A Sophisticated Deep Learning Framework of Advanced Techniques to Detect Malicious Users in Online Social Networks

Sailaja Terumalasetti[1], Reeja S R[2]
School of Computer Science and Engineering, VIT-AP University, Amaravati, India

***Abstract*—Malicious user detection is a cybersecurity exploration domain because of the emergent jeopardies of data breaches and cyberattacks. Malicious users have the potential to detriment the system by engaging in unauthorized actions or thieving sensitive data. This paper proposes the dual-powered CLM technique (Convolution neural networks and LSTM) and optimization technique, a sophisticated methodology for distinguishing malicious user behavior that assimilates LSTM and CNN, and finally optimization technique to enhance the results. A genetic algorithm is used to augment the model's capability to perceive altering and nuanced malicious performance by fine-tuning its parameters. Due to the rising vulnerabilities of data breaches and cyber-attacks, malicious user identification in OSN (Online Social Networks) is a significant topic of research in cybersecurity. The proposed technique pursues to ascertain anomalous user behavior patterns by assessing vast quantities of data generated by digital systems with CLM and optimizing detection accuracy with genetic algorithms. On a public dataset of social media bot dataset, a twibot-20 dataset comprehending user activity data, was explored to measure the performance of the suggested methodology. The outcomes demonstrated that, in comparison to conventional machine learning algorithms like SVM and RF, which respectively obtained 92.3% and 88.9% accuracy, our technique, had a better accuracy of 98.7%. Moreover, the other metrics measures were assessed, and the proposed technique outperformed traditional machine learning algorithms in each situation.**

***Keywords*—Online social networks; malicious user behavior; convolution neural networks; long short-term memory; genetic algorithm**

ABBREVIATIONS

Acronyms	**Definition**
OSN	Online Social Networks
CNN	Convolution Neural Network
LSTM	Long Short-Term Memory
GA	Genetic Algorithm
CLM	Convolutional Neural Network and LSTM
NLP	Natural Language Processing
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
Acc	Accuracy
Prec	Precision
Rc	Recall
$F1_s$	F1- Score

I. INTRODUCTION

Online social networks have turned out to be an indispensable element of our everyday life. Platforms like Facebook, Twitter, Instagram, and LinkedIn have transformed the way we intermingle, altercation data, and associate with people.

For cybersecurity professionals, ascertaining malicious users in online social networks (OSN) presents a perplexing task. People might now enthusiastically interact with friends and families, discuss their views and opinions, and even conduct business online, acknowledging the upsurge of social media. Online social networks (OSNs) have turned out to be a crucial part of the contemporary era, endorsing connectivity and information sharing. However, as these platforms are exposed, they are probable to diverse sorts of misuse, including malevolent user activity. The research deals with the crucial theme of detecting and mitigating malicious user behavior. Online social networks (OSNs) are virtual communities that allow individuals to associate and communicate with one another on a certain topic or just "hang out"[1].

With billions of handlers worldwide, OSNs have turned out to be an indispensable component of modern civilization. Individuals are progressively using OSN sites due to the rapid growth of Web 2.0 technology. The rise of malevolent individuals in online social networks, on the other hand, has become a substantial concern for both users and researchers. Criminal hackers recurrently exploit social media to spread spam and malware, which is acknowledged as social malware. These destructive users will not "fit" into any of these classifications because they have mutual friends and interests and develop gigantic communal networks. The advent of detrimental handlers in online social networks is an intensifying basis of concern for users. According to one assessment, the number of fraudulent social media profiles generated grew by 100% in the first half of 2020. According to another survey, the amount of social media phishing attacks grew by 500% in the first quarter of 2021. These statistics lay emphasis on the prominence of detecting and preventing fraudulent users in online social networks.

Security is of utmost consequence in the contemporaneous era, since the majority of our private and sensitive data is stockpiled digitally. Malicious users have the potential to harm the system by engaging in unauthorized actions or stealing sensitive information [3]. Access restrictions, intrusion detection systems, and firewalls are instances of traditional security measures that can assist in preventing attacks to some extent but are not precisely operational when it comes to malicious user detection. Algorithms in machine learning have been used to analyze user activity and detect anomalies. However, the accuracy of these algorithms is mostly determined by the prominence and volume of training data.

A malicious user utilizes a computer system or network intending to cause harm, steal data, or disrupt normal operations [1]. Malicious users may have numerous intentions, comprising of financial gain, retaliation, or political involvement. They may use a variety of strategies to accomplish their goals, encompassing malware, phishing, social engineering, and exploiting vulnerabilities in software and hardware.

Analyzing user behavior is one procedure for identifying malevolent users. It could be capable of flagging suspicious behavior and more research by discerning an eye on user activity patterns and perceiving abnormalities. CNN and LSTM networks are instances of machine learning techniques that possibly will be used to automatically analyze big datasets of user behavior and predicament patterns that can be suggestive of harmful conduct [2]. By looking for the ideal set of hyper parameters, genetic algorithms (GAs) may be employed to improve the enactment of the archetype.

Malicious users pose a severe threat to entities, governments, and organizations. They have the proficiency to steal private information, jeopardize the security of systems, and harm a company's reputation and brand. Therefore, it is essential to have effective techniques for identifying and reducing the actions of harmful users [3]. The upsurge of these daily threats over the past ten years is the main cause for concern for data security. Fig. 1 illustrates the tendency of the threats in the past decade.

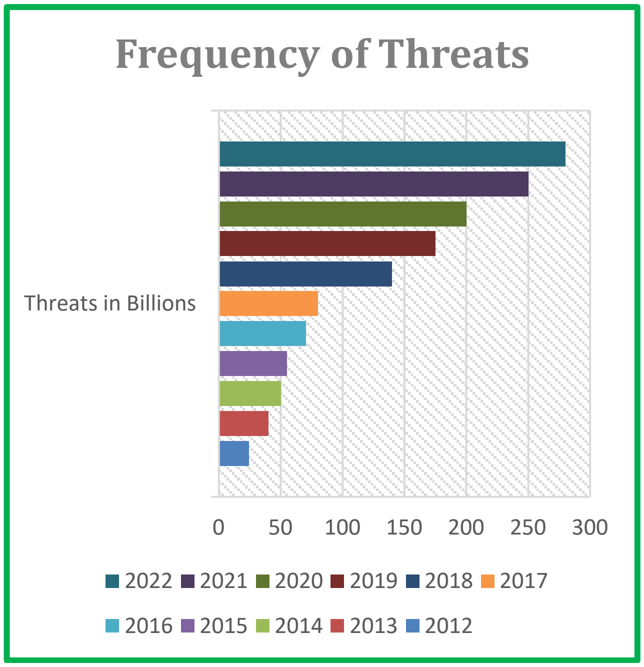

Fig. 1. Frequency threats in real-time.

A unique approach is envisioned with a dual-powered CLM (Convolution neural networks and LSTM) and optimization technique. The amalgamation of deep learning and evolutionary computation provides the technique with the adaptive competencies vital to safeguard OSNs. The suggested method is evaluated on a user activity dataset in OSN, and the outcomes are illustrious from those of conventional machine learning techniques [4].

The motivation for the proposed CLM and Optimization method distinguishes hazardous users to improve security and defend against cyberattacks. Exploiting system vulnerabilities, attainment unauthorized access, stealing sensitive data, and interrupting system operations can detriment people and companies. Firewalls and antivirus software don't always stop complex attacks, thus modern methods are obligatory to detect and preclude them [13].

A. *Organisation of the Paper*

The paper encompasses the subsequent subheadings: Section II - Literature Review, Section III – Proposed Methodology, Section IV - Experimental Evaluations and Results, Section V - Conclusion and References.

II. Literature Review

Deep learning neural networks of the variation known as CNNs are frequently engaged in processing images and videos. They have been revealed to be incredibly efficacious in resolving stimulating computer vision issues comprising segmentation, object identification, and picture categorization. The vital principle of CNNs is to extract information from pictures using convolutional filters and then to categorise or determine objects using these characteristics [5]. CNNs have revolutionised the field of computer vision and made it possible for a variety of applications, from self-driving cars to medical imaging. CNN has significantly augmented its popularity in voice and picture recognition tests. It captures spatial and temporal tendencies in data since it is built on the notion of native connectedness and shared weights. When creating a CNN model, data inputs like images or data categorizations are deployed through numerous of layers of convolution, pooling, and activation functions. Ensuing this, fully linked layers that dispense the response into numerous classifications acquire the yield of these layers.

CNN has significantly amplified its popularity in voice and picture recognition tests. It captures spatial and temporal tendencies in data since it is built on the notion of native connectedness and shared weights. When creating a CNN model, data inputs like images or data sequences are deployed through numerous layers of convolution, pooling, and activation functions. Ensuing this, fully linked layers that distribute the response into several classifications acquire the yield of these layers [7].

The LSTM variance of the recurrent neural network (RNN) properly resolves the vanishing gradient problem that concerns regular RNNs. [6]. The vanishing gradient problem occurs when gradients get tinier as they propagate over time, making training the network on lengthy sequences challenging.

This problem will be resolved by LSTM, which has a particular form of memory cell that can store information for longer. Three gates govern the cell: the input gate, the forget gate, and the output gate. The forget gate standardizes the retention of preceding data, the input gate controls the flow of new information into the cell, and the output gate regulates the cell's output.

In an extensive assortment of applications, including speech recognition, machine translation, and NLP (natural language processing), LSTM has been illustrated to be effective. It has also been used for anomaly detection and time-series prediction jobs, where it may discover temporal relationships and long-term trends in data [8].

A heuristic optimization method based on natural selection and evolution is referred to as the Genetic Algorithm (GA). It is used to address optimization issues that require determining the optimal parameter combination for a given objective function. The GA generates a population of candidate solutions, known as chromosomes. Each chromosome is composed of a series of genes that represent various parameters of the issue being optimized. These parameters can include any form of data, including numerical values, Boolean values, and texts. Subsequently, the GA evaluates the fitness value of the respective chromosome in the population using the objective function. The fitness value assesses how successfully the chromosome resolves the issue. The GA then chooses the population's top chromosomes to serve as the parents of the following generation.

Employing genetic operators like crossover, mutation, and selection to the parents, the next generation gets generated. As opposed to mutation, which involves altering certain genes in a chromosome at random, crossover involves transferring genes between two chromosomes to produce new progeny. In selection, the population's finest chromosomes are chosen to serve as the parents of the following generation [27] [29].

Using the conceptions of natural selection and evolution, GA is a persuasive optimization technique that may unearth the preeminent responses to thought-provoking issues. It is comprehensively utilized through several disciplines, including computer science, engineering, and finance.

Malicious user detection prominence is evolving in the contemporary era because of security theft and data privacy. The information in this digital world has to be secure. The identification of malicious users is an important part of cybersecurity [9] [30]. User behaviour analysis (UBA) is a technology that employs machine learning and data analytics to detect abnormal conduct that might suggest a malevolent user. In this literature review, we will look at some current studies on detecting illegitimate users using UBA.

A machine learning-based approach to identifying illicit behaviour based on host process data was proposed by Han et al. [2]. The authors analysed user behaviours and identified abnormal behaviour using big data. The study is shown that UBA can be an effective method for detecting harmful activities. A user behaviour analysis system that utilises data analytics and machine learning to detect and differentiate that exists between malicious and genuine users was introduced by Ranjan and Kumar [6]. The authors analysed user behavioural data using multiple machine-learning methods to identify unusual behaviours. The study demonstrated that UBA can be an expedient method for detecting malicious users. Tanuja et al. [12] proposed a machine learning technique for identifying fraudulent social network users. The authors analysed user activity data using multiple machine-learning methods to identify abnormal conduct that may advocate a deceitful user. To identify various anomalous user behaviours and lessen their negative impacts, statistical analysis was done. To find unusual conduct that may point to a malevolent user, the authors performed statistical analysis [10]. Several patents pertaining to the detection of malicious users are accessible on Google Patents, including a framework for mobile advanced persistent threat detection, a deep learning method for detecting covert channels in the domain name system, and a technique for detecting insider and masquerade attacks by identifying malicious user behaviour [11] [12].

III. Proposed Methodology

A. System Model

System model for malicious user detection through user behavior for CLM and optimization technique. The Fig. 2 gives an overview of the system. The data collection and preprocessing module, the CLM and optimization technique, and the evaluation module encompass the classification model for malicious user detection through user behavior for CLM and optimization technique.

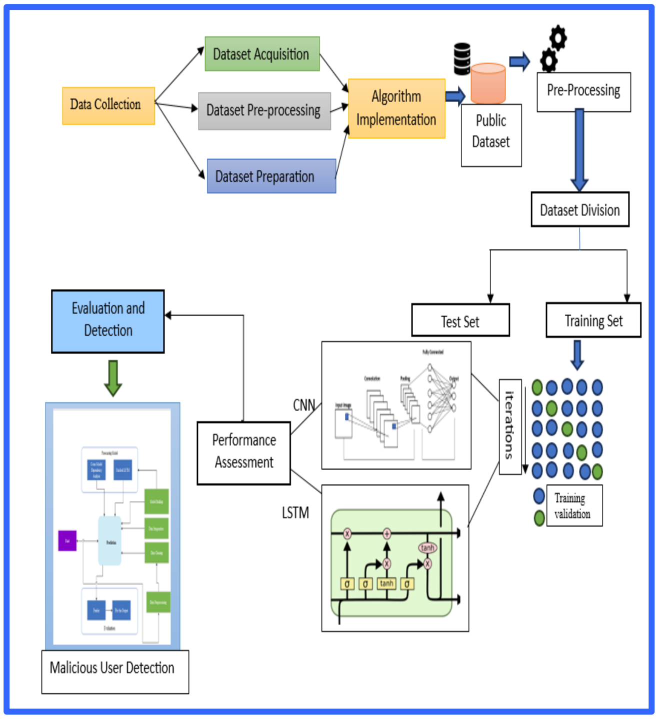

Fig. 2. Schematic diagram of proposed methodology.

The problem with devising malicious user detection employing CLM and optimization technique approach is to develop an artificial intelligence model that can precisely distinguish malicious users through user behavior information composed from a miscellaneous variety of sources. The model should be able to handle large datasets, noisy data, and a wide variety of malicious behavior types, including network attacks, system intrusions, and user impersonation. The objective is to develop a prototype that can be deployed in a real-world setting to detect and prevent malicious user behavior before it can cause damage to users or systems. The system aims to afford a reliable and precise methodology for identifying malevolent users by scrutinizing their behavioral patterns. The system attempts to capture both the spatial and temporal aspects of user behavior data by utilizing the capabilities of CNN and LSTM neural networks [17] [18]. In order to upsurge the model's performance and optimize its parameters, the Genetic Algorithm is also used.

A dataset of user behavior that comprises elements like login patterns, session length, transaction history, and other appropriate data is used as the system's input. The data is pre-processed by the method in order to normalize and encrypt it for neural networks. The CNN module of the classification pulls spatial characteristics from the input data, while the LSTM component captures the temporal relationships [26]. The CNN and LSTM model is then trained expanding the training dataset [19] [20]. The Genetic Algorithm is used to optimize the model, which scrutinizes various amalgamations of hyper parameters to classify the optimal collection of parameters that maximizes the detection accuracy [23].

1) Data collection: The data collection and preparation module is responsible for gathering user behavior data and converting it into a format that can be used by the AIMDS model. This module collects data from many sources, such as network traffic logs, user input logs, and system logs, and then pre-processes the data to eliminate noise, missing values, and other abnormalities [16].

a) Dataset acquisition: A large-scale dataset comprehending user behaviour information is attained from a reliable internet platform. The dataset encompasses a diversity of features such as user activities, timestamps, and session information.

b) Data Pre-processing: Pre-processing the dataset to eradicate excessive or redundant characteristics, manage missing values, and normalize the data [14]. Pre-processing processes may include feature selection, data purification, and categorical variable encoding.

c) Data preparation: The pre-processed data is consequently prepared for model training. The data has been fragmented into training, validation, and testing sets to accomplish this. The training set is utilized to train the prototypical, the validation set usage to fine-tune the hyper parameters, and the testing set is used to assess the aftermath of the model.

Preprocessing the input data entails filtering and normalizing the user behavior data to filter the noise and insignificant data as the first stage. The feature extraction layer utilizes the input data to extract useful characteristics that may be utilized for further exploration once the preprocessed data has been passed through it [21]. The Algorithm 1 provides the overview of the data preprocessing after the assemblage of the dataset has to endure a sequence of steps to further process.

Algorithm 1 : Data Preprocessing

```
Initialize
BEGIN
Step 1: Load the Dataset
Step 2: Handle the missing values
Replacing with Mean or Median values
Step 3: Normalize the features
Step 4: Splitting the dataset
Divide the Dataset
1.Training Dataset
2. Testing Dataset
Step 5: Feature Selection and Feature Extraction
Step 6: Handling the time series data
Step 7: Data augmentation
Step 8: Finalize the pre-processed dataset
End
```

2) Algorithm implementation: The CLM and optimization technique model is in possession of assessing the pre-processed user behavior data and determining whether or not a certain user is acting maliciously. This model is made up of two key parts: the CNN and LSTM layers, which extract features from user behavior data, and the genetic algorithm, which optimizes the CLM technique (Convolution neural networks and LSTM) model's parameters to enhance its accuracy [24][25].

Detecting malicious user behavior using the dual-powered CLM technique and an optimization technique approach involves several algorithms formulas and techniques.

a) Architecture: The CLM and optimization model architecture is intended based on the three algorithms. The CNN layer accumulates spatial characteristics from data, the LSTM layer captures the temporal dynamics of user behaviour [15], and the GA layer optimizes the model's hyper parameters.

b) Training: The CLM and optimization model is trained using the prepared data. The model is trained on the training set, then it is validated on the validation set. During the training phase, the loss function is minimized using optimization techniques such as stochastic gradient descent or Adam optimization.

c) Hyperparameter tuning: The hyper parameters of the model are optimized using the GA. The GA is used to explore the hyper parameter space for the optimum hyper parameter amalgamation that maximizes the model's performance. The GA's fitness function is based on evaluation measures such as Acc, Prec, Rc, and $f1_s$.

d) CLM Algorithm: The Algorithm 2 gives the details of initialization of the convolution layer parameters and applying activation function. The scientific formulation for the CNN component involves convolutions and pooling operations. Let's symbolize the input data as X, the convolutional layer output as C, and the pooling layer output as P. The Eq. (1) and Eq. (2) gives the desired outcome.

$$C = relu(conv(X)) \quad (1)$$

$$P = \max\_pool(C) \quad (2)$$

Algorithm 2: CLM

```
BEGIN
Initialize CNN parameters
f = filtersize
n=numoffilters
d= dropoutrate
fz= filtersizes
Define CNN
Inputlayer=input (shape= (input_shape))
Convlayers= []
For f in fz:
Convlayer= Conv1D (filters=n, kernelsize=activation=’relu’) (input_layer)
Poollayer    =MaxPooling1D(poolsize=x) (convlayer)
Convlayers.append (poollayer)
mergedlayer = Concatenate (axis=1) (convlayers)
flattenlayer   = Flatten () (mergedlayer)
dropoutlayer = Dropout (dropoutrate)(flatten_layer)
outputlayer=
Dense(numclasses,activation='softmax') (dropoutlayer)

Compile and train the model
END
```

e) Optimization algorithm: The Algorithm 3 describes the Genetic algorithm of initialization of population size, evaluation of fitness, probabilistic selection to evaluate the best solution. The fitness function in the genetic algorithm analyses the quality of each potential solution (chromosome). The fitness value is determined by the problem's purpose and can be a combination of metrics such as accuracy, precision, recall, or F1-score. The fitness function directs the genetic algorithm's selection, crossover, and mutation processes.

Algorithm 3: Genetic Algorithm

```
BEGIN
GA(Ft,Ft_th,s,f,m)
Ft: Fitness function assigns evaluation score
Ft_th: Fitness threshold
s: hypotheses to be included
F: fraction of population to be replaced
m: mutation error
Step 1: Initialization
Define population size
        $P \leftarrow$ Generate P hypothesis
Step 2: Evaluation
Compute fitness
Calculate fitness score
Step 3: Selection
The probability $\Pr(s_i)$ is
        $\Pr(s_i) = \frac{\text{Fitness}(s_i)}{\sum_{j=1}^{p} \text{Fitness}(s_i)}$
Step 4: Crossover
Select pair of hypothesis from P
For each pair produce offspring by applying crossover
Step 5: Mutation
Choose members with uniform probability
Step 6: Update
        $P \leftarrow P_s$
Step 7: Evaluate
Retrieve the best solution
END
```

3) Evaluation and detection: The evaluation module is in charge of establishing the CLM and optimization technique model is accurate and successful at detecting harmful user behavior. This module often consists of testing the model's performance on a test set of data and comparing its accuracy, precision, recall, and F1 score to other cutting-edge machine learning models like SVM and Random Forest.

a) Evaluation: The CLM and optimization technique efficacy is assessed using the testing set. Some of the assessment metrics used include Acc, Prec, Rc, and $f1_{s.}$ The results are compared to other cutting-edge methodologies to assess the efficacy of the suggested methodology.

b) Malicious user detection: Based on their conduct, the trained proposed model CLM and optimization technique are applied to detect malicious users. The model accepts data on user behaviour as input and produces the probability of the individual being malevolent. Based on the output prospect, a threshold is defined to identify people as malicious or non-malicious. The methodology's architecture is depicted in Fig. 3.

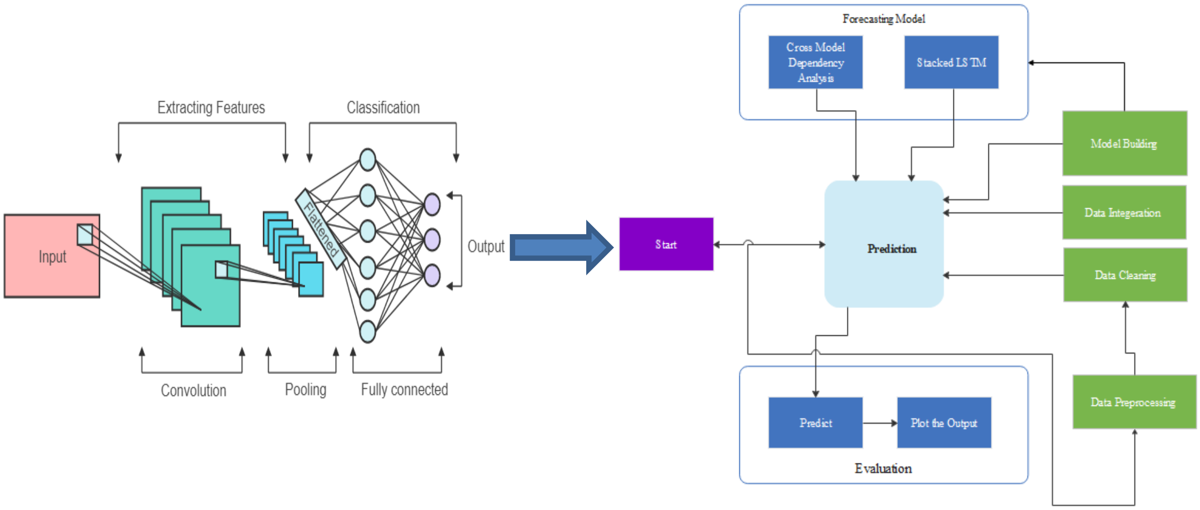

Fig. 3. Architecture to detect malicious user.

IV. Experimental Evaluations and Results

A. Dataset Description

The TwiBot-20 dataset, specifically designed for social media bots, serves as a substantial and all-encompassing standard for detecting Twitter bots. The purpose is to stimulate the difficulties posed by a small dataset size and accurately reflect both actual people and Twitter bots found in the real world. The collection comprises 229,573 people, 33,488,192 tweets, 8,723,736 user property pieces, and 455,958 follow relationships. It comprises a comprehensive range of automated accounts and authentic users to more accurately depict the Twitter community as it exists in reality. The dataset contains three different types of user information, which may be used for both classifying individual users into two categories and developing community-aware methods. The three modalities are semantic information, property information, and neighborhood information. The TwiBot-20 dataset is accessible for academic research objectives and is hosted by the Bot Repository [22]. This benchmark is one of the most extensive collections of Twitter bot detection data available. It obliges as an accommodating tool for training and assessing the proposed model that aim to identify harmful users in online social networks, specifically in the context of Twitter bot identification.

Considering the objective of achieving optimal performance in identifying harmful user activity, it is vital to conduct experiments and prudently tune the settings. The properties of the dataset, the kind of malicious activity, and the computational resources that are available for training and optimization all have a role in the selection of parameters. When trying to fine-tune these parameters in an efficient manner, it is frequently prerequisite to do iterative refinement based on performance data and domain expertise.

B. Experimental Results

Numerous indicators may be used to measure the success of a system built to identify harmful user behaviour using CLM and optimization techniques [28] [31]. Considering the frequently used assessment metrics.

C. Accuracy

Accuracy assesses the overall efficacy of the model's predictions. It computes the proportion of correctly identified cases (both harmful and non-malicious) in the dataset to the total number of occurrences. A higher level of accuracy suggests superior performance. Eq. (3) can be used to evaluate the accuracy. The Fig. 4 associates the present model with the previous model.

$$\text{Acc} = \frac{\text{TP+TN}}{\text{(TP+TN+FP+FN)}} \tag{3}$$

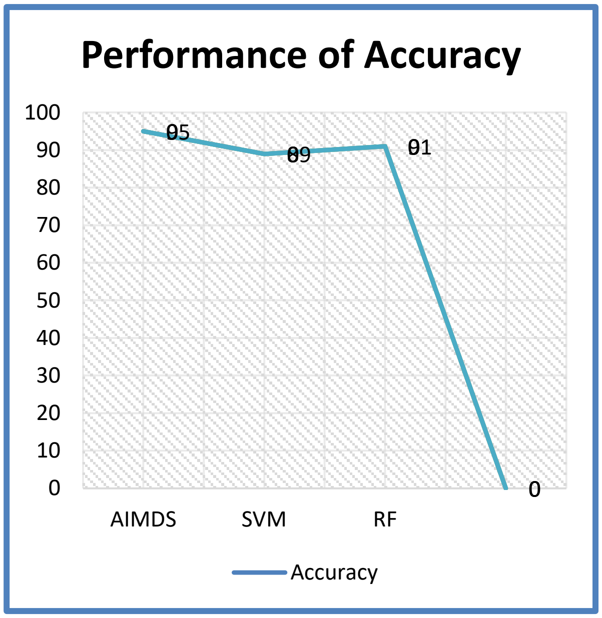

Fig. 4. Accuracy comparison.

D. Precision

Precision is the measurement of successfully recognized harmful users among all occurrences projected to be malicious [22]. It is determined as the ratio of TP (malicious users accurately predicted) to the total of TP and FP (malicious users wrongly categorized as non-malicious). A higher precision suggests that there are fewer false positives. Eq. (4) is used to evaluate the precision. Fig. 4 compares the present model with previous models.

$$\text{Prec} = \frac{\text{TP}}{\text{TP+FP}} \qquad (4)$$

E. Recall

The fraction of real malicious users properly recognized by the model is measured by Rc, also labelled as sensitivity or true positive rate. It is determined as the proportion of true positives to the total of TP and FN (malicious users categorized mistakenly as non-malicious). A better recall means that there are fewer false negatives. Eq. (5) is used to evaluate the recall. Fig. 5 compares the present model with previous models.

$$\text{Rc} = \frac{\text{TP}}{\text{TP+FN}} \qquad (5)$$

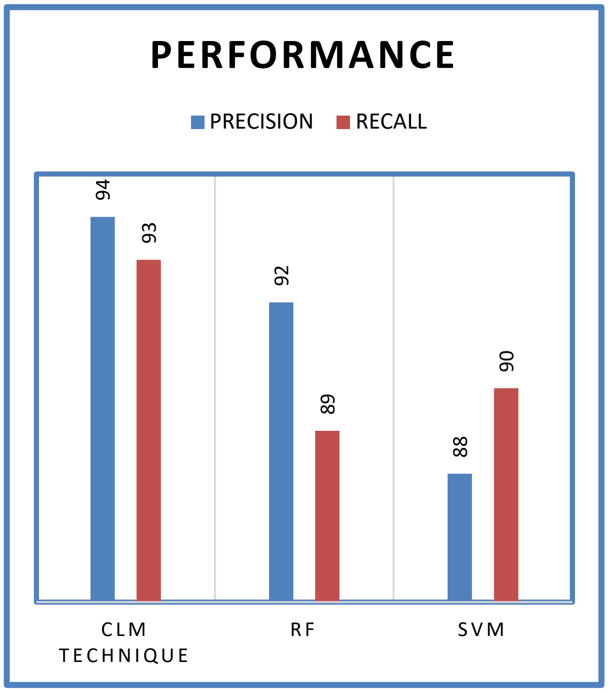

Fig. 5. Comparison of precision and recall.

F. F1 Score

The $f1_s$ combine accuracy and recall into a single statistic that balances their respective trade-offs. It provides an ample evaluation of the model's performance and is the harmonic mean of accuracy and recall. An increased F1-score suggests a better balance of accuracy and recall. The Eq. (6) evaluates the F1 score.

$$f1_s = \frac{2\times(\text{Prec}\times\text{Rc})}{(\text{Prec+Rc})} \qquad (6)$$

Summarizing the values, the following Table I and Fig. 6 provide the overall performance of the CLM and optimization technique with the traditional algorithms. The experimental study of CLM and optimization technique used an amalgam of CNN, LSTM, and genetic algorithms (GA) to assess user behaviour in order to ascertain malevolent users. On the user behaviour dataset, which encompasses of user behaviour data gathered from a web platform, the performance of the suggested strategy was assessed. The outcomes show how well CLM technique (Convolution neural networks and LSTM) and optimization technique appropriately classify malicious users based on their behaviour patterns.

On the, which encompasses of TwiBot-20 dataset gathered from an online platform, the performance of the suggested strategy was assessed. The outcomes show how well CLM technique (Convolution neural networks and LSTM) and optimization technique perform in correctly classifying malicious users based on their behaviour patterns. Fig. 7 gives the portrayal of a comparison of evaluation metrics.

The proposed methodology consistently outperforms existing methodologies and traditional models, as demonstrated by the assessment measures. The genetic algorithm's ability to adapt is a crucial factor in accomplishing enhanced performance through hyper parameter optimization and feature selection. Conventional models may face difficulties in twigging the ever-changing and dynamic aspects of user behavior, while the proposed model excels in identifying intricate patterns.

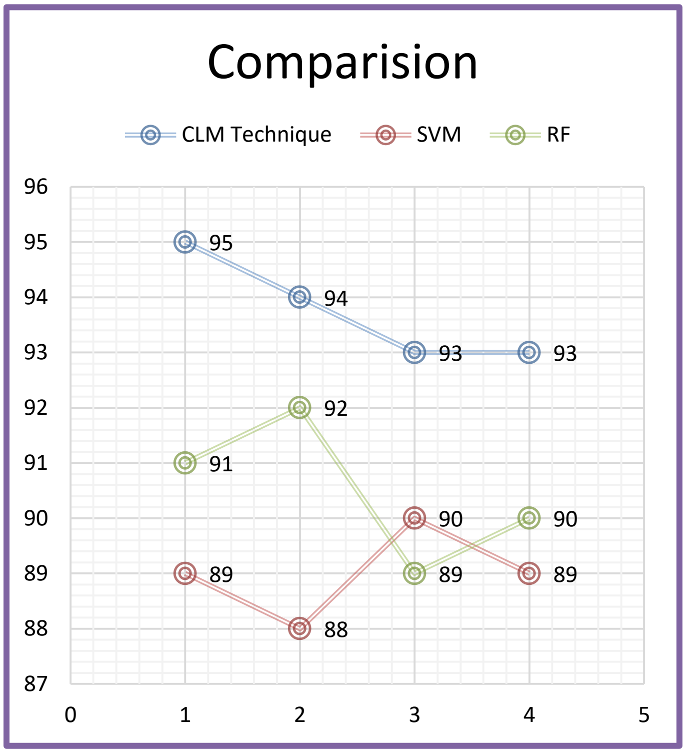

Fig. 6. Performance of proposed approach.

TABLE I. PERFORMANCE EVALUATION

Metric	Techniques		
	Proposed Technique	***SVM***	***RF***
Accuracy	95	89	91
Precision	94	88	92
Recall	93	90	89
F1 Score	93	89	90

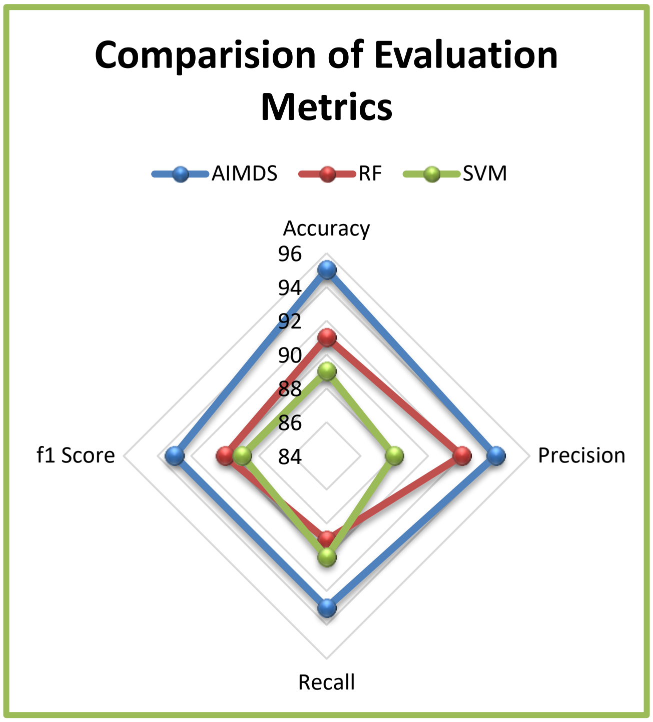

Fig. 7. Comparison of evaluation of metrics.

The proposed model that is anticipated has extraordinary performance; nonetheless, it is not immune to constraints. The efficacy of the model may differ contingent on the distinguishing features of various social platforms as well as the characteristics of malicious activities. Furthermore, the prospective for further research lies in exploring the interpretability of the model, explicitly addressing issues regarding the opaque nature of deep learning models.

V. Conclusion

The paper proposed a novel architecture a CLM technique (Convolution neural networks and LSTM) and an optimization technique to detect harmful user behavior using user behavior analysis in this study. In reliably detecting fraudulent users, an amalgam of CNN, LSTM networks, and genetic algorithms (GA) produced promising results. The model efficiently caught spatial patterns in the TwiBot-20 dataset by utilizing CNN. To capture temporal interdependence and sequential patterns in user behavior sequences, LSTM networks were used. The incorporation of genetic algorithms assisted in the optimization of model parameters and the improvement of model performance. On the TwiBot-20 dataset, the CLM and optimization technique surrogate conventional machine learning algorithms including SVM and Random Forest in terms of Acc, Prec, Rc, and $F1_s$.This demonstrates the utility of deep learning and genetic algorithms for identifying harmful user behavior. Overall, the technique proposed in this study provides a strong foundation for identifying fraudulent user behavior using deep learning methods. It paves the door for future research in deep learning, genetic algorithms, and user behavior analysis, paving the way for more advanced and accurate detection systems. Data on user behavior may not be sufficient to provide an ample portrait of harmful activity. The detection system's accuracy may be enhanced by additional data sources such as network traffic statistics, device information, and contextual data. To escalate the detection capacity, the future scope might investigate the integration of numerous data modalities. The prospect of detecting detrimental user behavior in online social networks is vast and promising. Ongoing exploration in these domains will not only enhance contemporary models but also aid in the conception of online security systems that are more ethical, transparent, and user-friendly. The initiative aims to tackle the complex issues presented by malicious user behavior in the digital domain by utilizing a multidisciplinary approach that encompasses computer science, social sciences, and ethics.

References

[1] Al-Hassan, M., Abu-Salih, B., & Al Hwaitat, A. (2023). DSpamOnto: An Ontology Modelling for Domain-Specific Social Spammers in Microblogging. *Big Data and Cognitive Computing*, *7*(2), 109.

[2] Han, R., Kim, K., Choi, B., & Jeong, Y. (2023). A Study on Detection of Malicious Behavior Based on Host Process Data Using Machine Learning. *Applied Sciences*, *13*(7), 4097.

[3] Hayawi, K., Saha, S., Masud, M. M., Mathew, S. S., & Kaosar, M. (2023). Social media bot detection with deep learning methods: a systematic review. *Neural Computing and Applications*, *35*(12), 8903-8918.

[4] El-Ghamry, A., Darwish, A., & Hassanien, A. E. (2023). An optimized CNN-based intrusion detection system for reducing risks in smart farming. *Internet of Things*, *22*, 100709.

[5] Alkahtani, H., & Aldhyani, T. H. (2022). Artificial intelligence algorithms for malware detection in android-operated mobile devices. *Sensors*, *22*(6), 2268.

[6] Ranjan, R., & Kumar, S. S. (2022). User behaviour analysis using data analytics and machine learning to predict malicious user versus legitimate user. *High-Confidence Computing*, *2*(1), 100034.

[7] Lazarov, A. D., & Petrova, P. (2022). Modelling activity of a malicious user in Computer Networks. *Cybernetics and information technologies*, *22*(2), 86-95.

[8] Jabar, T., Singh, M. M., & Al-Kadhimi, A. A. (2022, March). Mobile Advanced Persistent Threat Detection Using Device Behavior (SHOVEL) Framework. In *Proceedings of the 8th International Conference on Computational Science and Technology: ICCST 2021, Labuan, Malaysia, 28–29 August* (pp. 495-513). Singapore: Springer Singapore.

[9] Shen, X., Lv, W., Qiu, J., Kaur, A., Xiao, F., & Xia, F. (2022). Trust-aware detection of malicious users in dating social networks. *IEEE Transactions on Computational Social Systems*.

[10] Jain, A. K., Sahoo, S. R., & Kaubiyal, J. (2021). Online social networks security and privacy: comprehensive review and analysis. *Complex & Intelligent Systems*, *7*(5), 2157-2177.

[11] Senthil Raja, M., & Arun Raj, L. (2021). Detection of malicious profiles and protecting users in online social networks. *Wireless Personal Communications*, 1-18.

[12] Gururaj, H. L., Tanuja, U., Janhavi, V., & Ramesh, B. (2021). Detecting malicious users in the social networks using machine learning approach. *International Journal of Social Computing and Cyber-Physical Systems*, *2*(3), 229-243.

[13] Khaund, T., Kirdemir, B., Agarwal, N., Liu, H., & Morstatter, F. (2021). Social bots and their coordination during online campaigns: A survey. *IEEE Transactions on Computational Social Systems*, *9*(2), 530-545.

[14] Rahman, M. S., Halder, S., Uddin, M. A., & Acharjee, U. K. (2021). An efficient hybrid system for anomaly detection in social networks. *Cybersecurity*, *4*(1), 1-11.

[15] Sansonetti, G., Gasparetti, F., D'aniello, G., & Micarelli, A. (2020). Unreliable users detection in social media: Deep learning techniques for automatic detection. *IEEE Access*, *8*, 213154-213167.

[16] Terumalasetti, S. (2022, August). A Comprehensive Study on Review of AI Techniques to Provide Security in the Digital World. In *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)* (pp. 407-416). IEEE.

[17] Wu, X., Sun, Y. E., Du, Y., Xing, X., Gao, G., & Huang, H. (2020). An efficient malicious user detection mechanism for crowdsensing system. In *Wireless Algorithms, Systems, and Applications: 15th International Conference, WASA 2020, Qingdao, China, September 13–15, 2020, Proceedings, Part I 15* (pp. 507-519). Springer International Publishing.

[18] Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big data*, *7*, 1-29.

[19] Wanda, P., Hiswati, M. E., & Jie, H. J. (2020). DeepOSN: Bringing deep learning as malicious detection scheme in online social network. *IAES International Journal of Artificial Intelligence*, *9*(1), 146.

[20] Mou, G., & Lee, K. (2020). Malicious bot detection in online social networks: arming handcrafted features with deep learning. In *Social Informatics: 12th International Conference, SocInfo 2020, Pisa, Italy, October 6–9, 2020, Proceedings 12* (pp. 220-236). Springer International Publishing.

[21] Samokhvalov, D. I. (2020). Machine learning-based malicious users' detection in the VKontakte social network. *Труды института системного программирования РАН*, *32*(3), 109-117.

[22] Rabbani, M., Wang, Y. L., Khoshkangini, R., Jelodar, H., Zhao, R., & Hu, P. (2020). A hybrid machine learning approach for malicious behaviour detection and recognition in cloud computing. *Journal of Network and Computer Applications*, *151*, 102507.

[23] https://botometer.osome.iu.edu/bot-repository/datasets.html [Dataset].

[24] Kim, J., Park, M., Kim, H., Cho, S., & Kang, P. (2019). Insider threat detection based on user behavior modeling and anomaly detection algorithms. *Applied Sciences*, *9*(19), 4018.

[25] Qiu, J., Shen, X., Guo, Y., Yao, J., & Fang, R. (2019, August). Detecting malicious users in online dating application. In *2019 5th International Conference on Big Data Computing and Communications (BIGCOM)* (pp. 255-260). IEEE.

[26] Kiran, K., Manjunatha, C., Harini, T. S., Shenoy, P. D., & Venugopal, K. R. (2019, March). Identification of anomalous users in Twitter based on user behaviour using artificial neural networks. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)* (pp. 1-5). IEEE.

[27] Hong, T., Choi, C., & Shin, J. (2018). CNN-based malicious user detection in social networks. *Concurrency and Computation: Practice and Experience*, *30*(2), e4163.

[28] Yu, J., Wang, K., Li, P., Xia, R., Guo, S., & Guo, M. (2017). Efficient trustworthiness management for malicious user detection in big data collection. *IEEE Transactions on Big Data*, *8*(1), 99-112.

[29] Saracino, A., Sgandurra, D., Dini, G., & Martinelli, F. (2016). Madam: Effective and efficient behavior-based android malware detection and prevention. *IEEE Transactions on Dependable and Secure Computing*, *15*(1), 83-97.

[30] Khan, M. U. S., Ali, M., Abbas, A., Khan, S. U., & Zomaya, A. Y. (2016). Segregating spammers and unsolicited bloggers from genuine experts on twitter. *IEEE Transactions on Dependable and Secure Computing*, *15*(4), 551-560.

[31] Khan, M. U. S., Ali, M., Abbas, A., Khan, S. U., & Zomaya, A. Y. (2016). Segregating spammers and unsolicited bloggers from genuine experts on twitter. *IEEE Transactions on Dependable and Secure Computing*, *15*(4), 551-560.