A Screening System for COVID-19 Severity using Machine Learning

Abang Mohd Irham Amiruddin Yusuf [1], Marshima Mohd Rosli[2], Nor Shahida Mohamad Yusop[3]
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450, Selangor, Malaysia[1, 2, 3]
Institute for Pathology, Laboratory and Forensic Medicine, University Teknologi MARA[2]
47000 UiTM, Sungai Buloh, Selangor, Malaysia[2]

***Abstract*—COVID-19 disease can be classified into various stages depending on the severity of the patient. Patients in severe stages of COVID-19 need immediate treatment and should be placed in a medical-ready environment because they are at high risk of death. Thus, hospitals need a fast and efficient method to screen large numbers of patients. The enormous amount of medical data in public repositories allows researchers to gain information and predict possible outcomes. In this study, we use a publicly available dataset from Springer Nature repository to discuss the performance of three machine learning techniques for prediction of severity of COVID-19: Random Forest (RF), Naïve Bayes (NB) and Gradient Boosting (GB). These techniques were selected for their good performance in medical predictive analytics. We measured the performance of the machine learning techniques using six measurements (accuracy, precision, recall, F1-score, sensitivity and specificity) in predicting COVID-19 severity. We found that RF generates the highest performance score, which is 78.4, compared with NB and GB. We also conducted experiments with RF to establish the critical symptoms in predicting COVID-19 severity, and the findings suggested that seven symptoms are substantial. Overall, the performance of various machine learning techniques to predict severity of COVID-19 using electronic health records indicates that machine learning can be successfully applied to determine specific treatment and effective triage.**

Keywords—Severity prediction; COVID-19; random forest; Naïve Bayes; gradient boosting

I. INTRODUCTION

Research has shown that electronic health records (EHR) are becoming increasingly valuable to predict health outcomes or disease diagnoses [1]. For many years, researchers have been analysing EHR with statistical and machine learning techniques for prognostic evaluations. Statistical techniques are designed to determine relationships between variables, and machine learning techniques are designed to make the most accurate predictions based on EHR. Although both techniques play an important role in research, previous analysis has shown that machine learning outperforms statistical techniques because of the recent advancement of tools for data analytics and large quantity of data humanity has access to since the information explosion [2]–[4].

A significant number of prediction models utilising EHR have been proposed in the literature over past years, including in the recent COVID-19 pandemic [1]. The COVID-19 pandemic presented massive data for examining social, behavioural, public health, and economic impacts [5]. Many recent studies have conducted data mining analysis using various available datasets that focus on the occurrence of confirmed, fatal and recovered COVID-19 cases globally to understand the threats and predict the subsequent planning of containment activities [6]–[9].

The severity of COVID-19 disease varies from mild to critical stages. Screening of positive COVID-19 patients at primary care clinics is used as the initial triage to determine the severity stage and admission to hospital [10]. With use of the conventional methods, the process to manage COVID-19 patients in the initial triage is not efficient because of long waiting times for screening [11]. It has previously been observed that limited hospital resources and staff during the COVID-19 pandemic makes it important to decide which patients require more urgent treatment and which patients that can wait [12]. Therefore, a timely clinical decision is important to help in early detection of serious disease and provide effective treatment for the individual patient, which is important for reasonable allocation of medical resources.

In recent years, machine learning techniques have attained popularity in the health area because of their capability to deal with enormous, complex and unbalanced data, and yield outcomes such as prediction [6]. Many machine learning techniques have been employed in forecasting the potential spread of COVID-19 [5], [11], [13], [14]; however, few studies have reported on the severity prediction of COVID-19 patients [15], [16].

This study discusses the initial prediction of the severity of the COVID-19 stage by clinical information in EHR using machine learning techniques. Early prediction of severity of the COVID-19 stage allows the healthcare organization to develop an effective disease management approach, which may help prevent the progression of the disease. It also helps overcome the limited number of staff in the hospitals by enabling frontliners to help the screening process. By classifying COVID-19 patients into the stage of severity, it helps doctors to prioritized critical patients thus making the screening process more efficient. Further, it may improve the quality of life of the patient. This study also examines the significance of the 21 variables of COVID-19 in the study population. The scope of this study covers the design of the screening system application and the development of machine learning model for the system.

II. Methodology

A. Project Methodology

We used a hybrid methodology (Fig. 1) which combine waterfall and agile methods. The hybrid methodology model consists of five phases: Requirement, Design, Develop, Test and Evaluation. The requirements and evaluation phases are from the waterfall methodology and the design, test and develop phases are from the agile methodology.

In the requirement phase, we collected the project resources such as dataset, tools and machine learning techniques. We designed the high-level system designs, flowchart and graphical user interfaces in the design phase. In the develop phase, we developed the predictive models using machine learning techniques and we test the predictive models until the best results were achieved. We evaluated the models in the evaluation phase using performance metrics such as accuracy, precision and sensitivity.

B. Model Training Flow

Fig. 2 shows the model training flow that we implemented in the development phase to train and test the machine learning techniques. We used the raw data for data preprocessing, feature selection, train model with training and testing data, and evaluate the model. The data preprocessing involves removing some irrelevant features that does not relevant, removing missing values, removing outliers, data transformation and data balancing. After obtaining the cleaned data, it will be split into two parts which is training data and testing data. The model will learn from the training data while testing data will be used to evaluate the model performance.

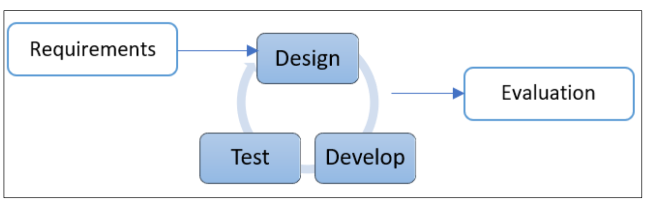

Fig. 1. Hybrid Methodology Model.

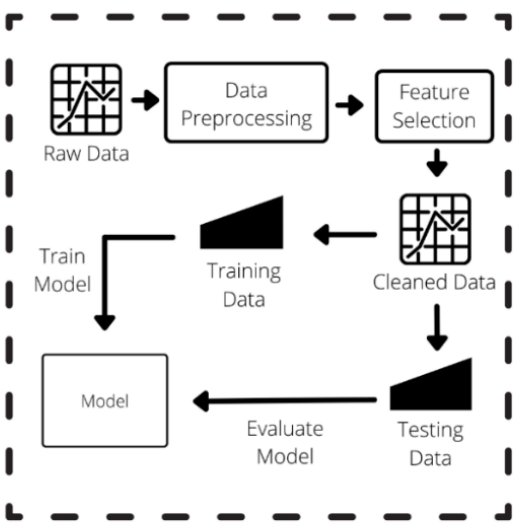

Fig. 2. Model Training Flow.

C. High-Level Screening System Design

The screening system (Fig. 3) consists of two main components: a) system and b) model. The system component requires an input from user to make a screening prediction using the predictive model derived from the model component. The system will response with output (prediction results) and store the results in the database.

D. System Flowchart

The flowchart of the screening system describes the way how the proposed system should work. As shown in Fig. 4, the flow begins with user login. Users need to register their username and password to use the system. When they enter their credentials, the system will verify the user and only allows authenticated users to login into the system. If they fail to verify themselves, the user must input the username and password again. There are two types of users for this system which is the doctor and the nurse. When nurse login, they can enter patient's information and predict the severity of the COVID-19 patient. The system will display the predicted outcome and the score of the severity prediction. The input entered by the nurse will be stored in the database so that doctors can view the data. When the doctor login, the system will show the record of patients and their status. Doctors can keep track of patients' status so that they know which patients are still waiting to be examined and which have been examined.

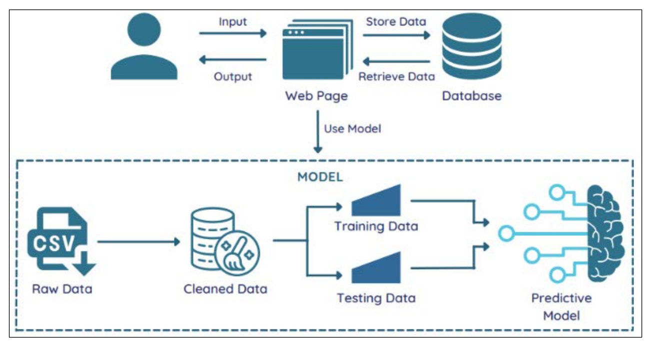

Fig. 3. High-Level Screening System Design.

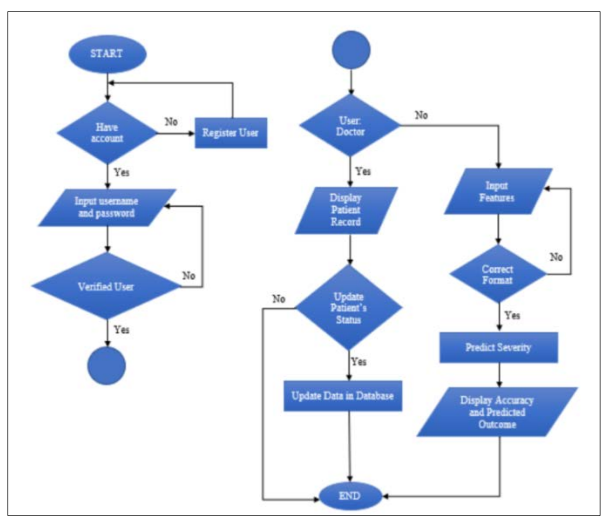

Fig. 4. System Flowchart.

E. Graphical user Interface Design

Fig. 5 shows graphical user interface (GUI) for the screening system that will serve as a way for medical practitioners to interact with the system. The GUI consists of seven web pages with forms for inputs and tables and graph for outputs.

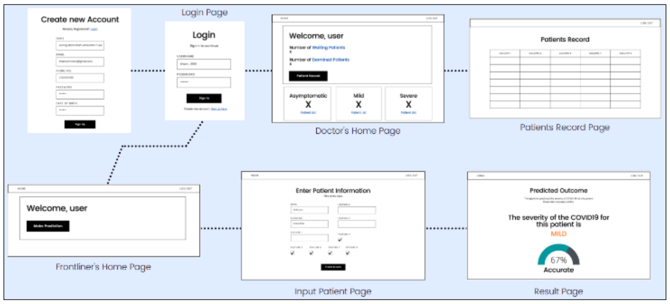

Fig. 5. Proposed Graphical user Interface.

III. Development of Machine Learning Model

A. Dataset

The dataset was obtained from the Springer Nature data repository and consists of 478 patients (https://springernature.figshare.com/articles/dataset/Data_associated_with_the_article_Epidemiological_and_clinical_characteristics_of_imported_cases_with_COVID-19_infection_a_multicentre_study/12159918). The dataset contains the epidemiological and clinical characteristics of COVID-19 cases in China, and consists of 21 variables, including the patient's demographic profile, epidemiological characteristics, clinical data, contact history, case cluster, outcome and type of case. The performance of machine learning models depends upon the quality of the data. Thus, pre-processing was conducted on the dataset to handle missing values, outliers and data generalisation.

After the pre-processing, only eight independent variables that is relevant to COVID-19 symptoms that will be used to predict the dependent variable (type) with three class labels. Table I illustrates the description for each variable.

The original dataset distributions of each class in the target variable are disproportionate which may result in a poor predictive accuracy over the minority class [17] and bias towards the majority class. Thus, to rebalance the data, this study used the synthetic minority oversampling technique (SMOTE) which will increase the minority class to match with majority class. This data balancing technique aims to avoid imbalanced classification in developing predictive models on the dataset. The results before resample and after resample using the SMOTE technique are shown in Fig. 6.

B. Tools

We developed the predictive models using python programming language that provide useful libraries such as Scikit-learn for feature extraction and model training and testing. We designed the GUI forms with text fields and buttons using HTML and CSS. We stored the data in a MySQL database.

C. Pre-processing and Feature Extraction

We analysed and compared the accuracy of three feature selection methods to identify the features or variables that most likely contribute to the severity of illness. The feature selection methods are Pearson Correlation, Random Forest Importance and Recursive Feature Elimination (RFE). Fig. 7 shows the accuracy comparison of feature selection methods applied with the Random Forest model.

TABLE I. Variable Description

Variable	**Description**
Gender	The gender of the patient.
Age	The patient's age in years.
Fever	Whether the patient have fever.
Cough	Whether the patient have cough.
Fatigue	Whether the patient have fatigue.
Dyspnea	Whether the patient have dyspnea.
Headache	Whether the patient have headache.
H-Temperature	The patient's highest temperature.
Type	The type of cases or the level of severity for the COVID-19 case. (1=Asymptomatic, 2=Mild, 3=Severe)

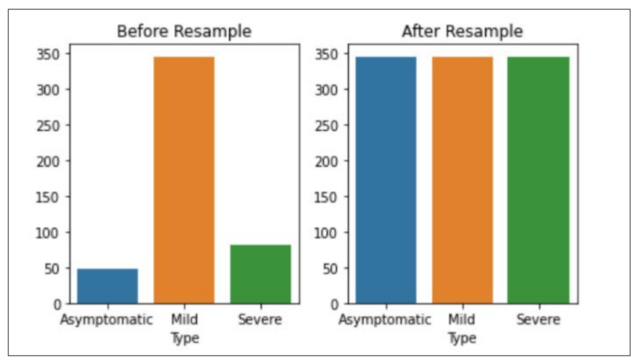

Fig. 6. Original Dataset vs Synthetic Minority Oversampling Technique Dataset.

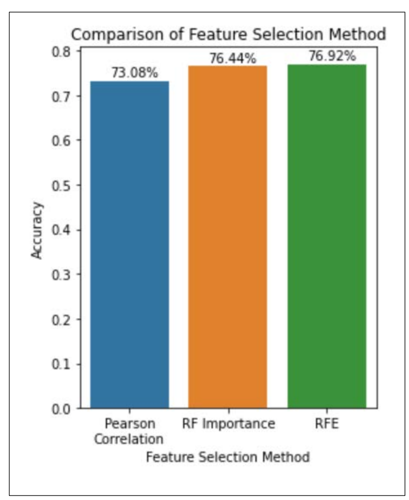

Fig. 7. Comparison of Feature Selection Methods.

The results show that the Random Forest model produces the highest accuracy of 76.92% when using RFE, compared with 76.44% for Random Forest Importance and 73.08% for Pearson Correlation. Thus, RFE was used to determine the importance of the eight features or variables according to the contribution in COVID-19 illness. We used RFE with cross-validation to remove features or variables iteratively and find the most relevant features or variables that contribute to the severity of illness and have strong correlations between the selected features or variables. Fig. 8 shows the accuracy of the model obtained with cross-validation using different numbers of features.

Based on the line chart in Fig. 8, the accuracy of the model is at the highest (78.0%) when the number of features selected is seven. Thus, we used those seven features in this study, which are Gender, Age, Fever, Cough, Fatigue, Dyspnoea and H-Temperature.

D. Model Training

We applied three machine learning techniques: Random Forest (RF), Naïve Bayes (NB) and Gradient Boosting (GB). The brief methodology, working procedure and differences are provided below.

1) Random forest: RF is an ensemble method that combines multiple decision trees (DTs) together in a single forest. It is a powerful supervised machine learning algorithm that is capable of performing both classification and regression tasks with high accuracy [18]–[20]. Various studies in the medical area have used RF algorithms to, for example, diagnose diabetes mellitus [21], [22], identify cervical cancer [23]–[25], or predict the risk of severity for COVID-19 patients at hospital admission [26].

In predicting COVID-19 severity, the RF technique will choose random samples from a given dataset and build an individual decision tree for each sample. One single RF model will have many DTs. Each DT will generate a prediction based on RF technique. As can be seen in Fig. 9, each prediction result contains a vote and the majority votes will be chosen as the final result.

2) Naïve bayes: The NB algorithm works on the Bayesian theorem whereby the probability of a class depends on the probabilities of its variables [27]. The NB classifier is used to maximise the probability of the target class given the features. Previous studies reported that NB can be used to solve the classification problems with multiple classes [28], [29]. In this study, NB can be utilised in the type of COVID-19 severity. Fig. 10 shows the equation used in NB.

3) Gradient boosting: GB is a method to develop classification models by optimising known techniques, such as DT, by adding new learners in a gradual sequential manner [30], [31]. This algorithm is also helpful as a prediction model as it has been used in past research [32], [33] to predict COVID-19. Individually, DT might give a weak prediction ability. However, when combined in an ensemble method, it can improve the accuracy by sequentially upgrading its performance. Fig. 11 shows how boosting in an ensemble method works.

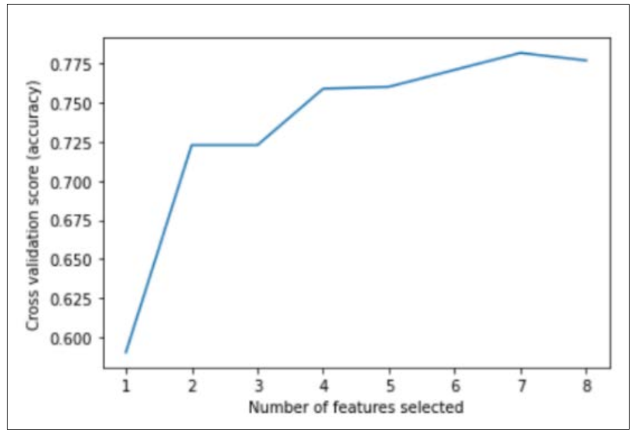

Fig. 8. Accuracy of Model by the Number of Features Selected.

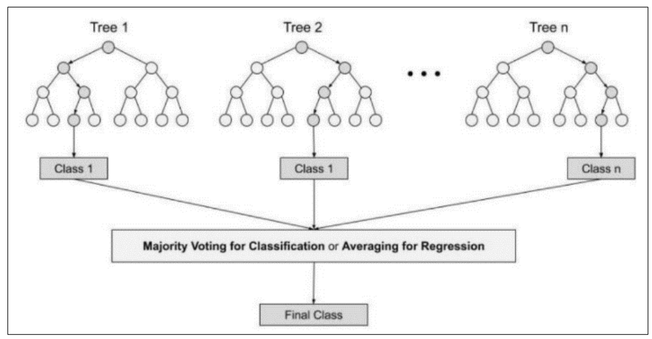

Fig. 9. RF Tree.

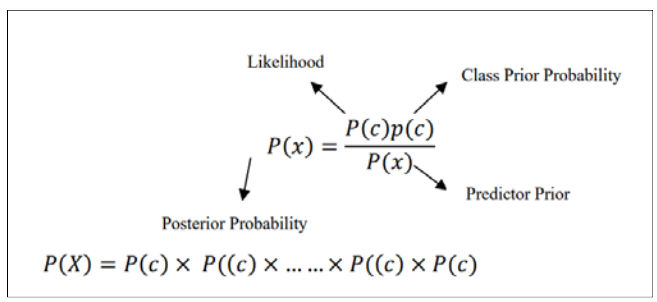

Fig. 10. Equation in NB Classifier [27].

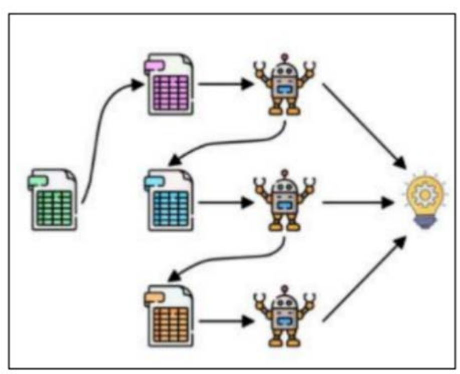

Fig. 11. Boosting Method.

As shown in Fig. 11, although the GB classifier is part of an ensemble method, this method is different from RF in that it does not take the majority vote from each tree. In contrast, it generates the results at a single tree and improves them sequentially until it obtains a good result. The RF, NB, and GB models were run using python software with the abovementioned independent variables or features (Gender, Age, Fever, Cough, Fatigue, Dyspnoea and H-Temperature) and dependent variable (Type), with three class labels.

E. Evaluation Method

We evaluated the performance of machine learning techniques according to evaluation metrics: accuracy, precision, F1-score, recall, sensitivity and specificity. These evaluation metrics used four basic attributes based on confusion matrix which are: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). We calculated the accuracy values using Eq. (1).

$$\text{Accuracy} = TP + TN/(TP + FP + TN + FN) \quad (1)$$

We used precision, recall, F1-score, sensitivity and specificity to support the accuracy values. The method for these metrics is described in Eq. (2), Eq. (3) , Eq. (4) , Eq. (5) and Eq. (6), respectively.

$$\text{Precision} = TP/(TP + FP) \quad (2)$$

$$\text{Recall} = TP/(TP + FN) \quad (3)$$

$$\text{F1-score} = 2 \times (\text{Recall} \times \text{Precision})/(\text{Recall} + \text{Precision}) \quad (4)$$

$$\text{Sensitivity} = TP/(TP + FN) \quad (5)$$

$$\text{Specificity} = TN/(TN + FP) \quad (6)$$

IV. Results

Our main objective is to compare the three machine learning models based on the accuracy performance in this section. In order to analyze the differences, we compare the performance accuracy using the five-fold cross-validation with stratification as a testing method to derive the best predictive model for optimal results. We measure the performance using various metrics including classification accuracy, precision, F1-score, recall, specificity and sensitivity to ensure the predictive model was fit to produce accurate results. Table II shows the performance of the three machine learning models that have been investigated in this study.

Overall, results show that all the machine learning models can be used for predicting the severity of COVID-19 patients. However, the RF model produced the highest accuracy value of 78.4. This may be because of advantages of the RF model, such as building each tree independently and averaging the votes. The averaging method may be advantageous when dealing with multiclassification that aggregates individual predictions into a collective prediction.

The RF and GB models recorded the same values for precision, specificity and sensitivity, which are 91.0, 94.7 and 98.6, respectively. The NB model obtained perfect recall and sensitivity values of 100.

TABLE II. Results of RF, NB and GB for COVID-19 Severity Prediction

Performance measures	Machine learning model		
	Random Forest	**Gradient Boosting**	**Naïve Bayes**
Accuracy	**78.4**	**77.5**	**76.1**
Precision	91.0	91.0	89.0
F1-score	95.0	95.0	94.0
Recall	99.0	99.0	100
Specificity	94.7	94.7	93.3
Sensitivity	98.6	98.6	100

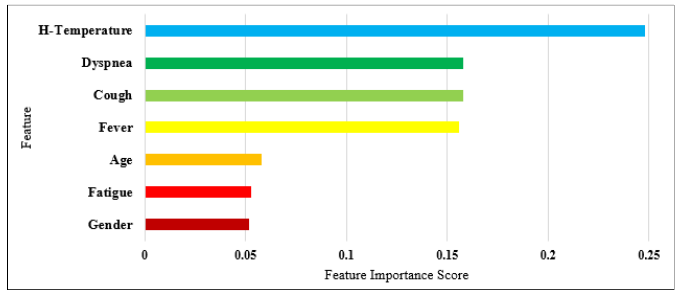

Fig. 12. Visualising Significant Features for COVID-19 Severity Stage Prediction Produced by RF Model.

We conducted another experiment using the RF model to determine the significant features or symptoms in predicting the severity stage of COVID-19 patients. The result is illustrated in Fig. 12, in which the most significant feature is H-Temperature.

V. Discussion

This study set out to report on the performance of machine learning techniques in predicting the severity of COVID-19 patients by using a publicly available dataset from the Springer Nature repository that contains clinical information of COVID-19 patients. Machine learning techniques provide an additional effective way of early screening of patients and do not replace the clinical evaluation.

In our study, results showed that the three machine learning techniques had similar average predictive accuracy in classifying severity of COVID-19 patients (accuracy >0.75). This is consistent with prior findings using an RF model in clinical data [15] and other prior findings using RF, GB, and ensemble learning algorithms in primary care [34]. The RF model outperformed the other machine learning models with the highest accuracy to predict the severity stage of COVID-19.

Despite the similar accuracy, this study found minimum variations for other performance values between the machine learning models. The specificity values were regularly high for RF and GB models, which indicate that the proportion of patients without severe COVID-19 symptoms was correctly classified. The sensitivity and precision values were also consistently high for RF and GB models, indicating that the models identified most of the COVID-19 patients with mild, severe and critical disease severity stages.

The findings of this study have important implications for developing a COVID-19 severity screening system to assist doctors to manage COVID-19 patients before they are examined. Since the outbreak of COVID-19, hospitals have been struggling to keep up with the number of patients because of limited staff. Patients need to be screened so that doctors can prioritise the severe patients as they are at high risk of death. Therefore, more efficient ways to screen the patients are needed because manual screening can be time consuming. By using a COVID-19 severity screening system powered by machine learning models, hospitals can deploy frontliners who are not medical experts to help screen the patients before they undergo actual examination by the doctors. Thus, the time to screen the patients can be shortened.

VI. Conclusion

Machine learning techniques help to reduce the effort and time for medical practitioners to conduct early predictions for healthcare management purposes. As the number of deaths due to COVID-19 increases, a COVID-19 severity screening system is essential to reduce the progression of the disease. In this study, a comparison of three machine learning techniques, RF, NB, and GB were performed, and RF was found to achieve the best performance score. The results of our study show the potential of applying machine learning techniques in the early predictions of a COVID-19 severity screening. Using RF algorithm, body temperature (H temperature) has been found to be important criteria for diagnosing COVID-19 severity level. This may call further investigation to explore temperature data visualization for temperature monitoring of COVID-19 patients. Other than that, a hybrid of machine learning techniques with optimisation algorithms with more data will be examined to further improve accuracy.

Acknowledgment

The authors would like to thank the Ministry of Education Malaysia and Universiti Teknologi MARA (UiTM) for their financial support to this project under UiTM Internal REI Grant 600-IRMI/REI 5/3 (001/2019). We would also like to thank the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Selangor, Malaysia for all the supports.

References

[1] M. E. Hossain, A. Khan, M. A. Moni, and S. Uddin, “Use of Electronic Health Data for Disease Prediction: A Comprehensive Literature Review,” IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 18, no. 2, pp. 745–758, Mar. 2021, doi: 10.1109/TCBB.2019.2937862.

[2] A. J. Steele, S. C. Denaxas, A. D. Shah, H. Hemingway, and N. M. Luscombe, “Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease,” PLoS One, vol. 13, no. 8, 2018, doi: 10.1371/journal.pone.0202344.

[3] S. A.G. et al., “Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma,” American Journal of Gastroenterology, vol. 108, no. 11. 2013.

[4] A. Guo, N. R. Mazumder, D. P. Ladner, and R. E. Foraker, “Predicting mortality among patients with liver cirrhosis in electronic health records with machine learning,” PLoS One, vol. 16, no. 8 August, 2021, doi: 10.1371/journal.pone.0256428.

[5] A. S. Adly, A. S. Adly, and M. S. Adly, “Approaches Based on artificial intelligence and the internet of intelligent things to prevent the spread of COVID-19: Scoping review,” Journal of Medical Internet Research, vol. 22, no. 8. 2020, doi: 10.2196/19104.

[6] K. C. Santosh, “AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data,” J. Med. Syst., 2020, doi: 10.1007/s10916-020-01562-1.

[7] L. S. Khoo, A. H. Hasmi, M. A. Ibrahim, and M. S. Mahmood, “Management of the dead during COVID-19 outbreak in Malaysia,” Forensic Sci. Med. Pathol., 2020, doi: 10.1007/s12024-020-00269-6.

[8] K. Mizumoto and G. Chowell, “Transmission potential of the novel coronavirus (COVID-19) onboard the diamond Princess Cruises Ship, 2020,” Infect. Dis. Model., 2020, doi: 10.1016/j.idm.2020.02.003.

[9] H. Estiri, Z. H. Strasser, J. G. Klann, P. Naseri, K. B. Wagholikar, and S. N. Murphy, “Predicting COVID-19 mortality with electronic medical records,” npj Digit. Med., vol. 4, no. 1, 2021, doi: 10.1038/s41746-021-00383-x.

[10] A. Alotaibi, M. Shiblee, and A. Alshahrani, “Prediction of severity of covid-19-infected patients using machine learning techniques,” Computers, vol. 10, no. 3, Mar. 2021, doi: 10.3390/computers10030031.

[11] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, “Artificial Intelligence (AI) applications for COVID-19 pandemic,” Diabetes Metab. Syndr. Clin. Res. Rev., 2020, doi: 10.1016/j.dsx.2020.04.012.

[12] J. R. Dietz et al., “Recommendations for prioritization, treatment, and triage of breast cancer patients during the COVID-19 pandemic. the COVID-19 pandemic breast cancer consortium,” Breast Cancer Research and Treatment, vol. 181, no. 3. 2020, doi: 10.1007/s10549-020-05644-z.

[13] H. Estiri, Z. H. Strasser, J. G. Klann, P. Naseri, K. B. Wagholikar, and S. N. Murphy, “Predicting COVID-19 mortality with electronic medical records,” npj Digit. Med., vol. 4, no. 1, Dec. 2021, doi: 10.1038/s41746-021-00383-x.

[14] A. Kumar, P. K. Gupta, and A. Srivastava, “A review of modern technologies for tackling COVID-19 pandemic,” Diabetes Metab. Syndr. Clin. Res. Rev., 2020, doi: 10.1016/j.dsx.2020.05.008.

[15] A. Alotaibi, M. Shiblee, and A. Alshahrani, “Prediction of severity of covid-19-infected patients using machine learning techniques,” Computers, vol. 10, no. 3, 2021, doi: 10.3390/computers10030031.

[16] W. A. Abbasi et al., “COVIDC: An expert system to diagnose COVID-19 and predict its severity using chest CT scans: Application in radiology,” Informatics Med. Unlocked, vol. 23, Jan. 2021, doi: 10.1016/j.imu.2021.100540.

[17] L. Wei, W. Luo, J. Weng, Y. Zhong, X. Zhang, and Z. Yan, “Machine Learning-Based Malicious Application Detection of Android,” IEEE Access, vol. 5, pp. 25591–25601, 2017, doi: 10.1109/ACCESS.2017.2771470.

[18] R. Srivatsan, P. N. Indi, S. Agrahari, S. Menon, and S. D. Ashok, “Machine learning based prognostic model and mobile application software platform for predicting infection susceptibility of COVID-19 using healthcare data,” Res. Biomed. Eng., vol. 38, no. 1, pp. 59–70, Mar. 2022, doi: 10.1007/s42600-020-00103-6.

[19] N. Amoroso, R. Cilli, T. Maggipinto, A. Monaco, S. Tangaro, and R. Bellotti, “Satellite data and machine learning reveal a significant correlation between NO2 and COVID-19 mortality,” Environ. Res., vol. 204, p. 111970, Mar. 2022, doi: 10.1016/j.envres.2021.111970.

[20] I. Alam, D. Md. Farid, and R. J. F. Rossetti, “The prediction of traffic flow with regression analysis,” in Advances in Intelligent Systems and Computing, 2019, vol. 813, doi: 10.1007/978-981-13-1498-8_58.

[21] X. Wang et al., “Exploratory study on classification of diabetes mellitus through a combined Random Forest Classifier,” BMC Med. Inform. Decis. Mak., vol. 21, no. 1, p. 105, Dec. 2021, doi: 10.1186/s12911-021-01471-4.

[22] N. Komal Kumar, D. Vigneswari, M. Vamsi Krishna, and G. V. Phanindra Reddy, “An Optimized Random Forest Classifier for Diabetes Mellitus,” in Advances in Intelligent Systems and Computing, vol. 813, 2019, pp. 765–773.

[23] K. P. Win, Y. Kitjaidure, K. Hamamoto, and T. Myo Aung, “Computer-Assisted Screening for Cervical Cancer Using Digital Image Processing of Pap Smear Images,” Appl. Sci., vol. 10, no. 5, p. 1800, Mar. 2020, doi: 10.3390/app10051800.

[24] S. F. Abdoh, M. Abo Rizka, and F. A. Maghraby, “Cervical Cancer Diagnosis Using Random Forest Classifier With SMOTE and Feature Reduction Techniques,” IEEE Access, vol. 6, pp. 59475–59485, 2018, doi: 10.1109/ACCESS.2018.2874063.

[25] D. N. Diniz et al., “A Hierarchical Feature-Based Methodology to Perform Cervical Cancer Classification,” Appl. Sci., vol. 11, no. 9, p. 4091, Apr. 2021, doi: 10.3390/app11094091.

[26] G. Wu et al., “Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: An international multicentre study,” Eur. Respir. J., vol. 56, no. 2, 2020, doi: 10.1183/13993003.01104-2020.

[27] D. Berrar, “Bayes’ theorem and naive bayes classifier,” in Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, vol. 1–3, 2018.

[28] L. Yahaya, N. David Oye, and E. Joshua Garba, “A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques,” Am. J. Artif. Intell., vol. 4, no. 1, 2020, doi: 10.11648/j.ajai.20200401.12.

[29] W. N. L. W. H. Ibeni, M. Z. M. Salikon, A. Mustapha, S. A. Daud, and M. N. M. Salleh, “Comparative analysis on bayesian classification for breast cancer problem,” Bull. Electr. Eng. Informatics, vol. 8, no. 4, 2019, doi: 10.11591/eei.v8i4.1628.

[30] N. Chakrabarty, T. Kundu, S. Dandapat, A. Sarkar, and D. K. Kole, “Flight arrival delay prediction using gradient boosting classifier,” in Advances in Intelligent Systems and Computing, 2019, vol. 813, doi: 10.1007/978-981-13-1498-8_57.

[31] A. Pina et al., “Virtual genetic diagnosis for familial hypercholesterolemia powered by machine learning,” Eur. J. Prev. Cardiol., 2020, doi: 10.1177/2047487319898951.

[32] M. Kukar et al., “COVID-19 diagnosis by routine blood tests using machine learning,” Sci. Rep., vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-90265-9.

[33] Y. Zoabi, S. Deri-Rozov, and N. Shomron, “Machine learning-based prediction of COVID-19 diagnosis based on symptoms,” npj Digit. Med., vol. 4, no. 1, 2021, doi: 10.1038/s41746-020-00372-6.

[34] R. F. Albuquerque Paiva de Oliveira, C. J. A. Bastos Filho, A. C. A. M. V. F. de Medeiros, P. J. Buarque Lins dos Santos, and D. Lopes Freire, “Machine learning applied in SARS-CoV-2 COVID 19 screening using clinical analysis parameters,” IEEE Lat. Am. Trans., vol. 19, no. 6, pp. 978–985, Jun. 2021, doi: 10.1109/TLA.2021.9451243.