5.7834095	3.6662875
8000	22.8946235	13.33819875	6.1803205
10000	44.13997175	25.37548475	16.696302
12000	75.97651125	43.29401675	27.9818585

The kernel performance is superior in the K20X, however, the main bottleneck when using these cards is to pass data between the host and the device.

It is observed that the performance behavior for double precision and single precision data processing is similar with CUDA. So the limitation of the cards is their memory size.

In this paper, symmetric matrices were deliberately used to study the partitioning of rows by matrices to apply Givens rotations. This type of partitioning allows the distribution of data to be very transparent in MPI and thus avoid a bottleneck in communication between MPI processes. However, this results in CUDA threads having to traverse the matrix by row, which generates a sparse cache problem. That is, the cache of the Streaming multiprocessor of the Nvidia GPUs is 46KB. By having multiple threads that must traverse the rows in reverse, there is no guarantee that the data for a CUDA thread will be available in the L1 cache. Which generates a poor performance in the execution of the kernel. In the case of the CUDA program with a card, the path of the CUDA threads is made per column, and thus each thread processes an element of the  $i$ -th row and a compact cache is generated, so there is a guarantee of the CUDA threads have the data with which they will operate in the L1 cache memory.

The effect of the use of the compact and dispersed cache is very noticeable in the experiments that were carried out. If the execution times of the CUDA programs of a card are compared with the CUDA multi GPU program, Table II and Table V, it is noted that the performance of the programs with a card is superior to that of the multi GPU version. At first glance one might think that it is due to the synchronization between the processes to send the  $i$ -th row to all the processes. But it has been observed that the execution of the multi GPU program with a single GPU consumes more than twice the time of the execution of the program designed for a single GPU.

A new version of the multi GPU program for QR factorization using Givens rotations must have a matrix partitioning per column. This approach would allow CUDA threads to work with compact cache memory, which would give CUDA kernel execution a good performance. Fortunately, there are ways to do a matrix scatter per column in MPI.

## VIII. CONCLUSIONS

This work presents the parallel implementation in CUDA for Givens factorization. This implementation is scalable to work with multiple GPUs when combined with MPI and CUDA. This work explore some strategies for working with large volumes of data combining MPI with CUDA. These strategies focus on partitioning the matrix to be factored. It can be seen that the program with escalation has less performance than the program for a card. But the performance is due to the way CUDA threads work with sub-matrices derived from row partitioning.

Communication between CPU memory and GPU card affects the performance of CUDA programs. The C2070 card, although it is from a previous generation to the K20X, presents better results than the K20X, because the data transfer times are shorter.

The main contribution of this work is that, for the QR factorization algorithm through Givens rotations, the relationship between matrix partitioning and the scattered or compact cache problems that occur at the GPU level by the how the CUDA kernel should process the sub array. In addition, it is proven that it is feasible to establish a combination of MPI and CUDA to have scalable algorithms that process large volumes of data on nodes that do not include the latest memory management technology from NVIDIA such as GPU Direct.

## ACKNOWLEDGMENT

The authors thank financial support given by the Mexican National Council of Science and Technology (CONACyT), as well as IPN for SIP Research Project number 20201079. The authors acknowledge both, the Center for Research and Advance Studies of the National Polytechnic Institute (CINVESTAV- IPN) and the Section of Research and Graduate Studies (SEPI) of ESCOM-IPN, for encouragement and facilities provided to accomplish this publication.

## REFERENCES

- [1] B. Alipourfard and J. X. Gao, "Solving all regression models for learning gaussian networks using givens rotations," *arXiv preprint arXiv:1901.07643*, 2019.
- [2] Y. Gan, B. Hu, W. Liu, S. Wang, G. Zhang, X. Feng, and D. Wen, "Endmember extraction from hyperspectral imagery based on qr factorisation using givens rotations," *IET Image Processing*, vol. 13, no. 2, pp. 332–343, 2018.
- [3] L. Marcellino and G. Navarra, "A gpu-accelerated svd algorithm, based on qr factorization and givens rotations, for dwi denoising," in *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, 2016, pp. 699–704.
- [4] C. Luo, K. Zhang, S. Salinas, and P. Li, "Efficient privacy-preserving outsourcing of large-scale qr factorization," in *2017 IEEE Trust-com/BigDataSE/ICSS*. IEEE, 2017, pp. 917–924.
- [5] P. Czarnul, J. Proficz, and K. Drypczewski, "Survey of methodologies, approaches, and challenges in parallel programming using high-performance computing systems," *Scientific Programming*, vol. 2020, 2020.
- [6] A. A. Serrano-Rubio, A. Meneses-Viveros, G. B. Morales-Luna, and M. Paredes-López, "Generic methodology for the design of parallel algorithms based on pattern languages," in *International Conference on Supercomputing in Mexico*. Springer, 2018, pp. 35–48.
- [7] I. Burylov, M. Chuvelev, B. Greer, G. Henry, S. Kuznetsov, and B. Sabanin, "Intel performance libraries: Multi-core-ready software for numeric-intensive computation." *Intel Technology Journal*, vol. 11, no. 4, 2007.
- [8] E. Wang, Q. Zhang, B. Shen, G. Zhang, X. Lu, Q. Wu, and Y. Wang, "Intel math kernel library," in *High-Performance Computing on the Intel® Xeon Phi?* Springer, 2014, pp. 167–188.
- [9] A. Kalinkin, A. Anders, and R. Anders, "Intel® math kernel library parallel direct sparse solver for clusters," in *EAGE Workshop on High Performance Computing for Upstream*. European Association of Geoscientists & Engineers, 2014, pp. cp-426.
- [10] Intel, "Math kernel library (mkl)," 2016. [Online]. Available: <http://software.intel.com/en-us/intel-mkl/>
- [11] E. Agullo, J. Demmel, J. Dongarra, B. Hadri, J. Kurzak, J. Langou, H. Ltaief, P. Luszczek, and S. Tomov, "Numerical linear algebra on emerging architectures: The plasma and magma projects," in *Journal of Physics: Conference Series*, vol. 180, no. 1. IOP Publishing, 2009, p. 012037.
- [12] A. Haidar, S. Tomov, P. Luszczek, and J. Dongarra, "Magma embedded: Towards a dense linear algebra library for energy efficient extreme computing," in *2015 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, 2015, pp. 1–6.

- [13] K. Rupp, P. Tillet, F. Rudolf, J. Weinbub, A. Morhammer, T. Grasser, A. Jungel, and S. Selberherr, "Viennacl—linear algebra library for multi- and many-core architectures," *SIAM Journal on Scientific Computing*, vol. 38, no. 5, pp. S412–S439, 2016.
- [14] C. Sanderson and R. Curtin, "Armadillo: a template-based c++ library for linear algebra," *Journal of Open Source Software*, vol. 1, no. 2, p. 26, 2016.
- [15] S. Eliuk, C. Upright, and A. Skjellum, "dmath: A scalable linear algebra and math library for heterogeneous gp-gpu architectures," *arXiv preprint arXiv:1604.01416*, 2016.
- [16] J. Dongarra, M. Gates, A. Haidar, Y. Jia, K. Kabir, P. Luszczek, and S. Tomov, "Portable hpc programming on intel many-integrated-core hardware with magma port to xeon phi," in *International Conference on Parallel Processing and Applied Mathematics*. Springer, 2013, pp. 571–581.
- [17] J. Dongarra, M. Gates, A. Haidar, J. Kurzak, P. Luszczek, P. Wu, I. Yamazaki, A. YarKhan, M. Abalenkovs, N. Bagherpour *et al.*, "Plasma: Parallel linear algebra software for multicore using openmp," *ACM Transactions on Mathematical Software (TOMS)*, vol. 45, no. 2, pp. 1–35, 2019.
- [18] A. H. Sameh and D. J. Kuck, "On stable parallel linear system solvers," *Journal of the ACM (JACM)*, vol. 25, no. 1, pp. 81–91, 1978.
- [19] —, "A parallel qr algorithm for symmetric tridiagonal matrices," *IEEE Transactions on Computers*, vol. 100, no. 2, pp. 147–153, 1977.
- [20] M. Cosnard, J.-M. Muller, and Y. Robert, "Parallel qr decomposition of a rectangular matrix," *Numerische Mathematik*, vol. 48, no. 2, pp. 239–249, 1986.
- [21] I. C. Ipsen, "A parallel qr method using fast givens' rotations." YALE UNIV NEW HAVEN CT DEPT OF COMPUTER SCIENCE, Tech. Rep., 1984.
- [22] H. Xu and W. Alexander, "Parallel qr factorization on a block data flow architecture," in *The 24th Southeastern Symposium on System Theory and The 3rd Annual Symposium on Communications, Signal Processing Expert Systems, and ASIC VLSI Design*. IEEE, 1992, pp. 332–336.
- [23] J. Dongarra, M. Faverge, T. Herault, M. Jacquelin, J. Langou, and Y. Robert, "Hierarchical qr factorization algorithms for multi-core clusters," *Parallel Computing*, vol. 39, no. 4-5, pp. 212–232, 2013.
- [24] T. Fukaya, Y. Nakatsukasa, Y. Yanagisawa, and Y. Yamamoto, "Choleskyqr2: a simple and communication-avoiding algorithm for computing a tall-skinny qr factorization on a large-scale parallel system," in *2014 5th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems*. IEEE, 2014, pp. 31–38.
- [25] S. N. Yeralan, T. A. Davis, W. M. Sid-Lakhdar, and S. Ranka, "Algorithm 980: Sparse qr factorization on the gpu," *ACM Transactions on Mathematical Software (TOMS)*, vol. 44, no. 2, pp. 1–29, 2017.
- [26] R. Andrew and N. Dingle, "Implementing qr factorization updating algorithms on gpus," *Parallel Computing*, vol. 40, no. 7, pp. 161–172, 2014.
- [27] C. Coti, "Scalable, robust, fault-tolerant parallel qr factorization," in *2016 IEEE Intl Conference on Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES)*. IEEE, 2016, pp. 626–633.
- [28] A. Buttari, S. Hauberg, and C. Kotsi, "Parallel qr factorization of block-tridiagonal matrices," 2019.
- [29] I. Corporation, "Intel xeon processor x5680." [Online]. Available: <http://ark.intel.com/products/47916/>.
- [30] N. Corporation, "Tesla k20x gpu accelerator," 2012. [Online]. Available: <http://www.nvidia.com/content/PDF/kepler/Tesla-K20X-BD-06397-001-v05.pdf>.
- [31] —, "Tesla c2075 and tesla c2070 computing processor board," 2010.
- [32] —, "Nvidia geforce gtx 460," 2011.

# VIPEye: Architecture and Prototype Implementation of Autonomous Mobility for Visually Impaired People

Waqas Nawaz<sup>1</sup>, Khalid Bashir<sup>2</sup>, Kifayat Ullah Khan<sup>3</sup>, Muhammad Anas<sup>4</sup>, Hafiz Obaid<sup>5</sup>, Mughees Nadeem<sup>6</sup>  
Dept. of Computer and Information Systems, Islamic University of Madinah, Al-Madinah, KSA<sup>1,2</sup>  
Fast Data Science (FDS) Lab, Dept. of Computer Science,  
National University of Computer and Emerging Sciences (FAST-NUCES), Islamabad, Pakistan<sup>3,4,5,6</sup>

**Abstract**—Comfortable movement of a visually impaired person in an unknown environment is non-trivial task due to complete or partial short-sightedness, absence of relevant information and unavailability of assistance from a non-impaired person. To fulfill the visual needs of an impaired person towards autonomous navigation, we utilize the concepts of graph mining and computer vision to produce a viable path guidance solution. We present an architectural perspective and a prototype implementation to determine safe & interesting path (SIP) from an arbitrary source to desired destination with intermediate way points, and guide visually impaired person through voice commands on that path. We also identify and highlight various challenging issues, that came up while developing a prototype solution, i.e. VIPEye - An Eye for Visually Impaired People, to this aforementioned problem, in terms of task's difficulty and availability of required resources or information. Moreover, this study provides candidate research directions for researchers, developers, and practitioners in the development of autonomous mobility services for visually impaired people.

**Keywords**—Safe and Interesting Path (SIP); Visually Impaired People (VIP); autonomous; mobility; computer vision; path guidance; VIPEye prototype; navigation; graph mining

## I. INTRODUCTION

Outdoor movement of a visually impaired person is limited and hard due to complete or partial short-sightedness [1]. Any non-impaired person can assist them while going out, however, this solution is not always viable. Fortunately, with the recent developments in technology, various innovative methods of assistance have been proposed based on ultrasonic devices, smart cameras, transplantation of robotic eyes, blind navigation system using GSM and RFID and voice navigation system among others [2]. One of the key essence of these developments is to improve the computer-vision technology to fulfill the visual needs of a blind person. To this end, various researchers are using vision based methods [3] focusing issues like sidewalk assessment, zebra crossing spotter, public cataloging of various objects like trees, detecting curb ramps, and public transit accessibility to facilitate smooth and safe walk of blind persons.

Computer vision methods [4]–[8] are also in use for a similar problem of robot vision where the task is to make the robot aware of its surrounding, to help it move around. This is a wide area of research known as Simultaneous Localization and Mapping (SLAM). SLAM uses a range measurement device which may be a laser scanner, sonar, vision or any

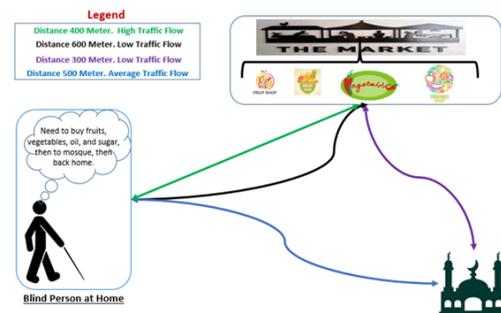


Fig. 1. A blind person wants to go market, then to mosque, and back home. There are two routes to the market. The green colored is short but busy road, whereas, black colored is longer but empty road. Similarly, there are dedicated fruit shops in the market and a combined shop of vegetables and fruits. To provide an optimal solution, the black colored path and a combined shop should be suggested as a navigation path.

other capable device. In case of vision based devices for range measurement, the problem is categorized as Visual SLAM (vSLAM) and is an extensively researched topic [5]. Combining vSLAM with topological representations of the environment [7] is an effort to yield improved and robust long-term navigation in real world environments for visually impaired people. However, these solution need improvement in the areas of 3D scene understanding, computations of dense flow scene representations for real time implementations and topological mapping of the environment.

Scene understanding is imperative for smooth navigation based on visual data [9]. This involves range measurement, detection and identification of objects from a scene. Smart phones are equipped with stereo cameras which enables calculating distances to objects [10] for range measurement. Object detection and identification is computationally expensive and traditional approaches fail to produce desired level of accuracy. Recent advances in Computer Vision using Deep Learning methods [11] provide good detection and recognition results but at the cost of computation resources. Another family of detection and recognition frameworks called the Unified (One Stage) frameworks also makes use of deep networks and provide significant improvement in terms of speed. We focus on the application of Unified Frameworks for navigation of the visually impaired. Therefore, these detection and recognition frameworks are not applicable to the problem under consider-

ation.

The integration of scene understanding solutions with proprietary or crowd-sourced geospatial mapping services like Google Maps or OpenStreetMap (OSM)<sup>1</sup> respectively can assist the movement of a blind person from a given source to destination(s) in the outdoor environment. However, we understand that such combination requires further developments in visual technology and investigation over path identification problem. For instance, Google Maps or similar applications suggest the shortest path(s) between source and destination without considering VIP friendly parameters such as higher traffic flow, rush of vehicles, and presence of obstacles. An example scenario is depicted in Fig. 1 for clarity. Finding a suitable path for a pedestrian, having visual disability and considering dynamics of the surroundings, is a challenging task [12]. The multi-criteria shortest path (MCSP) [13] problem and Skyline path computation [14] are similar to determine a suitable path considering various factors. Despite the effectiveness of MCPS and Skyline approaches for general mobility, it is not readily applicable to our problem due to involvement of scene understanding for guidance, which is hard to model.

In this article, we propose a system called SIP-vSUN (Safe and Interesting Path with Visual Scene Understanding) in order to help the visually impaired people smoothly navigate in the surroundings. The prospective solution is to be deployed on a smart phone equipped with the required commodity sensors. The computation heavy load of vision based methods are still not easily handled by smart phones, however, there are preliminary works showing the possibility of using vision based method on smart phones [15]. A real time object detection system called PeleeNet [16] has been proposed for mobile devices. Our proposed system relies on extending the existing works in object detection and recognition for its accuracy and speed and integrating it with range measurement of identified objects thus making it capable of working with required precision on hand held devices and additionally defines the safe and interesting paths for the visually impaired. Our SIP-vSUN system on a smart phone communicates with a blind person through off-the-shelf speech system, which supports voice commands for inputs and voice instructions for guidance and navigation. There are existing libraries like ARKit [17] and ARCore [18] on iOS and Android platforms respectively to realize the implementation of visual components of this system. PyTorch's [19] cross platform libraries will be used for the implementation of the deep network on mobile devices.

The potential contributions of this research work are as follows.

- We are aiming to envision the research community and practitioners on utilize existing tools and techniques, in graph mining and computer vision, towards the development of an effective autonomous mobility system. This system assists visually impaired people, who are deprived naturally or by accident, to walk easily in the surroundings.
- This study gives a clear direction to develop reasonable system, as an application on a smart phone,

which can be used autonomously by visually impaired persons.

- This study explores methods, balanced in their computational needs and performance, in the domain of object detection, recognition, and graph mining.
- The development of a prototype is the realization of our efforts in this domain, which has the potential to guide visually impaired people in a selected scenario and has many expects to be improved in near future.

The rest of the content is organized as follows. Section II explains the proposed methodology where we discuss the notion of safe and interesting path discovery and visual scene understanding for visually impaired people. The design and implementation perspective of the proposed application, i.e. VIPEye, is presented in Section III. Section III-C provided a brief comparative analysis of the the proposed application with the existing similar applications. The potential research directions in this domain are outlined in Section IV and conclusion is drawn in Section V.

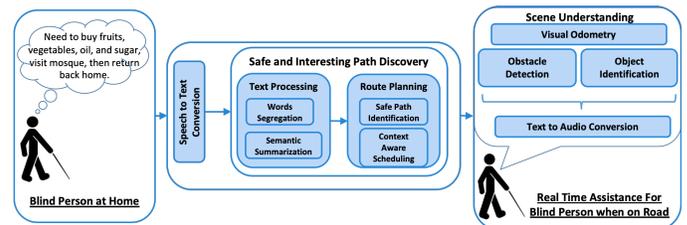


Fig. 2. Proposed System Architecture

## II. SAFE AND INTERESTING PATH WITH VISUAL SCENE UNDERSTANDING (SIP-vSUN)

In this section, we present the details of our proposed framework and Fig. 2 depicts the overall system architecture. The processing starts when a person plans to do something like buying items and/or visiting different places. The person talks to the system installed on his smart phone, where off-the-shelf speech to text component of SIP transforms the audio message into text. Next semantic summarization, route planning, and context-aware evaluation using textual information received from the speech engine are performed. The result is a path to be used by the system for navigation. The navigation is assisted through the visual component of our system in terms of real time object detection, recognition and range measurement. When the user is out in the open, the visual and navigation components interact with one another to update the path or guide in real time based on the environmental inputs.

### A. Safe and Interesting Path (SIP) Discovery

We now discuss the path planning component of our system. We present its example in Fig. 3 and 4 to deliver the overall idea.

Initially, we transform the obtained audio message into text, as guided in [7]. In order to obtain the action-items, we need to filter the important words from the message. Finally, we perform semantic summarization to group the contextually

<sup>1</sup><https://www.openstreetmap.org>

**Algorithm 1: SIP-vSUN Algorithm**

```

input : Destination  $d$ , List of Criteria for Path Selection  $C = \{c_1, c_2, \dots, c_n\}$ 
output: Comfortable navigation of a Visually Impaired Person on  $SIP$ 
1 Current location selected as starting point  $s$ ;
  // Candidate Path List
2 Retrieve set of candidate paths  $P = \{p_1, p_2, \dots, p_k\}$ 
  from  $s$  to  $d$  using navigation APIs;
  // Identify  $SIP$  from candidate paths
3 for each  $p_i \in P$  do
4   Evaluate  $p_i$  for  $\forall C$ ;
   //  $weight(p_i)$  is sum of scores for path  $p_i$ 
5   if  $weight(p_i) \leq weight(p_j)$  then
6     Choose  $p_i$  as  $SIP$ ;
7   end
8 end
  // Path Navigation
9  $vSUN$  for navigation on  $SIP$ ;

```

similar nouns together. For instance, fruits, vegetables, oil, sugar, stated in Fig. 3, are eatable items, whereas home and mosque are venues to find path for. In this way, we group the items together into relevant categories. Such grouping then helps to have an optimal path for smooth navigation. Here the challenging issues are Online Analytical Processing (OLAP) style aggregation and vocabulary of locality.

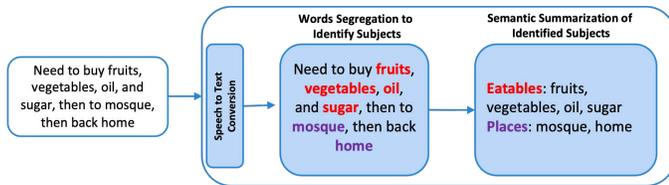


Fig. 3. Text Processing Module

The path identification component comes into action when we have items summarized into relevant grouping. Based on this information, the path discovery module identifies the current location of the person and the places to visit. During this processing, it communicates with 3rd party services (such as Google Maps, OpenStreetMap, MapBox) to determine safe and interesting path among possible set of paths between source and destination. The terms *safe* and *interesting* refer to least number of obstacles and hurdles for VIP mobility and most point of interests covered in short distance by the proposed path. This component also aims to perform clustering of actionable items like identifying the actions which can be performed together. For instance, fruits and vegetables can be purchased together from same shop. Similarly, oil and sugar can be bought together from same grocery store. On the other hand, mosque and home, from the example, are different workable options, so they should not be mixed with shopping agenda. In this way, the system aims to perform hierarchical clustering to identify an interesting path. In our case, we have challenge of hierarchical entity clustering.

The context-aware scheduling is very important in a sense

that a blind person is ignorant of the ground realities happening around. For instance, road construction and maintenance works are in progress, which makes it difficult and less interesting to walk. Similarly, the person is short of time to miss the prayer in the mosque, so he can adjust configuration. In this case, we have multiple criteria to fulfil prior to determine a best route, which is similar to skyline computation in literature. Therefore, the challenge is skyline and multi-criteria path formulation and its efficient computation. For initial experiments and simulations, we expect to utilize the GeoLife - GPS trajectories dataset by Microsoft Research Asia that contains outdoor movements of 17,621 trajectories of duration of 48,000 hours.

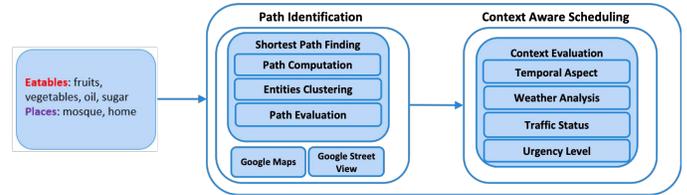


Fig. 4. Route Planning Module

**B. Visual Scene Understanding (vSUN)**

We present an overall architecture of the Visual Scene Understanding module in Fig. 5. All the inputs required are readily provided by the sensors already available in current smart-phones making it an excellent choice for this application. In addition to the Visual Odometry for range measurement, the proposed system is assisted by incorporating object detection and recognition for better scene understanding. In the following, we present our approach to address aforementioned challenges.

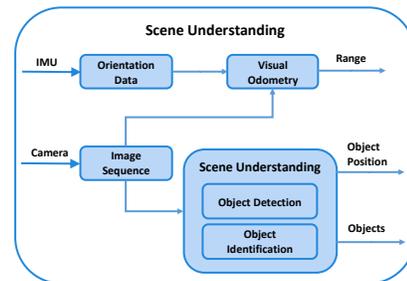


Fig. 5. The block diagram of Scene Understanding Module

Improving scene understanding, with regards to the specific problem of navigation for visually impaired people, involves identification of obstacles, moving objects, pedestrians, sidewalks, zebra crossings and roads. In addition to the identification of static and moving objects, our focus is on tracking moving objects in the scene to handle crowded environments. A better scene understanding [9] eventually benefits localization service for the visually impaired person within the environment.

Detection and identification of object and computation of bounding boxes around them in the scene is a challenging task to be completed in real time. In addition, there are problems associated with real world imagery giving spurious results. To this end, PeleeNet [16] is neural network based solution

for real time object identification and tracking. However, the problem with this approach is that it is trained on standard datasets which have generic classes and do not cover the range of objects including zebra crossing, sidewalks, roads, obstacles etc. In order to address this limitation, we propose to use *transfer learning* in two phases. In the initial phase with limited data, we propose to use the PeleeNet's convolutional network as a fixed feature. This essentially means removing the last fully connected layer and training a classifier on the convolutional codes received from the convolutional network. In the final phase when more data is gathered we propose to initialize the network with pre-trained weights and fine tune the convolutional neural network weights using our own dataset. This phase wise strategy provides the desired level of accuracy initially when there is not much data and improves even further as more data is accumulated. The fine tuning of the network weights can not be done in real time and should be done offline. PeleeNet not only provides us with an advantage of close to real time recognition of objects, it also has a model size which is two thirds of the model size used by similar solutions like MobileNet [20]. Objects identified are used by the Visual Odometry component for range measurement.

### III. VIPEYE PROTOTYPE DESIGN AND IMPLEMENTATION

In this section, we elaborate on the prototype [21] as guidance application (a limited version of the actual application) for VIPs. Initially, we discuss the design and implementation details such as the path planning and visual scene understanding towards autonomous mobility along with user interfaces. Afterwards, we provide an abstract comparative analysis on various existing systems and applications developed for VIPs in path guidance context.

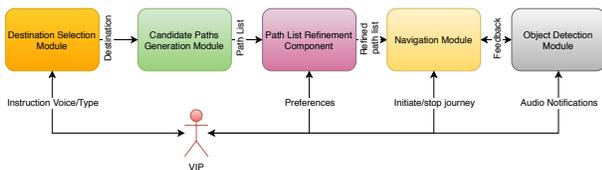


Fig. 6. Main components of our prototype application

#### A. Design Considerations

There are five main components involve in our implemented prototype application for VIPs. The block diagram shows the sequentially dependent modules in Fig. 6. Initially, the blind person specifies the destination through voice command or typing the name manually with the help of accessibility service in the application, which is the task of destination selection module. We assume that the application is started prior to this step and gets the current location of the user. Once, the destination is selected, the next module determines a set of candidate paths (i.e path list) from source to destination location. We can move from one location to another by following different paths. However, choosing one from the given set of paths is non-trivial and subjective task given the circumstances. Therefore, in the refinement module, we take the preferences from user to filter and choose an appropriate path from path list. Currently, the preferences are limited to various point of interests such as hospital, mosque,

school, university, etc. Multi-criteria shortest path finding is a well studied problem and is used as a candidate solution for path selection based on user preferences. Navigation and object detection modules can work and coordinate together once the user starts navigation on the selected path. These modules notify user in terms of standard audio messages similar to general purpose navigation applications. The object detection module covers limited set of objects in this prototype and subject to cover diverse kind of objects in next version.

The choice of an appropriate interaction interfaces for visually impaired users is also critical. Our prototype supports accessibility option with native support of the Android platform and designed simple interfaces to increase its usability for VIPs. User interfaces of our prototype with visual and non-visual aspects are presented in Fig. 8 and 7, respectively. Notice that our prototype application also allows the user to manually capture an image and run a deeper model on a better quality image for an improved scene understanding. The default operation of the application is to automatically capture images every  $1/4^{th}$  of a second and run it through a standard model to meet time constraints.

#### B. Development Perspective: Libraries and Technologies

We dedicate this section to elaborate on development aspects of our prototype system [21] as a mobile application implemented on Android platform.

The collection of static model elements such as classes, types, contents, and their relationships is presented in Fig. 9. The important functions of the project include *OnActivityResult* (to save the entered string as a destination), *getRoute* (to get the route that leads from source to destination), *getRouteCoordinates* (to save the whole route that leads from source to destination), *makeGeocodeSearch* (searches for places on route that leads from source to destination), *updateRoute* (updates the multiple routes leading from source to destination to a best route that leads from source to destination), *takePicture* (to take picture of objects that comes in user's path), *detectObject* (to check that the picture taken belongs to which category), and *speak* (To speak whatever it is passed to the function as an argument such as places names).

We briefly highlight important activities involved in our application that include path finding, navigation and object detection. In order to find path on user's way, the user enters the destination, the application checks whether the destination is valid or not. If valid then it asks to enter criteria and calculates best path for navigation. Else it asks the user again to enter the destination. To navigate on the path that leads to destination from source, the user selects the path that the application has returned in the activity called path finding. Our application tells user that path has been selected and also inform the estimated time to reach the destination. The user starts the journey and application guides the user in that journey on turn basis to reach the destination. The object detection activity involves the scene capturing and understanding. To detect objects, that comes in user's path, our application takes the picture of the scene with objects for every  $1/4^{th}$  of a second, categorize that picture and notifies the user if their is a detected obstacle on the way. User has an option of getting a more information about the obstacle by pressing the capture button as shown in Fig.



```
1 public void onClick(View view)
2 {
3     Intent intent = new PlaceAutocomplete.
4     IntentBuilder().build();
5     startActivityForResult(intent,
6     REQUEST_CODE_AUTOCOMPLETE);
7 }
```

Listing 1: OnClickListner pseudo code for the search box

- **Navigation Page:** The navigation page shows multiple paths leading to the destination (without any criteria). From here the user goes to the criteria page and enters up-to five criteria places in order of priority. Then the navigation page shows the best chosen path based on criteria.

```
1 private void getRoute(Point origin, Point
2 destination)
3 {
4     NavigationRoute.builder(this).origin(
5     origin).dest(dest).build().getRoute();
6 }
```

Listing 2: Getting routes from origin to destination

Then using the above function we also get the places on each route using makeGeocodeSearch function. By pressing the Start Navigation button the application guides the user to their destination. And by pressing the Start Walking button the user is be able to detect obstacles in their path. We achieved this functionality through MapBox libraries such as Mapbox LatLng, Mapbox MapView, Mapbox MapboxMap, Mapbox Style, Mapbox NavigationLauncher, Mapbox NavigationRoute, and Mapbox DirectionsRoute.

- **Navigation and Obstacle Detection Page:** The obstacles detection page shows the live stream, the current photo and the obstacles detected in the current photo. We used OtoliaStudios CameraView and Firebase FirebaseVisionObjectDetector for navigation and object detection purposes.

```
1 /*Live Stream*/
2 cameraView = findViewById(R.id.
3 id_cameraview);
4 cameraView.setLifecycleOwner(this);
5 //Take Picture (when user presses the
6 volume up key)
7 cameraView.takePicture();
8 cameraView.addCameraListener(
9 onPictureTaken(result){detectObjects(
10 result)});
11
12 /*Object Detection*/
13 void detectObjects(result)
14 {
15     objectDetector.processImage(image) {
16     public void onSuccess(List<
17     FirebaseVisionObject> detectedObjects){
18     // Displays the name(s) of the detected
19     object(s)
20     }};
21 }
```

Listing 3: Partial code on navigation and object detection

- **Non-visual Aspects:** It is obvious that VIPs are unable to see the interfaces of our proposed application rather communicate with it via non-visual senses such as speaking and hearing. Therefore, we provided an audio-based interaction and feedback mechanism implemented through existing libraries, i.e. AndroidSpeech TextToSpeech and AndroidSpeech's SpeechRecognizer, to covert from audio-to-text and vice verse.

```
1 /*Text-To-Speech*/
2 public void speak(String data)
3 {
4     textToSpeech.speak(data,TextToSpeech.
5     QUEUE_FLUSH,null);
6     // can be modified as per requirements
7 }
8 /*Speech-To-Text*/
9 Intent intent = new Intent(
10 RecognizerIntent.ACTION_RECOGNIZE_SPEECH)
11 ;
12 // more options can be added to this
13 function
14 startActivityForResult(intent, REQ_CODE);
15 void onActivityResult(int requestCode,
16 int resultCode, Intent data) {
17     // Check requestCode and resultCode for
18     confirmation
19     ArrayList result =
20     data.getStringArrayListExtra(
21     RecognizerIntent.EXTRA_RESULTS);
22     // Display or perform action according
23     to what was spoken
24 }
```

Listing 4: Non-visual interaction mechanism for VIPs

### C. Comparative Analysis

We describe few mostly related applications and systems used to assist in the navigation of the blind people as also highlighted in [22]. Recent surveys [23] [24] [25] [26] discussed various kinds of systems and applications developed for visually impaired people. These systems and applications are either the outcomes of research work in this domain as prototypes or entirely as business oriented products. These solutions are categorized into general purpose object recognition based [27]–[32], navigation related [33]–[41], and specialized systems and devices [42]–[46]. The discussion in this section highlights key aspects these systems cover in terms of guiding visually impaired people and also provide insights to improve them for better guidance.

We provide an overview of comparative analysis for various applications and systems with respect to predefined criteria in Table I. The criterion is selected based on suitability of these applications and systems for visually impaired people towards autonomous navigation, which is the main focus of this article. GPS criterion depicts whether the application or system supports location-aware services or not. These applications interact with the users through either audible instructions or haptic feedback. Not all applications and systems analyze images and video data to observe the surroundings and develop an understanding for the VIPs. The understanding of the surroundings depends upon effectively detecting and identifying the objects. Object detection is a key step towards scene understanding, where the majority of the techniques under

consideration detect various kinds of objects from individual images. However, few approaches detect objects continuously from a series of images taken through built-in camera. Most of the applications are available free of cost on either one or both platforms, i.e. iOS or Android. It is critical for any application or system to be autonomous so that a visually impaired person feels more independent. This is one of the major factors to influence the usability of any system or application, as depicted in the last column of the comparison table. It is interesting to know that existing solutions are either more focused on non-visual information, e.g. GPS signals and navigational semantics, or visual content to understand the surroundings. In our understanding, visual content analysis is a complex task compared with the analysis of non-visual information towards navigational guidance, therefore, many aspects in visual content analysis are still under consideration.

#### IV. RESEARCH DIRECTIONS

The potential research directions, while developing an effective solution for visually impaired people towards autonomous navigation, are as follows.

- 1) SIP prediction algorithm and repository: To develop an effective algorithm to predict or identify the safe and interesting paths, which are eventually stored in a repository for future correspondence. The identification of the SIPs through experience will result in a rich repository.
- 2) Intelligent path discovery approach: An approach needs to be developed to discover or explore paths from given source to destination(s). The path discovery problem emphasizes on various criterion defined or customized for impaired person. Defining the representative criterion for impaired person is also major concern.
- 3) Hazard detection and identification in real-time: We need an algorithm to detect various kinds of obstacle/hazards instantly to take respective measures by the impaired person while navigating known/unknown territory. This algorithm expected to utilize vital information (visual & non-visual) to make decisions quickly. The identification of such a useful information is itself a challenging task.
- 4) Context-aware navigation system: A novel approach is required to determine the context from the environment through visual and non-visual data. It helps the users to navigate with contextual information (knowledge of surroundings like passing by superstore, ATM, gas-station, etc.).
- 5) A comprehensive system with smart navigation: This product will be an integration of other essential components together to solve the navigation problem for visually impaired people. This system provides a comprehensive solution to smartly/intelligently navigate through known/unknown territories. This solution will be realized on smart devices (e.g. Smartphones) as these devices are already equipped with essential tools and commodity sensors.

#### V. CONCLUDING REMARKS

In this work, we have conceptualized an overall system as a candidate solution for mobility of visually impaired people in terms of visual scene understanding and graph mining approaches. A prototype implementation is presented and improvements are in progress on algorithms determining safe and interesting path by considering multiple factors associated with blind people. Additionally, We have highlighted various research directions for developers and practitioners to build effective services towards mobility of visually impaired people.

#### ACKNOWLEDGMENT

This work is done under the grant (first Tamayuz program of academic year 1439-1440 AH and research project number is 24/40) received from Deanship of research at Islamic University of Madinah(IUM), Saudi Arabia. We give special thanks to the administration of IUM for their support in every aspect of this work. We would like to thank and Acknowledge the work done by all the stakeholders of this project. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Islamic University of Madinah.

#### REFERENCES

- [1] R. R. Bourne, S. R. Flaxman, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas, J. Keeffe, J. H. Kempen, J. Leasher, H. Limburg *et al.*, "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis," *The Lancet Global Health*, vol. 5, no. 9, pp. e888–e897, 2017.
- [2] M. M. Islam, M. S. Sadi, K. Z. Zamli, and M. M. Ahmed, "Developing walking assistants for visually impaired people: A review," *IEEE Sensors Journal*, vol. 19, no. 8, pp. 2814–2828, 2019.
- [3] R. Jafri, S. A. Ali, H. R. Arabnia, and S. Fatima, "Computer vision-based object recognition for the visually impaired in an indoors environment: a survey," *The Visual Computer*, vol. 30, no. 11, pp. 1197–1222, 2014.
- [4] P. F. Alcantarilla, "Vision based localization: from humanoid robots to visually impaired people," *Electronics (University of Alcalá, 2011)*, 2011.
- [5] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual slam algorithms: a survey from 2010 to 2016," *IPSN Transactions on Computer Vision and Applications*, vol. 9, no. 1, p. 16, 2017.
- [6] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," *Pattern Recognition Letters*, 2018.
- [7] L. F. Rojo, L. Paya, F. Amoros, and O. Reinoso, "Using global appearance descriptors to solve topological visual slam," in *Advanced Methodologies and Technologies in Artificial Intelligence, Computer Simulation, and Human-Computer Interaction*. IGI Global, 2019, pp. 1127–1140.
- [8] E. Arnold, O. Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby, and A. Mouzakitis, "A survey on 3d object detection methods for autonomous driving applications," *IEEE Transactions on Intelligent Transportation Systems*, 2019.
- [9] B. Yang, Z. Lai, X. Lu, S. Lin, H. Wen, A. Markham, and N. Trigoni, "Learning 3d scene semantics and structure from a single depth image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 309–312.
- [10] Y. D. Salman, K. R. Ku-Mahamud, and E. Kamioka, "Distance measurement for self-driving cars using stereo camera," in *proceeding 6th Int. Conf. Comput. informations*. Universiti Utara Malaysia, 2017, pp. 235–242.
- [11] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *arXiv preprint arXiv:1809.02165*, 2018.

TABLE I. COMPARISON OF VARIOUS PATH GUIDANCE APPLICATIONS AND SYSTEMS IN TERMS OF THEIR CAPABILITIES AND THE QUALITY OF SUPPORTED FUNCTIONALITY.

Apps & Systems	GPS	Audio or Haptic	Image	Video	Object Detection	Free	Autonomous	Usability
NavCog3 [33]	No	Yes	No	No	No	Yes	Yes	MED
WayFindr [35]	No	Yes	No	No	No	Yes	Yes	MED
Be My Eyes [27]	No	Yes	Yes	Yes	No	Yes	No	MED
TapTapSee [28]	No	Yes	Yes	No	Yes	Yes	Yes	LOW
iIdentifi [30]	No	Yes	Yes	No	Yes	Yes	Yes	LOW
SeeingAI [31]	No	Yes	Yes	Yes	Yes	Yes	Yes	MED
BeSpecular [32]	No	Yes	Yes	No	No	Yes	No	LOW
Microsoft soundscape [34]	Yes	Yes	No	No	No	Yes	Yes	LOW
BlindSquare [37]	Yes	Yes	No	No	No	No	Yes	MED
Lazarillo [39]	Yes	Yes	No	No	No	Yes	Yes	LOW
Intersection Explorer [36]	Yes	Yes	No	No	No	Yes	Yes	MED
Lazzus [40]	Yes	No	No	No	No	Yes	No	LOW
Arianna Navigatio [38]	No	Yes	No	Yes	Yes	Yes	No	LOW
NaviLens [41]	No	Yes	Yes	No	Yes	Yes	Yes	MED
VIP-Eye (Prototype) [21]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	HIGH
Sunu [42]	No	Yes	No	No	Yes	No	Yes	LOW
OrCam [45]	No	Yes	Yes	No	Yes	No	No	LOW
Blaid [46]	Yes	Yes	Yes	No	Yes	No	Yes	MED
eSight [43]	No	Yes	Yes	No	Yes	No	Yes	MED
iBionics [44]	No	Yes	Yes	No	Yes	No	Yes	MED

- [12] J. E. Froehlich, A. Brock, A. Caspi, J. Guerreiro, K. Hara, R. Kirkham, J. Schönig, and B. Tannert, "Grand challenges in accessible maps," *Interactions, Association for Computing Machinery*, 2019.
- [13] N. Shi, S. Zhou, F. Wang, Y. Tao, and L. Liu, "The multi-criteria constrained shortest path problem," *Transportation Research Part E: Logistics and Transportation Review*, vol. 101, pp. 13–29, 2017.
- [14] K. Deng, X. Zhou, and H. Tao, "Multi-source skyline query processing in road networks," in *2007 IEEE 23rd international conference on data engineering*. IEEE, 2007, pp. 796–805.
- [15] T. Y.-H. Chen, L. Ravindranath, S. Deng, P. Bahl, and H. Balakrishnan, "Glimpse: Continuous, real-time object recognition on mobile devices," in *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2015, pp. 155–168.
- [16] R. J. Wang, X. Li, and C. X. Ling, "Pelee: A real-time object detection system on mobile devices," in *Advances in Neural Information Processing Systems*, 2018, pp. 1963–1972.
- [17] "Arkit 3: Get ready for the latest advances in augmented reality," 2020. [Online]. Available: {<https://developer.apple.com/augmented-reality/>}
- [18] "Arcore: With arcore, build new augmented reality experiences that seamlessly blend the digital and physical worlds." [Online]. Available: {<https://developers.google.com/ar/>}
- [19] "Pytorch: An open source machine learning framework that accelerates the path from research prototyping to production deployment." [Online]. Available: {<https://pytorch.org/>}
- [20] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.
- [21] "Visually impaired persons' eye (vip-eye): Safe and smooth navigation for visually impaired people using smart phone." [Online]. Available: {<https://github.com/ASHIR14/VIP-Eye>}
- [22] NoisyVision.org, "Apps and devices for blind and visually impaired people." [Online]. Available: {<https://www.noisyvision.org/2018/08/01/app-and-devices-for-blind-and-visually-impaired/>}
- [23] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," *Pattern Recognition Letters*, 2018.
- [24] W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions," *Sensors*, vol. 17, no. 3, p. 565, 2017.
- [25] S. Real and A. Araujo, "Navigation systems for the blind and visually impaired: Past work, challenges, and open problems," *Sensors*, vol. 19, no. 15, p. 3404, 2019.
- [26] A. Darvishy, "Assistive technologies: short overview and trends," in *1st Swiss conference on barrier-free communication, Winterthur, 15-16 September 2017*. ZHAW Zürcher Hochschule für Angewandte Wissenschaften, 2018, pp. 52–55.
- [27] H. J. Wiberg, "Be-my-eyes: Bringing sight to blind and low-vision people." [Online]. Available: {<https://www.bemyeyes.com/>}
- [28] "Taptapsee: Assistive technology for the blind and visually impaired." [Online]. Available: {<https://taptapseeapp.com/>}
- [29] "Cloudsight: Image processing library." [Online]. Available: {<https://cloudsight.ai/>}
- [30] "identifi: Mobile application for partially sighted and visually impaired people." [Online]. Available: {<http://getidentifi.com/>}
- [31] A. Seeing, "Talking camera app for those with a visual impairment," *Retrieved September*, vol. 15, 2018. [Online]. Available: {<https://www.microsoft.com/en-us/ai/seeing-ai>}
- [32] "Bespecular: Let blind people see through your eyes." [Online]. Available: {<http://www.bespecular.com/en/>}
- [33] D. Sato, U. Oh, K. Naito, H. Takagi, K. Kitani, and C. Asakawa, "Navcog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment," in *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 2017, pp. 270–279.
- [34] "Microsoft soundscape: A map delivered in 3d sound." [Online]. Available: {<https://www.microsoft.com/en-us/research/product/soundscape/>}
- [35] "Wayfindr: Indoor audio wayfinding and audio navigation." [Online]. Available: <http://www.wayfindr.net/wp-content/uploads/2018/08/Understanding-Audio-Wayfinding-and-Audio-Based-Navigation.pdf>
- [36] "Intersection explorer: Intersection explorer helps blind users explore neighborhoods." [Online]. Available: {<https://www.appbrain.com/app/intersection-explorer/com.google.android.marvin.intersectionexplorer>}
- [37] "Blindsquare: Pioneering accessible navigation – indoors and outdoors." [Online]. Available: {<https://www.blindsquare.com/>}
- [38] E. T. from the arXiv, "App turns smartphone into virtual cane for the blind." [Online]. Available: {<https://www.technologyreview.com/s/523401/app-turns-smartphone-into-virtual-cane-for-the-blind/>}
- [39] "Lazarillo: Orientation and autonomy for people with visual disability." [Online]. Available: {<https://lazarillo.cl/en/>}
- [40] "Lazzus: Navigate the city and explore your surroundings independently." [Online]. Available: {<http://www.lazzus.com/>}
- [41] "Navilens: A brand new system of bidimensional, or bidi, codes." [Online]. Available: {<https://www.navilens.com/>}
- [42] "Sunuband: It's your world. explore it with the sunu band." [Online]. Available: {<https://www.sunu.com/en/index.html>}
- [43] "esight: We make glasses that let the visually impaired see." [Online]. Available: {<https://www.esighteyewear.eu/>}
- [44] "ibionics diamond eye: How a local startup is building a better bionic eye." [Online]. Available: {<https://ottawacitizen.com/news/local-news/a-clear-view-how-a-local-startup-is-building-a-better-bionic-eye>}
- [45] "Orcam myeye 2: The most advanced wearable assistive technology for the blind and visually impaired, provides independence by conveying visual information, audibly." [Online]. Available: {<https://www.ocr.com/en/myeye2/>}
- [46] "Toyota project blaid: Wearable technology to help the blind and visually impaired." [Online]. Available: {<https://www.toyota.co.uk/world-of-toyota/stories-news-events/toyota-project-blaid>}

# A High Performance System for the Diagnosis of Headache via Hybrid Machine Learning Model

Ahmad Qawasmeh<sup>1</sup>  
Dept. of Computer Science  
The Hashemite University  
Zarqa, Jordan

Noor Alhusan<sup>2</sup>, Feras Hanandeh<sup>3</sup>  
Dept. of Computer Information System  
The Hashemite University  
Zarqa, Jordan

Maram Al-Atiyat<sup>4</sup>  
Department of Family Medicine  
Jordan University of Science  
and Technology, Jordan

**Abstract**—Headache has been a major concern for patients, medical doctors, clinics and hospitals over the years due to several factors. Headache is categorized into two major types: (1) Primary Headache, which can be tension, cluster or migraine, and (2) Secondary Headache where further medical evaluation must be considered. This work presents a high performance Headache Prediction Support System (HPSS). HPSS provides preliminary guidance for patients, medical students and even clinicians for initial headache diagnosis. The mechanism of HPSS is based on a hybrid machine learning model. First, 19 selected attributes (questions) were chosen carefully by medical specialists according to the most recent International Classification of Headache Disorders (ICHD-3) criteria. Then, a questionnaire was prepared to confidentially collect data from real patients under the supervision of specialized clinicians at different hospitals in Jordan. Later, a hybrid solution consisting of clustering and classification was employed to emphasize the diagnosis results obtained by clinicians and to predict headache type for new patients respectively. Twenty-six (26) different classification algorithms were applied on 614 patients' records. The highest accuracy was obtained by integrating K-Means and Random Forest with a migraine accuracy of 99.1% and an overall accuracy of 93%. Our web-based interface was developed over the hybrid model to enable patients and clinicians to use our system in the most convenient way. This work provides a comparative study of different headache diagnosis systems via 9 different performance metrics. Our hybrid model shows a great potential for highly accurate headache prediction. HPSS was used by different patients, medical students, and clinicians with a very positive feedback. This work evaluates and ranks the impact of headache symptoms on headache diagnosis from a machine learning perspective. This can help medical experts for further headache criteria improvements.

**Keywords**—High performance computing; Clinical Decision Support System (CDSS); machine learning; primary and secondary headache; performance analysis and improvement; headache diagnosis; open medical application

## I. INTRODUCTION

Headache is a common community physical discomfort, which has a negative impact on people's life especially in terms of work productivity and social relations. Headache is a pain in the various parts of head, which is categorized into two major types: primary headache and secondary headache. Primary headache consists of three main types:

- Tension-Type headache (TTH) [1]: it is a very common, mild to moderate head pain, which often feels like a tight band around the head but it can also be intense. Its causes are not understood very well.

- Cluster headache [2]: it is one of the most painful headache types. Patients may wake up at night because of the intensity in one side of head and/or around one eye.
- Migraine headache [3]: it is a recurring severe headache, which usually affects one side of the head accompanied with nausea, visual disturbances, and sound and light intolerance.

In contrast, secondary headache [4] is caused by or occurred secondarily to a long list of other conditions. The most common of which is medication-overuse headache. This type requires further medical examinations for better diagnosis.

Patients can buy over-the-counter headache pain medicines, which might be harmful in some cases. On the other hand, the increase of the numbers of patients put significant pressure on clinicians and healthcare facilities. This pressure may lead to unexpected medical errors. There is an increased risk of depression that may affect patients suffering from severe headaches. Moreover, some types of headache may cause silent death under certain circumstances because of different reasons such as lack of healthcare, wrong diagnosis, or getting large doses of painkillers. Regrettably, the process of headache diagnosis is not trivial because of the similarity in all headache types' symptoms. The short time spent by doctors in hospitals on each patient's case may arise medical errors [5] because of the large number of visiting patients. According to [6], [7], headache is one of the main reasons for medical consultation in the primary care units and neurological clinics. Furthermore, non-specialist people cannot detect the difference between headache symptoms, which could put their lives at risk. In some cases, such as headaches caused by high blood pressure or low blood sugar, painkiller is useless. A big challenge facing the healthcare industry is the quality of service. Quality of service means diagnosing the diseases correctly while providing effective treatment to patients. Celik et al. [5] showed that poor diagnosis can lead to disastrous unacceptable consequences.

Therefore, it is essential for research scientists to move towards more efficient diagnosis methods for the sake of better healthcare. It is necessary to find solutions to reduce the bad consequences of headache by increasing patient's awareness of his health.

The revolution of computer-based systems and technologies has led to the development of decision support systems

(DSS), which can give useful assistance in many fields. The Clinical Decision Support System (CDSS) [8] is an interactive software that helps specialists and others in the medical field to make correct decisions. CDSS has been successful in improving the decision of healthcare institutions and clinicians. Dong et al. [6] showed many existing evidences indicating that DSSs, mostly the guideline-based CDSSs, can be super useful in improving medical decision making.

The main motivation of this paper lies in overcoming some critical limitations in existing systems such as missing some headache types, lacking medical-patient tool, low prediction accuracy, using many attributes, lacking analysis of headache symptoms, etc. In order to overcome these limitations, we developed a user-friendly, safe and accurate high performance web-based Headache Prediction Support System (HPSS) focusing on the main headache types while reducing the number of classification attributes. HPSS is based on machine learning and relies on measurement attributes that can safely distinguish between the aforementioned main headache types. HPSS can be used by patients, medical students, and even clinicians for preliminary guidance with less time and effort, while obtaining high accuracy.

The major contributions of this work are listed below:

- Reduce the number of selected attributes, compared with the previous studies, through which we can accurately classify the four main headache types (Migraine, Tension Type, Cluster, and Secondary Headache), while maintaining high prediction accuracy.
- Collect a sufficient number of headache patients' records to create a sample that can be used to train a headache classifier.
- Employ a hybrid machine learning algorithm that consists of clustering and classification to train a classifier, while using 26 different classification algorithms.
- Create a Headache Prediction Support System (HPSS), which integrates a web-based user interface with a machine learning prediction model, so that HPSS can be easily used by patients, medical students and even clinicians as a preliminary guidance for better headache diagnosis.
- Test HPSS by allowing patients and clinicians to use it and check its usefulness and accuracy.

The rest of the paper is organized as follows: Section II describes the related research work in this area. Section III presents the methodology that was employed to build our Headache Prediction Support System (HPSS). Section IV explains the components of HPSS and demonstrates its user interface. Section V gives a comparison study of the different available headache diagnosis systems. Section VI provides concluding remarks and highlights future directions.

## II. RELATED WORK

Computerized CDSS has become an excellent choice to manage headaches because of its usefulness in solving complex medical problems, reducing medical errors [9], as well as improving patient-clinician cooperation [10]. The development of CDSS for the diagnosis of headache has long been a significant research topic. Most existing headache DSSs are based on

the International Classification of Headache Disorders (ICHD) criteria [11], [6].

Maizels et al. [12] developed an online Clinical Headache Assessment Tool (CHAT) with a focus on migraine headache. However, the tool targeted patients not clinicians. Yin et al. [13] developed a computerized headache guideline method using the SAGE (Standards-based sharable Active Guideline Environment) [14] module in addition to developing a DSS for the diagnosis of headache disease. Dong et al. [6] improved the previous work by developing a headache CDSS based on ICHD-3 beta and validating it via 543 headache patients. Although the CDSS gave high accuracy ratios in the diagnosis of some types of headaches, it still depended on the modeling method of SAGE, where the computerized clinical guideline representation module of headache diagnosis cannot be directly executed by computer. In the aforementioned two studies, the SAGE module was used to summarize the logical comparison expression in headache diagnosis. In contrast, this work depends on machine learning algorithms to diagnose main headache types. Krawczyk et al. [15] proposed a DSS for automatic classification of primary headaches by combining classification models with feature selection algorithms to better differentiate between TTH and migraine with aura or without aura. They used 6 machine learning algorithms for testing. The focus was on migraine, tension-type headache, and other headaches. The research was conducted on 1022 employees of both sexes between the ages of 20-65 via a questionnaire consisting of questions based on the ICHD-2 criteria. They found that headache is present in 579 (56.65%) of the employees, while the accuracy of classification was 81.02%. In comparison, our model is based on the latest ICHD-3 criteria [6] and was tested against 26 different classification algorithms while using 9 different attribute selection methods. Moreover, we developed an open-source and user-friendly web application that can be used by both patients and clinicians for better early diagnosis.

In case of reasoning, Yin et al. [16] developed a hybrid CDSS for primary headache disorder diagnosis by combining both rule-based reasoning (RBR) and case-based reasoning (CBR) models in order to simulate a headache specialist's thinking process. Their proposed approach achieved 95.24% for a probable migraine, 95% for probable tension-type headache, and 80% for probable cluster headache. In contrast, our work employs a hybrid model to diagnose migraine, tension-headache, cluster headache, in addition to secondary headache, while maintaining a better migraine accuracy of 99%. Yin et al. exploited their hybrid reasoning model by proposing a new CDSS based on CBR to solve the similarities between a probable migraine (PM) and probable tension-type headache (PTTH) [17]. In this research, 74 symptoms were selected for diagnosis, while the accuracy of PTTH was low 77%. In contrast, this work uses 19 symptoms only to diagnose migraine, tension-headache, cluster headache, and secondary headache, while maintaining high prediction accuracy.

Aljaaf et al. [18] created a synthetic dataset based on (ICHD-2), which means that they filled the dataset with the ideal symptoms for each instance. Their data set consists of 900 instances and 8 attributes, 66.67% of them are for patients diagnosed with primary headache disorders (migraine with and

without aura, tension type headache and cluster headache), while the others instances is for normal persons. Although they use 8 attributes, some of these attributes consist of multiple symptoms depending on the occurrence or nonoccurrence of all these symptoms together. This means that they actually have more than 8 real attributes. The study tested seven machine learning algorithms. The best obtained accuracy was 96.67%. In comparison, we collect our dataset by asking real patients to fill a headache questionnaire based on what they feel. In addition, from a medical point of view, we excluded migraine with aura to keep our application safe to be used by patients as migraine with aura has similar symptoms to stroke. Celik et al. [19] presented a website expert system that evaluates the classification accuracy of the ant colony optimization (ACO) algorithm for the primary headaches diagnoses via a questionnaire. This questionnaire consists of 40 attributes and is based on the ICHD-2 criteria and was filled by 850 headache patients under the supervision of a neurologist. The system was developed using MySQL database and PHP programming language. The overall accuracy was 96.9412%. The main disadvantage of this system was the large number of questions (40) that must be answered by patients, which affects the easiness of its usage. In addition, Celik et al. [5] created a summary table that consists of different proposed techniques for headache prediction such as [15], [20], and [21].

### III. METHODOLOGY

While it is possible to consult the International Classification of Headache Disorders, 3rd edition (ICHD-3) [4] for headache classification, the complexity of headache subtypes and diagnostic criteria may confuse even neurological specialists who are not familiar with the ICHD criteria. HPSS was also developed to be used by patients to increase their health awareness. Hence, we relied on machine-learning for constructing our Headache prediction Support System (HPSS). We briefly explain the main challenges encountered while constructing our machine learning model. We then describe the machine learning steps that were performed to build a reliable decision support system that can conveniently be used by patients and doctors in the form of web application.

#### A. Challenges in Machine Learning

Machine learning can be used to solve various problems. However, each problem is associated with some challenges and our problem is not an exception. In the following, we list the main encountered challenges:

- Sufficient training data had to be collected for accurate prediction while maintaining high levels of confidentiality. Communicating with real patients was done under the supervision of specialized clinicians after getting the required official approvals.
- The search space we initially had consists of many types of headache as well as many features (symptoms) according to ICHD-3. We reduced the number of symptoms by consulting specialized clinicians and employing different feature selection techniques while obtaining high accuracy. From a medical point of view, we only focused on the main headache types to maintain safety as HPSS can be used by patients.

#### B. Construction of Headache Prediction Support System (HPSS)

The core of our high performance web-based HPSS is based on machine learning techniques that include feature selection, data pre-processing, clustering, and classification algorithms. The basic idea of our approach is to train a machine classifier with a set of relevant records collected from real patients suffering from different types of headaches. Fig. 1 demonstrates the machine learning steps in HPSS.

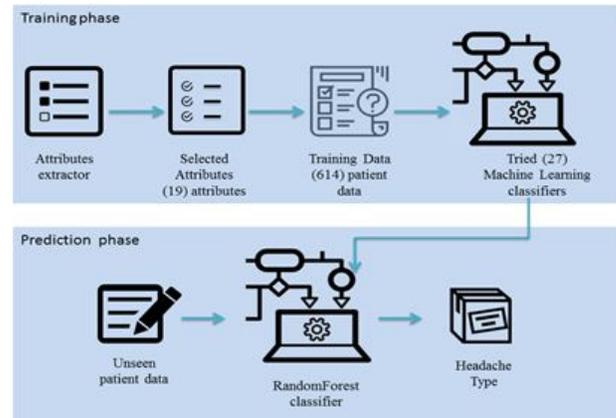


Fig. 1. Machine Learning Steps in HPSS

1) *Attribute selection:* The first step for constructing HPSS was to study the features that can be used to classify the main primary headache types (Migraine, Cluster Headache, Tension-Type Headache), while also distinguishing secondary headache. We relied on ICHD-3 criteria to extract HPSS attributes. The attributes express the symptoms that the patient feels when a headache occurs. Our main goal in this step was to reduce the attributes domain (number of attributes) in ICHD criteria while maintaining accurate classification. For this reason, we looked at the problem from two points of view: machine learning and medical. We first used 9 different attribute selection methods implemented in Weka tool [22] (CfsSubsetEval, GainRatioAttributeEval, InfoGainAttributeEval, OneRAttributeEval, Principal components, ReliefAttributeEval, SymmetricalUncertAttributeEval, WrapperSubsetEval, and CorrelationAttributeEval) to extract the highest ranked attributes that can categorize headache symptoms into one of the four chosen headache types. We applied these methods on a synthetic data set. Most methods showed that the attributes rank is convergent in importance. We then consulted medical specialists to strengthen our attribute selection findings and ensure the correctness and safety of using these attributes in HPSS for headache diagnosis and prediction. We ended up having 19 attributes (headache symptoms).

These attributes were used to create a questionnaire. The ministry of health in Jordan gave us the approval, upon checking the questionnaire, so that we can ask patients diagnosed with headache in public hospitals to fill it under the supervision of specialized clinicians. We collected a sample consisting of 614 records for patients (199 males and 415 females) with different ages. This sample was supervised, which means that each patient's record was assigned a headache type (migraine, cluster, tension, or secondary) according to the diagnosis of

his/her clinician. We then performed pre-processing on our sample. This step will be discussed in the next section.

Table I shows a sorted average rank for the 19 chosen attributes obtained by applying the aforementioned 9 different attribute selections on our new real data set. As can be seen, previous similar symptoms, Phonophobia, and Photophobia were the attributes with the highest ranks. Gender was found to have the least impact. We used the results of this table to validate our chosen attributes and check the impact of eliminating some attributes with low ranks on the overall accuracy. We show these results later in this paper.

TABLE I. THE AVERAGE RANK FOR THE 19 SELECTED ATTRIBUTES USING 9 DIFFERENT ATTRIBUTE SELECTION METHODS.

Number	Attribute name	Average rank
1	Previous similar symptoms	80.9968
2	Phonophobia	75.7321
3	Photophobia	73.9364
4	Sensation of fullness in the ear	70.4729
5	Nausea and/or vomiting	66.1487
6	Miosis and/or ptosis	71.3452
7	Sense of restlessness or agitation	71.0957
8	Aggravation by routine physical activity	66.0399
9	Forehead and facial sweating	69.4715
10	Eyelid edema	68.4178
11	Pulsating quality	65.1123
12	Nasal congestion and/or rhinorrhea	67.3729
13	Severe intensity	62.5464
14	Conjunctival injection and/or lacrimation	66.5132
15	Moderate intensity	60.835
16	Age	59.2182
17	Unilateral location	57.0289
18	Duration	57.4817
19	Gender	55.9611

2) *Data Pre-processing*: Data preprocessing was applied on our 614 records sample collected from patients to extract a clean data set. Fig. 2 describes the data preprocessing steps.



Fig. 2. Data Preprocessing

Data cleaning was the first step in which we corrected the unconscionable data. In data normalization, the attributes' values were scaled down. In data transformations, we converted the data set format to Attribute-Relation File Format (ARFF) to ensure that all attributes contain legal values. Later, some missing values, such as gender, were filled. Finally, data integration and noise identification were performed to create a

clean real data set, which was later used as a training set for classification.

3) *Clustering and Classification*: We employed a hybrid model that integrates K-Means clustering with Random Forest classifier. The initial diagnosis was done by clinicians. The clustering phase of this model was exploited to emphasize the initial medical diagnosis and confirm the efficiency of chosen attributes in headache prediction from a computerized point of view. K-Means results remarkably matched the initial medical diagnosis. However, for the few unmatched predictions, we had further discussions with our specialized clinicians and modified the data accordingly. We later passed K-Means clustering output to the classification phase.

Classification is a two-step process, where the first is to train a classifier (training phase) while the second is to use the trained classifier for classifying unknown instances (prediction phase). We applied 26 different classification algorithms, with the default number of folds (10), on our supervised real data set generated from the previous phases. These algorithms differ in terms of speed, accuracy, scalability, robustness, and interpretability. We chose the most accurate classifier in the prediction phase of HPSS. Table II demonstrates the results of the classifiers.

Depending on the results from Table II, the RandomForest gave the highest accuracy among all classifiers with (92.7%) by applying the default number of folds (10) and the number of tree was (500). Usually, the higher the number of trees the better to learn the data but adding large number of trees can slow down the training process. The time spent to build the model for the 614 records was (0.45) seconds. The correctly classified instances were 569, while 45 were incorrectly classified. The number of folds means that the classifier divides the data into (10) sections (folds). The classifier uses nine folds for training data and one fold for testing data. This process is repeated in each classifier stage.

Table III illustrates the performance of the Random Forest algorithm through its confusion Matrix.

The first row in the confusion matrix indicates that 329 instances out of the 332 that were actually classified as migraine were correctly predicted. The second row indicates that 162 out of 187 instances were correctly predicted as secondary. Third row indicates that 46 instances out of 56 were correctly predicted as tension, and the fourth row indicates that 32 instances out of 39 were correctly predicted as cluster. The recall values refer to the recall rate of the classifier in each class. The precision values in the confusion matrix refer to the precision of the classifier at predicting each class. The number of instances of each class significantly affects the accuracy of its prediction. We collected 332 instances of migraine, 187 instances of secondary, 39 instances of tension and 56 instances of cluster. We treated a small number of patients suffering from tension-type headache, although it is the most common type of headache. This is due to the fact that patients who suffer from tension-type headache rarely visit clinics. Moreover, cluster headache is considered a very rare headache to happen, so we got few patients suffering from cluster headache. The accuracy should notably improve by increasing the number of patients diagnosed with cluster and tension type headaches.

TABLE II. CLASSIFICATION RESULTS USING 26 DIFFERENT CLASSIFICATION ALGORITHMS

Algorithm	TP Rate	FP Rate	Precision	Recall	F-Measure	Accuracy
NaïveBayes	0.793	0.132	0.795	0.793	0.793	79.32%
NaïveBayesMultinomialText	0.541	0.541	0.292	0.541	0.38	54.07%
Logistic	0.814	0.116	0.813	0.814	0.813	81.43%
MultilayerPerception	0.855	0.093	0.854	0.855	0.854	85.50%
SimpleLogistic	0.822	0.124	0.819	0.822	0.819	82.25%
SMO	0.821	0.125	0.815	0.821	0.817	82.08%
IBK	0.832	0.105	0.832	0.832	0.832	83.22%
KStar	0.809	0.119	0.811	0.809	0.808	80.94%
LWL	0.643	0.266	0.558	0.643	0.594	64.33%
AdaBoostM1	0.614	0.244	0.565	0.614	0.575	61.40%
AttributeSelectedClassifier	0.878	0.078	0.876	0.878	0.875	87.79%
Bagging	0.86	0.108	0.858	0.86	0.856	85.99%
ClassificationViaRegression	0.866	0.096	0.864	0.866	0.863	86.65%
IterativeClassifierOptimizerused	0.85	0.104	0.847	0.85	0.847	85.02%
MultiClassClassifier	0.836	0.119	0.833	0.836	0.832	83.55%
RandomCommittee	0.878	0.087	0.876	0.878	0.876	87.79%
RandomSubSpace	0.829	0.151	0.835	0.829	0.811	82.90%
DecisionTable	0.849	0.107	0.848	0.849	0.845	84.85%
JRip	0.845	0.089	0.846	0.845	0.845	84.53%
PART	0.866	0.075	0.865	0.866	0.865	86.65%
HoeffdingTree	0.793	0.132	0.795	0.793	0.793	79.32%
J48	0.879	0.082	0.879	0.879	0.877	87.95%
LMT	0.875	0.065	0.878	0.875	0.876	87.45%
RandomForest	0.926	0.075	0.905	0.926	0.903	92.67%
RandomTree	0.793	0.12	0.793	0.793	0.793	79.32%
REPTree	0.793	0.12	0.793	0.793	0.793	79.32%

TABLE III. RANDOM FOREST CONFUSION MATRIX

Recall	Migraine	Secondary	Tension	Cluster	<- classified as
99.1%	329	3	0	0	Migraine
86.6%	23	162	1	1	Secondary
82.1%	8	2	46	0	Tension
82.05%	1	6	0	32	Cluster
	91.1%	93.6%	97.8%	96.9%	<b>Precision</b>

TABLE IV. THE ACCURACY AFTER EXCLUDING THE LOWEST RANK ATTRIBUTES

# of Attributes	Eliminated attributes	Accuracy
18	Gender	92.50%
17	Gender and Duration	91.25%
16	Gender, Duration, and Unilateral	89.94%
15	Gender, Duration, Unilateral, and Age	85.01%

Our last step in the training phase was to test our classifier while reducing the number of attributes. This reduction was performed through different stages by eliminating some low rank attributes at each stage. Table IV shows the accuracy after each stage. We used this experiment to check the impact of eliminating the low rank attributes on the overall accuracy. As can be obtained, while reducing the number of attributes, we were still able to obtain high accuracy. Moreover, while eliminating more attributes, the accuracy decreased because it becomes harder for the machine to classify patients.

#### IV. HPSS GRAPHICAL USER INTERFACE (GUI)

In order to enable patients, medical students or even specialists to use our system in a convenient way, we developed our system as a web-based software called Headache Prediction Support System (HPSS). HPSS was programmed with Java Server Page (JSP). We used bootstrap to build a responsive mobile website, while using Hypertext Markup Language (HTML), Cascade Style Sheet (CSS) and JavaScript (JS) for the front end. The integrated development environment (IDE) was eclipse EE.

### A. HPSS Work Mechanism

Fig. 3 demonstrates the work mechanism of HPSS in details. The user is first asked to fill an online questionnaire. The answers are saved in an ARFF file as a testing data. This file is then tested against our training set using Random Forest classification. Based on this testing, the system displays the predicted headache type based on the questionnaire's answers and gives a guidance message to the user to take the required action. HPSS also provides useful tips and links about the different types of headache.

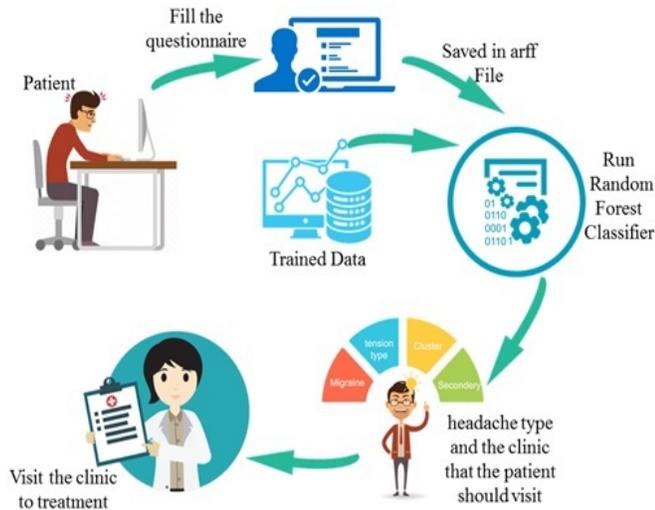


Fig. 3. HPSS Mechanism

### B. HPSS Graphical User Interface (GUI)

HPSS does not only classify headache types, but also offers other helpful pages to the different types of users. We explain each page in the following. The home page, depicted in Fig. 4, contains an image that includes four hyperlink photos to other pages (Headache Prediction Page, Headache Types Page, Frequently Asked Questions Page, and About Us Page).

The prediction page, seen in Fig. 7, consists of some check-box questions that represent the 19 chosen headache symptoms. When the user checks a specific box, this means that he/she has this particular symptom. Then, when the user completes choosing the symptoms he/she has, he/she must click on the apply button. The results are saved in an ARFF file. The Random Forest algorithm reads the user symptoms and tests them against the existing training set. The headache type is then displayed to the user on the screen depending on the classification results as a message on the top of the screen. This message can be one of the following based on the predicted headache type:

- 1) You most likely have Cluster Headache. Please check with your healthcare provider. If your age is above 50 and/or you have these symptoms for the first time, it is highly recommended to check with your clinician for further evaluation.
- 2) You most likely have a Tension-Type Headache. Please check with your healthcare provider. If your age is above 50 and/or you have these symptoms for the first time, it is

highly recommended to check with your healthcare provider for further evaluation.

- 3) You most likely have Migraine. Please check with your healthcare provider. If your age is above 50 and/or you have these symptoms for the first time, it is highly recommended to check with your healthcare provider for further evaluation.
- 4) You most likely have Secondary Headache. Please check with your healthcare provider to find out the cause of your headache to receive appropriate treatment.

We used the tooltip class in CSS to show a brief description about every symptom when the user moves the cursor over the corresponding checkbox. A prediction message example is shown in Fig. 8. The Headache Types Page is shown in Fig. 5. This page has four clickable photos, where each one of them explains one particular headache type classified in HPSS. The explanation gives an overview about the headache type, its symptoms, triggers, causes, and preventive measures.

The frequently asked questions page, shown in Fig. 6, contains some general questions/answers related to headaches. A specialized clinician was consulted to create these pages. Finally, the (About Us) page, seen in Fig. 9, explains the purpose of HPSS and its main concepts. It also gives contact information in the case of any queries about the usage of HPSS.

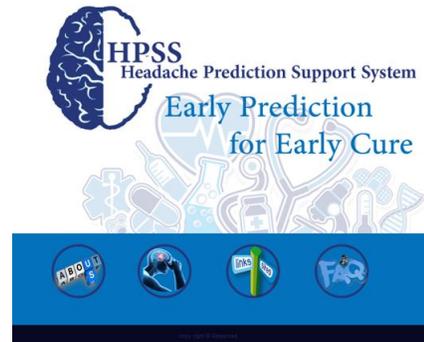


Fig. 4. HPSS Home Page

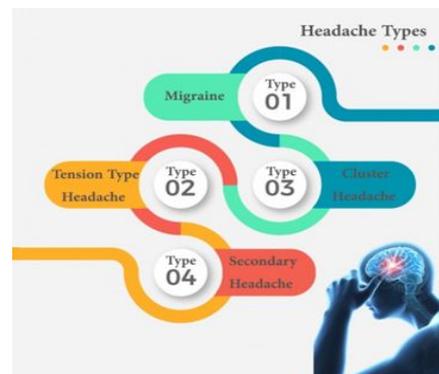


Fig. 5. Headache Types Page

## V. DISCUSSION AND ANALYSIS

There have been many studies targeting headache diagnosis. In this comparative study, we mainly focused on the most recent and related systems, which provide clear prediction

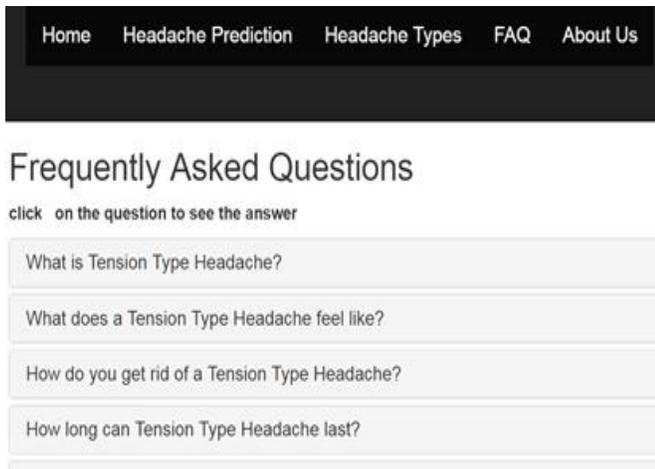


Fig. 6. Frequently Asked Questions (FAQ) Page

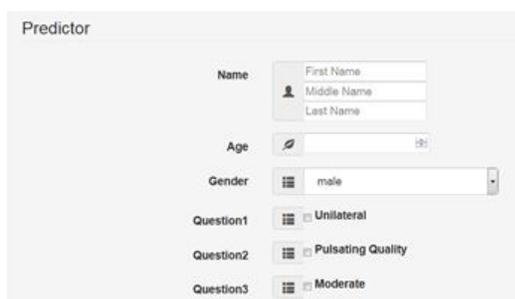


Fig. 7. HPSS Headache Prediction Page

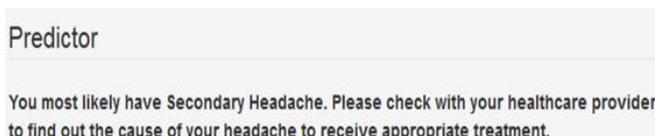
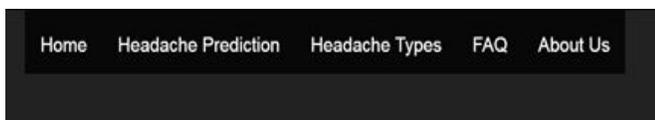


Fig. 8. HPSS Prediction Message Example



## HPSS

HPSS provides preliminary guidance for patients, medical students and even clinicians for better headache diagnosis. HPSS does not give any prescriptions or medications.

Fig. 9. HPSS About Us Page

accuracy results as shown in Table V. Table V highlights the usage of these systems via 9 different metrics. The missing values in the table were either unexplained or unclear in some of these systems.

Many systems have achieved significant prediction performance. However, they were either using synthetic data [18], missing secondary headache [5], [23], [19], [16], [24], lacking

medical-patient interaction system [25], [20], using many attributes [25], [5], or obtaining relatively low accuracy [24]. Moreover, almost all existing systems do not analyze or rank headache symptoms from a computerized perspective. They also use previous versions of headache criteria for headache diagnosis.

In contrast, our model is based on a hybrid solution that consists of clustering and classification. It was tested with 26 different classification algorithms while using 9 different attribute selection methods. We relied on the latest headache ICHD-3 criteria to extract HPSS attributes. We reduced the attribute's domain to 19 attributes while predicting different types of primary headache in addition to secondary headache with high accuracy of 99.1% for migraine prediction and overall accuracy of 93%. HPSS provides a web-based system for patients, medical students, and clinicians for better early diagnosis. Patients can safely use HPSS due to its helpful advices and guidance without giving any prescriptions.

Our study also ranks headache attributes, as shown in Table I. This can assist clinicians and medical experts to better diagnose headaches while saving much time and effort. This can also help for further headache criteria improvements.

## VI. CONCLUSION AND FUTURE WORK

This paper presented our experiences in developing a hybrid, reliable, and web-based prediction support system for the diagnosis of primary and secondary headaches according to the latest ICHD-3 criteria. This work aimed to improve the process of headache diagnosis via machine learning model accompanied with graphical user interface called HPSS. HPSS targets different end-users that include patients, medical students and clinicians for initial safe headache diagnosis. HPSS provides an easy-to-use interface for predicting both primary and secondary headaches with an overall accuracy of 93% among 614 real patients with 19 attributes only, while obtaining a migraine accuracy of 99.1%.

HPSS showed that a hybrid machine learning model that combines unsupervised with supervised learning can improve headache prediction accuracy compared with a singular model consisting of classification only.

HPSS was tested by different patients, medical students, and clinicians. All of them gave us a very positive feedback about its convenient interface, usefulness, and accuracy.

### A. Future Work

We believe that examining more patients' records should lead to further accuracy improvement. The scope of prediction can also be extended to cover other subtypes of headache.

## ACKNOWLEDGMENT

The authors of this paper would like to thank the Hashemite University for providing the required computing resources. We also thank the Ministry of Health (MoH) in Jordan for permitting the collection of medical records from patients under the supervision of their clinicians.

TABLE V. COMPARISON STUDY

Authors	Real data	Attributes	Headache types	Methods	Accuracy	Patients tool	Medical tool	Hybrid	Attr Rank
Andrew et al. [26]	68 + 54		Migraine, Tension-Type, Other	Structured Headache Diagnosis Interview	91	no	IBM only.	no	no
Pryse-Philips et al. [25]	461	45	Migraine	Classification and Regression Trees	91	no	no	no	no
Maizels and Wolfe [12]	135		Episodic Migraine, Chronic Tension-Type, Episodic Tension-Type, Episodic Cluster	Initial branch points determined by headache frequency and duration	75-100	Limited	yes	no	no
Yin et al. [13] [17]	676	24	Probable Migraine, Probable tension	Case-based reasoning, Genetic algorithms and K- Nearest neighbors	93.14, 89.36	no		no	no
Aljaaf et al. [18]	900 synthetic	8	Migraine, Tension, Cluster, Other	Naive Bayes, Artificial Neural Network, Decision Tree, Zero R Classifier, Support Vector Machines, k-Nearest Neighbors, Logistic Regression	92.67, 96.11, 97, 66.67, 96, 96.22, 95.33	no	no	no	no
Walters [20]	1829	4	Migraine	Four Item Migraine Screening Algorithm	93	no	no	no	no
Celik et al. [5] [23]	850	40	Migraine, Tension Type, Cluster	Immunos-1, Immunos-2, Immunos-99, AIRS1, AIRS2, AIRS2-Parallel, ClonalG, CSCA	94.47, 71.65, 95.65, 99.29, 98.82, 99.65, 98.71, 99.18		yes		no
Celik et al. [19]	850	40	Migraine, Tension Type, Cluster	Ant colony optimization-based	98.2%, 92.4%, 98.2%	Yes web-based	yes	no	no
Yin et al. [16]	511	17	Migraine, Tension Type, Cluster	Rule-based and case-based reasoning	97.2	no		yes	no
Khayamnia et al. [24]	190	12	migraine, tension, Infection headache, IICP headache	(LFE) algorithm, MLP and SVM	88% 90%	no	no	yes	no

REFERENCES

[1] K. Linde, G. Allais, B. Brinkhaus, Y. Fei, M. Mehring, B.-C. Shin, A. Vickers, and A. R. White, "Acupuncture for the prevention of tension-type headache," *Cochrane Database of Systematic Reviews*, no. 4, 2016.

[2] A. Bahra, A. May, and P. J. Goadsby, "Cluster headache: a prospective clinical study with diagnostic implications," *Neurology*, vol. 58, no. 3, pp. 354–361, 2002.

[3] T. J. Steiner, L. J. Stovner, and T. Vos, "Gbd 2015: migraine is the third cause of disability in under 50s," 2016.

[4] H. C. C. of the International Headache Society (IHS), "The international classification of headache disorders, (beta version)," *Cephalalgia*, vol. 33, no. 9, pp. 629–808, 2013.

[5] U. Çelik, N. Yurtay, E. R. Koç, N. Tepe, H. Güllüoğlu, and M. Ertaş, "Diagnostic accuracy comparison of artificial immune algorithms for primary headaches," *Computational and mathematical methods in medicine*, vol. 2015, 2015.

[6] Z. Dong, Z. Yin, M. He, X. Chen, X. Lv, and S. Yu, "Validation of a guideline-based decision support system for the diagnosis of primary headache disorders based on ichtd-3 beta," *The journal of headache and pain*, vol. 15, no. 1, p. 40, 2014.

[7] F. Ahmed, "Headache disorders: differentiating and managing the common subtypes," *British journal of pain*, vol. 6, no. 3, pp. 124–132, 2012.

[8] L. Theodoros and T. Eleftherios, "Assessment of a new decision support expert system in headaches cases," *Homeopathy*, vol. 103, no. 01, 2014.

[9] A. Belard, T. Buchman, J. Forsberg, B. K. Potter, C. J. Dente, A. Kirk, and E. Elster, "Precision diagnosis: a view of the clinical decision support systems (cdss) landscape through the lens of critical care," *Journal of clinical monitoring and computing*, vol. 31, no. 2, pp. 261–271, 2017.

[10] M. Y. Smith, J. D. DePue, and C. Rini, "Computerized decision-support systems for chronic pain management in primary care," 2007.

[11] S. D. Silberstein, J. Olesen, M. Bousser, H. Diener, D. Dodick, M. First, P. Goadsby, H. Göbel, M. Lainez, J. Lance *et al.*, "The international classification of headache disorders, (ichd-ii)—revision of criteria for 8.2 medication-overuse headache," *Cephalalgia*, vol. 25, no. 6, pp. 460–465, 2005.

[12] M. Maizels and W. J. Wolfe, "An expert system for headache diagnosis: the computerized headache assessment tool (chat)," *Headache: The Journal of Head and Face Pain*, vol. 48, no. 1, pp. 72–78, 2008.

[13] Z. Yin, Z. Dong, S. Yu, X. Lu, G. Feng, and H. Duan, "A guideline-based decision support system for headache diagnosis," *Studies in health technology and informatics*, vol. 192, pp. 1022–1022, 2013.

[14] S. W. Tu, J. R. Campbell, J. Glasgow, M. A. Nyman, R. McClure, J. McClay, C. Parker, K. M. Hrabak, D. Berg, T. Weida *et al.*, "The sage guideline model: achievements and overview," *Journal of the American Medical Informatics Association*, vol. 14, no. 5, pp. 589–598, 2007.

[15] B. Krawczyk, D. Simić, S. Simić, and M. Woźniak, "Automatic diagnosis of primary headaches by machine learning methods," *Open Medicine*, vol. 8, no. 2, pp. 157–165, 2013.

[16] Z. Yin, L. Min, X. Lu, and H. Duan, "A clinical decision support system for primary headache disorder based on hybrid intelligent reasoning," in *Biomedical Engineering and Informatics (BMEI), 2014 7th International Conference on*. IEEE, 2014, pp. 683–687.

[17] Z. Yin, Z. Dong, X. Lu, S. Yu, X. Chen, and H. Duan, "A clinical decision support system for the diagnosis of probable migraine and probable tension-type headache based on case-based reasoning," *The journal of headache and pain*, vol. 16, no. 1, p. 29, 2015.

- [18] A. J. Aljaaf, D. Al-Jumeily, A. J. Hussain, P. Fergus, M. Al-Jumaily, and N. Radi, "A systematic comparison and evaluation of supervised machine learning classifiers using headache dataset," in *Advanced Intelligent Computing Theories and Applications - 11th International Conference, ICIC 2015, Fuzhou, China, August 20-23, 2015. Proceedings, Part III*, 2015, pp. 101–108. [Online]. Available: [https://doi.org/10.1007/978-3-319-22053-6\\_12](https://doi.org/10.1007/978-3-319-22053-6_12)
- [19] U. Celik and N. YURTAY, "An ant colony optimization algorithm-based classification for the diagnosis of primary headaches using a website questionnaire expert system," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 25, no. 5, pp. 4200–4210, 2017.
- [20] A. B. Walters and T. A. Smitherman, "Development and validation of a four-item migraine screening algorithm among a nonclinical sample: The migraine-4," *Headache: The Journal of Head and Face Pain*, vol. 56, no. 1, pp. 86–94, 2016.
- [21] W. Yanping and D. Huilong, "An application of ant colony optimization clustering approach for primary headache diagnosis," in *Ubiquitous Computing Application and Wireless Sensor*. Springer, 2015, pp. 643–648.
- [22] Weka, "Contents attribute evaluators," 2008.
- [23] U. Celik, N. Yurtay, E. R. Koc, N. Tepe, H. Gulluoglu, and M. Ertas, "Migraine, tension-type and cluster-type of headaches classification by using immunos algorithms," *Journal of Medical Imaging and Health Informatics*, vol. 6, no. 5, pp. 1173–1177, 2016.
- [24] M. Khayamnia, M. Yazdchi, A. Heidari, and M. Foroughipour, "Diagnosis of common headaches using hybrid expert-based systems," *Journal of medical signals and sensors*, vol. 9, no. 3, p. 174, 2019.
- [25] W. Pryse-Phillips, M. Aubé, M. Gawel, R. Nelson, A. Purdy, and K. Wilson, "A headache diagnosis project," *Headache: The Journal of Head and Face Pain*, vol. 42, no. 8, pp. 728–737, 2002.
- [26] M. E. Andrew, D. B. Penzien, J. C. Rains, G. E. Knowlton, and R. D. McAnulty, "Development of a computer application for headache diagnosis: The headache diagnostic system," *International journal of bio-medical computing*, vol. 31, no. 1, pp. 17–24, 1992.

# Efficient Cache Architecture for Table Lookups in an Internet Router

Hayato Yamaki

The University of Electro-Communications  
Chofu, 182-8585, Japan

**Abstract**—Table lookup is the most important operation in routers from the aspects of both packet processing throughput and power consumption. To realize the table lookup at high throughput with low energy, Packet Processing Cache (PPC) has been proposed. PPC stores table lookup results into a small SRAM (static random access memory) per flow and reuses the cached results to process subsequent packets of the same flow. Because the SRAM is accessed faster with significant lower energy than TCAM (Ternary Content Addressable Memory), which is conventionally used as a memory for storing the tables in routers, PPC can process packets at higher throughput with lower power consumption when the table lookup results of the packets are in PPC. Although the PPC performance depends on the PPC hit/miss rates, recent PPCs still show high PPC miss rates and cannot achieve sufficient performance. In this paper, efficient cache architecture, constructed of two different techniques, is proposed to improve the PPC miss rate more. The simulation results indicated that the combined approach of them achieved 1.72x larger throughput with 41.4% lower energy consumption in comparison to the conventional PPC architecture.

**Keywords**—Router architecture; table lookup; Packet Processing Cache (PPC); Ternary Content Addressable Memory (TCAM)

## I. INTRODUCTION

Recent increase in internet communication traffic amount is remarkable owing to the spread of data consuming applications such as cloud services, video streaming, and internet-of-things (IoT) applications. Accordingly, requirements of the throughput in routers become more and more serious. It reaches 400 Gbps or 1 Tbps in recent years. The increase in internet traffic also induces a problem of the power consumption in routers [1], [2]. The reports [3], [4] emphasized that several percentages of the energy generated in the world is consumed by network devices. Thus, routers must consider both the packet processing throughput and energy efficiency. This demand becomes more serious in core routers, which are placed in core networks and handle huge amount of internet traffic.

For routers, table lookup operation is the most slow and power consumed operation in packet processing [5], [6], [7]. When a packet arrives at router, the router needs to lookup tables (e.g., a routing table and ACL (access control list) to obtain the information required to process the packet, such as the output port and filtering decision. These tables are often stored into ternary content addressable memories (TCAMs) to gain high lookup performance, especially in core routers, which are placed at core networks and required to process huge amount of packets. The TCAM can lookup tables at one cycle by comparing all data stored in the TCAM simultaneously. However, owing to this power consuming operation,

a TCAM significantly consumes energy in comparison to a same-sized SRAM [8]. According to [9], [10], 40% of the power consumption in routers is due to the TCAMs. TCAM is also insufficient from the viewpoint of the lookup performance. To achieve over 400 Gbps, it is required for routers to process packets every 1.25 nano seconds if the shortest length packets come continuously. However, recent TCAM products show the access latency of approximately 5 nano seconds. Thus, to increase table lookup throughput with reducing the energy, improving the table lookup is needed.

PPC (packet processing cache) is an attractive approach to meet the requirement [11], [12]. PPC can reduce the number of TCAM accesses by storing table lookup results into a small SRAM and reusing the stored results to process subsequent packets. PPC can finish the table lookups of a packet at high speed with low energy consumption when the corresponding TCAM lookup results of the packet are in the cache (i.e., a cache hit). Thus, the PPC performance depends on the number of cache hits/misses, and reducing the number of PPC misses is the most important issue to improve the PPC performance. However, PPC still cannot satisfy this requirement because of the following two problems: the high average PPC miss rate and the low attack tolerance. This study proposes a novel efficient cache architecture, which constructed of Port-aware Cache and Victim IP Cache, for high-throughput and low-power table lookup.

This paper extends the previous work of Yamaki et al [13]. Different from [13], this paper newly adds the more detailed analysis (Fig. 3) and the explanation of the concrete hardware (Fig. 6) in Sec. 4. In addition, to resolve the problem that the performance of Port-aware Cache depends on the network configurations, this study newly proposed Semi-static Port-aware Cache, which decides the best mix of each entry sizes by prior trials at a boot process of a router. Moreover, this study newly investigated the various sizes of Victim IP Cache and evaluated them in Sec. 6. The writing and figures in the manuscript are also improved for easy understanding as a whole. The main contribution of the paper is summarized below.

- This paper indicated that HTTP and DNS packets impacted on the PPC performance significantly from the perspective of the number of packets and flows.
- Port-aware Cache, one of the proposed approach in this paper, can not only prevent increases in PPC misses caused by attacks (8.64% improvement) but also reduce the number of PPC misses caused by HTTP packets (9.02% improvement).

- 64KB Victim IP Cache, the other approach of this paper, can save the 85.5% of all the packets missed in PPC by caching them to a victim cache per destination IP address.
- The simulation results showed that the combination of the two approaches can achieve 1.72x larger throughput with 41.4% lower energy in comparison to the conventional PPC.

The rest of the paper is organized as follows. First, more detailed architecture of PPC and the problems are shown in Section 2. Section 3 introduces the relative works of reducing the cache miss in PPC. In Sections 4 and 5, the two proposed architecture, Port-aware Cache and Victim IP Cache, are proposed, respectively. Section 6 evaluates the proposed architecture, and finally we conclude the work in Section 7.

## II. PACKET PROCESSING CACHE

### A. Outline

PPC has been proposed as a supplemental approach of TCAM lookup and can realize high-throughput and low-power table lookup by reducing the number of TCAM accesses. In PPC, a flow is defined based on the five tuples (i.e. source/destination IP addresses, source/destination port numbers, and protocol number) of packets. PPC stores TCAM lookup results per flow into a small SRAM and references the stored results to process subsequent packets of the same flow. Because the packets of the same flow are processed using the same TCAM lookup results, PPC enables to process packets using the SRAM without accessing TCAM when PPC has the TCAM lookup results of the flow.

Fig. 1 shows the table lookup flow with PPC. Conventionally, PPC entries are constructed of the 13 bytes tags (i.e., five tuples) and 15 bytes data (i.e., TCAM lookup results). The TCAM lookup results include the routing table lookup result of 1 byte, ARP table lookup result of 12 bytes, ACL lookup result of 1 byte, and QoS table lookup result of 1 byte. PPC entries are addressed using the hash value of the five tuples as the index. Typically, a 32KB small SRAM (i.e., approximately 1,024 entries) is used as PPC considering the SRAM latency. It is because there is few latency gap between two memories (i.e., TCAM and PPC), unlike processor caches. For example, L2 cache latency of microprocessors is almost the same of that of TCAM (approximately 5 nano seconds).

### B. Problems

The table lookup performance with PPC is mainly determined by a PPC miss rate because PPC accesses are significantly faster with lower energy consumption than TCAM, and they are almost negligible. Thus, achieving low PPC miss rate is the most important issue for PPC. However, PPC has two problems to meet this requirement.

1) *Tolerance to Attacks:* PPC has little attack tolerance because it may register a large number of attack flows in PPC when one-packet-based attacks, such as port-scan attacks, pass through a router. Consequently, many useful PPC entries are evicted by attack flows, and it causes the significant degradation in the PPC hit rate. The attacks induce two disadvantages to PPC. First, attack flows created by such attacks never hit

in PPC because they are mainly constructed of one packet. Second, useful PPC entries are evicted by attack flows.

2) *A large number of TCAM accesses:* If a PPC miss occurs, a router must access TCAMs several times (four times in the case of Fig. 1) to obtain each table lookup result. Thus, the number of PPC misses significantly impact on the table lookup performance. However, the state-of-the-art PPC still remains the PPC miss rate of 30% [14], [15]. It indicates that 30% of all packets still access to TCAM. Especially considering the power consumption, further improvement in the PPC miss rate is important. The most effective approach to meet this requirement is to increase the PPC entries: However, it is not reasonable from the following two reasons. First, PPC cannot increase the capacity largely due to the access latency, as mentioned before. As the PPC capacity, the size like L1 caches in microprocessors (i.e. 32KB) is acceptable. Second, PPC easily increase the capacity because of the large PPC entry size (28KB per entry).

## III. RELATED STUDIES

In this section, related studies of this work are introduced. Although there are many studies focusing on PPC, the attack tolerance of PPC was not considered in all studies. Thus, this section shows studies of improving the average PPC miss rate.

One approach to improve the PPC miss rate is that reducing the cache tag information and increasing the PPC entries without increasing the capacity. As mentioned in Section 2, the cache tag of PPC is 13-byte flow information, and it is one of the reason that PPC has a few entries. Digest Cache was proposed Chang et al. [16]. This method uses the hash values calculated from the five tuples as cache tags instead of the five tuples. Likewise, Ata et al. proposed the cache which used the three tuples (i.e., source/destination IP addresses and smaller port number) as cache tags instead of the five tuples [17]. These works effectively reduce the TCAM accesses by increasing the number of stored flows. However, the compressing tags cause cache conflicts, and thus, an extra hardware for avoiding the cache conflicts is required.

As other approaches, there are studies of reducing the PPC misses by improving the hash conflicts. The paper [18] emphasized that CRC hash function is not appropriate as the cache indexes of PPC and it caused many cache conflicts. They proposed a novel PPC indexing method, which split the cached area of PPC into two areas and used two different hash functions. They showed two universal hash functions [19] are effective to reduce the cache conflicts. The one problem of this method is that the implementation cost of the universal hash functions is high due to the large input-data size (i.e., 13-byte five tuples).

Improving the cache replacement policy is one of the effective approach to reduce the PPC misses. Kim et al. pointed out that LRU does not fit for PPC because LRU determines the replaced entry based on only the last packet of the flow and cannot consider the flow characteristics [20]. They proposed two types of cache-replacement algorithms which utilize last two packets information of the flow to determine the replaced entry and reduced the PPC misses by several percentages compared with LRU. However, the hardware cost of storing last two-packet information, was not discussed. The increase

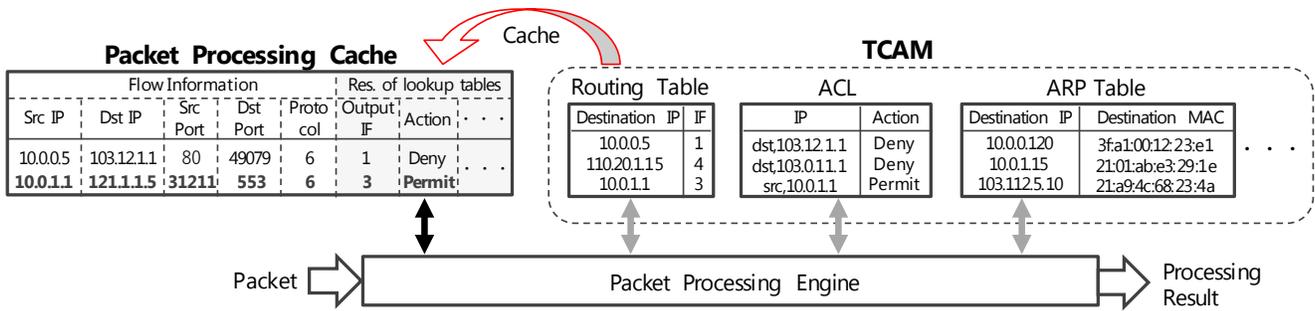


Fig. 1. Outline of table lookups with PPC.

in the PPC capacity due to this additional stores becomes a serious problem for the small-sized cache like PPC.

Yamaki et al. also proposed methods of reducing the PPC misses [12]. They focused on one-packet flows created applications such as DNS (domain name system), DoS (denial of service) attack, and port scan attacks because these flows never hit in PPC and proposed methods of denying packets from these flows. The simulation results showed that DNS-Aware Cache, one of the proposed methods, can reduce the number of PPC misses by 6%.

Although these approaches are effective to reduce the number of PPC misses, PPC still shows high PPC miss rate, as mentioned in the previous section. In addition, these approaches are not effective to prevent the attack influence.

#### IV. PORT-AWARE CACHE

This study first proposes Port-aware Cache to improve the PPC miss rate and reduces a negative impact of attacks. Port-aware Cache stores flows per application group by assigning different cache areas to each application group. As a result, an increase in PPC misses caused by attacks can be avoided by isolating the impact of each application group. Furthermore, it also improves the average PPC miss rate by assigning suitable number of entries in each application group.

##### A. Motivation

To identify the construct of a flow (e.g., the number of packets composing a flow), the application of the flow is one of the important information. We explain it referring DNS (domain name system), HTTP (hypertext transfer protocol), and several types of attacks for examples. In general, a DNS flow consists of one packet because of the simple request-reply communication. In this case, DNS communication creates two one-packet flows (the request flow and the reply flow). On the other hand, an HTTP flow consists of a large number of packets because HTTP protocol requires 3-way handshake at first and sends internet contents subsequently. Like this, it is expected that one-packet-based attacks, such as vulnerability-based attacks and port-scan attacks, create a significant large number of flows with a small number of packets. Basically, the application is identified from the port numbers of the packet.

Fig. 2 shows the top 5 ports of sending packets in a network and the amount of packets and flows. In this measurement and

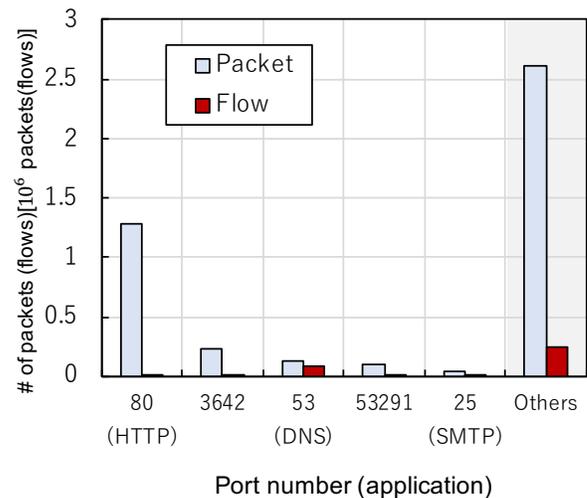


Fig. 2. Top 5 ports of sending packets in a network and the amount of packets and flows.

following analyses, An in-house traffic-analysis program and core-network traffic in Japan, called WIDE traffic, are used. Details of them are described in Sec. 6.

As shown in Fig. 2, although the HTTP packet amount is dominant in the network (30%), the HTTP flow amount is not a large portion in the network (4.8%). On the other hand, although the DNS flow amount is dominant in the network (25%), the DNS packet amount is not a large portion in the network (2.8%). This is caused by the difference in the application protocol, as mentioned above. This result also mean that HTTP and DNS especially impact on the PPC miss rate because of a large amount of packets or flows. More specifically, PPC miss rate greatly depends on the number of PPC misses caused by HTTP packets, and DNS flows affect on PPC entries due to the considerable insertion.

Attack flows also have individual characteristics. As mentioned in Sec. 2.2, some attack flows are constructed of a few packets and sent to internet drastically in a short period. Consequently, useful PPC entries are evicted and polluted. Fig. 3 and 4 shows the impact of attacks on the PPC counts. These figures show the breakdown of PPC hits and PPC misses from the perspective of applications with two sizes

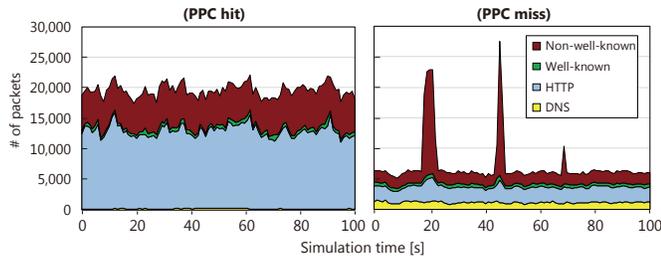


Fig. 3. The breakdown of PPC hits (the left figure) and PPC misses (the right figure) from the perspective of applications with 1K PPC entries.

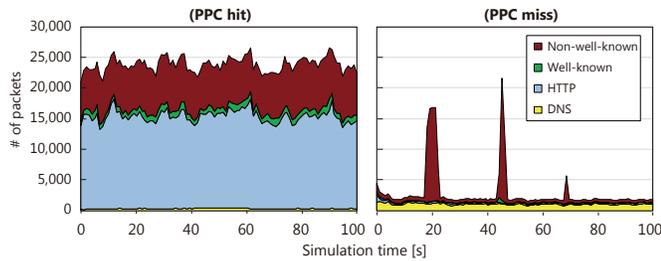


Fig. 4. The breakdown of PPC hits (the left figure) and PPC misses (the right figure) from the perspective of applications with 1M PPC entries.

of PPC. The non-well-known in the graph is defined as the packets whose port numbers are not well-known ports. In this measurement, three attacks were observed at 19s, 43s, and 68s. These figures indicate that a drastic increase in PPC misses is caused by attack flows. In addition, attack flows indirectly cause an increase in PPC misses caused by HTTP and other-application packets when the PPC size is small. It is because attack flows evict a large number of useful flows of HTTP and other applications when attacks occur. Note that PPC misses caused by DNS packets are not affected by attacks and the number of PPC entries because DNS packets hardly hit in PPC. On the other hand, most packets of HTTP and other applications have potential to hit by preparing a large number of entries.

### B. Architecture of Port-aware Cache

From the discussion in Section 4.1, this study proposes Port-aware Cache. In Port-aware Cache, PPC area is divided into 3 ranges: the DNS, HTTP, and other-application ranges. From the results of Fig. 3 and 4, the well-known ports and non-well-known ports are not distinguished because the flows of the well-known ports hardly impact on the PPC hits/misses.

The whole architecture of Port-aware Cache is depicted in Fig. 5. For the access to each application range, the address of PPC is recalculated from the CRC hash value of the five tuples and the smaller port number. This calculation is processed in the Modifier module. Details of the Modifier module is shown in Fig. 6. Offset 1, offset 2, and offset 3 depicted in Fig. 6 show the number of entries in DNS, HTTP, and other-application ranges, respectively. The modifier module adequately selects the address based on the smaller port number.

There are two advantages of Port-aware Cache. First, it is expected that Port-aware Cache reduces the number of PPC

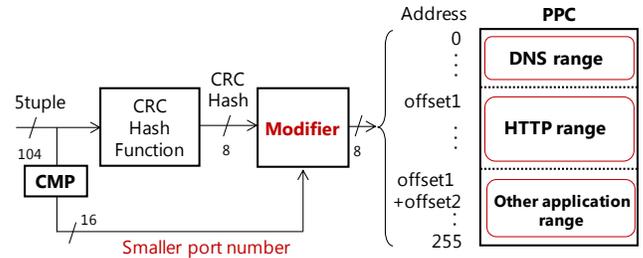


Fig. 5. Architecture of Port-aware Cache.

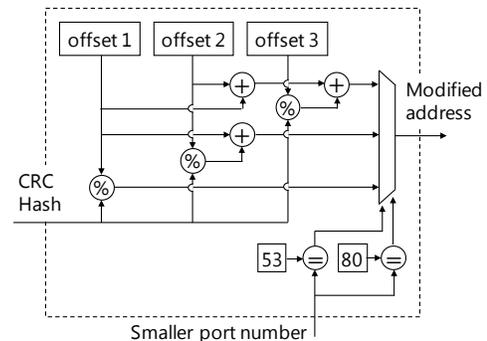


Fig. 6. Details of Modifier.

misses by assigning appropriate number of PPC entries to each application range. As mentioned in Section 4.1, DNS packets not only rarely hit in PPC but also disturb PPC entries. From this reason, assigning a few PPC entries to the DNS range is better. Unlike DNS, HTTP packets have many opportunities to hit in PPC. Consequently, assigning a large portion of PPC entries to the HTTP range is better. Second, Port-aware Cache can suppress the negative impact of attacks to the PPC miss rate by isolating each application range. As shown in Fig. 3 and 4, attack packets not only cause PPC misses of their own but also evicts useful flows, such as HTTP flows, and impede PPC hits. This problem is resolved in Port-aware Cache by isolating flows to each application range.

### C. Semi-static Port-aware Cache

In this section, Semi-static Port-aware Cache, which is an improvement in Port-aware Cache, is introduced. In Port-aware Cache, the entry sizes of each application range (i.e., the offsets 1, 2, and 3 in Fig. 6) are an important factor to decide the cache performance. In addition, it is considered that the best mix of the entry sizes of each application area varies depending on networks. From these reasons, Semi-static Port-aware Cache decides the best mix of the entry sizes from prior trial at the boot process of a router.

Fig. 7 shows the block diagram of Semi-static Port-aware Cache. Semi-static Port-aware Cache explores the best mix by trying various configurations of Port-aware Cache using a packet log of 1 minute captured at a boot process of a router. However, trying all the possible configurations is not realistic due to the explosive combinations. Thus, Semi-static Port-aware Cache uses heuristic approach to explore the best

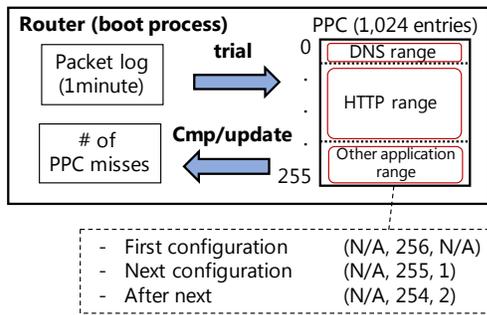


Fig. 7. Block diagram of Semi-static Port-aware Cache. The configuration (A, B, C) means the assigned addresses in the DNS range, the HTTP range, and the other-application range, respectively.

mix of the entry sizes. The basic idea is to assign PPC entries to the HTTP range as many as possible.

More specifically, as the beginning of the trial, all the PPC entries are assigned to the HTTP range, as shown in the first configuration depicted in Fig. 7. The next process is as follows. (1) The number of PPC misses in the case of PPC with the first configuration is measured using the packet log, and the result is stored to a register. (2) The PPC configuration is updated by subtracting one to the HTTP range and adding one to the other-application range. (3) The number of PPC misses is newly measured using the same packet log, and the result is compared to the registered result. (4) If the new result is smaller than the registered one, the new result is stored to the register, and the process is returned to (2). (4)' On the other hand, if the registered result is smaller, the process is returned to (2), and the DNS range is added by one instead of adding one to the other-application range hereafter. (5) If the registered result is smaller again, the current configuration is decided as the best mix. By applying Semi-static Port-aware Cache, the maximum number of trials to decide the best mix of the entry sizes becomes the maximum number of indexes (i.e., 256 times in the case of 4-way PPC with 1,024 entries).

## V. VICTIM IP CACHE

Victim IP Cache is also proposed in this study to further improve the TCAM access rate. It is placed between PPC and TCAMs and accessed if PPC misses occur. When a packet misses in PPC, the packet accesses Victim IP Cache before accessing TCAM.

### A. Motivations

As discussed in Section 2, a router with PPC still requires to access TCAMs because of a large PPC miss rate. Fig. 8 depicts the comparison of the energy consumption of the table lookup with TCAM only approach and that with PPC-based approach. We also showed the breakdown of the energy consumption in Fig. 8. Details of the method for calculating the energy consumption of the table lookup is explained in Section 6. The graph indicates that PPC significantly improves the energy consumption of the table lookup by reducing the number of TCAM accesses. However, the TCAM still consumes a large portion of the energy consumption of the table lookup, while the energy consumed by the SRAM is

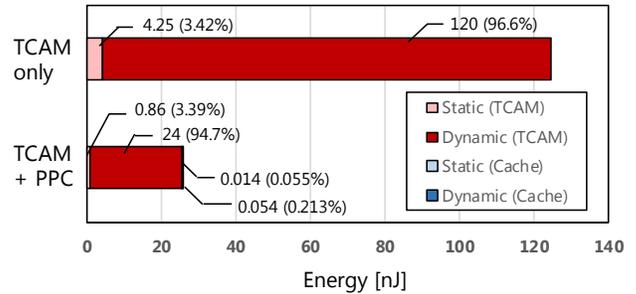


Fig. 8. Breakdown of energy consumption of the table lookups per packet.

almost negligible. For this, further improvement in the number of TCAM access is required.

To realize this, this study focused on IP Cache, which is another study of caching packets. Unlike PPC, IP Cache stores only routing table lookup (and ARP table lookup) results and reuses them by using destination IP addresses of packets as the lookup keys. Past studies show that IP Cache can achieve the PPC hit rate of more than 90% [6]. In this paper, IP Cache and PPC are combined to reduce the whole number of TCAM accesses more.

### B. Architecture of Victim IP Cache

Victim IP Cache supports PPC by storing the some table lookup results of PPC miss flows. Different from PPC, which uses five tuples as cache tags, Victim IP Cache caches only the results of the routing table and ARP table, and thus, it uses only the destination IP address as cache tags. When a packet misses in PPC, the packet accesses Victim IP Cache. At this time, if a cache hit occurs in Victim IP Cache, the TCAM accesses for searching the routing table and ARP table are omitted. As explained in the previous section, because IP Cache has a possibility to obtain higher cache hit rate than PPC owing to the high temporal locality, Victim IP Cache may save a large number of packets which miss in PPC.

Fig. 9 shows the outline of the table lookups with Victim IP Cache. Victim IP Cache entries are constructed of 4-byte destination IP address as cache tags and 1-byte output port information and 6-byte destination MAC address as cache data, namely, 11 bytes per entry. This low entry size makes the number of Victim IP Cache entry size larger than PPC, and thus, Victim IP Cache has a possibility to obtain higher cache hit rate. Different from a flow-based victim cache, the entries of Victim IP Cache are not shifted to PPC when cache hits occur in Victim IP Cache.

## VI. EVALUATIONS

This section provides the evaluation of the proposed cache architecture using an in-house PPC simulator and packet traces captured in real networks. We first show the evaluation of Port-aware Cache from the perspectives of the PPC miss reduction and the attack tolerance. Next, Victim IP Cache is evaluated by comparing to a typical victim cache. Finally, the combination of Port-aware Cache and Victim IP Cache is evaluated from the perspectives of the throughput and energy consumption.

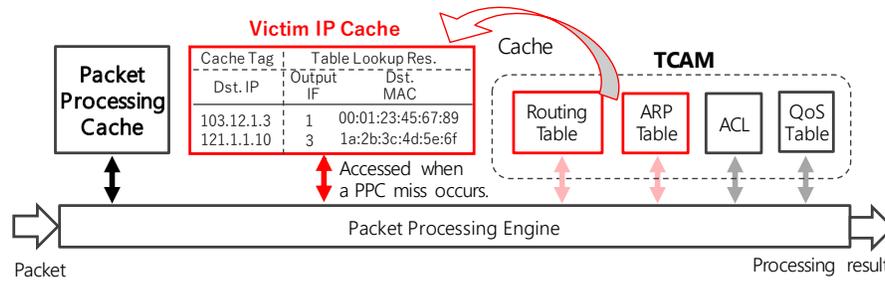


Fig. 9. Outline of table lookups with Victim IP Cache.

TABLE I. SIMULATOR PARAMETERS.

Parameter	Value
Clock frequency	2 GHz
Cache	
Number of entries	1,024
Number of ways	4
Replacement policy	LRU
Access latency	0.5 ns
TCAM	
Access latency	5 ns

TABLE III. SIMULATION RESULT OF SEMI-STATIC PORT-AWARE CACHE.

	UFL	WIDE	APN	TXG
(Conv. PPC) PPC miss rate [%]	26.6	25.5	39.2	16.9
(Static) PPC miss rate [%]	28.8	23.2	45.0	30.9
DNS range [entries]	12	12	12	12
HTTP range [entries]	704	704	704	704
Other-app. range [entries]	308	308	308	308
(Semi-static) PPC miss rate [%]	26.4	23.2	38.2	16.7
DNS range [entries]	4	12	44	16
HTTP range [entries]	380	704	36	88
Other-app. range [entries]	640	308	944	920

TABLE II. DETAILS OF NETWORK TRACES.

Trace Name	Packets/sec.
IPLS [22]	99,264
UFL [22]	51,319
MRA [22]	41,372
FRG [22]	32,722
CNIC [22]	31,023
WIDE [23]	24,657
PSC [22]	22,807
APN [22]	20,793
TXG [22]	11,610
COS [22]	7,972
BUF [22]	7,827

### A. Experimental Setup

For the simulation, an in-house PPC simulator, which was written in C++, and 11 types of packet traces were used. Details of the parameters set in the simulator and packet traces are summarized in Tables I and II. We configured the associativity of PPC to 4-way set associative and the total number of PPC entries to 1,024 entries, which were the typical configuration of PPC. The SRAM latency was set to 0.5 nano seconds from the estimation of CACTI 6.5 [21], which was a major tool for simulating the memory. The packet traces used in this simulation were captured in various universities or laboratories networks (obtained from RIPE Network Coordination Centre [22]) and a core network in Japan (obtained from WIDE MAWI WorkingGroup [23]).

The PPC simulator can simulate the table lookup operations in a router with PPC. First, the flow information (i.e., the five tuples) of a packet is extracted from a trace file in accordance with the timestamp of the packet. After reading a packet, the packet is sent to PPC and judged whether a PPC hit or miss. If the packet hits in PPC, the simulator finishes processing of the packet. On the other hand, if the packet misses in PPC, the packet is sent to a TCAM module. After passing the TCAM access latency, the processing of the packet is finished, and

a new PPC entry is added. The simulator measures the PPC miss rate every seconds.

### B. Evaluation of Port-aware Cache

1) *Usefulness of Semi-static Port-aware Cache:* First, Semi-static Port-aware Cache was evaluated. Table III shows the PPC miss rates of Semi-static Port-aware Cache measured in four networks. Note that this paper showed the results of the four networks because of the same trend. For comparison, this paper also showed the PPC miss rates of conventional PPC and naive Port-aware Cache (referred to as Static). In naive Port-aware Cache, the entry sizes of the DNS range, HTTP range, and other-application range were set to 12, 704, and 308 entries, respectively, which are the best mix derived from the WIDE trace.

Table III indicates that Semi-static Port-aware Cache can improve the PPC miss rates of all networks compared to conventional PPC, while naive Port-aware Cache improves the PPC miss rate of only the WIDE trace. It means that the best mix of the entry sizes varies depending on the networks and that Semi-static Port-aware Cache fits this demand.

2) *Attack Tolerance:* The attack tolerance of Port-aware Cache was evaluated. To evaluate this, the WIDE trace was used because eight attacks were observed in this trace at 19 second, 46 second, 69 second, 286 second, 465 second, 723 second, 737 second, and 783 second. Fig. 10 shows the PPC miss rates of Port-aware Cache and conventional PPC. When compared to conventional PPC, Port-aware Cache can reduce the number of PPC misses by 8.64% on average against attacks.

Moreover, we analyzed the breakdown of PPC misses in the WIDE trace from 0 second to 100 seconds and showed it in Fig. 11. In comparison to Fig. 3, Port-aware Cache can not

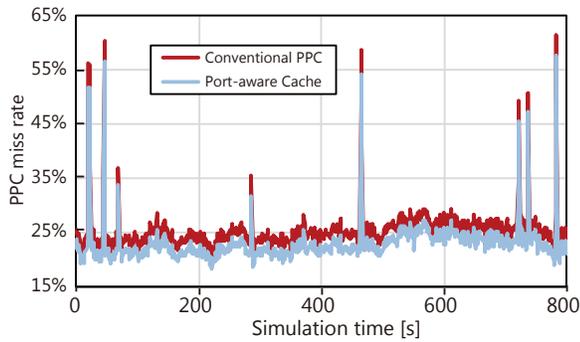


Fig. 10. PPC miss rates of Port-aware Cache and conventional PPC in WIDE trace.

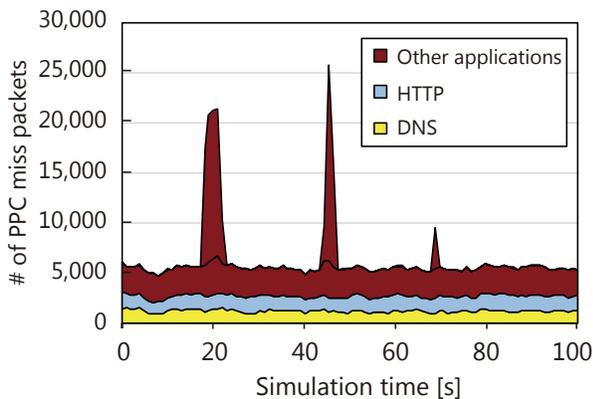


Fig. 11. The breakdown of PPC misses in the case of Port-aware Cache.

only prevent increases in PPC misses caused by attacks but also reduce the number of PPC misses caused by HTTP packets. However, PPC misses caused by other-application packets are still remained largely, and thus, reducing them is an important issue for further improvement.

### C. Evaluation of Victim IP Cache

1) *Improvement in avg. PPC miss rate:* To reveal the usefulness of Victim IP Cache, we also implemented a typical flow-based victim cache and compared the PPC miss rates. The typical flow-based victim cache stores packets missed in PPC per flow. The PPC miss rates are summarized in Table IV. The number of victim cache entries was varied from x1 (compared to PPC, i.e., 32KB) to x8 (i.e., 256KB). As shown in the table, the number of PPC misses can be reduced significantly by Victim IP Cache. For example, 64KB Victim IP Cache improved the PPC miss rate by 78.7% compared to conventional PPC, while the 64KB typical flow-based victim cache improved them by 32.9%. The results showed that Victim IP Cache is more effective than a typical flow-based victim cache. As a future work, there is a room for further improvement in the PPC miss rate by combining the typical flow-based victim cache and Victim IP Cache.

2) *Throughput and Energy:* As evaluated in the previous section, Victim IP Cache can significantly improve the cache miss rate. However, different from PPC, the cache miss rate of

Victim IP Cache does not directly represent the throughput and energy of the table lookup because packets hit in Victim IP Cache must access TCAMs to search the ACL and QoS table. This subsection introduces throughput and energy models of the table lookup with PPC and Victim IP Cache and estimated them based on the calculation. The throughput and energy models are already considered in [14], and this study extends them to evaluate Victim IP Cache.

First, the throughput model was extended. The table lookup throughput obtained by PPC and Victim IP Cache, represented as  $T$ , is calculated as (1).

$$T = \frac{l}{t_{avg}} = \min \left( \frac{l}{t_{ppc}}, \frac{l}{t_{vic}m_{ppc}}, \frac{l}{t_{icam}(2m_{diff}/n + m_{vic})} \right) \quad (1)$$

Here,  $t_{ppc}$ ,  $t_{vic}$ ,  $t_{icam}$ , and  $t_{avg}$  represent the PPC, Victim IP, TCAM, and average lookup latency, respectively, and  $m_{ppc}$  and  $m_{vic}$  represent the cache miss rates of PPC and Victim IP Cache, respectively. Moreover,  $m_{diff}$  represents the gap of the cache miss rates between  $m_{ppc}$  and  $m_{vic}$ . The variables  $n$  and  $l$  in (1) represent the number of tables in a router and the packet length, respectively. In this study, we supposed four tables and 64 bytes as  $n$  and  $l$ , respectively. Equation (1) means that the table lookup throughput is restricted by the minimum throughput among PPC, Victim IP Cache, and TCAM. In comparison to conventional PPC, Victim IP Cache can achieve higher throughput by increasing the achievable throughput of TCAM.

Next, we extended the energy model of the table lookup in a router. The energy consumed by the table lookup with PPC and Victim IP Cache, represented as  $E$ , is calculated as (2).

$$E = D_{ppc} + D_{vic}m_{ppc} + D_{icam}(2m_{diff} + nm_{vic}) + (S_{ppc} + S_{vic} + S_{icam})t_{avg} \quad (2)$$

Here,  $D_{ppc}$ ,  $D_{vic}$ , and  $D_{icam}$  represent the dynamic energy of PPC, Victim IP Cache, and TCAM per access, while  $S_{ppc}$ ,  $S_{vic}$ , and  $S_{icam}$  represent the summation of the static power of each memories. Equation (2) means that introducing Victim IP Cache increases the static power and dynamic energy of Victim IP Cache although it can reduce the dynamic energy of TCAM.

Based on (1) and (2), the throughput and energy consumption of the table lookup with PPC and Victim IP Cache were estimated. The latency and energy consumption of each memory were estimated using CACTI. The calculated values of the throughput and energy consumption are summarized in Tables V and VI. As shown in Table V, Victim IP Cache realizes 1.65x higher throughput in comparison to conventional PPC. However, it also shows that the TCAM access still restrict the table lookup throughput. Furthermore, as shown in Table VI, the total energy consumption of the table lookup can be reduced by 39.1% using Victim IP Cache in comparison to the conventional PPC while typical flow-based victim cache reduce it by 32.6%. Thus, introducing the victim caches is effective to reduce the energy consumption of the table lookup although it additionally consumes the energy of the victim caches.

TABLE IV. AVERAGE PPC MISS RATES OF VICTIM IP CACHE IN VARIOUS NETWORKS.

	IPLS	UFL	MRA	FRG	CNIC	WIDE	PSC	APN	TXG	COS	BUF	Mean
Conventional PPC	26.5%	26.2%	39.3%	5.14%	8.21%	23.4%	11.3%	36.4%	15.3%	11.6%	20.7%	17.3%
w/ typical victim cache												
x1	22.7%	17.7%	32.7%	3.60%	5.22%	19.5%	8.59%	30.3%	11.0%	9.17%	19.7%	13.4%
x2	19.4%	13.8%	24.3%	3.35%	4.43%	16.6%	7.71%	25.9%	9.54%	8.32%	19.4%	11.6%
x4	14.6%	9.71%	17.0%	3.14%	3.48%	14.3%	6.87%	22.3%	7.78%	7.34%	19.2%	9.58%
x8	9.78%	6.30%	11.6%	2.94%	2.66%	12.7%	5.76%	19.4%	5.95%	6.60%	19.0%	7.73%
w/ Victim IP Cache												
x1	15.9%	17.6%	28.6%	2.71%	4.23%	11.2%	3.11%	23.2%	2.94%	1.16%	0.42%	5.37%
x2	10.9%	11.7%	23.1%	2.24%	2.44%	8.86%	1.75%	17.4%	1.49%	0.88%	0.29%	3.68%
x4	5.69%	6.07%	14.4%	1.89%	1.09%	7.34%	1.04%	13.4%	0.76%	0.79%	0.27%	2.42%
x8	2.79%	2.71%	6.66%	1.65%	0.51%	6.41%	0.78%	10.6%	0.53%	0.77%	0.27%	1.65%

TABLE V. ACHIEVABLE THROUGHPUT OF PPC, VICTIM CACHES, AND TCAM.

	Conventional	w/ typical victim	w/ Victim IP Cache
PPC	1,024 Gbps	1,024 Gbps	1,024 Gbps
Victim Cache	N/A	2,964 Gbps	2,964 Gbps
TCAM	<b>593 Gbps</b>	<b>886 Gbps</b>	<b>978 Gbps</b>

TABLE VI. ENERGY CONSUMPTION AND THE BREAKDOWN.

	Conventional	w/ typical victim	w/ Victim IP
TCAM (dynamic)	20.7 nJ (96.3%)	13.9 nJ (95.7%)	12.6 nJ (95.7%)
(static)	0.73 nJ (3.41%)	0.49 nJ (3.39%)	0.45 nJ (3.39%)
Caches (dynamic)	0.03 nJ (0.16%)	0.06 nJ (0.39%)	0.06 nJ (0.43%)
(static)	0.02 nJ (0.07%)	0.08 nJ (0.55%)	0.07 nJ (0.55%)
Total	<b>21.5 nJ</b>	<b>14.5 nJ</b>	<b>13.1 nJ</b>

#### D. Combined Approach

Finally, the combined approach of Port-aware Cache and Victim IP Cache was evaluated. Table VII shows the estimation of the average PPC miss rate, throughput, and energy consumption. The combined approach of Port-aware Cache and Victim IP Cache can achieve 1.72x higher throughput of the table lookups with 41.4% smaller energy per packet. We showed the combined approach reaches 1Tbps with remarkable small energy consumption compared to conventional PPC.

### VII. CONCLUSION

PPC has been proposed to realize high-throughput and low-energy table lookup in routers. PPC shows significant impact on the table lookup throughput and the energy consumption; however, there is a room for improvement. To further improve PPC performance, this paper proposed two novel cache architecture, called Port-aware Cache and Victim IP Cache.

Port-aware Cache divides the cache space into three areas: DNS, HTTP, and other-application areas. it can isolate the influence of flows, such as attack flows, and improve the PPC miss rate by assigning appropriate number of PPC entries to each area. The simulation result indicated that Semi-static Port-aware Cache can decide the best mix of each entry sizes by prior trials at a boot process of a router and improve the PPC miss rate by 9.02% and the number of PPC misses in attacks by 8.64%. For further improvement, it is required to reduce the number of PPC misses caused by applications without HTTP and DNS.

Victim IP Cache supports PPC by caching the routing table and ARP table lookup results of packets which miss in PPC.

Because IP Cache has a possibility to achieve higher cache hit rate than PPC, it may save a large number of PPC miss packets. The simulation results showed that 64KB Victim IP Cache can further improve the cache miss rate by 85.5%. As a result, the energy consumption of the table lookup can be reduced by 39.1% compared to conventional PPC.

Finally, the combined approach of Port-aware Cache and Victim IP Cache was considered. The simulation results also showed that both the highest throughput and lowest energy consumption can be achieved by the combined approach. It showed 1.72x higher table-lookup throughput with 41.4% smaller energy per packet in comparison to conventional PPC.

### ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP18K18022.

### REFERENCES

- [1] The Ministry: Tabulation and Estimation of Internet Traffic in Japan, 2016, [http://www.soumu.go.jp/main\\_content/000462459.pdf](http://www.soumu.go.jp/main_content/000462459.pdf)
- [2] METI: Green IT Initiative in Japan, <http://www.meti.go.jp/english/policy/GreenITInitiativeInJapan.pdf>.
- [3] Fan, J., Hu, C., He, K., Jiang, J., and Liuy, B.: Reducing power of traffic manager in routers via dynamic on/off-chip scheduling, 2012 Proc. IEEE INFOCOM, pp.1925-1933, Orlando, FL (2012).
- [4] Zheng, X. and Wang, X.: Comparative study of power consumption of a NetFPGA-based forwarding node in publish-subscribe Internet routing, Computer Communications, vol. 44, pp.36-43 (2014).
- [5] Gamage, S. and Pasqual, A.: High performance parallel packet classification architecture with PoPular Rule Caching, 18th IEEE International Conference on Networks (ICON), pp.52-57, Singapore (2012).
- [6] Talbot, B., Sherwood, T., and Lin, B.: IP caching for terabit speed routers, IEEE Global Telecommunications Conference (GLOBECOM'99), vol.2, pp.1565-1569, Brazil (1999).
- [7] Guinde, N. B., Rojas-Cessa, R., and Ziavras, S. G.: Packet classification using rule caching, 2013 Fourth International Conference on Information, Intelligence, Systems and Applications (IISA 2013), pp1-6, Piraeus (2013).
- [8] Agrawal, B. and Sherwood, T.: Ternary CAM Power and Delay Model: Extensions and Uses, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol.16, no.5, pp.554-564 (2008).
- [9] Nawa, M. et al.: Energy-efficient high-speed search engine using a multi-dimensional TCAM architecture with parallel pipelined subdivided structure, 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), pp.309-314, Las Vegas, NV (2016).
- [10] Hewlett-Packard Development Company.: Energy Efficient Networking - Business white paper, <http://h17007.www1.hp.com/docs/mark/4AA3-3866ENW.pdf>.
- [11] Okuno, M. and Nishi, H.: Network-Processor Acceleration-Architecture Using Header-Learning Cache and Cache-Miss Handler, The 8th World Multi-Conference on Systemics, Cybernetics and Informatics (SCI 2004), pp.108-113 (2004).

TABLE VII. SUMMARY OF PPC MISS RATE, THROUGHPUT, AND ENERGY CONSUMPTION.

	Conventional	w/ Port-aware Cache	w/ Victim IP Cache	w/ combination
Miss rate (PPC) [%]	17.3%	16.4%	17.3%	16.4%
(victim) [%]	N/A	N/A	3.68%	3.65%
Throughput [Gbps]	592 Gbps	624 Gbps	978 Gbps	1,021 Gbps
Energy [nJ]	21.5 nJ	20.4 nJ	13.1 nJ	12.6 nJ

- [12] Yamaki, H. and Nishi, H.: An Improved Cache Mechanism for a Cache-based Network Processor, In Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'12), pp.1-7, Las Vegas, NV (2012).
- [13] Yamaki, H.: Efficient Cache Architecture for Packet Processing in Internet Routers, In Proceedings of the 2020 Future of Information and Communication Conference (FICC 2020), Vol.1, pp.338-352, CA, USA (2020).
- [14] Yamaki, H. and Nishi, H.: Line Replacement Algorithm for L1-scale Packet Processing Cache, In Adjunct Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing Networking and Services (MOBIQUITOUS 2016), pp.12-17, Hiroshima, Japan (2016).
- [15] Yamaki, H., Nishi, H., Miwa, S., and Honda, H.: Data Prediction for Response Flows in Packet Processing Cache, 2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC), pp.1-6, San Francisco, CA (2018).
- [16] Chang, F., Feng, W. C., and Li, K.: Efficient Packet Classification with Digest Caches, In Proceedings of Third Workshop Network Processors and Applications (NP-3), 2005.
- [17] Ata, S., Murata, M., and Miyahara, H.: Efficient cache structures of IP routers to provide policy-based services, IEEE International Conference on Communications (ICC 2001), vol.5, pp. 1561-1565, Helsinki (2001).
- [18] Liao, G., Yu, H., and Bhuyan, L.: A new IP lookup cache for high performance IP routers, In Proceedings of the 47th Design Automation Conference (DAC), pp.338-343, Anaheim, CA (2010).
- [19] Carter, J. L., and Wegman, M. N.: Universal classes of hash functions (Extended Abstract), In Proceedings of the ninth annual ACM symposium on Theory of computing (STOC '77), pp.106-112, New York, NY (1977).
- [20] Kim C., Caesar M., Gerbear A., and Rexford J.: Revisiting Route Caching: The World Should Be Flat, In Proceedings of the 10th International Conf. on Passive and Active Network Measurement (PAM '09), pp.3-12 (2009).
- [21] Muralimanohar, N. et al.: Optimizing NUCA organizations and wiring alternatives for large caches with CACTI 6.0, In Proceedings of the 40th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO 40), pp.3-14, Chicago, USA (2007).
- [22] RIPE Network Coordination Centre: Réseaux IP Européens Network Coordination Centre RIPE NCC, <http://www.ripe.net/>.
- [23] WIDE MAWI WorkingGroup: MAWI Working Group Traffic Archive, <http://mawi.wide.ad.jp/mawi/>.

# A Robust Scheme to Improving Security of Data using Graph Theory

Khalid Bekkaoui<sup>1</sup>, Soumia Ziti<sup>2</sup>, Fouzia Omary<sup>3</sup>  
Intelligent Processing and Security of Systems (IPSS)  
Faculty of Sciences, Mohammed V University  
Rabat, Morocco

**Abstract**—With the incredible growth of using internet and other new telecommunication technologies, cryptography has become an absolute necessity for securing communications between two or more entities, particularly in the case of transferring confidential data. In the literature, many encryption systems have been proposed against attack threats. These schemes should normally overcome the concerns by ensuring confidentiality, integrity and authenticity of transmitted data. However, several of them have shown weaknesses in terms of security and complexity. Hence the need for a robust and powerful non-standard encryption algorithm to prevent any traditional opportunity to sniff data. In this work, we propose a new encryption system that perfectly meets the security requirements. The scheme is based essentially on the principles of graph theory which are very promising at plain text representations. Our approach proposes another use of the concept of Hamiltonian circuit and adjacency matrix using a shared key and a pseudo-random generator. After analysis of the experimental results, which were very promising, the technique was found to be both efficient and robust.

**Keywords**—Cryptography; encryption; security; graph theory; Hamiltonian circuit; adjacency matrix

## I. INTRODUCTION

Cryptography is a branch of cryptology that relies on a set of techniques and methods that transform a clear or readable message into a completely unintelligible one. This discipline deals with several security issues such as the confidentiality of transmissions through unsecured channels, the privacy of individuals, the archiving of data on unsecured media, and so on. Cryptography thus allows the study and analysis of data encryption systems aimed at minimizing the reach of hackers and limiting, as much as possible, their unauthorized access to such data, while preserving the key concepts of information security that are confidentiality, integrity, authentication and finally non-repudiation [1]. The purpose of cryptography is therefore the construction of protection schemas that provide ironclad assurances of confidentiality or authenticity of transmitted messages when dealing with malicious attempts to access them.

Confidentiality is an essential aspect of security. It can be guaranteed via an encryption mechanism, through which data becomes unintelligible to any unauthorized party attempting to access it. The role of encryption algorithms is to transform a clear message into an encrypted one so that only authorized people can retrieve the message in its original, clear form by performing the reverse-process to encryption, namely decryption. By design, decryption should be made as difficult as possible to any unqualified, unauthorized party attempting to carry it out.

Over the years, cryptography has steadily evolved and gradually become indispensable to modern society. Any and all contributions to this field of work have always been of great interest. According to the literature, there are three types of cryptography. The first is symmetric-key cryptography, which involves using a secret key, such as DES [2], IDES [3], AES [4], or others for the purposes of encryption. The second is asymmetric cryptography, which is based on the use of two different keys (one private and the other public) [1], such as RSA [5], ElGamal [6], Diffie-Hellman [7], and so on. The third and last type is what's called hybrid cryptography, and it combines the previous two encryption methods.

Today, modern cryptology is able to make use of a considerable set of mathematical tools. This has led to greater gains in performance and efficiency. Graph theory in particular is an area that is seen as being potentially promising in this respect, as it introduces concepts that might help solve problems in all areas related to networks.

Graph theory in mathematics and computer science is the study of graphs. Generally, a graph can be used to represent the structures and connections that form a complex set while using clear representations of the elements and expressing the relationships between them in a more tangible way, that is by defining communication networks for instance, as well as road networks, electrical circuits, and so on. Graphs therefore offer a way of thinking that allows for the modeling of a wide variety of problems using edges and vertices to represent them. The Seven Bridges of Königsberg (1736) [8] is a mathematical problem known for having laid the foundations for graph theory. Since the beginning of the 20th century, it has developed into a full-fledged branch of mathematics, thanks in part to the work of König, Menger, Cayley, Berge, and Erdős. (References must be added)

Graph theory has become a key element in many applications within computer science. It's a relatively recent concept that has been successfully integrated and has allowed for the development of more powerful encryption algorithms that have proven difficult to crack even for modern software solutions. This is essentially a matter of modeling encryption problems by representing them in graph form so that they become problems in graph theory to which solutions are generally known or more accessible. Solutions to graph problems can be relatively easy and efficient (the time it takes to process them computationally can be fairly reasonable given their polynomial dependence on the number of vertices in the graph). The solutions can also be quite difficult (where processing time increases exponentially) in which case a heuristic a practical

problem-solving approach is used to find the optimal solution.

As a relatively new yet quite powerful tool, graph theory is recognized by government agencies and organizations that have a vested interest in security as having made considerable contributions. It is used in the development of various encryption techniques as well as in sophisticated data communications. This has led to the application of the concepts introduced in graph theory in cryptography on a broad scale, seeing as many NP-hard problems stem from this theory.

Given all of the above, there seemed to be a great need for a new encryption system based on graph theory to be developed, that would ensure a high degree of security while requiring relatively simple resource processing. In this paper, an application of the principles related to this theory in the field of cryptography is presented; its aim being the development of a communications scheme that is both efficient and secure. This proposal makes use of disjoint Hamiltonian circuits for the presentation of data, and of the divide-and-conquer paradigm to simplify processing and facilitate encryption.

The remainder of the paper is arranged as follows. Section 2 presents preliminary knowledge. A literature review of related works is explained in Section 3. Section 4 describes the proposed scheme. Experimental results and analyses are detailed in Section 5, and finally, the conclusion is given in Section 6.

## II. PRELIMINARY KNOWLEDGE

- **Graph** : Graph theory is a branch of applied mathematics. Fundamentally, a graph consists of a set of vertices, and a set of edges, where an edge is something that connects two vertices in the graph. A graph is a pair  $(V, E)$ , where  $V$  is a finite set and  $E$  is a binary relation on  $V$ .  $V$  is called a vertex set whose elements are called vertices.  $E$  is a collection of edges, where an edge is a pair  $(u,v)$  with  $u,v$  in  $V$ . Graph  $G = (V, E)$  is a collection of  $V$  nodes connected by  $E$  links [1].
- **Simple graph** : Undirected graph that has no loops (edges connected at both ends to the same vertex) and no more than one edge between any two different vertices.
- **Path** : A path is a simple graph whose vertices can be ordered so that two vertices are adjacent if and only if they are consecutive in the list [1].
- **Undirected Graph** : A graph in which each edge symbolizes an unordered, transitive relationship between two nodes. Such edges are rendered as plain lines or arcs [1].
- **Cycle** : Refers to a chain where the initial and terminal node is the same and that does not use the same link more than once is a cycle.
- **Hamiltonian Path** : A path that visits each vertex exactly once in an undirected graph.
- **Hamiltonian Circuit** : A Hamiltonian cycle (or Hamiltonian circuit) is a Hamiltonian Path such that there is an edge (in graph) from the last vertex to the first vertex of the Hamiltonian Path.

- **Adjacency Matrix** : Given a graph  $G$  with  $n$  vertices (ordered from  $v_1$  to  $v_n$ ). The Adjacency Matrix  $M$  of size  $n \times n$  related to  $G$  can be defined by:

$$\begin{cases} M_{ij} = p & \text{if There exists a path from } v_i \text{ to } v_j \\ M_{ij} = 0 & \text{otherwise.} \end{cases} \quad (1)$$

Where  $p$  is the weight of the edge  $(v_i, v_j)$ .

- **Divide and Conquer** : An algorithmic strategy which is mainly based on dividing an initial problem into several roughly equal sub-problems, and then solve the sub-problems separately before combining their results. This strategy is able to considerably reduce the complexity of mathematical problems that require a lot of processing.
- **Blum Blum Shub (BBS)** : [9] A simple unpredictable pseudo-random number generator that was proposed in 1986, and whose mathematical equation is described as follows:

$$x_{n+1} = x^2 \pmod{M} \quad (2)$$

where  $M = p.q$  is the product of two prime numbers  $p$  and  $q$ . The security of this generator fully depends on the complexity of factoring  $M$ , which means that the two primes must be properly chosen to ensure a certain robustness. BBS is a good choice for many applications, especially those related to cryptography as it can generate unpredictable sequences.

## III. RELATED WORKS

Nowadays, graph theory has contributed greatly to the development of various encryption techniques. A review of the relevant literature reveals several methods that have been put forward for such purposes.

Yamuna [10] proposed an encryption mechanism using Hamiltonian paths. The data is represented using a Hamiltonian path, and the complete graph is constructed by weighting the remaining vertices to increase the level of security.

Al Etaiwi [11] put forward a new encryption algorithm based on graph theory. His paper presents a new symmetric encryption algorithm that uses the concepts of complete graph and minimum spanning tree to strengthen security.

Yamuna [12] showed that Hamiltonian circuits could be used to represent multiple messages through a single graph and that encryption could be done using time-dependent functions.

Yamuna and al. [13] used musical notation to represent the secret key (musical note) and graph theory properties to generate keys. This approach is based on the Propagating Cipher Block Chaining (PCBC) mode for encrypting binary messages. In 2014, the same authors proposed a PIN-code encryption method in the form of a digraph [14].

In [15], the authors have proposed a graph based modified DES (GMDES) algorithm which is more secure than the classical DES algorithm [2]. The proposed graph is not fully depended on secret key, and for the same plain text it produces

different cipher text using the same secret key which reduces the probability of various attacks.

Agarwal and Uniyal [16] proposed an encryption scheme based on transforming ASCII values into prime numbers using an encryption key of similar size to the clear message. The authors then randomly generated a prime weighted graph by taking into account the prime number weights assigned to each of the edges.

The system that Amounas [17] put forward handles the original data using graph theory and some of its properties. The main concept being the generation of the complete weighted graph. More specifically, this approach offers a new way of labeling the edges of a graph. It subsequently applies the matrix approach based on ECC operations to generate strong encrypted text.

Recent work in the literature includes the technique proposed in [18], where each character of the data has been encrypted into an Euler Graph. they used the Hamiltonian Circuit as key to secure the data.

Selim G. Akl designed in [19] an encryption process to transmit a secure message. The author employed three distinct graphs constructed successively, and based on an unconventional mapping, conjectured to be a trapdoor one-way function, and which is conceived especially for graph structures.

In [20], two graph based public key cryptosystems have been proposed for protecting valuable information. The first method is purely based on matrix properties, and the second is based on graphical codes.

#### IV. THE PROPOSED APPROACH

The system put forward in this paper uses the fundamental concepts of graph theory to facilitate the handling of raw data. The basic idea is to generate weighted graphs by using Hamiltonian circuits in a novel way.

The results that have been achieved in this work appear to be promising, especially in terms of complexity and speed when it comes to processing the clear message. The proposed mode of operation is essentially based on the divide-and-conquer design paradigm which involves breaking down an initial problem into smaller sub-problems and then dealing with each sub-problem separately.

We conceive this approach primarily to address complexity concerns, as most existing works represent a clear message using a graph of similar size. Which, in turn, becomes an adjacency matrix used to process and handle data. It automatically follows that in such cases, the larger the matrix, the more complex the linear operations will be.

With that in mind, using the divide-and-conquer design paradigm has allowed us to reduce complexity by dividing the message into smaller blocks instead of processing it in its entirety. Each block is represented by disjoint Hamiltonian circuits to reduce the size of the graph associated with the block.

To describe fully how the new encryption technique actually works, we will illustrate it in two main algorithms (Encryption in Algorithm 1 and Decryption in Algorithm 2).

Each algorithm includes some functions that we will define and explain their functioning.

The input of our Graph Encryption algorithm is a clear message with  $n$  characters. The function **ASCII\_Transformation** converts each character of message into its ASCII value. It returns an array of integer belonging to the interval  $[0,255]$ .

The second function **Decomposition\_Block** decomposes the array into several  $k'$  blocks  $BlockSet_{k'}$  using the following formula:

$$n = 25k + r \quad (3)$$

Such that  $n$  is the size of the message,  $r$  (belonging to the interval  $[0, 24]$ ) is the remainder of the division of the message by 25, and  $k$  its quotient. If the division is accurate then  $k' = k$ , otherwise  $k' = k + 1$ .

The third function **Decomposition\_Key** generates from KEK (Key Encryption Key) master key of size  $m = 256$ ,  $K_i$  sub-keys as square matrices of order 13 (where  $i = 0 \dots k' - 1$ ) which will be used to encrypt each block (Fig. 1 depicts this process).

The generation of sub-keys is carried out in two steps. First, a size  $m = 13k'$  Key is generated from a parameter  $m = 256$  size KEK master key. Then,  $k'$  other sub-keys are generated from that Key as square matrices of order 13. For each  $Block_i$ , the ASCII value of the first character is used to specify a digit of the KEK master key (using the position that is supposed to fill in the range  $[0,255]$ ). This digit is normally used to generate the seed  $S_i$  of size 13. Then we use the vector  $S_i$  to generate the sub-key  $K_i$  in the form of square matrices of order 13.

Additionally, each size-25 block (minimum block size) is partitioned into two size 13 and 12 sub blocks respectively. Given that in a complete graph with  $n$  vertices, there is  $(n-1)/2$  disjoint Hamiltonian circuits if  $n$  is an odd number strictly greater than 3 [21].

A graph of size 13 can contain 6 disjoint Hamiltonian circuits. It follows that we represent each block by a graph of size 13 containing two disjoint Hamiltonian circuits. Moreover, we convert the odd size sub block into an Eulerian cycle using the ASCII values of its characters, thus representing the weights of the edges. Thereafter, we represent the second sub block by using one of the other five Hamiltonian circuits by filling the missing values with the ASCII code of the null character (**Block\_Graph**). Then we use the adjacency matrix to represent the resulting graph (**Graph\_Matrix**). This representation is very advantageous, not only in terms of the complexity of the processing but also compared to the traditional representation of the message, which would normally take place within a single Eulerian cycle.

##### A. Encryption Mode

We encrypt each block using CBC (Cipher Block Chaining). In this mode, an 'exclusive OR' (XOR) operation is applied to each  $Block_i$  using the preceding  $block'$ 's encryption before encrypting the current block itself using the same

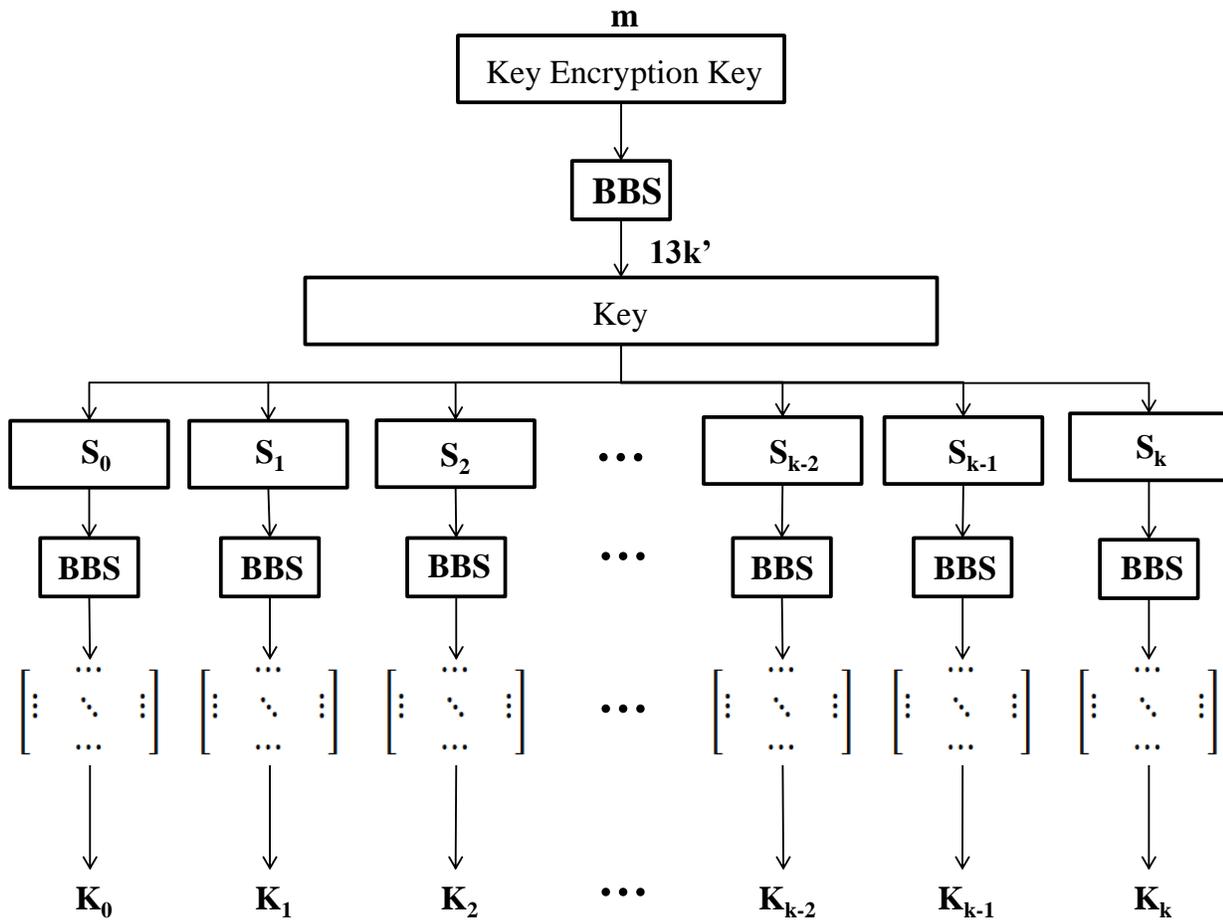


Fig. 1. Key Generator

process. The binary operator XOR concerns in our case the  $M_i$  and the  $C_{i-1}$  matrices as follows:

$$M'_i = C_{i-1} \oplus M_i \quad (4)$$

The result is then used in another XOR operation with the  $K_i$  sub-key produced by the BBS pseudo-random generator to obtain the  $C_i$  of the current block,

$$C_i = M'_i \oplus K_i \quad (5)$$

The encryption of the first block ( $M_0$ ) is performed after passing a randomly generated initialization matrix ( $IM$ ) through an XOR gate. Each encrypted block will be represented by a vector  $V_i$  of size  $13^2$  by concatenating the lines of the adjacency matrix using the function **ConcatenateLines**. Finally, the resulting vectors are concatenated by the function **ConcatenateVec** resulting a single vector of size  $13^2 k'$  which will represent the encrypted message  $EM$  sent back out, in addition to the vector  $FCB$  containing the positions used to generate the Key. Fig. 2 clearly illustrates the encryption mode.

### B. Decryption Mode

The input of our Graph Decryption algorithm is a encryption message  $EM$  of size  $m$  ( $13^2 k'$ ). The function **Decomposition\_Vector** decomposes the Vector  $V$  of size  $13^2 k'$  into several  $k'$  encrypted block  $EMatrixSet_{k'}$ . To decrypt one message of size  $m$ . The number of blocks is calculated as follows:

$$k' = m \div 13^2 \quad (6)$$

We generate also one key of size  $13k'$  from the master key using vector  $FCB$ . Sub-keys are then generated to decrypt each block  $C_i (i = 1 \dots k')$ . Using the formula

$$M_i = C_{i-1} \oplus M'_i \quad (7)$$

Where

$$M'_i = C_i \oplus K_i \quad (8)$$

and

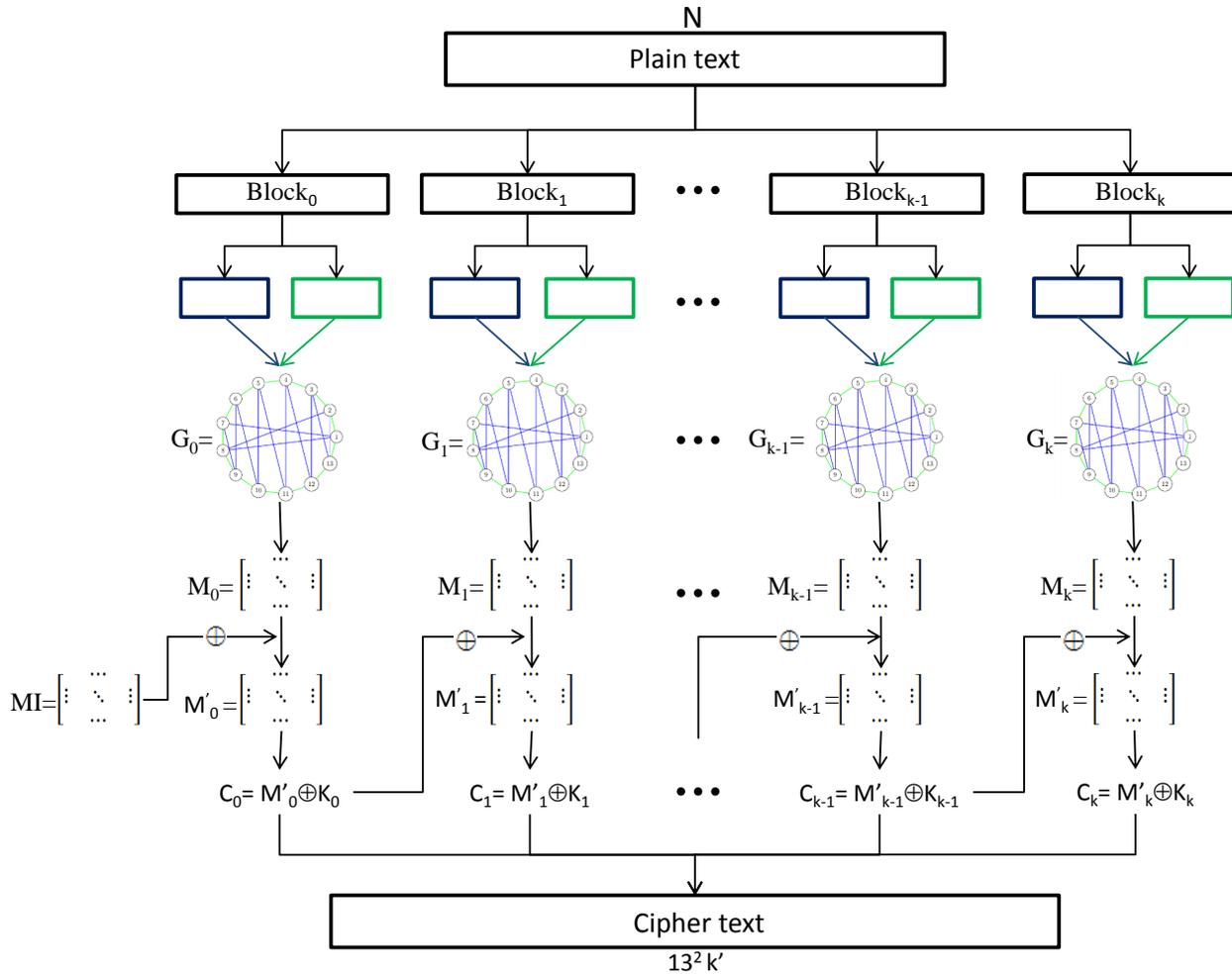


Fig. 2. Encryption Scheme

$$M_0 = IM \oplus M'_0 \quad (9)$$

After this stage, each block is decrypted then the disjoint circuits are extracted and the blocks  $Block_i$  (**Graph\_Block**) are reconstructed. Moreover, we convert the ASCII values into characters for each block. Finally, the character blocks are concatenated (**Concatenate\_Block**) to form the clear message as shown in Fig. 3.

## V. EXPERIMENTAL RESULTS AND ANALYSES

The evaluation of the Encryption technique includes the performance and efficiency of the algorithm, and on the other hand how the scheme can react in terms of robustness against certain attacks such as the Brute-force attack.

### A. Statistical Tests

In this part, the DIEHARD test [22] is used to analyze the quality of the random generation of the proposed block cipher. The main purpose of this test is to establish that our algorithm is able to withstand statistical attacks. In other words, a secure

block cipher output should be statistically indistinguishable from a random output via the encryption function. For this test to be carried out, a sequence of randomly generated ciphers is first converted into binary to produce a bit-stream larger than 10 MB. Then, this bit-stream is statistically analyzed by subjecting it to the DIEHARD tests. The DIEHARD tests verify the p-value of the randomly generated numbers, where the p-value is in the interval [0.025, 0.975]. The mean values that were obtained are summarized in Table I. Results show that the bit-stream generated using our proposed method has passed all DIEHARD tests. What is more, our encryption system displays satisfactory random and statistically indistinguishable behavior.

### B. Brute-Force Attack

Brute-force attacks are a way to find all potential key arrangements using a fast prediction tool. On the assumption that a high-quality machine that takes  $10^{-10}$  seconds to test the validity of each key is used, and that the numbers used in the master key are between 1 and 100, our algorithm has  $100^{256}$  possible keys. A brute-force attack would take about  $10^{-10} \times 100^{256}$  seconds to obtain the correct key. Thus, a

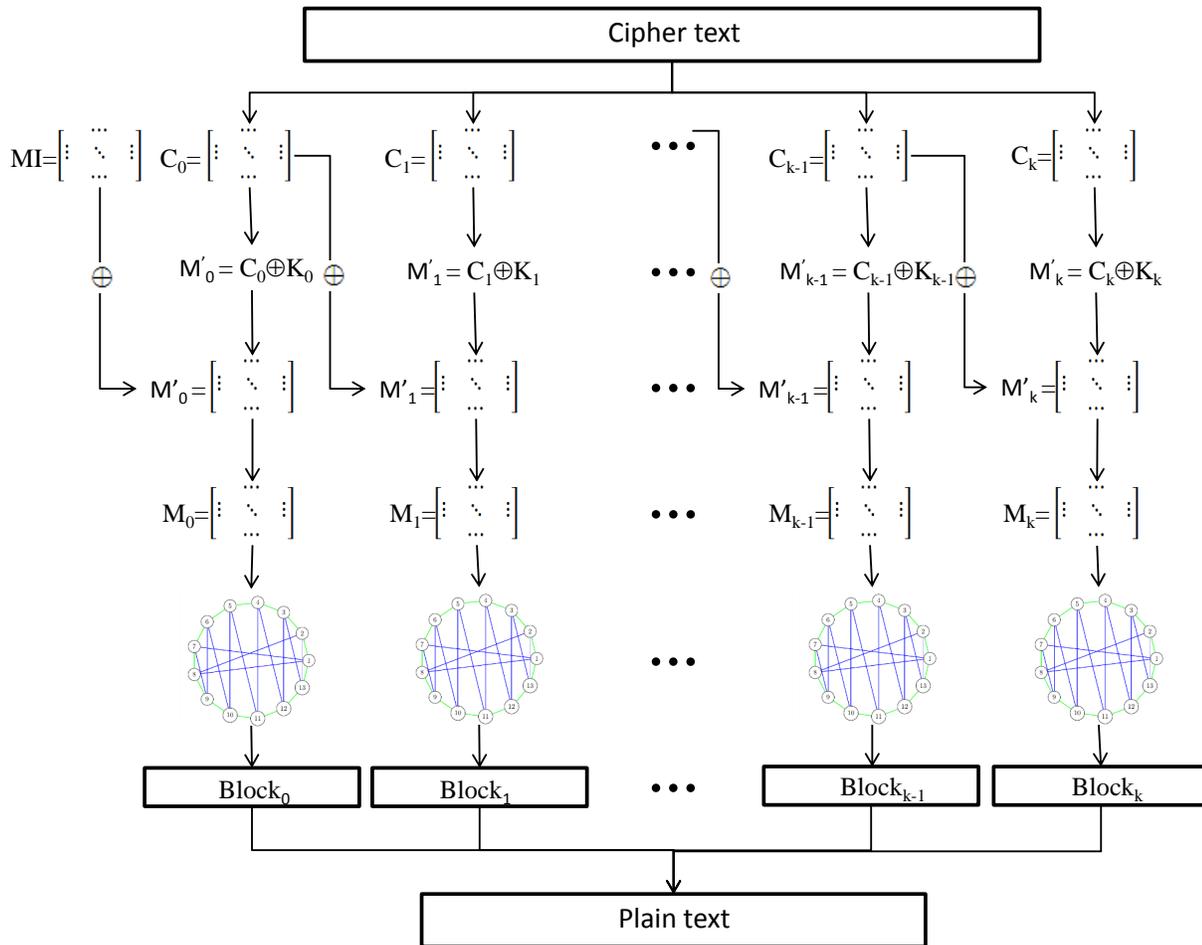


Fig. 3. Decryption Scheme

brute-force attack with an exhaustive key search is impossible within a reasonable timeframe.

To reveal a 25-character message, that is in the case where only one block is used, there are 100 possibilities to determine one of the master key's numbers, which will represent the seed for the BBS generator that's used to generate the vector  $S_0$ . That said, the prime numbers used as input parameters for the generator are difficult to determine (because of factorization issues). Therefore, it is almost impossible to find the sub-key if the pq product is sufficiently large.

## VI. CONCLUSION

This paper puts forward a new block-based encryption scheme that utilizes the divide-and-conquer paradigm as well as the fundamental concepts of graph theory to simplify and facilitate processing. Various statistical tests have been carried out to prove that this new algorithm is secure. All of those tests have confirmed that this algorithm resists statistical attacks. Moreover, the BBS-based generator has been used to generate encryption keys for our algorithm, which has further improved key strength. As future work, we aim to design our own pseudo-random generator in order to provide pseudo-keys. We

also aim to exploit other graph theory properties for a more robust representation of data.

## REFERENCES

- [1] A. J. Menezes, J. Katz, P. C. Van Oorschot, and S. A. Vanstone, *Handbook of applied cryptography*. CRC press, 1996.
- [2] P. FIPS, "81, des modes of operation," *Issued December*, vol. 2, p. 63, 1980.
- [3] W. Meier, "On the security of the idea block cipher," in *Workshop on the Theory and Application of Cryptographic Techniques*. Springer, 1993, pp. 371–385.
- [4] V. Rijmen and J. Daemen, "Advanced encryption standard," *Proceedings of Federal Information Processing Standards Publications, National Institute of Standards and Technology*, pp. 19–22, 2001.
- [5] N. P. Smart, "The "naive" rsa algorithm," in *Cryptography Made Simple*. Springer, 2016, pp. 295–311.
- [6] —, "Public key encryption and signature algorithms," in *Cryptography Made Simple*. Springer, 2016, pp. 313–347.
- [7] A. J. Menezes, J. Katz, P. C. Van Oorschot, and S. A. Vanstone, *Handbook of applied cryptography*. CRC press, 1996.
- [8] G. Alexanderson, "About the cover: Euler and königsberg's bridges: A historical view," *Bulletin of the american mathematical society*, vol. 43, no. 4, pp. 567–573, 2006.

TABLE I. DIEHARD TEST

Test name	p-value	Assessment
diehard birthdays	0.26375543	PASSED
diehard operm5	0.37541747	PASSED
diehard rank 32x32	0.95699200	PASSED
diehard rank 6x8	0.09885031	PASSED
diehard bitstream	0.93471890	PASSED
diehard opso	0.73703729	PASSED
diehard oqso	0.17764052	PASSED
diehard dna	0.81870913	PASSED
diehard count 1s str	0.97971001	PASSED
diehard count 1s byt	0.49404053	PASSED
diehard parking lot	0.27155245	PASSED
diehard 2dsphere	0.35355239	PASSED
diehard 3dsphere	0.58482092	PASSED
diehard squeeze	0.91687701	PASSED
diehard sums	0.15611362	PASSED
diehard runs	0.64253136	PASSED
diehard craps	0.63765229	PASSED
marsaglia tsang gcd	0.74963931	PASSED
sts monobit	0.92126172	PASSED
sts runs	0.28893885	PASSED
sts serial	0.50145071	PASSED
rgb bitdist	0.69014502	PASSED
rgb minimum distance	0.57112646	PASSED
rgb permutations	0.59422228	PASSED
rgb lagged sum	0.59927829	PASSED
rgb kstest test	0.10026759	PASSED
dab bytedistrib	0.49551450	PASSED
dab dct	0.08738803	PASSED
dab filltree	0.22157233	PASSED
dab filltree2	0.22430630	PASSED
dab monobit2	0.15458308	PASSED

```

input : Clear message of n characters  $CMC_n$ , KEK
         master key, Random square Matrix  $IM$  of
         size 13
output: Encrypted message  $EM$ (Vector  $V$  of size
          $169k'$ ), the vector  $FCB$ 

1 begin
2    $CMAN=ASCII\_Transformation(CMC_n,n)$ ;
3   if  $n\%25 == 0$  then
4     |  $k'=n \div 25$  ;
5   else
6     |  $k'=(n \div 25) + 1$  ;
7   end
8    $BlockSet_{k'}=Decomposition\_Block(CMAN,k')$ ;
    $K_{k'}=Decomposition\_Key(KEK,k',FCB)$ ;
9   for element  $Block_i$  of the set  $BlockSet_{k'}$  do
10    |  $G_i=Block\_Graph(Block_i,13)$ ;  $M_i =$ 
      | Graph\_Matrix( $G_i$ );
11    | if  $i=0$  then
12      |  $M'_0 = IM \oplus M_0$ ;
13    | else
14      |  $M'_i = C_{i-1} \oplus M_i$ ;
15    | end
16    |  $C_i = M'_i \oplus K_i$ ;
17    |  $V_i=Concatenate\_Lines(C_i)$ ;
18  | end
19  |  $EM=Concatenate\_Vec(V_{k'})$ ;
20 end

```

Algorithm 1: Encryption Algorithm

```

input : Encrypted message  $EM$ (Vector  $V$  of size
          $169k'$ ), KEK master key, the vector  $FCB$ 
output: Clear message of n characters  $CMC_n$ , KEK
         master key, Random square Matrix  $IM$  of size
         13

1 begin
2    $k'=m \div 13^2$ ;
3    $EMatrixSet_{k'}=Decomposition\_Vector(EM,k')$ ;
    $K_{k'}=Decomposition\_Key(KEK,k',FCB)$ ;
4   for element  $C_i$  of the set  $EMatrixSet_{k'}$  do
5     |  $M'_i = C_i \oplus K_i$ ;
6     | if  $i=0$  then
7       |  $M_0 = IM \oplus M'_0$ ;
8     | else
9       |  $M_i = C_{i-1} \oplus M'_i$ ;
10    | end
11    |  $G_i=Matrix\_Graph(M_i)$ ;
12    |  $Block_i=Graph\_Block(G_i)$ ;
13  | end
14  |  $BlockSet_{k'}=Concatenate\_Block(Block_{k'},k')$ ;
15  |  $CMC_n=Concatenate\_Block(BlockSet_{k'},k')$ ;
16 end

```

Algorithm 2: Decryption Algorithm

[9] L. Blum, M. Blum, and M. Shub, "A simple unpredictable pseudo-random number generator," *SIAM Journal on computing*, vol. 15, no. 2, pp. 364–383, 1986.

[10] M. Yamuna, M. Gogia, A. Sikka, and M. J. H. Khan, "Encryption using graph theory and linear algebra," *International Journal of Computer Application. ISSN*, pp. 2250–1797, 2012.

[11] W. M. Al Etaiwi, "Encryption algorithm using graph theory," *Journal of Scientific Research and Reports*, pp. 2519–2527, 2014.

[12] M. Yamuna, k. D. Mukesh, K. Tushar, and R. Tnujoy, "Encryption of multiple messages using hamiltonian circuit and time dependent function," *International Journal of Advanced Scientific Research and Technology*, vol. 3, pp. 2249–6149, 2013.

[13] M. Yamuna, A. Sankar, S. Ravichandran, and V. Harish, "Encryption of a binary string using music notes and graph theory," *International Journal of Engineering and Technology*, vol. 5, no. 3, pp. 2920–2925, 2013.

[14] M. Yamuna, A. Suwathi, and N. Krishnan, "Four digit pin number as a digraph," *International Journal of Computer Application*, no. 4, pp. 100–107, 2014.

[15] D. Sensarma and S. S. Sarma, "Gmdes: A graph based modified data encryption standard algorithm with enhanced security," *Int J Res Eng Technol*, vol. 3, no. 3, pp. 653–60, 2014.

[16] S. Agarwal and A. S. Uniyal, "Prime weighted graph in cryptographic system for secure communication," *International Journal of Pure and Applied Mathematics*, vol. 105, no. 3, pp. 325–338, 2015.

[17] F. AMOUNAS, "An innovative approach for enhancing the security of amazigh text using graph theory based ecc," *International journal of scientific research in science, engineering and technology*, 2016.

[18] P. Amudha, A. C. Sagayaraj, and A. S. Sheela, "An application of graph theory in cryptography," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 13, pp. 375–383, 2018.

[19] S. G. Akl, "The graph is the message: design and analysis of an unconventional cryptographic function," *From Parallel to Emergent Computing, Adamatzky, A. et al., Eds., Taylor & Francis, CRC Press, Boca Raton, Florida*, 2019.

[20] D. Sensarma and S. S. Sarma, "Application of graphs in security," *Inter-*

*national Journal of Innovative Technology and Exploring Engineering*,  
2019.

- [21] N. Deo, *Graph theory with applications to engineering and computer science*. Courier Dover Publications, 2017.
- [22] G. Marsaglia, "Diehard test suite," *Online: [http://www. stat. fsu. edu/pub/diehard](http://www.stat.fsu.edu/pub/diehard)*, vol. 8, no. 01, p. 2014, 1998.

# 2D/3D Registration with Rigid Alignment of the Pelvic Bone for Assisting in Total Hip Arthroplasty Preoperative Planning

Christian A. Suca Velando<sup>1</sup>, Eveling G. Castro Gutierrez<sup>2</sup>  
Escuela Profesional de  
Ingeniería de Sistemas  
Universidad Nacional de San Agustín  
Arequipa, Perú

**Abstract**—In Total Hip Arthroplasty preoperative planning requires the definition of medical parameters that help during the intraoperative process; these parameters must be allocated with accuracy to make an implant to the patient. Currently, preoperative planning carries out with different methods. It can be by using a prosthesis template (2D) projected on x-ray images or by using a computed tomography (CT) in order to set a 3D prosthesis. We propose an alternative developing preoperative planning through reconstructed 3D models using 2D x-ray images, which help to get the same precise information such as a CT. On this paper it has proposed to test the framework from the authors Bertelsen A and Borro D, it is an ITK-Based Framework for 2D-3D Registration between x-ray images and a computed tomography. We used the approach of this paper using two fixed images (reference images) and one moving image (image to transform) to do a intensity registration. This method uses a ray casting interpolator to generate a Digitally Reconstructed Radiograph (DRR) or virtual x-ray. We also applied a normalized gradient correlation for comparing the patient x-ray image and the virtual x-ray image optimized by a nonlinear conjugate gradient, both metric and optimizer are useful to update rigid transformation parameters which have an additional scale parameter which produced better results such as 0.01855mm on the alignment of relocated reference volume and 15.5915mm on the alignment of deformed and relocated reference volume of Hausdorff distance between both models (reference volume and transformed volumetric template).

**Keywords**—Digitally Reconstructed Radiograph (DRR); intensity registration; rigid transformation; 3D models; Total Hip Arthroplasty (THA); preoperative planning; Computed Tomography (CT)

## I. INTRODUCTION

The Total Hip Arthroplasty (THA) surgery is applied mostly to older people, which has some pains in joints, commonly in the hip, and it is the part most affected of the body. THA surgery requires accurate data from the patient to set correct prosthesis location and size. Another type of data is a preoperative digital x-ray image (2D Planning), which uses specialized software, for instance, “Orthocad”, which is a tool where established the correct size of a prosthesis. There exists a planning tool using CTs, where the measures are more precise, and it creates a 3D prosthesis on the volume generated from the CTs. As mentioned Sariali in [1], there exists some assessment where he made a comparing between planning in 2D using radiographs and

3D templates using CT scans. It fixes position parameters such as femoral offset, the center of head femur, distances from great trochanter to the center of the femoral head, and distance from lesser trochanter to the neck osteotomy. These parameters let us set the size and the position of the prosthesis. So Sariali shows that 3D planning is better in 96% than 2D plan, which only gets 16% of accuracy to establish stem and cup. In 2d planning, the hip anatomy is not accurate because of the torsion, and bone hardness does not exist on the 2d plan, and Surgeons need it to avoid lower limb discrepancy.

There exist different works in order to use conventional x-ray images in order to establish precisely data of prosthesis, building a volume with a volumetric template and the x-ray images with intensity registration technique [2], [3], [4], [5], [6], [7], [8], [9], [10], it results in less cost and not exposure radiation to the patient [11]. Exist another type of reconstruction using characteristics or landmarks [12], [13], [14], [15], [16], [17] this work required previous manual segmentation of the evaluated structured. In the case of using the intensity information, it doesn't require a previous segmentation, the image is entered in original form so it is an advantage in order to establish a correct volume.

It is crucial to know what types of views of the patient's x-ray images are suitable for this job, so specialists required some kinds of x-ray in a total hip arthroplasty, depending on the symptoms required in different views in order to recognize the illness condition. There exist multiples views [18]; for instance, important views are Anteroposterior, Outlet, Inlet, Alar, and Obturator view. In this case, it needs to know the correct angle of the selected view in order to generate the virtual x-ray; these were analyzed on [19]. There exist another impressive view, which is a lateral view using in THA preoperative planning in [20]. In some cases, the Judet view or antero-posterior view (AP view) the specialist doesn't view fractured of the acetabulum. In the posterior column of the pelvis, there exist some displacements or discontinuity. It only requires setting up the patient in the lateral decubitus position with the affected side toward the x-ray cassette. The x-ray centers in the posterior of the hip. In the case of surgeries components, the patient can be rotated 10 to 20 degrees away from the right lateral. So the lateral view (LT view) is easy to obtain, and we use this configuration to generate the virtual

x-ray. Nevertheless, it could have sensitive changes on tilt and rotation parameters, as mentioned Tannast in in [21].

It tested two implementations, which make 2d-3d registration. This work was from ITK journal publications<sup>1</sup>.

The first paper is called: ITK-based implementation of two Projection 2D-3D Registration Method with an application in patient setup for external Beam Radiotherapy by Jian Wu [22]. It talks about radiotherapy using registration process between preoperative patient volume and the current volume in radiation process with the aim to adapts rays to the tumor form applying transformations made with a C-ARM tomography.

The second paper is by Bertelsen and Borro [23]. It uses a medical linear accelerator, which is a type of machine that produces CT (anthropomorphic cranial phantom). It appends to the base of ITK registration some approaches. The first approach is the metric, which is the Normalized Gradient Correlation from two images to one image. Nevertheless, it only uses the value metric, and it does not use derivative value. Another important point in the generation of digitally reconstructed radiographs (DRR), this method includes image orientation, which is essential to generate virtual images in any direction; this new interpolator was called Patched Ray Cast. It used a Euler transformation, but in this case, we used a similarity transformation to include scale parameter.

The paper is organized as follows: Section 2 presents the methods with subsections of the preparing the virtual x-rays and the volumetric template and the general properties that is used for registration process. In Section 3, we present the two experiments with relocated volume and deformed relocated volume. The validating process on 3d model, transformation parameters, and cost function is showed in Section 4. Finally in Section 5 gives the conclusions.

## II. METHODS AND PROCEDURES

The whole process is following the new scheme of work based on Bertelsen A., Borro D work which is an ITK-Based Framework for 2D-3D Registration with Multiple Fixed Images.

The first process consists of deformed and alignment with a volumetric template, which obtains from the Visible Human Project from the Iowa University (moving image) [24] with the x-ray image (fixed image). The model should transform in a 2d image, this image interpolates using ray casting interpolation to produce a virtual x-ray, in other work such as “simulating x-ray images from deformable shape and intensity models on the GPU” by Moritz Ehlike [25], using statistically shape and intensity model (SSIM) [26]. They applied ray-casting by using some intensity volume which contains intensity information in the tetrahedral faces from the volume. In this work, it uses the interpolator with the method “Siddon Jacobs Ray Tracing” by Filip Jacobs in [27].

The second process registers the moving image to the fixed image through the intensity registration process. It gets neighbor intensities from the current pixel evaluated to establish the closest intensity value to achieve the intensity of the fixed image, there exist some works which use intensity registration

such as [28] where use registration on fluoroscopic images in order to set the hip joint kinematics or using non rigid registration which use b-splines method [29].

To insert the criteria of an image comparing it requires a metric value, in this case, it uses a normalized gradient correlation (NGC) using a neighborhood operator called Sobel operator applying in every axis or dimensions of the images, such as mentioned Penney on this work [30] the NGC metric filters soft tissues and focuses only on bones data. Nevertheless, this metric does not avoid the high intensities such as surgery instruments on x-ray images.

It is encapsulated by the optimization process, which with a cost function called Fletcher-Reeves-Polak-Ribiere optimizer [31], it gets the correct transformation parameters using the derivatives or changes between every pixel of the images. The optimization has the number of iterations, the step length, and the step tolerance, which allows establishing boundary conditions to stop the optimization and return the actual transformation parameters.

This process is iterative in the work of getting the best transformation parameters, the type of transformation used is a Similarity Transformation [32] where use a versor form on rotation parameters [33] (which avoid a gimbal lock) and isometric scale parameter (which is equal in all axis), and these parameters transformed the volumetric template, interpolating the 3d image with these parameters. It generated a new DRR in every resolution and made the same process until achieving the minimum threshold error or until complete the number of iterations.

### A. Establishing the Virtual X-ray Fixed Images

To establish the DRR, we use an interpolator with a ray casting method [27]. It consists of traverse a ray into the volume, then integrates the different intensity values which intersect the current ray, it is saved on the 2d virtual image and use a correct threshold to work only with the specific data (bones) of the volume data.

Parameters were the next: Pixel Spacing, Dimension output, Distance of ray source, Translation and Rotation Parameter of the camera, 2d projection normal position, the center rotation relative to the center of volume and threshold.

In order to get a virtual x-ray with specific data, the volumetric template needs preparing its label map obtained by the semiautomatic segmentation process by the “3D slicer” tool, it used arithmetic filter for multiplying label map with the whole original volume. So it gets a segmented volume with only the region of interest, in this case, the pelvic without sacrum but with the intensity information. The result could be observed in Fig. 1.

### B. Selected Properties in Registration Process

For virtual x-ray image generation it used an isocenter which sometimes is equals to the center of the volume. So according to the size of the “Test Volume”, every axis were divide in two parts so the result is the center of the volume. As size of volume is (333, 181, 245)vx so the center is configured in (166, 90, -122)vx but there exist some displacement in order to project the image and center is modified according to the

<sup>1</sup><http://insight-journal.org>

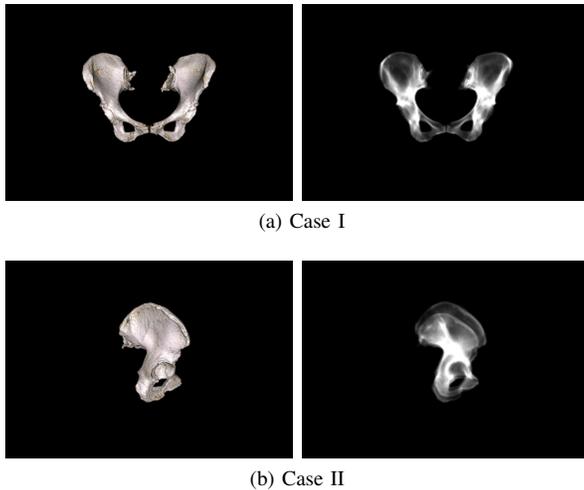


Fig. 1. In Fig. 1a Simulated X-ray from Volumetric Template. It has a size of (333,245) mm, resolution of (1,1) mm, Focal Point (0,-1000,0) mm, Distance from Volumetric Template to Simulated X-ray (-124) mm and Orientation (90, 0, 0) degrees. In Fig. 1b. Simulated X-ray from Volumetric Template. It has a size of (181,245) mm, Resolution of (1,1)mm, Focal Point (-1000,0,0) mm, Distance from Volumetric Template to Simulated X-ray (200) mm and Orientation (90, 90, 0) degrees

angle of projection. In case of AP view It used 'Y' axis so in keeps its original value and the isocenter for AP view is (166, -124, -122)vx. In the other hand, in case of LT view, 'X' axis was used given an isocenter (-200, 90, -122)vx. So It compared with the original projection on 2D plane with the 3D render in order to view the same image before applying the registration process. It has different focal distance because in every projection because the input volume has different number of voxels in every dimension. 200mm for AP view and -124mm for LT view.

In the AP view the first value is the corner (166mm) on x axis, the second (-124mm) is the corner on y axis (is the depth with respect the volume) and last (-122mm) is the corner on z axis. In the LT view (-200) is the depth and (90, -122) are the y and z axis.

In generation of virtual images there exist different parameters in order such as size, resolution(spacing), center of the image (2dcx), distance from source to the center of the volume (SCD), center of the volume(isocenter) and the projection angle (RP). There exist other important parameter which is the distance from volume to x-ray plane. the relation according to default example was 200mm from volume to the image and 1000mm from volume to the source. So the virtual fixed images resulted were translated 200mm from the volume for both views AP and LT initially.

The AP view was separated from the template volume 200 mm of the internal center. Its origin has LIA orientation so it is located in left inferior anterior corner to the volume. The origin is given by (166 -124 -122)mm. The pixels quantity is (333, 245)px with an spacing (1,1)mm. The result image will be only one slice in sagittal view. The focal point where the observer or the ray begins will be in order to project the virtual image is located in (0, -1000, 0)

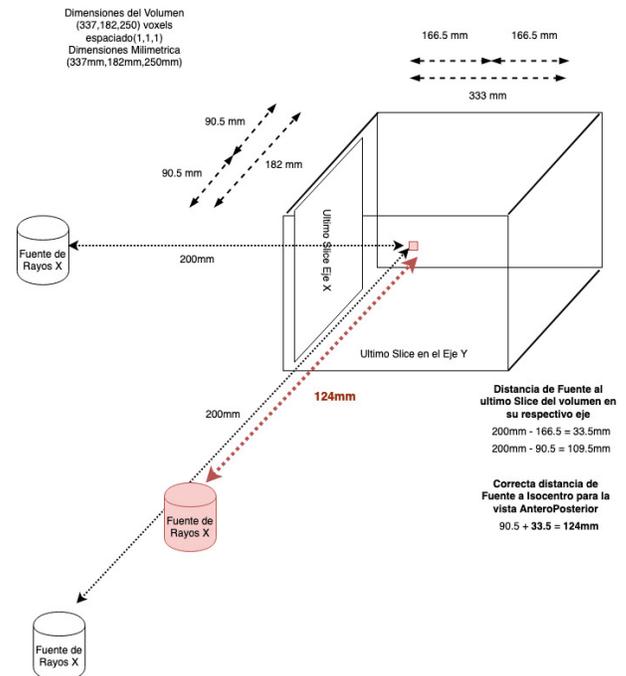
The LT view is separated 124mm from the center left to

the template volume. It considering the angle of projection 90 positive degrees. It also has and origin with PIL orientation. This center is located in the posterior inferior left corner. So the origin was located in (-200 90 -122)mm. The number of pixels on the image is (181,245)px with spacing of (1 1)mm. The result was only one slice in coronal view. The focal point where the observer or the ray begins will be in order to project the virtual image is located in (-1000, 0, 0)

Many times, It was established a general setting for x-ray image generation for 2 views, but it is necessary a particular setting for every view, it is due to different projection, it sometimes will be close or so far to the volume because the volume has different size in every axis which gives this unbalancing in the metric.

In order to solve this problem, it varied the distance from the source to the center of volume, it could be observed better in Fig. 2, this correction is applied in the AP view because this view has a little contribution to the metric, the variation is limited when the distance from the source to last slice of volume will be the same. In this case the difference is 109mm and 33.5mm for every view respectively, it uses the minimum difference 33.5mm (LT view) and replies in the AP view. If it adds half distance in the axis 'y' to minimum difference, it will give 124mm from source to the isocenter.

The first criteria is doing this balancing according to the cost function. Furthermore the histogram which helps to view the intensity values from different views.



Fix distance from source to isocenter

Fig. 2. Fix AP View with the Correct Focal Distance

### III. EXPERIMENTS

It generates 20 tests with synthetic data, first set of registrations use a reference relocated template like a moving

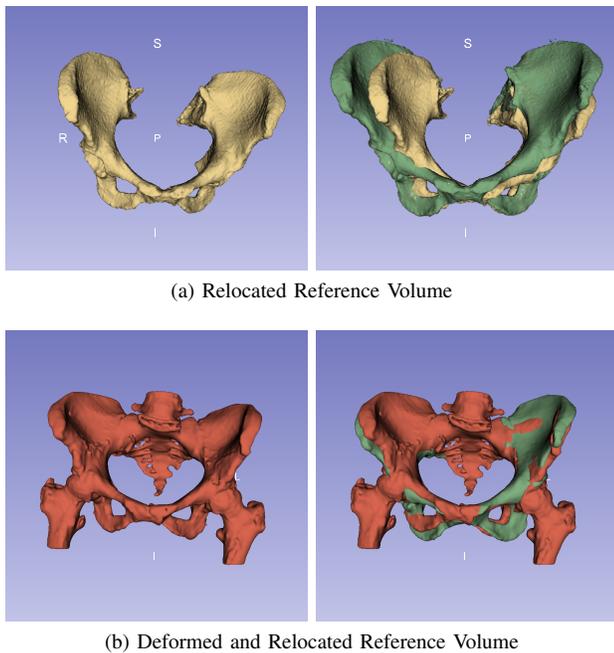


Fig. 3. In Fig. 3a is the Relocated Reference Volume is a Source of Simulated X-ray Images with Rotation (-6.12, -7.92, -5.44) degrees, Translation (-7.68, 7.00, 8.57) mm, Scale (1.18) units. In Fig. 3b. is the Deformed and Relocated Volume as Source of Simulated X-ray Images with Rotation (0.0006, -0.0382, -0.0155) degrees, Translation (2.7677, 0.4437, -2.2839) mm and Scale (1.0352) units

image, the second set of registrations use a reference volume which is deformed and relocated like a moving image. In the case of the fixed image is the volumetric template used for both sets of registrations.

In these experiments, it tests sensibility checking the difference of seven transformation parameters between the reference volume and the volumetric template. It analyzes the error difference of every transformation parameter. Finally, it checks the final volumes comparing its overlapping and getting the Hausdorff distance between them.

#### A. Registration Test with Reference Relocated Volume

This test gives a close view than how well the registration process is developing because it used the same information for the fixed and moving image. It only transformed the volumetric template to obtain a reference relocated volume. It chooses the worst registration (18th registration) of the set of registrations (20 in total) to analyze the bounds of the registration process.

The reference relocated volume could be observed in Fig. 3. This volume compared with the volumetric template is rotated  $-6.12^\circ$  to the left side,  $-7.92^\circ$  rotated in its z-axis,  $-5.44^\circ$  rotated forward, In translation it was translated  $-7.68\text{mm}$  away from the origin,  $7\text{mm}$  up to the origin,  $8.57\text{mm}$  translated to the left, the scale was reduced  $1.18\text{mm}$ .

The difference or subtraction between projected volumetric template in gray color and projected reference volume in white

color could be observed in Fig. 4, in the middle of the first row on the left.

After the registration process, it observed in Fig. 4, in the middle of the second row on the left, the difference between the projection of the volumetric template and relocated reference volume, as it can see, the subtraction shows perfectly the overlapping between both volumes. Further, it obtained  $0.0230085$  of Hausdorff distance. When it considered the 20 tests, an average value Hausdorff distance of  $0.01855\text{mm}$  was obtained.

It also shows the respective error values in Table I. It shows the most difference in translation Y, but with close values, the target was  $7.01\text{mm}$ . Meanwhile, the register value was  $6.89\text{mm}$ . Finally, it could observe the final overlapping between the transformed volumetric template and the relocated reference volume in Fig. 4, in the middle of the last row on the left.

#### B. Registration Test with Reference Deformed and Relocated Volume

It has chosen a test that has a vast amount of error to describe what conditions produce this and how to resolve the problem. This test was the eleventh test between 20 tests. All of these tests are crucial because they have extra information (femur and sacrum bone) on the fixed images.

The input volume or "Reference Volume" was deformed in an interval  $[-11,11]\text{mm}$  with a random BSpline transform, and it was relocated for the volumetric template, as observed in Fig. 3 in the second row. The new position is in rotation  $[0.0006^\circ, -0.04^\circ, 0.02^\circ]$  for x,y,z-axis respectively. It was translated with  $[2.77, 0.44, -2.28]\text{mm}$ , and it is scaled with  $1.03$  units of the original size volumetric template.

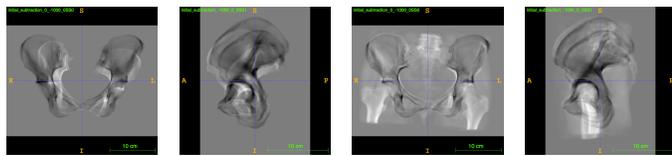
The difference between fixed images and the moving image could be observed in Fig. 4 in the middle of the first row on the right, the darkest part is for the volumetric template, and the brightest part is for reference volume.

When the registration process completes, it can observe those final subtractions between reference volume and volumetric template in Fig. 4 in the middle of the second row on the right, which has bright areas that are not covered for the volumetric template. It is due to the volumetric template is entirely different from the reference volume by initial deformation. If the reference volume were only deformed, it would show the initial difference, which does not allow the complete superposition.

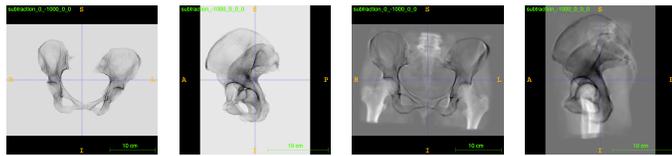
The metric with the same information moving and fixed images could achieve as maximum value  $2.0$ , but when the metric works with different information images with fewer parts or structures (sacrum, left, and right femur), it reduces the metric, so it achieves as maximum value  $1.0$ . This value was observed in the experiments. So if the metric is closed to  $1.0$  is a useful metric.

It could be observed that there are more errors in rotation 'Y' and 'Z' and in translation 'Y' and translation 'Z' and

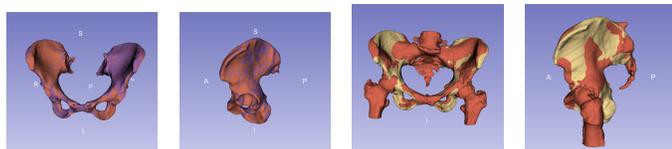
slightly in rest parameters. It could be observed error values in Table I. Finally, a comparative between volumes is made by using superposition, which is observed in Fig. 4, in the middle of the last row on the right. it obtained 17.39mm of Hausdorff distance on this particular test but when it considered the 20 tests, an average value Hausdorff distance were 15.5915mm was obtained.



(a) Subtraction before Registration using Relocated Volume and Deformed Volume



(b) Subtraction after Registration using Relocated Volume and Deformed Volume



(c) Overlapping using Relocated Volume and Deformed Volume

Fig. 4. Comparing Reference Volume and Volumetric Template applying Image Subtraction and Overlapping

#### IV. VALIDATING PROCESS

##### A. 3D Models Validation

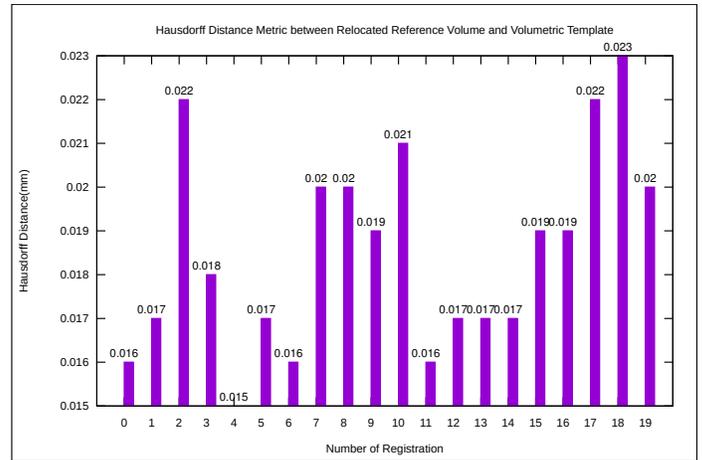
To compare the difference between the two volumes, it used the Hausdorff distance. It helps to set the appropriate metric for registration accuracy.

It could be observed in Fig. 5, the Hausdorff distance of every registration using the relocated reference volume and volumetric template, as it can see, the best registration process was the 4th registration with 0.015mm of accuracy and the worst registration was the 18th registration with 0.023mm of accuracy. The average Hausdorff distance in this set was 0.01855mm

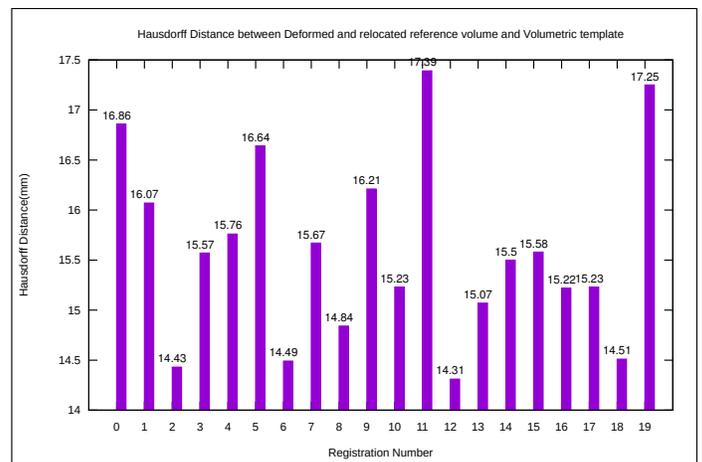
On the other hand, in Fig. 5, the second set of registration, using the relocated and deformed reference volume and volumetric template, the best registration was 12th registration with 14.31mm of accuracy, and the worst registration was 11th registration with 17.39mm of accuracy. The Hausdorff distance average in this set was 15.5915mm.

##### B. Transform Parameters Validation

In Fig. 6, it shows the difference of transformation values, according to the deformed and relocated reference volume. The least quantity of error distribution was for rotation and scale parameter. In the case of rotations, for X rotation, its



(a) Case I



(b) Case II

Fig. 5. Hausdorff Distance of the First Set of 20 Registrations. 5a Using Relocated Volume and Template 5b Using Deformed Volume and Template

median was 0.30° degrees and a maximum value of 0.69° degrees, it has a balanced distribution. In Y rotation, its median was 0.27° degrees and a maximum value of 0.68° degrees, its distribution is very close to X rotation, and its distribution is more frequent below the median. In the case of Z rotation, it gives more error quantity. The maximum error achieves 75% of all data was 1.75° degrees, and its median was 1.49° degrees. It has an outlier of 2.59° degrees.

Similarly, the translation had more error, it was Z translation with a maximum error of 1.57mm, and its median was 1.18mm, it has a compact distribution because from its median to 75% of data. In the case of Y translation, its median was 0.62mm and a maximum error of 1.64mm, which is another outlier, but with 0.77mm because it is more probable distribution. In the case of X translation, it has a median of 0.40mm and a maximum error of 1.12mm. It has a distribution more probable below its median, and it achieves 25% of data the error of 0.22mm.

The last option, in scale parameter, represents the less

TABLE I. TRANSFORMATION VALUES IN 18TH AND 11TH REGISTRATION

Parameters	Transformation Values of Registered Volume and Relocated Volume			Transformation Values of Registered Volume and Deformed Volume		
	Reference Vol.	Registered Vol.	Error	Reference Vol.	Registered Vol.	Error
X Rotation	-6.119503	-6.113421	0.006083	-4.373908	-4.133677	0.240231
Y Rotation	-7.920640	-7.892612	0.028028	-1.779127	-2.210442	0.431315
Z Rotation	-5.440008	-5.441351	0.001343	0.137158	-1.167670	1.304828
X Translation	-7.681812	-7.480782	0.201030	2.767714	2.984251	0.216537
Y Translation	7.007667	6.886352	0.121315	0.443732	1.062340	0.618608
Z Translation	8.570599	8.526593	0.044005	-2.283939	-0.711641	1.572298
General Scale	1.178369	1.178348	0.000021	1.035213	1.028790	0.006423

quantity of error, it has a maximum error of 0.013 units. Its median was 0.006 units, and it has a balanced distribution.

### C. Cost Function Validating

After registration process, it is possible to analyze the value of metric varying the value of transformation parameter, it is alter with interval of transformation values, it also applied a step size and view the contribution for every fixed image or combine views (AP and LT views). The contribution varying one transformation parameter on 11th Registration using Relocated Volume could be observed in the next Fig. 7a, 7b.

In every diagram it established the metric value of registered volume and metric value of reference volume. It can be observed that in the translation 'y' and 'z' and rotation 'z' there is a wide error because metric value of reference volume is far from metric value of registered volume. In the rest of parameters are good because both metric values are similar.

As mentioned before, this difference is due to exchange of roles between scale and translation, when the registration updating the seven parameters it updates equally these parameters according to the metric, but sometimes there exist some issues with scale and translation because there exists a competition between them. If some of the translation parameters achieve faster its optimal value after scale parameter, it produces a big error because changes in translation on the 'z' or 'y' axis are right changes because it decreases the size of the projected image and the scale takes other wrong value. If change optimizer parameters or increase the number of iterations in order to achieve a right value of the scale, it will be in vain, the scale parameter never achieves its right value because of the translation parameters already have its right value before the scale.

So the correct form to use the similarity transformation is that scale has the right value more quickly than the translation parameter, It is achieved by making the scale have more variation of change in the initial scale parameters and so the scale achieves its value more quickly than rest of parameters.

## V. CONCLUSIONS

The present work shows that rigid registration is basic in order to initial alignment because it not always our template will be in the correct magnification or center in the center according to the center of the radiographic. This article helps to give a guide in what parameters are required on fixed images when it uses synthetic data. It set the relation between template volume (moving image) and the x-ray images (fixed images) by establishing a correct focal point. Also, the dataset

and the new framework can be evaluated from here:

<https://github.com/chiconasa3000/2D-3D-RegistrationTool>.

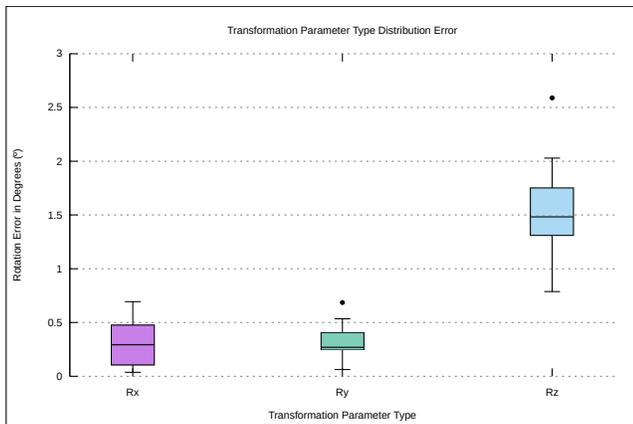
The rigid process was used with CPU on Workstation Intel Xeon E5-2643 v3 (30 MB Cache, 3.40 GHz), and it lasted 579.64 seconds or 9.66 minutes for 11th registration.

## ACKNOWLEDGMENT

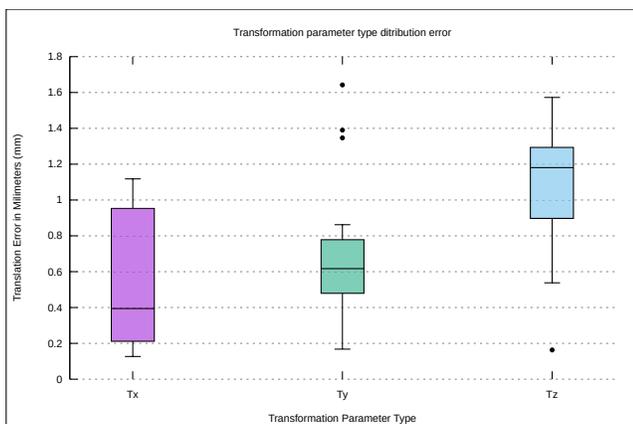
This research was supported from Universidad Nacional de San Agustín de Arequipa Contract : IBA-0012-2016, part of the project "Construction of a 3D model to assist in hip joint replacement surgery from x-ray images, for the treatment of osteoarthritis and osteoporosis, at the UNSA Hospital of Arequipa". Thanks to the CiTeSoft Contract: EC-0003-2017-UNSA for the equipment and the resources bring to the project.

## REFERENCES

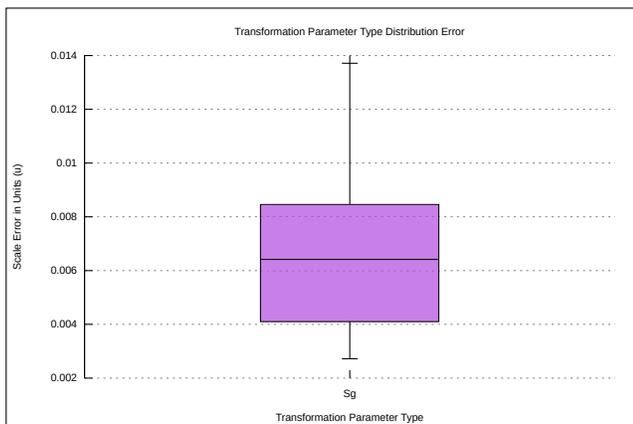
- [1] E. Sariali, R. Mauprivez, F. Khiami, H. Pascal-Moussellard, and Y. Catonné, "Accuracy of the preoperative planning for cementless total hip arthroplasty. a randomised comparison between three-dimensional computerised planning and conventional templating," *Orthopaedics & Traumatology: Surgery & Research*, vol. 98, no. 2, pp. 151–158, 2012.
- [2] W. Yu, C. Chu, M. Tannast, and G. Zheng, "Fully automatic reconstruction of personalized 3d volumes of the proximal femur from 2d x-ray images," *International journal of computer assisted radiology and surgery*, vol. 11, no. 9, pp. 1673–1685, 2016.
- [3] N. Baka, B. L. Kaptein, M. de Bruijne, T. van Walsum, J. Giphart, W. J. Niessen, and B. P. Lelieveldt, "2d–3d shape reconstruction of the distal femur from stereo x-ray imaging using statistical shape models," *Medical image analysis*, vol. 15, no. 6, pp. 840–850, 2011.
- [4] G. Zheng, "Statistically deformable 2d/3d registration for estimating post-operative cup orientation from a single standard ap x-ray radiograph," *Annals of biomedical engineering*, vol. 38, no. 9, pp. 2910–2927, 2010.
- [5] —, "Statistical shape model-based reconstruction of a scaled, patient-specific surface model of the pelvis from a single standard ap x-ray radiograph," *Medical physics*, vol. 37, no. 4, pp. 1424–1439, 2010.
- [6] W. Yu and G. Zheng, "2d-3d regularized deformable b-spline registration: Application to the proximal femur," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*. IEEE, 2015, pp. 829–832.
- [7] W. Yu, M. Tannast, and G. Zheng, "Non-rigid free-form 2d–3d registration using a b-spline-based statistical deformation model," *Pattern recognition*, vol. 63, pp. 689–699, 2017.
- [8] C. L. Chan, C. Anitescu, Y. Zhang, and T. Rabczuk, "Two and three dimensional image registration based on b-spline composition and level sets," *Communications in Computational Physics*, vol. 21, no. 2, pp. 600–622, 2017.
- [9] M. Nejati and H. Pourghassem, "Multiresolution image registration in digital x-ray angiography with intensity variation modeling," *Journal of medical systems*, vol. 38, no. 2, p. 10, 2014.



(a) Difference on Rotation Parameters between Template and Reference Volume



(b) Difference on Translation Parameters between Template and Reference Volume



(c) Difference on Scale Parameter between Template and Reference Volume

Fig. 6. Transformation Parameters Error Distribution using Deformed Volume

[10] I. Van der Bom, S. Klein, M. Staring, R. Homan, L. W. Bartels, and J. P. Pluim, "Evaluation of optimization methods for intensity-based 2d-3d registration in x-ray guided interventions," in *Medical Imaging 2011: Image Processing*, vol. 7962. International Society for Optics and Photonics, 2011, p. 796223.

[11] H. Almohiy, "Paediatric computed tomography radiation dose: a review of the global dilemma," *World journal of radiology*, vol. 6, no. 1, p. 1, 2014.

[12] N. Baka, C. Metz, C. J. Schultz, R.-J. van Geuns, W. J. Niessen, and T. van Walsum, "Oriented gaussian mixture models for nonrigid 2d/3d coronary artery registration," *IEEE transactions on medical imaging*, vol. 33, no. 5, pp. 1023–1034, 2014.

[13] K. Vikas and R. Bhamudi, "Application of laplacian surface deformation and self-organizing maps to calculate shape correspondence for statistical shape modeling," in *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*. IEEE, 2014, pp. 369–372.

[14] V. Karade and B. Ravi, "3d femur model reconstruction from biplane x-ray images: a novel method based on laplacian surface deformation," *International journal of computer assisted radiology and surgery*, vol. 10, no. 4, pp. 473–485, 2015.

[15] B. Jian and B. C. Vemuri, "Robust point set registration using gaussian mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1633–1645, 2011.

[16] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. Kuhn, "Landmark-based elastic registration using approximating thin-plate splines," *IEEE Transactions on medical imaging*, vol. 20, no. 6, pp. 526–534, 2001.

[17] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[18] L. J. Fernández Palomo, "Imagenología en las fracturas de la pelvis," *Ortho-tips*, vol. 4, no. 4, pp. 242–249, 2008.

[19] G. C. Polesello, T. S. Nakao, M. C. d. Queiroz, D. Daniachi, W. Ricioli Junior, R. P. Guimarães, E. K. Honda *et al.*, "Proposal for standardization of radiographic studies on the hip and pelvis," *Revista brasileira de ortopedia*, vol. 46, no. 6, pp. 634–642, 2011.

[20] N. J. Giori and A. O. Sidky, "Lateral and high-angle oblique radiographs of the pelvis aid in diagnosing pelvic discontinuity after total hip arthroplasty," *The Journal of arthroplasty*, vol. 26, no. 1, pp. 110–112, 2011.

[21] M. Tannast, G. Zheng, C. Anderegg, K. Burckhardt, F. Langlotz, R. Ganz, and K.-A. Siebenrock, "Tilt and rotation correction of acetabular version on pelvic radiographs," *Clinical orthopaedics and related research*, vol. 438, pp. 182–190, 2005.

[22] J. Wu, "Itk-based implementation of two-projection 2d/3d registration method with an application in patient setup for external beam radiotherapy," 12 2010.

[23] A. Bertelsen and D. Borro, "An itk-based framework for 2d-3d registration with multiple fixed images," 05 2011.

[24] B. Lorensen, "Ie9 for windows phone 7: Adobe flash, demos and development." [Online]. Available: <https://mri.radiology.uiowa.edu/>

[25] M. Ehlke, H. Ramm, H. Lamecker, H.-C. Hege, and S. Zachow, "Fast generation of virtual x-ray images from deformable tetrahedral meshes," *IEEE Transactions on Visualization and Computer Graphics*, 2013.

[26] H. Lamecker, "Variational and statistical shape modeling for 3d geometry reconstruction," Ph.D. dissertation, Freie Universität Berlin, 2008.

[27] F. Jacobs, E. Sundermann, B. De Sutter, M. Christiaens, and I. Lemahieu, "A fast algorithm to calculate the exact radiological path through a pixel or voxel space," *Journal of computing and information technology*, vol. 6, no. 1, pp. 89–94, 1998.

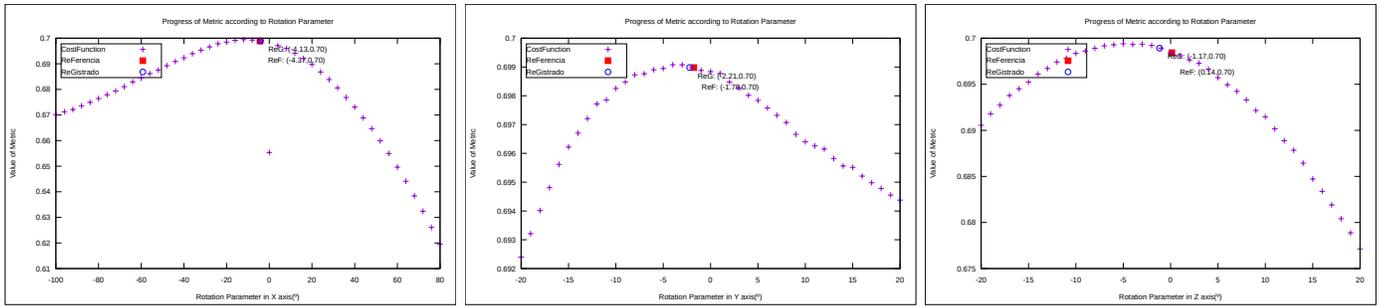
[28] F. D'Isidoro, P. Eschle, T. Zumbun, C. Sommer, S. Scheidegger, and S. J. Ferguson, "Determining 3d kinematics of the hip using video fluoroscopy: guidelines for balancing radiation dose and registration accuracy," *The Journal of arthroplasty*, vol. 32, no. 10, pp. 3213–3218, 2017.

[29] L. A. Schwarz, "Non-rigid registration using free-form deformations," *Technische Universität München*, 2007.

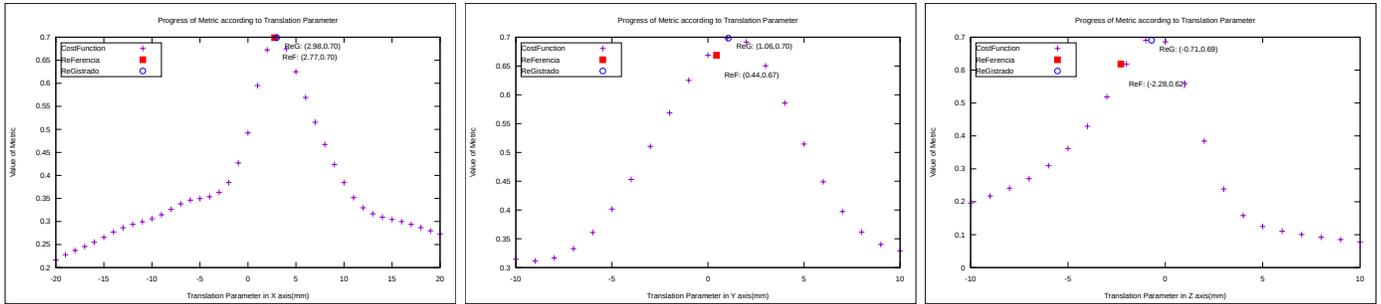
[30] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. Hill *et al.*, "A comparison of similarity measures for use in 2-d-3-d medical image registration," *IEEE transactions on medical imaging*, vol. 17, no. 4, pp. 586–595, 1998.

[31] E. Polak and G. Ribiere, "Note sur la convergence de méthodes de directions conjuguées," *Revue française d'informatique et de recherche opérationnelle. Série rouge*, vol. 3, no. 16, pp. 35–43, 1969.

[32] H. J. Johnson, M. McCormick, L. Ibáñez, and T. I. S. Consortium, *The ITK Software Guide*, 3rd ed., Kitware, Inc., 2013, *In press*.



(a) Value of Cost Function between Reference Volume and Registered Volume Changing the Rotation Parameters



(b) Value of Cost Function between Reference Volume and Registered Volume Changing the Translation Parameters

Fig. 7. In Fig. 7a the Rotation z has a Considerable difference it is because of the Initial difference between Deformed and Relocated Volume it always appears even if the Registration was Developing well. In Fig. 7b. Translation 'Y' and 'Z' also shows a Big difference due to Exchange of Roles between Scale and Translation.

[33] W. R. Hamilton, *Elements of quaternions*. Longmans, Green, & Company, 1866.

# Evaluating Contact Detection, Size Recognition and Grasping State of an Object using Soft Elastomer Gripper

Kazi Abaul Jamil

Department of Automation

School of Electronics, Information and Electrical Engineering

Shanghai Jiao Tong University

**Abstract**—Object handling process of the sensitive or fragile object is critical to preserve its quality. In this domain, soft robotics has gained a lot of attention. However, limitations of detecting contact and grasping behavior is still a challenging task due to the non-linear behavior of soft gripper. Moreover, to regulate grasping behavior, exact real-time contact feedback is a crucial task. To improve the contact detection accuracy a gradient-based algorithm is proposed with the feedback from a simple resistive flex sensor and a pressure sensor. For that purpose firstly, the resistive flex sensor is embedded into the gripper to get the real-time gripper's finger position. Secondly, solenoid valves and pressure sensors are used to control the pneumatic pressure from the pump, and finally, a closed-loop control system is developed for controlling the grasping process. The proposed contact detection algorithm can provide contact feedback with an accuracy of  $\pm 3mm$ , which is implemented to perform the size recognition of the sphere-shaped objects. A real-time experimental setup has been developed which can successfully perform a pick and place of fruits and vegetables. The key benefits of the proposed algorithm are less complexity and better accuracy.

**Keywords**—Grasping process; contact feedback; size recognition

## I. INTRODUCTION

In the recent few years, soft robotics has gained a lot of consideration due to its unrivaled agility, compliance and safe interactions with objects [1] [2]. Robotic manipulation without degrading the quality of sensitive, fragile or deformable objects quality is a challenging task. The soft gripper has wide applications in automated industries such as safe human-robot interaction [3], minimally invasive surgery [4], harvesting and food handling [5], [6]. Various soft actuators are introduced using pneumatic, cables, fluid, motor or magnetic fields. Embedding additional sensors to this type of actuator is challenging and makes the system bulky and incompact. Accurate observation of contact detection and grasping behavior by image processing is perplexing task [7], [8]. Moreover, grasp target information i.e size or position remains unidentified due to opaqueness or obstacles between the gripper. Such grippers are difficult to regulate the excessive contact force for sensitive objects [6]. The contact detection is very important for sensitivity analysis during grasping. The interface of contact detection with an object may perform different grasping related operations i.e. object geometry recognition, size recognition and pose estimation. A deep convolutional network algorithm is proposed where a soft robot grasps the

object using camera [9]. In [6], [7], [10] the concept of contact detection is discussed, whereas real implementation is still challenging. A clustering method is implemented to identify a set of irregular shape objects based on the internal state of the finger [11]. This method is limited to contact detection or size recognition of the grasped object. In order to experience the real-time proprioceptive grasping, flexible and stretchable sensors embedded with soft bodies are fabricated and designed [12] [13]. Implemented pressure and curvature sensing of soft actuators in a closed loop system represents the concept of soft haptic perception based on the piezoresistive effect [14].

The characteristics of linearity, robustness, compactness, low power consumption, the flexibility of commercially available resistive flex sensors suits to integrated with the soft actuators [15]. Whenever soft actuator bends the resistivity of flex sensor changes. This technique is utilized to estimate the internal state of the fingers of the gripper and perform the haptic identification of grasped objects [16]. The data-driven approach was implemented to estimate the bending angle of the soft actuator based on flex sensor value and onboard pressure sensor value [17]. However, this method needs to collect a lot of experimental data for calibration of each actuator. Thus, to overcome this problem, in [6] a linear response of object size and flex sensor reading are analysed from real-time raw feedback sensor data for evaluating contact and size of objects. In [18] contact detection and size recognition of round shape objects are performed comparing bend sensor and pressure sensor data. The bending state has been measured by comparing current bend value with free bend value which is difficult to calibrate when source pressure value is varying.

To the best of the author's knowledge, there is a need for an alternative approach that can overcome the limitations of real-time monitoring of the bending state. The prime contributions of this study are listed below:

- A contact detection algorithm has been proposed based on the gradient response of a commercially available embedded resistive flex sensor. The contact detection algorithm is validated using the touch sensor in the experiments.
- A size recognition algorithm has been proposed considering sphere-shape objects based on raw sensory contact feedback.
- The grasping state of the object is analysed using

raw sensory feedback ensuring whether the object is grasped or released, which helps to reduce the grasp uncertainty or delay in soft robust grasping.

- Proposed algorithm is capable to monitor the real-time bending state of the actuator during grasping of the target object.
- Key benefits of the proposed setup and algorithm are less complexity, better accuracy, and lower cost.

The rest of the paper is organized as follows: Section II describes the experimental design including the fabrication of soft gripper, hardware setup, and noise filter design. The methods of the proposed algorithm are discussed in Section III. Section IV discusses the experiments performed using the proposed method. The results of the performed experiment are discussed in Section V. In Section VI, an automated fruit sorting environment is implemented to perform a real-time sorting process. Section VII concludes how the proposed methods are used to preserve the quality of sensitive or fragile objects.

## II. EXPERIMENTAL DESIGN

### A. Design of Soft Robotic Gripper

The design of the gripper is considered from open source design (<https://softroboticstoolkit.com>) with customized soft gripper, a size of 150\*20\*10 mm is prepared by using smooth on Smooth-On Ecoflex-30 material as shown in Fig. 1. Additionally, a flex sensor of length 11.5cm is embedded into the base of the soft gripper as shown in Fig. 2 to perform the contact detection and size recognition. During validation of the proposed contact detection algorithm, a touch sensor is embedded at the fingertip of the soft gripper, whose measurement is considered as a ground truth.

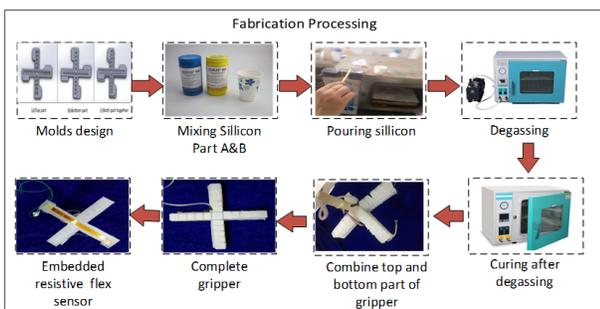


Fig. 1. Fabrication Process

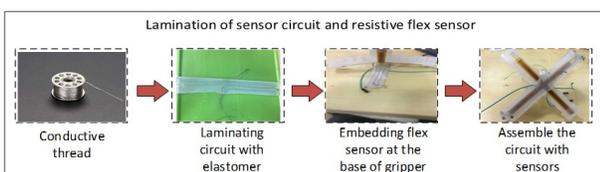


Fig. 2. Embedding Resistive Flex Sensor

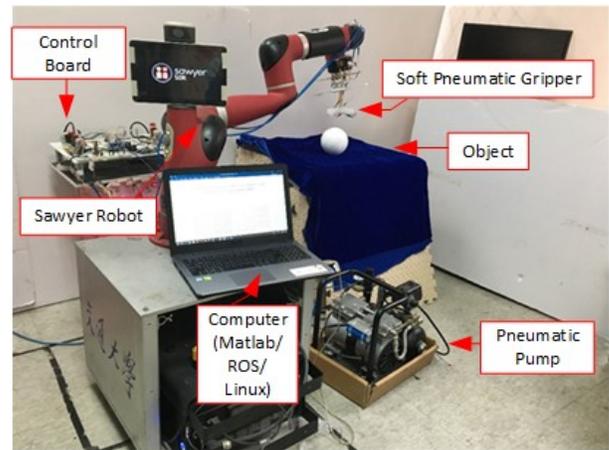


Fig. 3. Diagram of the Experiment Setup

### B. Hardware Setup

An experimental setup has been constructed systematically to perform the grasping test as shown in Fig. 3.

- Soft gripper: Gripper with embedded flex sensor.
- Sawyer robot: To perform pick and place of an object.
- 3D printed support: A 3D printed support is mounted at the end-effector of sawyer robot to facilitate the pneumatic soft gripper.
- Control board: A closed loop control is implemented to obtain the required pressure for grasping operation. It consists of a microcontroller board (Arduino Uno), solenoid valves (Delixi solenoid valve, 4v410 DC 24V), MOSFET switch module (SZH-AT021, SZH), and pressure sensor (100PGAA5, Honeywell).
- Pneumatic supply: Pneumatic pump with 0.82 gal cylinder is used for pneumatic supply.
- Computer: Linux OS, Arduino, Matlab software, robotic operation system (ROS).

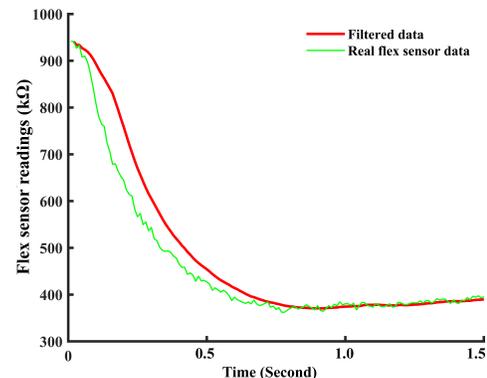


Fig. 4. Implementing Moving Average Filter Data

Implementation of small pneumatic tank as a pneumatic capacitor and a sponge inside the pneumatic channel as a resistors

are introduced to smoothen the real-time data obtained from flex sensor [17].

As the hardware setup was not enough to obtain smooth data, a moving average filter is introduced given by equation (1).

$$y[i] = \frac{1}{m} \sum_{j=0}^{m-1} x[i+j] \quad (1)$$

Where,  $m$  = window size;  $x$  = sensor value array;  $i$  = filtered sample point;  $j$  = real sensor sample point As shown in Fig. 4,  $m=20$  samples are considered as a window size in each iteration.

### III. METHODS

In this section, the flex sensor response at different source pressure are discussed. A contact detection algorithm with flex sensor is proposed and a size recognition algorithm is described to perform the sorting task of sphere-shaped objects.

#### A. Characteristics of Flex Sensor

As the pneumatic pressure increases the flex sensor embedded inside the finger bends and gives a continuous change in resistance. To analyse the characteristics of the flex sensor, two different types of grasping are performed, i) free grasping and ii) object grasping. In free grasping no object is kept between the fingers and the sensor reading with free grasping is shown in Fig. 5 at different inflation pressure from 137 kPa to 193 kPa, respectively. Keeping the source pressure constant, a free

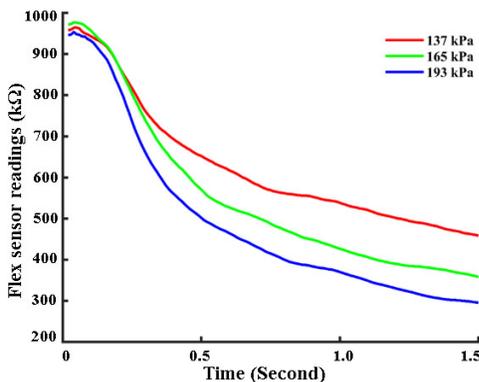


Fig. 5. Flex Sensor Readings during Free Grasping

grasping and an object grasping are performed in Fig.6 which shows the deviation of sensor response of object grasping from free grasping.

#### B. Contact Detection

In this study, it is observed the change in slope of flex sensor reading helps to identify the contact with the target object. For a function of  $N$  variables,  $F(x, y, z, \dots)$ , the gradient is defined in equation (2),

$$\nabla F = \frac{\partial F}{\partial x} \hat{i} + \frac{\partial F}{\partial y} \hat{j} + \frac{\partial F}{\partial z} \hat{k} + \dots + \frac{\partial F}{\partial N} \hat{n} \quad (2)$$

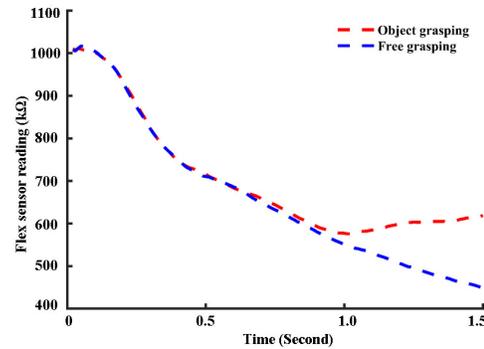


Fig. 6. Comparison of Object Grasping and Free Grasping

Based on gradient theory, contact point detection on one dimensional flex sensor data is used to derive from equation (2), where  $\nabla F$  defined as the change in flex sensor value  $\partial F$  to sample number  $\partial x$  as in equation (3).

$$\nabla F = \frac{\partial F}{\partial x} \hat{i} \quad (3)$$

---

#### Algorithm 1: Contact Detection Algorithm

---

```

Initialization;
Apply pneumatic pressure (open valve);
for Read flex sensor value do
    Calculate the numerical gradient by using
    equation (3);
    if numerical gradient +ive for a batch of sample
    point? then
        Contact Detected;
        Close pneumatic pressure (close valve);
        Record flex sensor reading;
        break;
    else
        Continue applying pressure;
    end
end
* batch of sample point = considering 15 consecutive
flex sensor readings;

```

---

#### C. Size Recognition Algorithm

Through contact detection algorithm it can be analyzed that different size objects results a unique value of flex sensor, which shows the capability of size recognition of object. To demonstrate the size recognition ability of SPG, objects of different sizes and materials are grasped. A human hand during grasping an object, firstly it forms enveloping grasping and analyze the shape and dimension of object and then power grasps to pick the object. When the tip of the SPG touches the object the whole gripper is ready to grasp in envelope type. A suitable pneumatic pressure is provided within working pressure range to confirm grasping stability, and grasping state are maintaining for a period of time, and then deflate the pressure to release the gripper. It is noteworthy that in power grasping process the hand pose is invariable. Mimicking the human behavior and based on previously described contact

detection algorithm size recognition algorithm are introduced in Algorithm [2].

---

**Algorithm 2:** Online Size Recognition Algorithm

---

```

Initialization;
Apply pneumatic pressure (open valves);
for all object to be grasped do
    if Object grasp and contact with the object using
        Algorithm 1 then
        Record flex sensor reading;
        break;
    else
        Record NULL;
        Continue applying pressure;
    end
end
while Contact is acheived do
    Solve the recorded value using curve fitting
    equation (4);
    Predict size by solving equation;
    if object grasped successfully then
        Close valves;
        Do pick and place operation;
    end
end

```

---

IV. EXPERIMENTS

A. Contact Detection and Validation

In the proposed contact detection algorithm (*Algorithm1*), a series of contact detection experiments are performed and analyzed by considering the numerical gradient for a sample set of flex sensor readings. It has been noted that during contact occurrence with the object the the numerical gradient of bending value changes drastically. In literature, contact detection was performed by comparing with free grasping but there is always a chance of nonlinear data that may give a wrong contact detection. Moreover, the proposed contact detection algorithm is validated using a touch sensor attached to the fingertip of the gripper. An envelope grasping is performed using a flex sensor and touch sensor simultaneously. When the fingertip contacts with the object, the touch sensor gives a high digital signal as shown in Fig. 7. It is observed that the contact point achieved by the proposed contact detection algorithm and touch sensor are approximately the same.

B. Size Recognition Implementation

The exclusive behavior of flex sensor response with different size objects not only demonstrates repeatability of contact detection but also shows the capability of size recognition. A set of 3D printed PLA spheres (error±0.01mm) of diameter from 50mm to 104mm with an interval of 6mm is considered as shown in Fig. 8. Mimicking the human grasping behavior [19], envelope grasping is performed using *Algorithm2*.

By considering a constant value of source pressure (i) envelope grasping are performed (ii) contact point flex sensor value of these 9 objects is collected using the *Algorithm1* (iii) curve is fitted to flex sensor readings and measured object size.

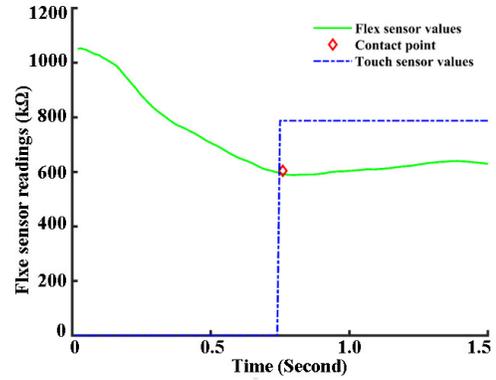


Fig. 7. Contact Detection Validation using Touch Sensor



Fig. 8. Set of Different Size PLA Spheres

Fig. 9 depicts the curve fitting graph of polynomial *degree2* using solving equation (4):

$$f(x) = 0.0004x^2 - 0.09528x + 37.98 \quad (4)$$

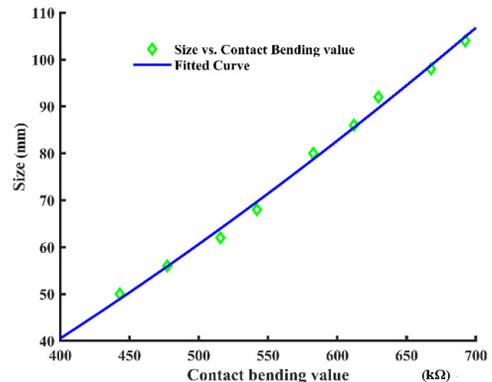


Fig. 9. Curve Fitting of Polynomial Degree 2

Where  $x$  is the data from the contact flex sensor obtained with the aid of *Algorithm1*. The fitted polynomial curve of *degree2* obtained from the proposed method, provides better R-square value as compared to previous work [18]. By solving the equation (4), the size of the object can be predicted.

### C. Grasping State

Sensitive objects handling is a crucial task in the automated industry because during grasping it can degrade the quality of the objects. Feedback of grasping state i.e. whether the object is grasped or not is a challenging task. In this section, the grasping state has been analyzed using raw data of sensory feedback. Fig. 10 shows the response of the successful and unsuccessful grasping state of an object. During successful grasping, after contact occurrence, the numerical gradient tends to increase constantly whereas during unsuccessfully grasping gradient value is decreased abruptly. This type of information can be useful to the robot as a feedback grasping confirmation of the object. Furthermore, it can be implemented in human-robot interaction where a robot may acquire the information of the object is taken by humans or not which may save the delay in the interaction process.

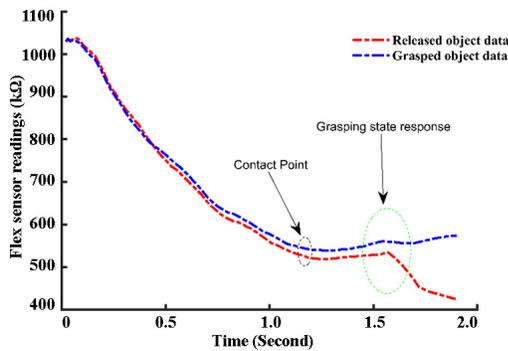


Fig. 10. Comparison of Successful and Unsuccessful Grasp State of the Target Object

## V. TEST RESULTS

### A. Size Recognition

A suitable source pressure (172 kPa) is provided to perform the testing of size recognition. When the soft pneumatic gripper contacts with the objects, it starts grasping. Based on *Algorithm2* online size recognition has been performed by utilizing the curve fitting equation. A PLA sphere of 74mm diameter is chosen to perform envelope grasping, which demonstrates the repeatability of the size recognition algorithm. It can be noticed from Table I, that the prediction error of the real size and predicted size is within  $\pm 3mm$ . Moreover, the size recognition algorithm is tested with sensitive fruits, vegetables and daily life objects of three different sizes respectively as depicted in Fig. 11. The prediction error is presented in Table II, which shows the diverse application of the proposed size recognition algorithm regardless type of material.

From the above table it can be analysed that proposed contact detection and size recognition algorithm performs well compared with [18]. Proposed work is a less complex and easy to implement.

TABLE I. REPEATABILITY ANALYSIS

ExperimentNo	Error analysis and repeatability analysis of an object		
	Real Size(mm)	Predicted Size(mm)	Errors(mm)
1	74	73.65	0.35
2	74	74.54	0.54
3	74	72.98	1.02
4	74	73.43	0.57
5	74	73.83	0.1
6	74	72.54	1.46
7	74	73.20	0.8
8	74	72.76	1.24



Fig. 11. Test Object for Size Recognition

## VI. APPLICATION

### A. Fruit Sorting Process

After successful testing of size recognition algorithm, an automated fruit sorting environment is created (i) a sawyer robot for performing pick and place operation (ii) different size of tomatoes, oranges or apples are kept on a transport belt for handling and sorting process (iii) SONAR sensor is used to detect the object at picking position on transport belt (iv) A



Fig. 12. Online Size Recognition and Fruit Sorting Process

TABLE II. ONLINE SIZE RECOGNITION

Experiment No	Size recognition and error analysis of test objects			
	Test object	Real size (mm)	Predicted size (mm)	Errors (mm)
1	Tomato	79.03	80.42	1.39
2	Tomato	73.53	75.34	1.81
3	Tomato	50.20	48.33	1.87
4	Orange	86.69	89.12	2.43
5	Orange	66.29	65.75	0.54
6	Orange	76.48	78.51	2.03
7	Apple	86.54	88.40	1.86
8	Apple	70.70	69.23	1.47
9	Apple	77.68	78.65	0.97
10	Tennis Ball	64.08	65.97	1.89
11	Led Bulb	78.59	76.58	2.01

FSR sensor is attached at the center position of the lower base of the gripper (v) ROS with Matlab is used to implement the proposed algorithm. The fruits placed on the transport belt are continuously moving toward the robot picking position. When the SONAR sensor detects the object, the transportation belt is stopped and the manipulator starts moving towards the objects to be picked. When FSR sensor touches the object to be picked, manipulator stops and grasping operation is performed [18]. The grasping is performed by sending command from Matlab as a rosnode and also monitors the real-time feedback of the flex sensor. After the successful online implementation of the proposed contact detection and size recognition algorithm, the robot manipulator is planned to place the object into their respective bins. This process is performed for a series of objects as shown in Fig. 12.

## VII. CONCLUSION

Sensitive objects handling is a crucial task in the automated industry because during grasping it can degrade the quality of the objects such as fruits, vegetables, and daily life items. This paper demonstrates the implementation of simple sensory feedback during grasping which can be used in the handling process of objects. A numerical gradient-based algorithm is proposed and validated to detect the contact occurrence of the target object. Based on the contact detection algorithm, a size recognition algorithm is proposed which is implemented to perform an automated object sorting process. An experimental setup is developed using the sawyer robot controlled using ROS and MATLAB to implement the proposed technique. It has been noted that the proposed contact detection method has shown better accuracy ( $\pm 3mm$ ) in terms of size recognition as compared with previous work. Moreover, successful and unsuccessful grasping state of objects has been analysed using raw flex sensor data which can be utilized in human-robot interaction to increase efficiency in the automation process. The key benefits of the proposed algorithm are less complexity, better accuracy, and lower cost. This research can be extended using a wider set of objects with more variation and grasp orientation. Multi-segment flex sensor may provide better sensitivity and accuracy.

## REFERENCES

[1] "Guest editorial focused section on soft actuators, sensors, and components (sasc)," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 1, pp. 1–4, Feb 2019.

[2] Z. Wang, Y. Torigoe, and S. Hirai, "A prestressed soft gripper: Design, modeling, fabrication, and tests for food handling," *IEEE Robotics and Automation Letters*, vol. 2, no. 4, pp. 1909–1916, Oct 2017.

[3] P. Polygerinos, N. Correll, S. A. Morin, B. Mosadegh, C. D. Onal, K. Petersen, M. Cianchetti, M. T. Tolley, and R. F. Shepherd, "Soft robotics: Review of fluid-driven intrinsically soft devices; manufacturing, sensing, control, and applications in human-robot interaction," *Advanced Engineering Materials*, vol. 19, no. 12, p. 1700016, 2017.

[4] J. Avery, M. Runciman, A. Darzi, and G. P. Mylonas, "Shape sensing of variable stiffness soft robots using electrical impedance tomography," in *2019 International Conference on Robotics and Automation (ICRA)*, May 2019, pp. 9066–9072.

[5] "A dual-mode soft gripper for food packaging," *Robotics and Autonomous Systems*, vol. 125, p. 103427, 2020.

[6] K. Elgeneidy, G. Neumann, S. Pearson, M. Jackson, and N. Lohse, "Contact detection and size estimation using a modular soft gripper with embedded flex sensors," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 498–503.

[7] A. D. Marchese, C. D. Onal, and D. Rus, "Autonomous soft robotic fish capable of escape maneuvers using fluidic elastomer actuators," *Soft Robotics*, vol. 1, no. 1, pp. 75–87, 2014.

[8] Z. Zhang, A. Petit, J. Dequidt, and C. Duriez, "Calibration and external force sensing for soft robots using an rgb-d camera," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2356–2363, 2019.

[9] C. Choi, W. Schwarting, J. DelPreto, and D. Rus, "Learning object grasping for soft robot hands," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2370–2377, 2018.

[10] Q. Hao, Z. Li, H. Yan, G. Li, and B. Su, "A natural orifice soft robot with novel driven method for minimally invasive surgery (mis)," in *2017 2nd Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*. IEEE, 2017, pp. 289–294.

[11] B. S. Homberg, R. K. Katschmann, M. R. Dogar, and D. Rus, "Haptic identification of objects using a modular soft robotic gripper," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 1698–1705.

[12] J. A. Rogers, T. Someya, and Y. Huang, "Materials and mechanics for stretchable electronics," *science*, vol. 327, no. 5973, pp. 1603–1607, 2010.

[13] S. Li, H. Zhao, and R. F. Shepherd, "Flexible and stretchable sensors for fluidic elastomer actuated soft robots," *MRS Bulletin*, vol. 42, no. 2, pp. 138–142, 2017.

[14] J. M. Butt, H. Wang, and R. Pathan, "Design, fabrication, and analysis of a sensorized soft robotic gripper," in *2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, July 2018, pp. 169–174.

[15] G. Saggio, F. Riillo, L. Sbernini, and L. R. Quitadamo, "Resistive flex sensors: a survey," *Smart Materials and Structures*, vol. 25, no. 1, p. 013001, 2015.

[16] C. Schlagenhauf, D. Bauer, K.-H. Chang, J. P. King, D. Moro, S. Coros, and N. Pollard, "Control of tendon-driven soft foam robot hands," in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2018, pp. 1–7.

[17] K. Elgeneidy, N. Lohse, and M. Jackson, "Bending angle prediction and control of soft pneumatic actuators with embedded flex sensors—a data-driven approach," *Mechatronics*, vol. 50, pp. 234–247, 2018.

[18] Y. Chen, S. Guo, C. Li, H. Yang, and L. Hao, "Size recognition and adaptive grasping using an integration of actuating and sensing soft pneumatic gripper," *Robotics and Autonomous Systems*, vol. 104, pp. 14–24, 2018.

[19] X. Chen, Z. Li, Y. Wang, and J. Liu, "Effect of fruit and hand characteristics on thumb-index finger power-grasp stability during manual fruit sorting," *Computers and electronics in agriculture*, vol. 157, pp. 479–487, 2019.

# Ensemble Methods to Detect XSS Attacks

PMD Nagarjun<sup>1</sup>, Shaik Shakeel Ahamad<sup>2\*</sup>

Department of CSE, K L University, Vijayawada, India<sup>1</sup>

College of Computer and Information Sciences (CCIS)

Majmaah University, Al Majmaah, Kingdom of Saudi Arabia<sup>2</sup>

**Abstract**—Machine learning techniques are gaining popularity and giving better results in detecting Web application attacks. Cross-site scripting is an injection attack widespread in web applications. The existing solutions like filter-based, dynamic analysis, and static analysis are not effective in detecting unknown XSS attacks, and machine learning methods can detect unknown XSS attacks. Existing research to detect XSS attacks by using machine learning methods have issues like single base classifiers, small datasets, and unbalanced datasets. In this paper, supervised ensemble learning techniques trained on a large labeled and balanced dataset to detect XSS attacks. The ensemble methods used in this research are random forest classification, AdaBoost, bagging with SVM, Extra-Trees, gradient boosting, and histogram-based gradient boosting. Analyzed and compared the performance of ensemble learning algorithms by using the confusion matrix.

**Keywords**—Cross-site scripting; machine learning; ensemble learning; random forest; bagging; boosting

## I. INTRODUCTION

Machine learning algorithms are useful in detecting unknown and new XSS attacks in Web Applications. Ensemble methods are a combination of different base models, and the ensemble learning models can give optimal results compared to base models [1]. In XSS attacks, the attacker can steal victim's session cookie, sensitive data of victim, implement keyloggers at browser, and damage the reputation of a trusted Website.

A common problem in existing XSS prevention techniques are the incapability of detecting unknown or new XSS attacks [2]. Highly effective XSS detection models can be built by using ensemble learning techniques. AdaBoost, bagging, Extra-Trees, gradient boosting, random forest, histogram-based gradient boosting are ensemble methods, which uses base models like decision trees, etc.

Cross-site scripting injection attacks are categorized into three types, and they are persistent (stored), non-persistent (reflected), and DOM-based attacks. Many existing solutions primarily focused on preventing only one type of XSS attack, and there are only a few solutions to avoid all types of attacks [3]. The proposed ensemble learning models can detect all types of attacks by proper implementation at the server and client-side.

Ensemble methods use different algorithms to achieve better prediction rate. Usually, ensemble learning involves the same base learning algorithm. The limitation in ensemble methods is that these require more computations compared to a single model. In ensemble learning base models are combined in three ways.

**Bagging:** In bagging (bootstrap aggregation) weak learning algorithms applies on a small sample dataset and takes an average of all learners prediction. Bagging will decrease the variance.

**Boosting:** It is an iterative method, in boosting sample weights are adjusted based on the previous classification. Boosting will decrease bias error.

**Stacking:** In this output of one model is given as input to another model. Stacking will decrease variance or bias based on models used.

The purpose of this paper is to investigate and compare the prediction accuracy of machine learning ensemble methods in detecting Cross-site scripting attacks in Web Applications.

The paper is organized as follows: Section 2 contains related work. We prepared XSS data for training and testing in Section 3. We implemented the ensemble learning models in Section 4. We analyzed the performance of proposed ensemble models in Section 5. Sections 6 and 7 contains conclusion and future work.

## II. RELATED WORK

Rodriguez et al. [4] analyzed 67 documents related to XSS attacks. According to their research, most of the researches use browser tools or web page analysis methods to prevent XSS attacks, very few researches on machine learning algorithms to prevent these attacks. Based on their research most common issues in existing researches are detecting only one type of XSS attacks, low attacks data, only restricted to one programming environment like PHP, same data for different researches, methods not scalable, high false positives, methods work on only one browser, few methods proposed to use artificial intelligence, etc.

S. Gupta and B. B. Gupta [5] did a study on defense mechanisms of XSS attacks, and they stated that safe input handling is one of the essential techniques to mitigate XSS attacks. A good XSS defensive technique needs to differentiate malicious code and legitimate JavaScript code automatically.

Hydara et al. [6] studied 115 research papers on XSS attacks. Based on their study, non-persistence XSS attacks are popular, and there is a need for solutions to remove XSS vulnerabilities from the source code.

Shanmugasundaram et al. [7] stated that developers lack knowledge on implementing existing XSS solutions in their web applications.

\*Corresponding Author

Aliga et al. [8] study showed that most of the XSS prevention solutions are client-side, and they are unable to detect new XSS attacks, and these solutions lack self-learning capabilities. They reviewed 15 XSS prevention techniques, and out of 15, only two techniques have self-learning capabilities.

Nunan et al. [9] used supervised ML methods like Naive Bayes and SVM to detect XSS attacks. Their total data set size 216054, and among them, 15366 are XSS attacks. They evaluate the algorithms based on accuracy, detection, and false alarm rates, etc. Their results show that compared to Naive Bayes, the SVM achieved the best performance. They selected the following features for classification of XSS attacks Obfuscation of code, the number of domains, URL Length, duplicate special characters, Schemes, etc.

Mereani and Howe [10] developed Random Forest, kNN, and SVM models to detect XSS malicious code, and they used labeled data in training. They trained using 2000 samples and used 13000 for testing. In their experiments, they reached accuracy up to 99.75%. They extracted Structural features contain a set of special characters in malicious JavaScript, and Behavioural Features includes function and commands used in malicious JavaScript code, a total of 59 features from both categories.

Rathore et al. [11] developed an ML method for Social networking services (SNSs) to detect XSS attacks. In their method, extracted Webpage features, URL features, and SNSs features from web pages and used this data to train models. Some of the features include domains in a URL, URL length, Iframes count, external link counts, and malicious JavaScript codes in SNSs webpage, etc. 1000 SNSs pages used to build a dataset for testing and used different classifiers in their testing. They achieved 97.2% accuracy in their tests.

Akaishi and Uda [12] used a combination of classifiers to detect XSS attacks in their research. Their data set contain balanced 10000 samples where attack data in URL format. They divided the attack sentence into words, co-occurrence, and frequency of words used in their classification. They used word2vec based model in their research to transform words into vectors, and used those vectors is classification algorithms. According to them, CNN and SVM are the best filters for real-world problems.

Mokbal et al. [13] proposed a Multilayer perceptron based model to detect XSS attacks. Their model achieved an accuracy of 99.32% in detecting attacks. Their dataset contains a total of 138569 samples, and among them, 38569 are attack samples. They extracted URL based, HTML based, and JavaScript-based features form content and used these features in training proposed models. Some of the features like URL length and special characters in URL, HTML tags, JavaScript events, etc.

Wang, Cai, and Wei [14] proposed a deep learning-based framework to detect malicious JavaScript. The structure contains logistic regression, deep learning method, and sparse random projection. They extracted features from JavaScript code by using Stacked denoising autoencoders (SdA). These features used to train SVM or logistic regression models. Classification of malicious code done by logistic regression.

Their labeled dataset contains 14783 malicious JavaScript codes and 12320 benign samples. Their model achieved 94.9% accuracy.

### III. DATA COLLECTION AND DATA PREPROCESSING

For this research, collected XSS vectors by using popular XSS tools like XSSstrike, XSSER [15] and from different sources collected thousands of attack vectors. The dataset contains 154626 unique samples with labels. Half of this dataset is XSS attack vectors, and another half (77313) of the dataset is safe vectors. XSS attack vectors and Safe vectors are maintained at 128 characters, and longer sequences are split into 128 character chunks. Fig. 1 shows safe vector generator, by using this, generated safe vector samples. These randomly generated safe vectors are three types with length ranges from 40 to 126 those are, string with only uppercase or lowercase alphabets, strings with all alphabets and digits, and strings with all alphabets, digits and special characters. The below examples show different types of safe vectors:

1. kikDfuPLasVpSDqfKLMUTbyDAssjedEhphsOSPUnxO  
OHwDUkdHxLyJGPoMRIVERzJwuTVmbCwwYjVTtQ  
TfApXparHUUEEiidfUWBfjNUnVovFYNIbTJJ
2. aLcmHRaDMXwMmOmzQDhbEfeSYcZTRsPNkbjcoCa  
YauezgpthiPEvrUGfOXHGljqgZSDiArGKshBDvmcYm  
OdOYIpDsfbfGoPrwQXikjltIqImReZGbeVFwABEJZg  
Sn
3. BqAoxOrvaovydRv8QuQmQvoAk6hUbTaUFx18al7jYZ  
XBWvf1GWHllbwgYd1qR2mx
4. x54fQrSJicA8f2KInEibadR3NrAVwkTgKdFn8WqBpqB  
KcufKJZ1zPpqybBPPQCu0LcWHjkRqvEgnJHUolgRLiZ  
ebe13wt7b6S1uY23cWkbleU7dzbKyQMysra18u
5. Y0P/U#Y\_Dk#NNZ?p>B]6Ndb[&.^iMI=~ts8Depf\*C`aQ  
>!d[:p02LzJ,`5"hVCqAPXonVtrQ]L9`JBD=8L<c"TI-  
?PASb7bs/[.lXXMyQ:7av`q?m-@XV7"xm(
6. 2\{k@1\WMNXMi/3[1=mo#UHv5Da@-PzvG%\*t(h-  
f[L25+{IU3#2Y\_[msZ8h\_^QP\$@E4quPS~.~JddH"G3.+2  
)1~+svNQ.HPuCT5eKZVV\*[Ej]\*x5

The number of safe vectors generated depends on XSS attack vectors, to maintain the balance between XSS and safe samples of the dataset. This balanced dataset used to train and test the models. The below examples shows sample XSS attack vectors.

1. 
2. <script\x20type="text/javascript">javascript:alert(19);</sc  
ript>

To prepare input for models, converted the character sequence of XSS attacks, and Safe vectors into Unicode integer format, Fig. 2 shows sample data in Unicode format. The dataset is standardized by using sklearn's [16] StandardScaler function, Standardization of a dataset will improve the performance and accuracy of machine learning algorithms. Fig. 3 shows a sample data after standardization without the output column. The preparing process of dataset for model training shown in Fig. 4.

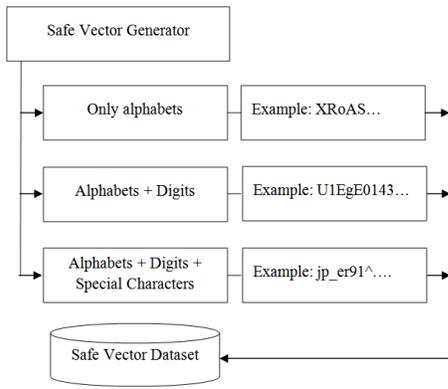


Fig. 1. Safe Vector Generator.

	0	1	2	3	4	5	6	...	122	123	124	125	126	127	128
86564	106	42	50	47	46	67	43	...	0	0	0	0	0	0	0
141242	51	36	91	126	65	59	55	...	0	0	0	0	0	0	0
74704	37	50	66	65	67	73	65	...	0	0	0	0	0	0	1
4185	60	65	47	43	47	79	110	...	0	0	0	0	0	0	1
47456	60	47	83	84	121	76	101	...	0	0	0	0	0	0	1
95944	101	67	118	77	90	99	117	...	0	0	0	0	0	0	0
61	60	83	84	89	76	69	62	...	0	0	0	0	0	0	1
147029	48	101	113	102	50	115	83	...	0	0	0	0	0	0	0
138884	71	72	107	119	85	66	89	...	0	0	0	0	0	0	0
81449	122	101	83	87	106	67	81	...	0	0	0	0	0	0	0

Fig. 2. Sample Data from the Dataset in Unicode Format.

	0	1	2	...	125	126	127
0	0.180603	-0.114946	-1.220119	...	-0.086933	-0.04473	-0.043924
1	-0.146358	-0.443607	-0.561638	...	-0.086933	-0.04473	-0.043924
2	0.088645	-0.402525	0.058109	...	-0.086933	-0.04473	-0.043924
3	0.241908	1.405115	1.336337	...	-0.086933	-0.04473	-0.043924
4	-0.146358	-1.881503	-1.491258	...	-0.086933	-0.04473	-0.043924
5	-0.146358	-1.306345	-0.019359	...	-0.086933	-0.04473	-0.043924
6	-0.146358	-0.443607	0.677856	...	-0.086933	-0.04473	-0.043924
7	0.201038	-1.347427	0.019375	...	-0.086933	-0.04473	-0.043924
8	-0.146358	-1.306345	1.220134	...	-0.086933	-0.04473	-0.043924
9	-0.146358	1.035371	1.258868	...	-0.086933	-0.04473	-0.043924

Fig. 3. Sample Data from the Dataset after Standardization.

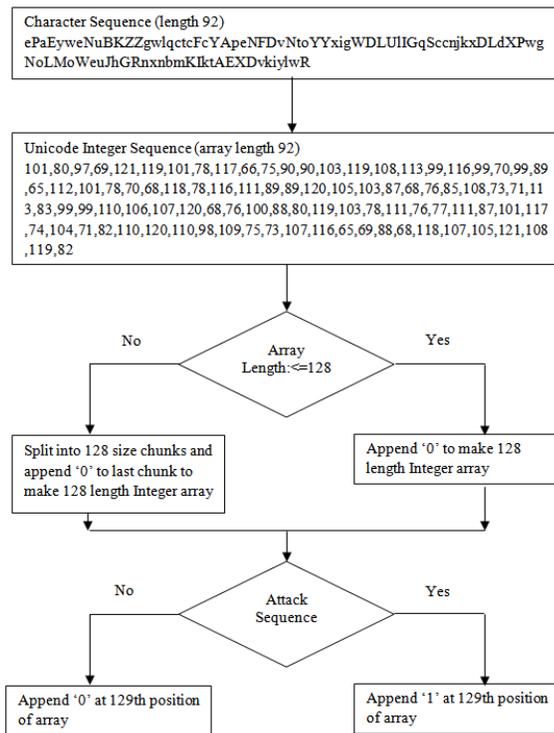


Fig. 4. Process of Preparing the Dataset.

#### IV. IMPLEMENTATION OF ENSEMBLE METHODS

In this research supervised ensemble machine learning methods are used to detect XSS attacks. The ensemble learning methods [17] used are random forest classification, AdaBoost, bagging with SVM, Extra-Trees, gradient boosting, and histogram-based gradient boosting. These ensemble classification methods are effective in detecting XSS attacks compared to base models.

Google Colab [18] is used to build and test these models. The working environment includes Python 3.6.9, scikit-learn, TensorFlow 2.1.0 (includes Keras) [19], etc.

The dataset contains balanced unique 154626 samples, 77313 are Safe vectors and 77313 are XSS attacks. Total samples divided into 8:2 ratio for training (123700) and testing (30926) samples.

Fig. 5 shows the confusion matrix. The confusion matrix values are used to compare and evaluate the models.

Confusion matrix [20] used to calculate performance metrics of a model, by using the confusion matrix one can calculate the following values.

$$\text{Recall} = (\text{TRUE POSITIVE}) / (\text{TRUE POSITIVE} + \text{FALSE NEGATIVE})$$

$$\text{Precision} = (\text{TRUE POSITIVE}) / (\text{TRUE POSITIVE} + \text{FALSE POSITIVE})$$

$$\text{F-measure} = (2 \times \text{Recall} \times \text{Precision}) / (\text{Recall} + \text{Precision})$$

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

##### A. Random Forest Classifier (RF)

Random forest classifier contains a collection of decision trees, and each decision tree fits on a subset of the dataset. Based on the output of all decision trees, the random forest classifier decides the final class of an input object. Table I shows the confusion matrix of the random forest model, and Table II shows the recall, precision, F-measure, and accuracy of the random forest model in detecting XSS attacks and safe vectors. Cross-validation scores of random forest classifier model are 0.99803, 0.99774, 0.99787, 0.99796, 0.99822 and the mean is 0.99796. In the random forest model, reached accuracy up to 0.99822.

		Predicted Values	
		TRUE POSITIVE [TP]	FALSE NEGATIVE [FN]
Actual Values	FALSE POSITIVE [FP]		TRUE NEGATIVE [TN]
	TRUE POSITIVE [TP]		FALSE NEGATIVE [FN]

Fig. 5. Confusion Matrix.

TABLE I. RANDOM FOREST MODEL CONFUSION MATRIX

	Safe samples (Predicted)	XSS samples (Predicted)
Safe samples (Actual)	15463	0
XSS samples (Actual)	62	15401

TABLE II. RECALL, PRECISION, F-MEASURE, AND ACCURACY OF RANDOM FOREST MODEL

	Recall	Precision	F-measure
Safe samples	1.00000	0.99601	0.99800
XSS samples	0.99599	1.00000	0.99799
Accuracy = 0.99800			

B. AdaBoost Classifier (AB)

Boosting algorithms are used to reach high accuracy, AdaBoost (Adaptive Boosting) is a popular ensemble boosting algorithm works on decision trees. AdaBoost combines multiple low performing classifiers to get high performing classifier. In AdaBoost's every iteration, weak classifiers are tweaked (weighted data) based on the accuracy of previous training. The confusion matrix of the AdaBoost classifier model is shown in Table III, and Table IV shows the recall, precision, F-measure, and accuracy of the AdaBoost classifier model in detecting XSS attacks and safe vectors. Cross-validation scores of AdaBoost classifier model are 0.9977, 0.99793, 0.99735, 0.99764, 0.99832 and the mean is 0.99779. In the AdaBoost model, reached accuracy up to 0.99832.

C. Bagging Classifier with SVM (BC)

Bootstrap aggregating (or Bagging) is an ensemble method in machine learning. SVM is used as a base classifier for the bagging model. In bagging, the base classifiers are trained (fits) on a randomly selected subset data of the original dataset, and the final prediction depends on individual base classifiers predictions. The confusion matrix of the bagging model is shown in Table V, and Table VI shows the recall, precision, F-measure, and accuracy of the bagging model in detecting XSS attacks and safe vectors. Cross-validation scores of bagging classifier model are 0.98192, 0.98228, 0.98186, 0.98264, 0.98276 and the mean is 0.98229. In the bagging model, reached accuracy up to 0.98276.

D. Extra-Trees Classifier (ET)

Extra-Trees (Extremely Randomized Trees) method is an ensemble method similar to the random forest classifier. In Extra-Trees classifier, decision trees are constructed randomly in the forest, and these decision trees trained (fits) on subsets of data. The final prediction depends on all decision trees predictions. The confusion matrix values of the Extra-Trees classifier model is shown in Table VII, and Table VIII shows the recall, precision, F-measure, and the accuracy of the Extra-Trees classifier model in detecting XSS attacks and safe vectors. Cross-validation scores of Extra-Trees classifier model are 0.99049, 0.99088, 0.99069, 0.99192, 0.99175 and the mean is 0.99115. In the Extra-Trees classifier model, reached accuracy up to 0.99192.

E. Gradient Boosting Classifier (GB)

Gradient boosting classifier is an ensemble boosting algorithm, where a weak classifier is modified into a strong classifier. In the gradient boosting classifier, decision trees are base classifiers, and loss function is optimized while adding a new tree. Table IX shows the confusion matrix of the gradient boosting model, and Table X shows the recall, precision, F-measure, and the accuracy of the gradient boosting model in detecting XSS attacks and safe vectors. Cross-validation scores of gradient boosting classifier model are 0.99618, 0.99573, 0.99644, 0.99609, 0.99648 and the mean is 0.99618. In the gradient boosting model, reached accuracy up to 0.99648.

TABLE III. ADABOOST MODEL CONFUSION MATRIX

	Safe samples (Predicted)	XSS samples (Predicted)
Safe samples (Actual)	15440	23
XSS samples (Actual)	46	15417

TABLE IV. RECALL, PRECISION, F-MEASURE, AND ACCURACY OF ADABOOST MODEL

	Recall	Precision	F-measure
Safe samples	0.99851	0.99703	0.99777
XSS samples	0.99703	0.99851	0.99777
Accuracy = 0.99800			

TABLE V. BAGGING MODEL CONFUSION MATRIX

	Safe samples (Predicted)	XSS samples (Predicted)
Safe samples (Actual)	15243	220
XSS samples (Actual)	354	15109

TABLE VI. RECALL, PRECISION, F-MEASURE, AND ACCURACY OF BAGGING MODEL

	Recall	Precision	F-measure
Safe samples	0.98577	0.97730	0.98152
XSS samples	0.97711	0.98565	0.98136
Accuracy = 0.99800			

TABLE VII. EXTRA-TREES CLASSIFIER MODEL CONFUSION MATRIX

	Safe samples (Predicted)	XSS samples (Predicted)
Safe samples (Actual)	15463	0
XSS samples (Actual)	272	15191

TABLE VIII. RECALL, PRECISION, F-MEASURE, AND ACCURACY OF EXTRA-TREES CLASSIFIER MODEL

	Recall	Precision	F-measure
Safe samples	1.00000	0.98271	0.99128
XSS samples	0.98241	1.00000	0.99113
Accuracy = 0.99800			

TABLE IX. GRADIENT BOOSTING MODEL CONFUSION MATRIX

	Safe samples (Predicted)	XSS samples (Predicted)
Safe samples (Actual)	15406	57
XSS samples (Actual)	79	15384

TABLE X. RECALL, PRECISION, F-MEASURE, AND ACCURACY OF GRADIENT BOOSTING MODEL

	Recall	Precision	F-measure
Safe samples	0.99631	0.99490	0.99561
XSS samples	0.99489	0.99631	0.99560
Accuracy = 0.99800			

F. Histogram-based Gradient Boosting Classification (HGBC)

Histogram-based gradient boosting classification is an ensemble boosting algorithm, which is better compared to gradient boosting for large datasets. HGBC can handle missing values. In HGBC, decision trees are base classifiers. Table XI shows the confusion matrix of the HGBC model, and Table XII shows recall, precision, F-measure, and the accuracy of the HGBC model in detecting XSS attacks and safe vectors. Cross-validation scores of HGBC model are 0.99874, 0.99877, 0.99851, 0.99871, 0.9989 and the mean is 0.99873. In the HGBC model, reached accuracy up to 0.9989.

TABLE XI. HGBC MODEL CONFUSION MATRIX

	Safe samples (Predicted)	XSS samples (Predicted)
Safe samples (Actual)	15447	16
XSS samples (Actual)	32	15431

TABLE XII. RECALL, PRECISION, F-MEASURE, AND ACCURACY OF HGBC MODEL

	Recall	Precision	F-measure
Safe samples	0.99897	0.99793	0.99845
XSS samples	0.99793	0.99896	0.99845
Accuracy = 0.99800			

V. RESULTS AND DISCUSSION

This research evaluated the XSS detection rate in ensemble learning techniques. AdaBoost, bagging with SVM, Extra-Trees, gradient boosting, random forest classification, and histogram-based gradient boosting models are trained on a large labeled dataset and evaluated these methods performance based on their accuracy, recall, precision, and the F-measure. Table XIII compares the performance metrics of all models, and Table XIV compares the cross-validation scores of all models, and Fig. 6 shows the mean score of cross-validations of models. From the results, it is concluded that all ensemble methods performed well and reached an accuracy of more than 98% in all models.

Form all tested ensemble machine learning algorithms, the histogram-based gradient boosting classification model is the best performed model with the highest possible accuracy of 0.9989.

TABLE XIII. COMPARISON OF PERFORMANCE METRICS

Model	Recall		Precision		F-measure		Accuracy
	Safe	XSS	Safe	XSS	Safe	XSS	
RF	1.00000	0.99599	0.99601	1.00000	0.99800	0.99799	0.99800
AB	0.99851	0.99703	0.99703	0.99851	0.99777	0.99777	0.99777
BC	0.98577	0.97711	0.97730	0.98565	0.98152	0.98136	0.98144
ET	1.00000	0.98241	0.98271	1.00000	0.99128	0.99113	0.99120
GB	0.99631	0.99489	0.99490	0.99631	0.99561	0.99560	0.99560
HGBC	0.99897	0.99793	0.99793	0.99896	0.99845	0.99845	0.99845

TABLE XIV. COMPARISON OF CROSS-VALIDATION SCORES

Model	Fold 1 score	Fold 2 score	Fold 3 score	Fold 4 score	Fold 5 score	Mean score
RF	0.99803	0.99774	0.99787	0.99796	0.99822	0.99796
AB	0.99777	0.99793	0.99735	0.99764	0.99832	0.99779
BC	0.98192	0.98228	0.98186	0.98264	0.98276	0.98229
ET	0.99049	0.99088	0.99069	0.99192	0.99175	0.99115
GB	0.99618	0.99573	0.99644	0.99609	0.99648	0.99618
HGBC	0.99874	0.99877	0.99851	0.99871	0.9989	0.99873

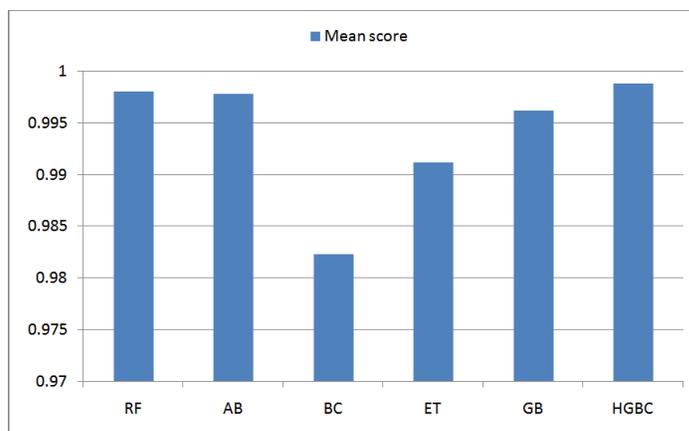


Fig. 6. Mean Score of Ensemble Learning Algorithms.

VI. CONCLUSION

We developed and analyzed supervised ensemble machine learning methods to detect XSS attacks in Web applications. Ensemble learning techniques are a collection of base classifiers, and these ensemble methods perform better than single classifiers. Existing solutions to detect XSS attacks by using machine learning methods have issues like single base classifiers, small datasets, and unbalanced datasets. We trained and evaluated proposed models on a large balanced dataset, and in this research, we detect XSS attacks in data submitted by the user. In this work, we evaluated the performance of random forest classification, AdaBoost, bagging with SVM,

Extra-Trees, gradient boosting, and histogram-based gradient boosting models in detecting XSS attacks and safe vectors. We compared the performance of models by using the confusion matrix metrics. The results show that all ensemble learning models performed exceptionally well in detecting XSS attacks and safe vectors. We reached the highest accuracy of 0.9989 in the histogram-based gradient boosting classification model.

## VII. FUTURE WORK

In future, the work can be extended to detect other Web application attacks like SQL injection. The models can be tested by integrated into real world applications to detect attacks.

## REFERENCES

- [1] Y. Zhou and P. Wang, "An ensemble learning approach for XSS attack detection with domain knowledge and threat intelligence," *Comput. Secur.*, vol. 82, pp. 261–269, 2019.
- [2] U. Sarmah, D. K. Bhattacharyya, and J. K. Kalita, "A survey of detection methods for XSS attacks," *J. Netw. Comput. Appl.*, vol. 118, pp. 113–143, 2018.
- [3] L. K. Shar and H. B. K. Tan, "Defending against cross-site scripting attacks," *Computer (Long. Beach. Calif.)*, vol. 45, no. 3, pp. 55–62, 2011.
- [4] G. E. Rodríguez, J. G. Torres, P. Flores, and D. E. Benavides, "Cross-site scripting (XSS) attacks and mitigation: A survey," *Comput. Networks*, vol. 166, p. 106960, 2020.
- [5] S. Gupta and B. B. Gupta, "Cross-Site Scripting (XSS) attacks and defense mechanisms: classification and state-of-the-art," *Int. J. Syst. Assur. Eng. Manag.*, vol. 8, no. 1, pp. 512–530, 2017.
- [6] I. Hydar, A. B. M. Sultan, H. Zulzalil, and N. Admodisastro, "Current state of research on cross-site scripting (XSS) - A systematic literature review," *Inf. Softw. Technol.*, vol. 58, no. July 2015, pp. 170–186, 2015.
- [7] G. Shanmugasundaram, S. Ravivarman, and P. Thangavellu, "A study on removal techniques of Cross-Site Scripting from web applications," in *2015 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC)*, 2015, pp. 436–442.
- [8] A. P. Aliga, A. M. John-Otumu, R. E. Imhanhahimi, and A. C. Akpe, "Cross Site Scripting Attacks in Web-Based Applications," *J. Adv. Sci. Eng.*, vol. 1, no. 2, pp. 25–35, 2018.
- [9] A. E. Nunan, E. Souto, E. M. Dos Santos, and E. Feitosa, "Automatic classification of cross-site scripting in web pages using document-based and URL-based features," in *2012 IEEE symposium on computers and communications (ISCC)*, 2012, pp. 702–707.
- [10] F. A. Mereani and J. M. Howe, "Detecting cross-site scripting attacks using machine learning," in *International Conference on Advanced Machine Learning Technologies and Applications*, 2018, pp. 200–210.
- [11] S. Rathore, P. K. Sharma, and J. H. Park, "XSSClassifier: An Efficient XSS Attack Detection Approach Based on Machine Learning Classifier on SNSs.," *JIPS*, vol. 13, no. 4, pp. 1014–1028, 2017.
- [12] S. Akaishi and R. Uda, "Classification of XSS Attacks by Machine Learning with Frequency of Appearance and Co-occurrence," in *2019 53rd Annual Conference on Information Sciences and Systems (CISS)*, 2019, pp. 1–6.
- [13] F. M. M. Mokbal, W. Dan, A. Imran, L. Jiuchuan, F. Akhtar, and W. Xiaoxi, "MLPXSS: An Integrated XSS-Based Attack Detection Scheme in Web Applications Using Multilayer Perceptron Technique," *IEEE Access*, vol. 7, pp. 100567–100580, 2019.
- [14] Y. Wang, W. Cai, and P. Wei, "A deep learning approach for detecting malicious JavaScript code," *Secur. Commun. networks*, vol. 9, no. 11, pp. 1520–1534, 2016.
- [15] I. M. Babincev and D. V. Vuletić, "Web application security analysis using the kali Linux operating system," *Vojnoteh. Glas.*, vol. 64, no. 2, pp. 513–531, 2016.
- [16] S. Raschka and V. Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing Ltd, 2019.
- [17] R. Polikar, "Ensemble learning," in *Ensemble machine learning*, Springer, 2012, pp. 1–34.
- [18] E. Bisong, "Google Colaboratory," in *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, Springer, 2019, pp. 59–64.
- [19] M. Abadi et al., "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.
- [20] V. Labatut and H. Cherifi, "Accuracy measures for the comparison of classifiers," *arXiv Prepr. arXiv1207.3790*, 2012.

# Evaluation of a Model Maximizing the Quality Value of Selected Software Components in a Library

Koffi Kouakou Ive Arsene<sup>1</sup>, Samassi Adama<sup>2</sup>, Kouamé Appoh<sup>3</sup>  
Ecole Doctorale Polytechnique, Institut National Polytechnique (INP-HB)  
Yamoussoukro, Côte d'Ivoire

**Abstract**—Reusable software components are selected from libraries by developers and integrated into existing software systems to improve their quality. In this article, we evaluate a mathematical model based on an approach of optimization of the selection of the software components according to their quality. This is a linear programming model with constraints. It takes into account the quality characteristics of the components based on standard ISO / IEC 9126, the financial cost and the adaptation time. The experience with the ILOG Cplex Studio optimization tool gave satisfactory results.

**Keywords**—Software component quality; reuse; reusable components; reusability; mathematical model; simulation; validation; maintenance effort

## I. INTRODUCTION

The development of modern and complex software systems involves the use of reusable software components. These components selected and integrated into the systems must first be evaluated and tested. This has the consequence of strengthening user confidence by integrating reusable software components into their software systems. The choice of the selection of these components depends first on the needs and functional requirements of customers and users. Thus in the thesis of YAHLALI Mebarka, the researcher argues that quality is not often the essential of the development process [1]. Yet users are often confronted with the vastness of available offer in the various software libraries but also the multiplicity of software that can insure same services. In addition to the functional properties, it is also necessary and important to know how software studied render their services. In our selection process, the non-functional requirements related to the quality of the software component are taken into account. The addition of the evaluation of the quality criteria of the software components to their functional properties in the selection process so conditions the selection of the most appropriate components, better adapted and at reduced cost. For economic reasons, studies have shown that since 2006 software development based on software components exceeds 40% of total software systems developed [2], [3]. Beyond the economic reasons, other researchers have noticed technological and scientific interests [4], [3], [5]. In [6], the researchers affirm that the developers give a particular interest to technologies related to software components and especially when these intervene in the development of complex and large-scale applications. The aim is to improve productivity and speed up the time of marketing for the products developed. In [7], the authors proposed an optimization approach of software components selection based on their

quality. The developed model, score based, evaluates the quality value of the selected software components. This score is calculated from the quality characteristics with values associated in relation to quality services rendered, financial cost and predicted adaptation efforts. This allows us to select the component best suited to the functional needs and quality needs on demand. To determine the quality attributes and the factors that affect the selection and reuse of reusable components, we formulate the following research question: Can the financial cost and maintenance effort affect the selection and reuse of the selected software components? In order to respond to the concerns raised, we propose, in this article an automatic method for selecting reusable software components. This method makes it possible to maximize the selection process by taking into account the quality characteristics of the component, the financial cost and the maintenance effort on the one hand and moreover, assess the quality of the reusable software components selected according to the indicators and quality needs desired by users or companies. This research work is summarized in these following points:

- Identification of problems relating to the selection and reuse of software components.
- Identification of methods from the literature for solving problems related to the reuse of software components.
- Proposal for a new model for automatic selection of reusable software components. This model establishes a link between the financial cost, the maintenance effort and the quality indicators of the software component defined by the ISO / IEC 9126 standard. It is based on linear programming by constraints [7]. It also takes into account the selection of components in a large repository with many characteristics requirements.
- Evaluation of the proposed method. This evaluation consists first of all in defining the decision variables of our model. These variables relate to the financial cost of the component, the adaptation effort and the possibility of choosing the given component from a set of software components. Then we determined the constraints related to the defined variables. At last, we calculated the quality value of the selected components with an optimization tool. This quality depends on the values of the quality characteristics of the software component and on other factors which impact the selection intervening in our model.

- Comparison of results obtained with previous work by researchers.

This paper is organized as follows. Section 2 presents previous work. In Section 3, we propose our mathematical model and its constraints. We have in section 4 the results of our simulations. In section 5, we compared our results with the methods and models proposed in [2] [8] and [9]. We conclude in Section 6.

## II. STATE OF THE ART

Several works have been carried out for the selection of software components taking into account the functional and non-functional properties.

Some authors have based on surveys to identify important attributes that influence the reusability of selected components [10] [11] [12]. They then validated their work by empirical manner.

In [11], the research concerns a survey of 22 cases next to experts and practitioners in equitable manner to understand how the choice of software components from various sources is effected for industrial practice. The objective of this study is to understand how to adapt research solutions to the needs of the industry. This allows to facilitate industry decision-making regarding the reuse of ready to use software components. The results showed that solutions of expert are deterministic and based on optimization approaches. While those of practitioners are non-deterministic and seem better suited to decision-making for component selection in the industry.

In [9], the researchers developed a metric between the *time-effort* and the reusability of the software component. They then validated their work from a survey next to experts. The results showed that the time taken to understand the software component has a correlation with the value of reusability. They concluded that if the time taken to understand the software component is too much, the lower the value of reusability. Therefore this software component will be used less.

Other authors have developed optimization models based on quality criteria and attributes defined by user [8] [2] [13]. These models have been sometimes simulated and validated.

In [2], the research carried on the process of selecting software components taking into account the attributes defining the dependencies between the quality criteria. The method used is based on matrix calculations of vectors and eigenvalues. The goal is to help developers to understand component details based on quality criteria. This allows to facilitate decision making in the selection process.

In [13], the authors developed a model to automate the selection of software components. They defined a common format for the functional and non-functional properties of these components. They then calculated their *satisfaction index*. The objective is to evaluate the level of matching and conformity of the candidate components to be selected and those of the library Version 3.1 of substitute tool was used to measure the degree of satisfaction with a distribution weight.

Those works have made it possible to automate the selection of components with the aim of saving time.

In [8], the researchers developed a so-called reliability metric model that evaluates the quality of software components based on quality attributes of diverse dimensions. This model is based on linear programming by constraint. It allows to select relevant components relating to the solutions of the problems posed. Using computer experimentation with the Cplex solver version 12.2, those authors have found solutions to the problems posed. It is about selecting software components in libraries of large size with large number of requirements in a short time.

In [14], the authors developed an optimization algorithm for selection in large functionality models with multiple objectives. This algorithm called IVEA-II. It allows to do automatic researches to balance different objectives and improve existing methods.

In other works again, the researchers validated their model by the combination of led surveys and optimization algorithm.

Research in [15] has focused on an approach of identification high-quality software components from several sets. The authors modeled the problem of research and identification of reusable software components through multi-objective optimization. They defined three objective functions. The first objectives correspond to two metrics making it possible to simultaneously maximize cohesion based on the frequency of use of the components and the cohesion linked to the semantic relationship of these components. The last objective is a relation which makes it possible to minimize the coupling on the change about the changes historic from the evolutions of the underlying software system. By successively applying the different metrics above and the NSGA-III search algorithm which allows the grouping of different software components, the authors obtained better quality for different software components measured in terms of reusability characteristics. The results made it possible to identify highly cohesive and least coupled software components.

In [16], researchers worked on the efficient recovery of components from a large repository. They have shown that this technique is a new challenge in the process of selecting reusable components. This retrieval model is based on two (2) steps. The first is to select components that correspond the functional requirements. The second is an algorithm that allows to weight the quality requirements of software components and recommend to developers the highly weighted. The objective of those works is to involve the users in the choice of the quality attributes of the components during the construction of the project but also to organize into a hierarchy the quality attributes according to its needs.

In the next section, we present our model. This model is an optimization approach to select of software components according to their quality, based on linear programming with constraints. It makes it possible to establish an extension, a generalization of the model for evaluating the quality of the factors which influence the selection and reuse of the software components treated in [8]. We associated the maintenance effort with the model he developed.

### III. PRESENTATION OF THE DEVELOPED MODEL

Our job is to select a component and evaluate its quality. The mathematical model that we have developed makes it possible to maximize the value of the quality of the component according to the characteristics, the financial cost and the adaptation time of the component. Our model has two steps. The first is leaned on the prediction of the maintenance effort defined in [7]. According to the literature review, it is rare to find a component that can satisfy perfectly the requirements of users. In other words, the selected components may not fully the quality and service requirements expressed by the customer (see Fig. 1).

Fig. 1 designates N components, some features of which make the services perfect but others can do them partially. We will estimate the time required to improve the functionalities which partially render service at the level of defective components. We then applied the cosmic methods developed in [17], [18] and [19]. This allows us to determine the estimate of the maintenance and adaptation effort in accordance with equations (1) and (2).

$$\text{functional size of each processe } (i) = \sum_j^p \text{ functional size of process}(i, j) \quad (1)$$

$$\forall i \in Sc \text{ et } 1 \leq j \leq P \quad (2)$$

Where

Sc : set of the available components

Component (i) denotes the i<sup>th</sup> component of the set Sc

Process (i, j) denotes the process j of the i<sup>th</sup> component

Then, when we apply the estimate of the adaptation effort developed according to [17]. We obtain equation (3)

$$\text{Estimated Development effort} = \text{Component Size} * \text{Unit Cost} \pm \text{Predictive intervala} \quad (3)$$

This phase makes it possible to determine the adaptation time interval of each component to be predicted. This method then evaluates a financial cost and an adaptation time.

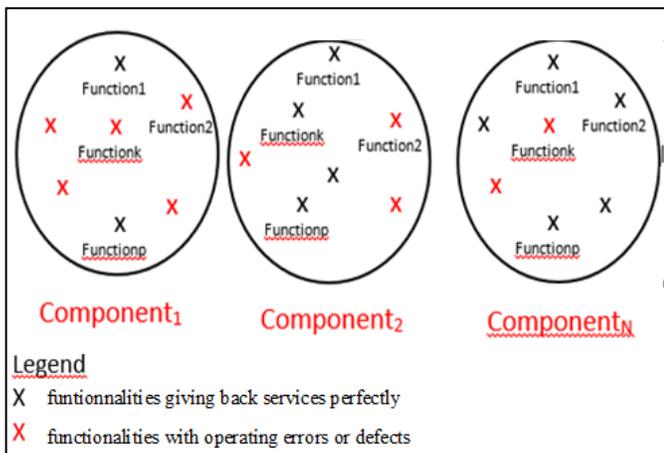


Fig. 1. Set of Selected Components with Some Malfunctions.

The second step allows us to define the objective function. This model is a score that calculates and evaluates the quality of software components based on quality characteristics, time predicted in equation (3), and financial cost. The following model (4) represents the objective function.

$$S_i = \sum_{h \in A} w_h q_{hi} x_i - [ac_i + (1 - a)t_i] x_i \quad (4)$$

and  $\forall i \in Sc$

Where

A: set of software quality characteristics;

Sc: set of available component;

$q_{hi}$  : the standard level of the quality attribute ;

$h \in A$  for component  $i$  ;

$C_i$ : Standardized cost of maintenance of the component  $i$ ;

$t_i$ : Standardized adaptation and maintenance time of the component  $i$ ;

$\alpha$  : Coefficient of adaptation

For the resolution of this model, we define the following constraints and decision variables.

#### Decision Variables

- A Boolean variable  $x_i$  for selection of the component  $i$  corresponding to 1 if the component is selected, otherwise 0;
- An real variable  $C_i$  designating the cost of the selected component  $i$ ;
- An real variable  $t_i$  designating the effort of adaptation of the selected component  $i$ .

#### Constraints

$$0 \leq a \leq 1$$

$$t_i \in \llbracket t_{min}; T_{max} \rrbracket$$

$$t_i = \frac{t_{i_{rel}}}{T_{max}} \text{ and } 0 \leq t_i \leq 1$$

$$T_{max} = 15 \text{ days} = 1.296 * 10^3 \text{ s}$$

$$C_i = \frac{C_{i_{rel}}}{C_{max}} \text{ et } 0 \leq C_i \leq 1$$

$$Q_{max} = 5$$

$$\sum_{h \in A} w_h = 1 \quad (5)$$

In [7], we defined a metric which evaluates the quality of the characteristics of the given software component. It designates the ability of the component to fulfill the criterion linked to functionality. This metric was associated with an ordinal variable of modalities taking the values in the set B.

$$B = \{Bad, Insufficient, Average, Good, Excellent\} \quad (6)$$

These modalities that we defined in (6), are associated respectively with the numerical values: 1; 2; 3; 4 and 5.

This allows us to determine the maximum value  $Q_{max} = 5$ ;

To optimize the cost and maintenance time parameters, we maximize the objective function. By considering equation (4) with constraints, we obtain the following equation (7)

After presenting our model we move to the validation and simulation phase.

$$\left\{ \begin{array}{l} \max(\sum_{h \in A} W_h q_{hi} x_i - [aC_i + (1-a)t_i]x_i) \forall i \in Sc \\ 0 \leq a \leq 1 \\ t_i = \frac{t_{i\_rel}}{T_{max}} \text{ and } 0 \leq t_{i\_rel} \leq T_{max} \\ C_i = \frac{C_{i\_rel}}{C_{max}} \text{ and } 0 \leq C_{i\_rel} \leq C_{max} \\ q_{hi} = \frac{q_{hi\_rel}}{Q_{max}} \text{ and } 0 \leq q_{hi\_rel} \leq Q_{max} \\ x = 1 \text{ if component selected else } x = 0 \end{array} \right. \quad (7)$$

Where

A: set of software quality characteristics;

Sc: set of available component;

$q_{hi}$  : the standard level of the quality attribute;

$h \in A$  for component  $i$ ;

$C_i$ : Standardized cost of maintenance of the component  $i$

$C_{i\_rel}$ : relative cost generated by component  $i$ ;

$C_{max}$ : maximum cost achieved by one of the selected components;

$t_i$ : Standardized adaptation and maintenance time of the component  $i$ ;

$t_{i\_rel}$ : Relative time, generated by component  $i$ ;

$T_{max}$ : is the maximum time achieved by one of the selected components;

$\alpha$  : Coefficient of adaptation.

#### IV. VALIDATION PHASE

Current software systems have become increasingly complex and large size [20], [21]. They can integrate a large number of software components during their construction. To benefit from a quality system, the best components of library must be selected. Indeed in [21], the authors notice the need for the management and use of quality software in order to avoid computer weakness in software systems that have become complex and large size. They developed a framework to assess the quality of software and to support decision-making in the engineering of large-scale critical systems.

Also those components must adapt to the needs and quality requirements of users. Our aim is to determine quality attributes that affect the reuse of software components. Our selection approach is described by the SlectCompo algorithm (See the algorithm Fig. 2).

#### A. Presentation of the Algorithm

Our algorithm allows to select from a set of available components (Cd), the optimized and selected component (Cos). The different steps of this algorithm are described in [7]. In order to show the practical utility of our model, we performed an experiment on a set of components with non-functional requirements.

SelectCompo Algorithm

1. *Input*: Set of available components (Cd)
2. *Output*: Optimized component and selected (Cos)
3. Begin
4. *While* (needs and requirements expressed in Cd) do
5. *For*  $i = 1$  to Component (Cd) do
6. *Select* (the component  $C_i$ )
7. Put in the list of selected components (Cs)
8. **Endfor**
9. **EndWhile**
10. *If* ((conditions Characterisks Filled) and (cost and relative time in intervals required) **then**
11. **For**  $i = 1$  to Component (Cs) **do**
12. *evaluate* (the quality value of the selected components)
13. **End if**
13. **If** (SatisfactionQuality) **then**
14. *Optimize* (the factors of cost and time of adaptation)
15. *Select* (the component (Cos))
16. *else* choose another component in the set Cs
17. **End if**
18. **End**

Fig. 2. Pseudo Code of SelectCompo.

#### B. Definitions of Characteristic Weights

Depending on the needs of the user, we define the importance of each characteristic in relation to the others. This allows us to determine the different weights of features defining quality criteria such as reliability, usability, security and maintainability using the Hierarchical Process Analysis (AHP) method. First we considered the binary comparison matrix of the different characteristics of the expert-defined component in Table I<sup>1</sup>.

TABLE I. BINARY COMPARISON MATRIX OF THE FEATURES OF THE SOFTWARE COMPONENTS

Binary Comparison Matrix	reliability	usability	security	Maintainability
reliability	1	3	2	0,25
usability	0,3333	1	0,3333	1
security	0,5	3	1	0,2
maintainability	4	1	5	1

<sup>1</sup><https://www.uqtr.ca/~gelinare/Logistique/ahp.doc>

Applying the AHP method, we obtain the weights of the different features which are quality attributes of the software component in Table II as follows:

TABLE II. WEIGHT OF CHARACTERISTICS

weights of	characteristics			
characteristics	reliability	usability	security	maintainability
weights	0,2221	0,2221	0,1656	0,4547

C. Selection of the Best Component

According to the AHP method, the sum total of the weight of the characteristics is equal to 1. We admit that we want to make the selection in a set of p components available in a library noted Cd. The user then defines his functional requirements and his non-functional requirements which are related to the quality of operation of the software components. The selection of the best component is done in two stages.

In the first part, the candidate components are selected based on the functional requirements. To do this, a matching between the requirements expressed by the client and the functionality of the library components is made.

We obtain a list (Cs) of k selected candidate components and verifying the defined condition (with  $k \leq p$ )(line 7 of the algorithm).

If we take for example,  $k = 5$  we define the following set:

$$Cs = \{ \text{component1, component2, component3, component4, component5} \}$$

The next step is to evaluate the quality of the components of the list (Cs) by binary comparison of their characteristics see Table I. For the realization of our experimentation, we considered the software components of the componentSource<sup>2</sup> platform to define the values of the different parameters financial cost and maintenance effort. This software component market has components whose dollar financial costs (Us dollars \$) are values in the interval [195.02 ; 3,000.00].Maintenance efforts must be made within a maximum of fifteen (15) days.

Normalized values are obtained by doing the ratio of the relative value and the maximum value. For any available component i of the library, we have:

$$t_i = \frac{t_{i\_rel}}{T_{max}} \text{ et } C_i = \frac{C_{i\_rel}}{C_{max}} \tag{8}$$

Where

$t_i$ : normalized time of the component i ;

$t_{i\_rel}$ : Relative time generated by component i;

$T_{max}$  : the maximum time achieved by one of the selected components;

$C_i$ : Normalized cost of the component i

$C_{i\_rel}$ : relative cost generated by component i;

<sup>2</sup><https://www.componentsource.com/fr/>

Cmax: maximum cost achieved by one of the selected components;

Relative and standardized financial costs then relative and standardized maintenance efforts of the components are grouped into Table III.

TABLE III. RELATIVE STANDARDIZED COST AND MAINTENANCE EFFORTS

	Componen nt1	Componen nt2	Componen nt3	Componen nt4	Componen nt5
Relative Costs in \$	975,1	343,18	1000	2240,1	2032,15
Relative Maintenance effort within days	12	5	9	10	2
Standardized cost	0,42396	0,14921	0,43478	0,97396	0,88354
Standardized maintenance efforts	0,8	0,33333	0,6	0,6667	0,13333

We search to maximize the quality values of the components by our model (see equation (8)). To do the experimentation, we will use an optimization tool.

D. Utilization a Solver

We move on to the practical phase of evaluating the model with an optimization tool. This type of software uses the combination of modeling capabilities and the power of Cplex.

Optimizers to quickly and easily solve important and difficult problems for users. This step gives two (2) possibilities of solution for the selection of software components.

- Using the CPLEX studio 12.8.0 solver from IBM and its OPL language, the solution proposed is:  $X=[0, 0, 0, 0, 1]$ ;  $C=[0.42396, 0.14921, 0.43478, 0.97396, 0.88354]$ ;  $T=[0.8, 0.33333, 0.6, 0.6667, 0.13333]$ . This means that component 5 is the best result and retained. It is therefore the best choice among this set of Cs elements. This means that component 5 is the best result and kept. So this is the best choice among these elements of the set Cs. We retain then  $Cos = \{ \text{component5} \}$ .
- Otherwise, we take back the selection in the list (Cs).

V. INTERPRETATION AND DISCUSSION

In the first part, we interpret our model and the results of our experimentation. Then in the second part we compare our results to the works of other authors cited above in this document.

A. Interpretation

We want our model to assess the quality of software components in a large repository and with a large number of requirements. It takes into account the quality characteristics

of ISO / IEC 9126, on the one hand and on the other hand factors of financial cost and maintenance effort.

In the case where the parameter  $\alpha$  is zero ( $\alpha=0$ ), the term containing the financial cost is canceled. The equation (4) becomes:

$$S_i = \sum_{h \in A} w_h q_{hi} x_i - [(1 - \alpha)t_i]x_i$$

and  $\forall i \in Sc$  (9)

We will be in the presence of a library of open source software components. Then the selection will be done in a set of open source components.

In the opposite case, if  $\alpha$  is 1 ( $\alpha = 1$ ), we have some ready-to-use components, that is the "cost" components. The equation (4) becomes:

$$S_i = \sum_{h \in A} w_h q_{hi} x_i - [ac_i]x_i$$

and  $\forall i \in Sc$  (10)

In this case, the selection depends only on quality characteristics and cost. This corresponds to the model developed in [8].

In short, our developed approach corresponds to a generalization of the selection of components in a large repository with many requirements taking into account the quality characteristics, cost factors and the maintenance effort. We present the results of our simulations with the graphical representations Fig. 3 and Fig. 4. Those different forms of graphs will allow us to better interpret and explain our results. Then we will make a comparative study of those results with previous works.

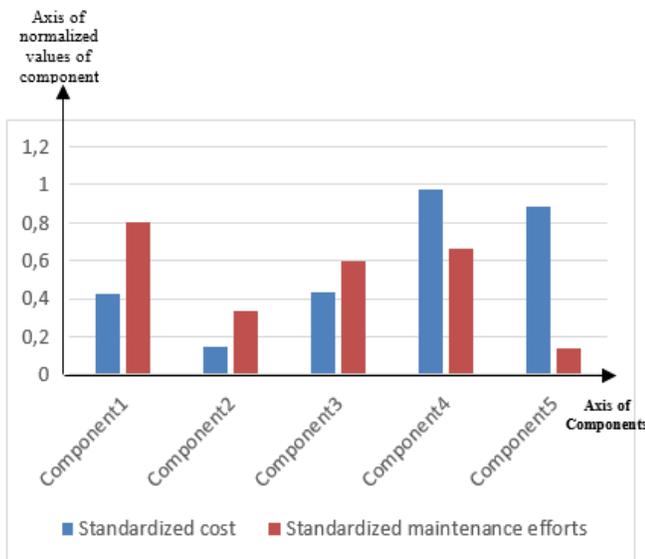


Fig. 3. Representation of Standardized Costs and Standardized Maintenance Efforts.

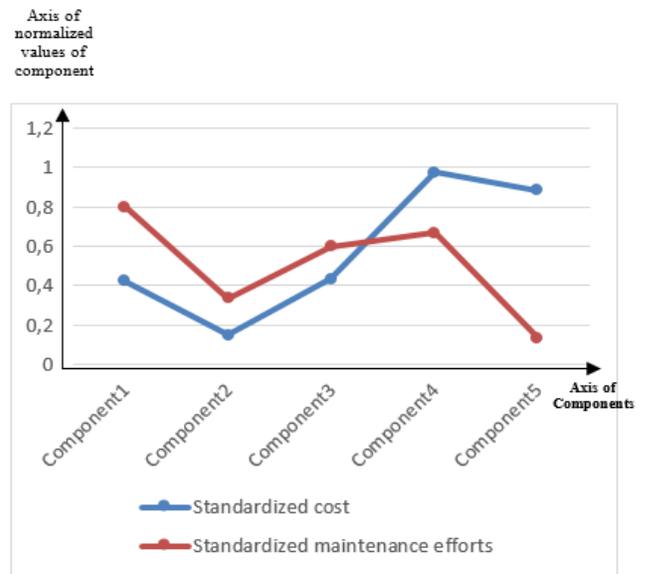


Fig. 4. Representation of Standardized Costs and Standardized Maintenance Efforts.

Our results show that the software component admitting a high maintenance effort, is not selected whatever the financial cost. Our study also argues that the maintenance effort must be of low value. So the components selected are those admitting a low maintenance effort like software component 5.

### B. Discussion

In this section, we compare our results with the work of the authors [2], [8] and [9]. In [2], the authors realized the selection of software components based on quality criteria. The method used is the analytical network process. This study does not take into account the particular qualities of the features of the component. The result of the experimentation is based on an overall judgment of the quality of the components by binary comparison between them. This technique uses manual calculations, very complex and very tedious. This results in huge waste of time to select the relevant components.

The effectiveness of works in [8] lies in the consideration of quality characteristics of ISO / IEC 9126 and the factor defining the financial cost in the selection process. Using the binary comparison technique of quality attribute, the authors added important elements in decision making to select software components. This technique made it possible to assess the functional and non-functional qualities of software components. Also, experimenting with the Cplex solver to automatically solve the developed model accelerates the results. Their works are saving of time a greatly compared to the work of the authors of [2].

Our work shows two important results. On the one hand, we obtain a saving time of less than a minute to automatically select the components with the same quality values as in [8]. On the other hand we have shown that the choice of software component selection depends on the maintenance effort and the financial cost.

So our results are in concordance with the work in [9]. Therefore, the maintenance effort parameter that we have associated in our model is an important factor and impacts decision-making for the selection of relevant software components in a library or in a component market. In addition, our results are in agreement with those of [9] where the authors have shown that if the time and effort required to understand a component increases, the reusability decreases.

## VI. CONCLUSION AND PERSPECTIVES

We have developed a model that allows to optimize the selection of software components based on their quality. Our approach corresponds to the generalization of the selection of components in a large repository with many requirements taking into account the quality characteristics, cost factors and the maintenance effort. In this paper we simulated our model. Our results show that the reusability of components depends on several factors. There are the quality characteristics, the financial cost and the maintenance effort. The lower the maintenance effort, the more the component is reused.

Our model shows that the maintenance effort is an important factor that influences decision-making in the process of selecting software components from component markets.

In the future work, we would like to develop a model that will allow us to assess the dependency ratio of the different factors having a correlation in the selection process of the software components of the libraries. We will also work on improving our selection process and our algorithm which support it to increase the quality of the selected components.

## REFERENCES

- [1] Yahlali Mebarka "Assemblage d'une application à base de composants : approche de calcul de qualité d'une composition", these, Université des sciences et de la technologie d'Oran Mohamed Boudiaf, 2015.
- [2] Shah Nazir, Sajid Anwar, Sher Afzal Khan, Sara Shahzad, Muhammad Ali, Rohul Amin, Muhammad Nawaz, Pavlos Lazaridis, John Cosmas, Software component selection based on quality criteria using analytic network process., Abstract and Applied Analysis Volume 2014.
- [3] A. Rawashdeh and B. Matalkah, "A new software quality model for evaluating COTS components," Journal of Computer Science, vol. 2, pp. 373-381, 2006.
- [4] Pfarr, T. and J.E. Reis., "The integration of COTS/GOTS within NASA's HST command and control system", 2002.
- [5] José P. Miguel , David Mauricio and Glen Rodríguez " A Review of Software Quality Models for the Evaluation of Software Products" International Journal of Software Engineering & Applications (IJSEA), Vol.5, No.6, November 2014.
- [6] Shabnam Gholamshahi Seyed Mohammad Hossein Hasheminejad, « Software component identification and selection: A research review », 31 October 2018, <https://doi.org/10.1002/spe.2656>.
- [7] Koffi Kouakou Ive Arsene, Samassi Adama, Kimou KouadioKonan Marcellin Brou" Proposal of Automatic Methods for the Reuse of Software Components in a Library ", International Journal of Advanced Computer Science and Applications 10(2) · January 2019, DOI: 10.14569/IJACSA.2019.0100208.
- [8] Pande, CJ Garcia, D Pant, "Optimal Component Selection for Component Based Software Development using Pliability Metric", ACM SIGSOFT Software Engineering Notes, January 2013.
- [9] Neha Sadana, Surender Dhaiya, Manjot Singh Ahuja, "A Metric for Assessing Reusability of Software Components", International Journal of Computer Application, R S. Publication, Vol. 1, Issue 4, pp. 98-108, Feb. 2014.
- [10] N. Sadana, S. Dhaiya, M.S. Ahuja, "An empirical review on factors affecting reusability of programs in software engineering," International Journal of Computing and Corporate Research, vol.3, July 2013.
- [11] Petersen K., Badampudi D., Shah S., Wnuk K., Gorschek T., Papatheocharous E., Axelsson J., Sentilles S., Crnkovic I. and Cicchetti A.: Choosing Component Origins for Software Intensive Systems: In-house, COTS, OSS or Outsourcing?--A Case Survey. IEEE Transactions on Software Engineering. Vol. PP, No. 99, March, pp.1-1, 2017.
- [12] N. Sadana, S. Dhaiya, MS Ahuja, "A Metric for Assessing Reusability of Software Components", International Journal of Computer Application, Issue 4, Volume 1, February 2014.
- [13] Bart George, R. Fleurquin, S. Sadou, H. Sahraoui « Un mécanisme de sélection de composants logiciels », juillet 2010.
- [14] Lian, X., Zhang, L., Jiang, J. and Goss, W., An approach for optimized feature selection in large-scale software product lines. Journal of System s and Software., pp. 636- 651, 2017.
- [15] Amit Rathee, Jitender Kumar Chhabra, "A multi-objective search based approach to identify reusable software components", Journal of Computer Languages 52, p.26-43 ,3 January 2019 <https://doi.org/10.1016/j.cola.2019.01.006>.
- [16] Sumit Sharma, Upinder Kaur, Pawanpreet Kaur « Component Recommender System Based on Collaborative Approach in Incremental Development », World Journal of Technology, Engineering and Research, Volume 2, Issue 1, 2017, p 51-59.
- [17] Cosmic : mesure de la taille fonctionnelle avec la méthode, <https://info.uqam.ca/midi-confs/2017-02-22-cosmic.pdf>.
- [18] Sylvie Trudel, mesure de la taille fonctionnelle avec la méthode cosmic (iso 19761): recherches récentes et applications industrielles », conférence du latece 2012.
- [19] Christof Ebert and Hassan Soubra, « functional size estimation technologies for software maintenance » IEEE software technology, November/December, 2014.
- [20] Rakesh Garg, R. K. Sharma, Kapil Sharma, Ramesh Kumar, « Identification, selection and evaluation of COTS selection criteria using Fuzzy Set Theory », International Journal of Computing and ICT Research, Vol. 9, Issue 2, December 2015, P 21-36.
- [21] Gabriella Carrozza, Roberto Pietrantuono, Stefano Russo, "A Software Quality Framework for Large-Scale Mission-Critical Systems Engineering", Information and Software Technology, 2018, doi: 10.1016/j.infsof.2018.05.009.

# Feature Selection and Performance Improvement of Malware Detection System using Cuckoo Search Optimization and Rough Sets

Ravi Kiran Varma P<sup>1</sup>, PLN Raju<sup>2</sup>, K V Subba Raju<sup>3</sup>, Akhila Kalidindi<sup>4</sup>  
MVGR College of Engineering  
Vizianagaram, AP, India

**Abstract**—The proliferation of malware is a severe threat to host and network-based systems. Design and evaluation of efficient malware detection methods is the need of the hour. Windows Portable Executable (PE) files are a primary source of windows based malware. Static malware detection involves an analysis of several PE header file features and can be done with the help of machine learning tools. In the design of efficient machine learning models for malware detection, feature reduction plays a crucial role. Rough set dependency degree is a proven tool for feature reduction. However, quick reduct using rough sets is an NP-hard problem. This paper proposes a hybrid Rough Set Feature Selection using Cuckoo Search Optimization, RSFSCSO, in finding the best collection of reduced features for malware detection. Random forest classifier is used to evaluate the proposed algorithm; the analysis of results proves that the proposed method is highly efficient.

**Keywords**—Cuckoo search; rough sets; feature optimization; malware analysis; malware detection; feature reduction; clamp dataset

## I. INTRODUCTION

From the past, many years' malware has become a significant security threat for systems and networks. Malware is defined as software or malicious code injected into a target system or network to make the system work abnormally [1]. Virus, Trojans, backdoors, worms, rootkits, spyware, adware, etc. are several forms of malware. In general, any malware is commonly termed as a virus, which was first framed by Fred Cohen [2] in the year 1983. Every malware is designed with a common intention of destroying or doing some illegitimate access on the system or gain access or retrieve some sensitive information from the system. The type of malware and the anti-malware or malware detection systems depends on the hardware/software platforms and the operating system. The main goal of attackers is to infect or morph malware to evade from the malware detectors.

The increase in the volume of the datasets has resulted in a decrease in performance and increased the complexity of the classification model, thereby resulting in need of feature reduction (FR). Feature reduction was defined by [3] as the "subset of features for enhancing the accuracy and at the same time decreasing the complexity of the classification model." A reduced subset is proved to be a useful subset if the number of features is reduced without the decrease of accuracy. In general, any FR techniques follow mainly two steps [4]. In the

initial step, the candidate subset is selected using a search technique, and in the final step, the selected subsets are evaluated using an objective function.

The existing feature selection algorithms are classified into two approaches based on the Objective function used. They are wrapper-based [5] and filter-based [6]. Filter-based techniques use statistical methods like the dependency degree or information measurement to evaluate the candidate subset and do not depend on the classification algorithm. In contrast, wrapper-based algorithms depend on the classification algorithm to evaluate the selected features. Hall & Lloyd, [7] applied both wrapper-based methods and filter-based methods and proved that filter based technique is faster and utilizes less CPU utilization. But the main drawback is that they did not provide any accuracy after implementing the techniques.

Shabtai et al. [8] extracted features from the Linux OS and compared a total of 10 datasets. The authors implemented a feature reduction method. He compared selection methods such as chi-square, fishers score, and information gain, and at last, they proved that information gain would produce better results and obtained an accuracy of 96.8%. It used a filter-based technique, which is the main advantage of that project. The work proposed by [9] makes use of Ant Colony Optimization and Rough Sets as a filter-based feature reduction for web phishing detection and achieved good results.

Rough sets (RS) is a mathematical approach that was first discovered by [10] in the year 1982. Because of its unique method, RS has become the most widely used technique in many fields of information technology [11]. The working strategy of RS is that it first generates all possible subsets, and from among those subsets, it selects the one with the minimum number of features and, at the same time, having a maximum dependency.

The author of [12] and [13] has included in their work the advantages and primary reasons for RS being used extensively. The authors of [14] have developed a malware dataset whose features are extracted from the API call sequences. The author used RS as the feature selection algorithm along with SVM as the classifier and has achieved more significant results. Many researchers have combined meta-heuristic algorithms along with RS to improve the accuracy and to obtain an effectively reduced subset.

Optimization of a dataset is also very essential to reduce the complexity of the classification model. Optimization algorithms are classified mainly into Traditional and heuristic methods. Most of the malware detection systems use heuristic methods. Meta-heuristic algorithms are divided again into different categories. One that is inspired by natural behaviors such as Ant colony optimization [15], Bee colony optimization [16] that is inspired by the natural behavior of bees, Cuckoo search (CS) [17] technique which is inspired from the natural behavior of Cuckoo bird. Second are evolutionary algorithms and, finally, logical search algorithms. A taxonomy of optimization algorithms is shown in Fig. 1.

Thanushkodi and Suguna [18] have combined rough sets along with Bee-Colony Optimization (BCO) for analyzing a medical dataset. The author also applied various combinations with rough sets and proved in the results that BCO, along with RS, work best for his dataset. Though BCO, along with RS would produce effective results than others, the only weakness is that it consumes more time in finding the reductant subset. Liang [19] have employed a Genetic algorithm, along with RS in the marketing application. The author concluded that this hybrid approach would work effectively for clustering datasets but failed to provide the results after the application of the algorithm. Rough sets can also be used along with other meta-heuristic algorithms such as Ant Colony Optimization, Cuckoo Search, Ant bee colony optimization, etc.

Cuckoo Search (CS) is one of the optimization technique which is extensively used by the present researchers. This Optimization technique was developed by Yang and Deb [20] in the year 2009, which is based on the reproduction strategy of a cuckoo bird. The first cuckoo search algorithm was modified by [17], works effectively for non-linear problems; they used the cuckoo search algorithm along with rough sets. The author

used different datasets and different optimization techniques to evaluate the datasets and provided the results. But the major drawback is that it is a wrapper based method and consumes longer run times. The author proved through his results that the modified algorithm, when used with SVM as the classifier, resulted in an average accuracy of 93.94%. Kumar and Shampa [21] used cuckoo search in the multi-reliable objective function and concluded that CS is effective when compared with that of other heuristic algorithms.

Ajit et al. [22] have used the clAMP dataset to create an integrated feature set. The author tried a total of 5 classifiers and compared against each other and obtained a result of 98% by using the J48 decision tree. However, not many features are reduced. Mouhammad & Samail [23] have also used the clAMP dataset and made a comparison of the results of all the classifiers. Among all the classifiers, this author has proved that the decision tree would give the best results. The author, at the end of his work, recommended that the extraction can be done much more effectively by using a genetic algorithm.

In this work the features are extracted from the PE header, which is present in all Windows executable files. The PE header has four sections embedded within it [24].

- The DOS header.
- PE file header or The Common object file format (COFF) header (also known as the File Header).
- The optional header.
- The section header.

The following are the features that can be extracted from different headers in the PE header.

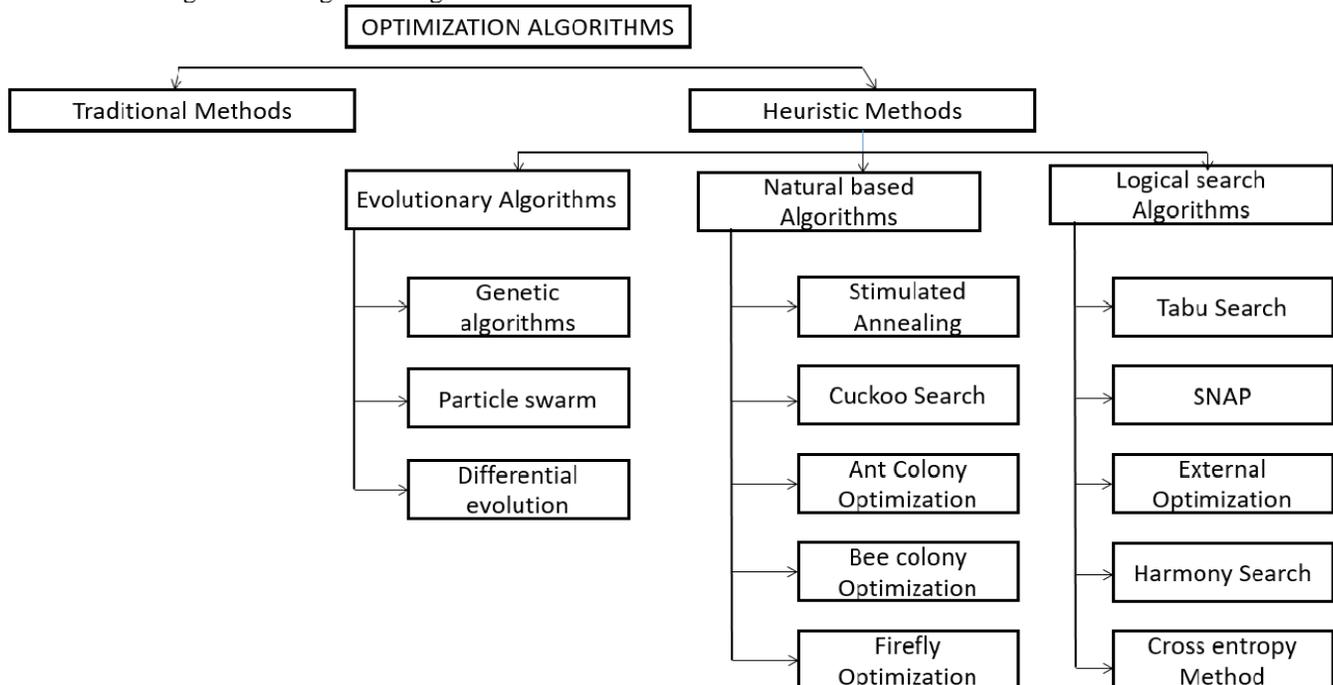


Fig. 1. A Taxonomy of Optimization Algorithms.

### A. DOS Header Features

Table I describes the features that can be identified with the help of the DOS header. The feature e\_magic is a fundamental feature that generally starts with the hex value 4D5A, which means 'MZ' [25] at the beginning and indicates that the file is an MS-Dos executable file.

### B. PE Header Features

This header is present in front of the object file or immediately after the signature of the image file [26]. Table II describes the features which can be extracted from the PE file header. The principal analysis lies in the feature TimeDateStamp. The people responsible for making a malware file first tries to change this feature.

### C. Optional Header Features

This header field describes the logical structure of the PE header. Every image file would have this optional header and would provide with the loader information. This header is not optional in case of image files whereas it is optional in case of other files. This header is again subdivided into the version and size attributes. The optional header contains a version attribute that is further subdivided. They are briefly described in Table III. The optional header contains a size attribute that is divided into the features as described in Table IV [27]. The optional header includes a location attribute that is subdivided into the features that are described in Table V.

### D. Section Header

The number of entries in the section table is predefined in the PE header by the Number of sections feature [28]. The features that can be extracted from a section header are mentioned in Table VI.

TABLE I. FEATURES EXTRACTED FROM DOS HEADER

Feature	Description	Type
e_magic	Magic number.	Numeric
e_cblp	Bytes on the last page of file	Numeric
e_cp	Number of pages in the file	Numeric
e_cparhdr	Header size in paragraphs	Numeric
E_maxalloc	Maximum extra number of paragraphs needed	Numeric
E_sp	Initial sp value	Numeric
E_lfanew	File address of new exe header	Numeric
e_csum	Checksum value	Numeric
e_minalloc	Minimum extra number of paragraphs needed	Numeric

TABLE II. FEATURES THAT CAN BE EXTRACTED FROM THE PE FILE HEADER

Feature	Description	Type
TimeDateStamp	Date and time of file creation	Numeric
Numberofsections	Size of the section table. Windows limits this size to 96	Numeric
Symbol attribute	Define location and size of COFF header	Numeric
DLL Characteristics	This field contains a combination of 16 different features	Numeric

TABLE III. OPTIONAL HEADER VERSION ATTRIBUTES FEATURES

Feature	Description	Type
MajorLinkVersion	Linkers major version number	Numeric
MajorLinkVersion	Linkers minor version number	Numeric
MajorOperatingSystemVersion	Major version number of OS	Numeric
MajorOperatingSystemVersion	Minor version number of OS	Numeric
Majorsubsystem	Major version number of Subsystem	Numeric
Majorsubsystem	Minor version number of Subsystem	Numeric

TABLE IV. OPTIONAL HEADER SIZE ATTRIBUTES FEATURES

Feature	Description	Type
SizeofCode	Size of code sections. If multiple sections are present then sum of all those sections	Numeric
SizeofInitializedData	Size of initialized data or if multiple data sections are present then the sum of all those sections	Numeric
SizeofuninitializedData	Size of uninitialized data or if multiple data sections are present then the sum of all those sections	Numeric
SizeofImage	Size of the image	Numeric
Sizeofheader	Size of all section headers	Numeric
SizeofStackreserve	Number of bytes reserved by stack	Numeric
SizeofStackcommit	Number of bytes required to commit the stack	Numeric
SizeofHeapreserve	Number of bytes reserved for heap	Numeric
SizeofHeapcommit	Number of bytes required to commit the heap	Numeric
SizeofOptionalHeader	Indicates the size of the optional header. This size is not fixed.	Numeric

TABLE V. SECTION HEADER FEATURES

Feature	Description	Type
Raw size	Size of section when stored on disk	Numeric
Virtual size	Size of section when stored on memory	Numeric
Virtual address	Address of virtual memory	Numeric
Physical address	Address of physical memory	Numeric
Entropy	This value is not present in PE file but calculated by an external header	Numeric

TABLE VI. FEATURES EXTRACTED FROM THE OPTIONAL HEADER

Feature	Description	Type
Section Alignment	Alignment of section loaded into memory	Numeric
File Alignment	Alignment of raw data section in the image file	Numeric
Baseofcode	A pointer at the beginning of code section	Numeric
BaseofData	Pointer at the beginning of data section	Numeric
Image Base	Address of first byte when image is loaded in the memory	Numeric

This work aims at reducing the feature set to optimize the Malware detection system. Through feature reduction we can automatically minimize the classification model complexity and lower the computational complexity.

## II. ROUGH SET FEATURE SIGNIFICANCE AND CUCKOO SEARCH OPTIMIZATION (RSFSCSO)

### A. Rough Set Theory (RST)

RST is an effective mathematical approach for selecting the best candidate feature subset [29]. RST has its advantages and disadvantages. RST is a pair of upper approximation and lower-approximation. These approximations can be calculated by using the equation.

$$\hat{a}p = \{p[p]_a \subseteq p\} \quad (1)$$

$$\hat{a}p = \{p[p]_a \cap p \neq \emptyset\} \quad (2)$$

The certainty of samples and the uncertainty of samples is defined by the positive region and negative region. The sum of both regions is defined by the bounded region.

$$RP(D) = \dot{U}_{a \in u \uparrow D} \hat{a}p \quad (3)$$

$$RN(D) = \dot{U} - U_{a \in u} \hat{a}p \quad (4)$$

$$RB(D) = U_{a \in u} \hat{a}p - \dot{U}_{a \in u \uparrow D} \hat{a}p \quad (5)$$

Before performing any feature selection algorithm, it is very important to calculate how much an attribute depends on another attribute. Dependency degree helps in calculating the amount of dependency of an attribute on another. In this work, the dependency degree plays a vital role in the calculation of the objective function.

$$\mathcal{D}d = \gamma(d) = \frac{|RP(D)|}{|\dot{U}|} \quad (6)$$

### B. Cuckoo Search Optimization Technique

Cuckoos have a different style of breeding behavior, as shown in Fig. 2. Cuckoo search is inspired by the natural response of the cuckoo bird. Cuckoo lay their eggs in host nests of other birds and depend on the host birds for hosting their eggs. Sometimes the host bird discovers the cuckoo bird eggs and either abandon them or change their nests to a new place. But cuckoos have the talent of producing eggs of the same color and shape as that of the host eggs. In general, cuckoo bird eggs hatch first and have more probability of getting more food. Levy's flight [4] is one mechanism that is used by the cuckoo bird to effectively search for their food. This levy's flight depends upon the levy's distribution function which is calculated by using the equation.

$$levy \sim U = v^{-\lambda} \quad (7)$$

Where  $\lambda$  ranges from  $0 < \lambda < 3$  and  $v$  are the step size.

In cuckoo search initially, the number of population nests and the iterations are initialized. After each iteration, the hybrid search is used to update the population of nests. It updates the population nest based on the conditions that the lowest quality nests are updated randomly based on the global search, and the remaining nests are updated locally using levy's flight. The new nest is updated using equation 8.

$$a^{t+1} = a^t + \alpha \odot levy(\lambda) \quad (8)$$

Where  $a^{t+1}$  is the updated solution

$\alpha$  is the step size which is equal to 1 in most of the cases.

$levy(\lambda)$  is given by the equation (9)

Binary cuckoo search [30] is a modified version of the cuckoo search which uses binary vectors in which 1 represents the selected features, and 0 represents the remaining. In this search, the nests represent the solution, and each egg represents a feature. The search initially starts by generating an initial population randomly, and then after each iteration, it updates the worst nests using levy's flight. To develop a binary vector, the equations (9) and (10) are used,

$$v(a_{x,y}^t) = \frac{1}{1 + e^{-a_{x,y}^t}} \quad (9)$$

$$a_{(x,y)}^{t+1} = \begin{cases} 1, & v(a_{x,y}^t) > \tau \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

Where  $\tau$  belongs to  $[0,1]$  and  $a_{(x,y)}^{t+1}$  represents new eggs.

Every cuckoo search algorithm makes use of a fitness function, which helps in evaluating the fitness of cuckoo as well as nests.

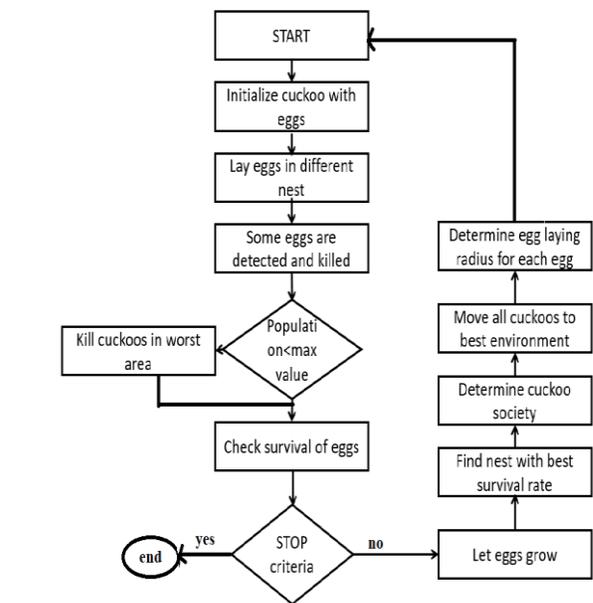


Fig. 2. Flow Chart of Cuckoo Breeding Behaviour.

The fitness function corresponding to this work is represented in equation (11)

$$F(R) = \frac{\mathcal{D}d}{|L|} \quad (11)$$

Where  $|L|$  represents the cardinality of the redundant set, and  $\mathcal{D}d$  is obtained from equation 7.

### C. Rough Set Feature Significance and Cuckoo Search Optimization (RSFSCSO)

In this algorithm,  $p_a$  value is taken as 0.25 which means that the final solution depends 75% on the global best solution

and 25% on the local best solution (lbest). The fitness function is calculated for every cuckoo and each value is compared with the global best solution (gbest). When the current fitness value is better than the global best solution then the gbest value is replaced and updated. This process is repeated iteratively until the stop criteria is met. The stop criteria in our algorithm is a maximum number of iterations which is given as input at the start of the algorithm. The max\_iteration value is taken as 5;

**RSFSCSO Algorithm**

Input:

- 1) number of nests N
- 2) Maximum number of iterations i(max\_iteration)
- 3) Step length  $\alpha=1$
- 4)  $p_a = 0.25$

Output:

- 5) An optimized subset of features(RS)

Procedure:

- 6) The population of N host nests are initialized as  $x_i$ , where  $i = 1,2,3 \dots N$
- 7)  $r=1$ ; //random number
- 8) While stopping criteria not met do
- 9) for (i=0 to N)
- 10) {
- 11) RS= feature subset corresponding to  $x_i$ .
- 12) // generate a new cuckoo( $x_i$ ) with the help of levy flight using equation 7.
- 13) //evaluate the fitness function  $f_{eval}(x_i)$  using the equation (11)
- 14) if(  $f_{eval}(x_i) < gbest$ )
- 15) {
- 16) lbest= $x_i$
- 17)  $gbest = (f_{eval}(x_i))$
- 18) RS=  $x_i$
- 19) Break;
- 20) } // end if;
- 21) } // end for;
- 22)  $r=r+1$ ;
- 23) Sort  $x_i$  by order of fitness function in descending order.
- 24) Pick a random nest j such that  $j! = i$ ;
- 25) for all Abandon a fraction of worst nests by comparing with  $p_a$  and update the nests using levy flight (equation 7)
- 26) Let the new egg generated is  $x_l$
- 27) RS= feature subset corresponding to  $x_l$
- 28) again evaluate the fitness function  $f_{eval}(x_l)$  using equation 11.
- 29) Sort the nests according to their fitness function in descending order.
- 30) Choose a random nest k
- 31) if( $f_{eval}(x_l) \geq f_{eval}(x_k)$ ) // solutions are ranked according to current best.
- 32) {
- 33)  $x_k = x_l$
- 34)  $f_{eval}(x_k) = f_{eval}(x_l)$

- 35) } //end if;
- 36) end for;
- 37) end while;

**III. EXPERIMENTAL RESULTS AND DISCUSSION**

In this work, the analysis has been carried out on the clamp dataset [31] comprising of 55 features in the raw dataset and 72 features in the integrated dataset that are extracted from the PE header of an executable file. The proposed algorithm was implemented using 5184 samples Table VII makes a comparison of classification accuracies and Fig. 3 shows a graphical representation of accuracies in both integrated and raw datasets.

The proposed algorithm was implemented on the Java platform, and WEKA 3.9 tool is used for classification. A Windows ten operating system with 8GB RAM is used for experimentation.

The dataset used in our work, before feature selection, is analyzed with other feature selection algorithms along with the proposed algorithm. The resulted accuracies are tabulated in Table VII, and the pictorial representation of these accuracies is represented in Fig. 3. Using random forest classifiers on both raw and integrated datasets, we have obtained an accuracy of 98.3% and 99.25%, respectively. Therefore random forest classifier is selected for evaluation of our feature reduction algorithm.

TABLE VII. COMPARISON OF CLASSIFICATION ACCURACY BEFORE FEATURE SELECTION

classifier	Accuracy (Raw Dataset)	Accuracy (integrated Dataset)
J48	97.2%	97.8%
Random forest	<b>98.3%</b>	<b>99.25%</b>
Naïve Bayes	62.78%	65.20%
Random Tree	96.27%	96.18%
Decision table	93.5%	95.0%
IBK	96.9%	97.98%

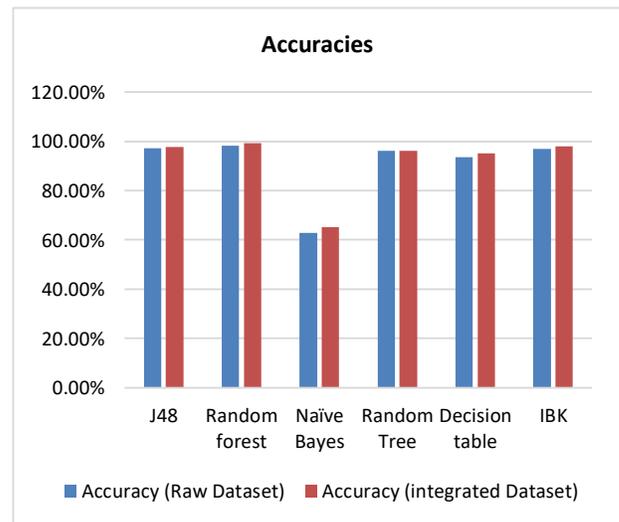


Fig. 3. Graph Representing Accuracies.

The proposed algorithm has just produced three features for the raw dataset and only two features for the integrated dataset. The critical features that were identified for the Integrated dataset are FH\_characteristics and OH\_DLL characteristics. The characteristics field of the file header is recognized as an essential feature by the proposed algorithm. Malware is differentiated by calculating the mean value of files. The mean value of malware files is lesser than that of benign data. The second feature identified is the DLL characteristic feature of the optional header. This feature is linked with the import table which consists of the names of files that are imported and exported. Malware files will have strange import tables when compared with regular data. Table VIII gives detailed results of accuracies after feature selection, and Fig. 4 shows the pictorial representation of the comparison of accuracies.

The features that were identified crucial for Raw dataset are characteristics, checksum and DLL characteristics. As mentioned above characteristics and DLL characteristics play a very effective role in the identification of malware file. The checksum is one other characteristic extracted from the DOS-Header. The checksum is a crucial feature because it validates at load time. It helps in preventing the entry of any damaged files or binaries. To evaluate and compare the performance of the proposed algorithm with other existing feature reduction algorithm, the same number of features are considered in all the cases. Table IX gives a brief description of other related works which used the same clamp dataset in their work. According to table IX, the proposed work produced better results than others.

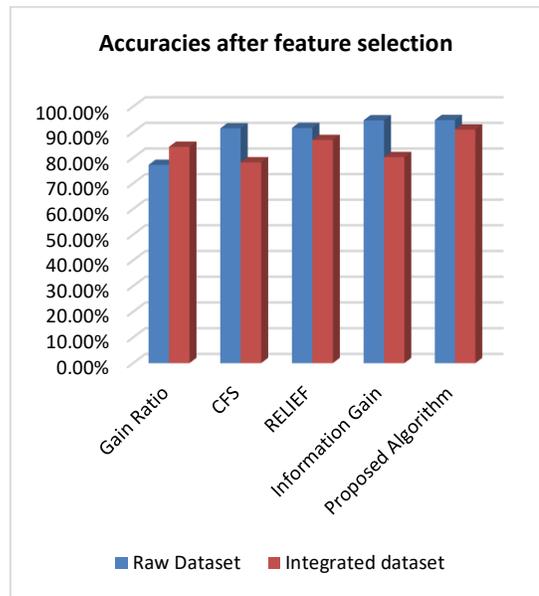


Fig. 4. Comparison of Accuracies after Feature Selection.

TABLE VIII. COMPARISON OF ACCURACIES ON RANDOM FOREST CLASSIFIER, AFTER FEATURE SELECTION

Feature selection algorithm	Accuracy (Raw Dataset)	Accuracy (Integrated Dataset)
Gain Ratio	77.19%	84.24%
CFS	91.45%	78.21%
RELIEF	91.62%	86.92%
Information Gain	94.6%	80.24%
Proposed Algorithm	<b>94.71%</b>	<b>91%</b>

TABLE IX. COMPARISON OF PROPOSED WORK WITH PREVIOUS WORKS

Dataset Reference	No of features after reduction	Feature selection Algorithm(if used any)	Accuracy (before feature selection)	Accuracy (after feature selection)
Mouhammad and Samail [23]	NA	NA	99.1%(integrate dataset)	NA
Kumar et al. [22]	15 (raw dataset)	ExtraTree Classifier	98.3% (raw dataset)	98.3%
Proposed algorithm	<b>3(raw dataset) 2 (Integrated Dataset)</b>	<b>Rough Set feature significance and Cuckoo search Optimization(RSFSCSO)</b>	<b>98.3%(Raw dataset) 99.25%(Integrated dataset)</b>	<b>94.71%(Raw dataset) 92%(Integrated dataset)</b>

## V. CONCLUSION

The PE file header features are extracted from Windows executables in the process of identifying malware using machine learning techniques. Feature reduction is a quite essential pre-processing phase in machine learning to improve the performance and reduce the space complexity. This paper presents the implementation of a rough-set based dependency degree as an objective function in cuckoo search optimization applied to the malware detection system. A massive 94.54% reduction of data size concerning raw dataset and 97.22% reduction of data size concerning the integrated dataset is achieved at a loss of marginal 3.59% and 7.52% accuracies for raw and integrated datasets, respectively. The advantage of RSFSCSO is that it is a filter-based feature reduction, and the final model does not depend on the classifier for feature reduction. However, since the cuckoo search optimization is a population-based solution selection, it normally takes more run time than dynamic search techniques like the ACO. A comparison of various optimization techniques with RS for feature selection can be made in future work.

## REFERENCES

- [1] M. Christodorescu, S. Jha, S.A. Seshia, D. Song and Bryant R.E, "Semantics-aware malware detection," in 2005 IEEE Symposium on Security and Privacy (S&P'05), Oakland, CA, USA, USA, 2005.
- [2] Sadia Noreen, Shafaq Murtaza, M. Zubair Shafiq and Muddassar Farooq, "Evolvable Malware," in 11th Annual conference on Genetic and evolutionary computation, Canada, 2009.
- [3] Koller Daphne and Sahami Mehran, "Toward Optimal Feature Selection," in Proceedings of the 13th International Conference on Machine Learning, 1996.
- [4] Ahmed F. Alia and Adel Taweel, "Feature Selection based on Hybrid Binary Cuckoo Search and Rough Set Theory in Classification for Nominal Datasets," I.J. Information Technology and Computer Science, vol. 9, no. 4, pp. 63-72, 2017.
- [5] Sebastián Maldonado and Richard Weber, "A wrapper method for feature selection using Support Vector Machines," Information Sciences, vol. 179, no. 13, pp. 2208-2217, 2009.
- [6] S L Shiva Darshan and C D Jaidhar, "Performance Evaluation of Filter-based Feature Selection Techniques in Classifying Portable Executable Files," Procedia Computer Science, vol. 125, pp. 346-356, 2018.
- [7] Mark A Hall and Lloyd A Smith, "Practical Feature Subset Selection for Machine Learning," in In C. McDonald(Ed.), Computer Science '98 Proceedings of the 21st Australasian Computer Science Conference ACSC'98, Perth, 1998.
- [8] Asaf Shabtai, Uri Kanonov, Yuval Elovici, Chanan Glezer and Yael Weiss, "'Andromaly': a behavioral malware detection framework for android devices," Journal of Intelligent Information Systems, vol. 38, p. 161-190, 2012.
- [9] Ravi Kiran Varma P and Padmaprabha K, "Web phishing detection: feature selection using rough sets and ant colony optimisation," International Journal of Intelligent Systems Design and Computing, vol. 2, no. 2, pp. 102-113, 2018.
- [10] Zdzisław Pawlak, "Rough sets," International Journal of Computer & Information Sciences, vol. 11, no. 5, p. 341-356, 1982.
- [11] Qinghua Zhang, Qin Xie and Guoyin Wang, "A survey on rough set theory and its applications," CAAI Transactions on Intelligence Technology, vol. 1, no. 4, pp. 323-333, 2016.
- [12] Mert Bal, "Rough Sets Theory as Symbolic Data Mining Method: An Application on Complete Decision Table," Information Science Letters, vol. 2, no. 1, pp. 35-47, 2013.
- [13] Rafael Bello and Rafael Falcon, "Rough Sets in Machine Learning: A Review," in Thriving Rough Sets, Studies in Computational Intelligence, vol 708, Cham, Springer, 2017, pp. 87-118.
- [14] Zhang Boyun, Yin Jianping, Wensheng Tang, Jinbo Hao and Dingxing Zhang, "Unknown Malicious Codes Detection Based on Rough Set Theory and Support Vector Machine," in The 2006 IEEE International Joint Conference on Neural Network Proceedings, Vancouver, 2006.
- [15] Vittorio Maniezzo and Antonella Carbonaro, "Ant Colony Optimization: An Overview," Essays and Surveys in Metaheuristics. Operations Research/Computer Science Interfaces Series, vol. 15, pp. 469-492, 2003.
- [16] Anil Kumar, Gaurav Kabra, Eswara Krishna Mussada, Manoj Kumar Dash and Prashant Singh Rana, "Combined artificial bee colony algorithm and machine learning techniques for prediction of online consumer repurchase intention," Neural Computing and Applications, vol. 31, pp. 877-890, 2017.
- [17] Mohamed Abd El Aziz and Hassanien Aboul Ella, "Modified cuckoo search algorithm with rough sets for feature selection," Neural Computing and Applications, vol. 29, p. 925-934, 2016.
- [18] N Suguna and K Thanushkodi, "A Novel Rough Set Reduct Algorithm for Medical Domain Based on Bee Colony Optimization," Journal of Computing, vol. 2, no. 6, pp. 49-54, 2010.
- [19] Wen-Yau Liang W, "The Genetic Algorithm Incorporates with Rough Set Theory—An Application in Marketing," International Journal of e-Education, e-Business, e-Management and e-Learning, vol. 3, no. 3, pp. 271-273, 2013.
- [20] Xin-She Yang and Deb Suash, "Cuckoo Search via Lévy Flights," in 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), Coimbatore, India, 2009.
- [21] Anil Kumar and Shampa Chakarverty S, "Design Optimization for Reliable embedded system using Cuckoo Search," in IEEE 3rd International Conference on Electronics Computer Technology, Kanyakumari, India, 2011.
- [22] Ajit Kumar, K.S. Kuppusamy and G. Aghila, "A learning model to detect maliciousness of portable executable using integrated feature set," Journal of King Saud University - Computer and Information Sciences, vol. 31, no. 2, pp. 252-265, 2017.
- [23] Mouhammd Alkasasbeh and Samail Al-Daleen, "Classification of malware based on file content and characteristics," in arXiv:1810.07252, 2018.
- [24] Yibin Liao, "PE-Header-Based Malware Study and Detection," Department of Computer Science, The University of Georgia, Athens, 2012.
- [25] Zatloukal Filip and Znoj Jiri, "Malware Detection Based on Multiple PE Headers Identification and Optimization for Specific Types of Files," Journal of Advanced Engineering and Computation (JAEC), vol. 1, no. 2, pp. 153-161, 2017.
- [26] R. Vyas, X. Luo, N. McFarland and C. Justice, "Investigation of Malicious Portable Executable File Detection on the Network using Supervised Learning Techniques," in IFIP/IEEE Symposium on Integrated Network and Service Management (IM), Lisbon, Portugal, 2017.
- [27] D. Devi and S. Nandi, "PE File Features in Detection of Packed Executables," International Journal of Computer Theory and Engineering, vol. 4, no. 3, 2012.
- [28] Jakub Ács, "Static detection of malicious PE files," Department of Computer Systems, Czech Technical University in Prague, Prague, 2018.
- [29] P.Ravi Kiran Varma, V. Valli Kumari and S. Srinivas Kumar, "A novel rough set attribute reduction based on ant colony optimisation," Int. J. Intelligent Systems Technologies and Applications, vol. 14, no. 3/4, pp. 330-353, 2015.
- [30] L. A. M. Pereira, D. Rodrigues, T. N. S. Almeida, C. C. O. Ramos, A. N. Souza, X.-S. Yang and J. P. Papa, "A Binary Cuckoo Search and Its Application for Feature Selection," Studies in Computational Intelligence, Springer, vol. 516, pp. 141-154, 2014.
- [31] "GitHub," [Online]. Available: <https://github.com/>. [Accessed 28 01 2019].