IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 4, December 2024, pp. 4472~4479
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i4.pp4472-4479

Mortality prediction of COVID-19 patients using supervised machine learning

Husnul Khuluq[1], Prasandhya Astagiri Yusuf[2], Dyah Aryani Perwitasari[3], Thang Nguyen[4]
[1]Department of Pharmacy, Faculty of Health Sciences, Universitas Muhammadiyah Gombong, Kebumen, Indonesia
[2]Department of Medical Physiology and Biophysics/Medical Technology Cluster IMERI, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia
[3]Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[4]Department of Pharmacology and Clinical Pharmacy, Faculty of Pharmacy, University of Medicine and Pharmacy, Can Tho, Vietnam

Article Info

Article history:

Received Jul 31, 2023
Revised Nov 6, 2023
Accepted Dec 12, 2023

Keywords:

Attribute evaluation
Confusion matrix
Data balancing
Synthetic minority over-sampling technique
Ten-fold cross-validation

ABSTRACT

Hospitalized patients with COVID-19 are at higher risk of mortality. Machine learning (ML) algorithms have been proposed as a possible strategy for predicting mortality rates among patients hospitalized with COVID-19. This study analyzed various ML algorithms and identified the best model to predict COVID-19 mortality based on demographic, clinical, and laboratory data collected at registration. Data from 4,314 eligible patients (3,384 survivors and 930 who died) was collected from the register of three hospitals in Yogyakarta province, Indonesia, based on the confirmed predictors. Next, ML algorithms were utilized to predict mortality. Finally, the confusion matrix was used to evaluate how effective the models performed. The best five predictors from 26 features were myocardial infarction, SpO_2, neutrophil, D-dimer, and creatinine. The results indicate that the random forest algorithm showed better performance than other ML algorithms in terms of accuracy, sensitivity, precision, specificity, and area under the curve (AUC), achieving values of 84.15%, 84.0%, 84.1%, 83.9%, and 90.02%, respectively. Implementing ML techniques can accurately predict the mortality rate associated with COVID-19. Therefore, this predictive model can help clinicians and hospitals predict COVID patients with a greater risk of death and effectively target more appropriate treatments.

Corresponding Author:

Husnul Khuluq
Department of Pharmacy, Faculty of Health Sciences, Universitas Muhammadiyah Gombong
St. Yos Sudarso 461, Gombong, Kebumen, Indonesia
Email: husnulkhuluq@unimugo.ac.id

1. INTRODUCTION

The new COVID-19 was first identified in Wuhan District, Republic of China, in December 2019 [1]. Since then, this infectious disease has rapidly spread worldwide. The World Health Organization (WHO) established the outbreak as a pandemic in January 2020 [2]. The COVID-19 virus showed various clinical outcomes, ranging from asymptomatic or mild symptoms to severe complications and, in specific cases, fatalities. The virus is highly transmissible and has rapidly spread worldwide, becoming a significant global health threat. The rapid transmission of COVID-19 caused a significant failure of medical resources and a decrease in frontline medical personnel [3]–[5].

Many COVID-19 patients rapidly get worse after having had initially mild symptoms. This demonstrates the importance of improved risk-strategy models. By using predictive models, clinicians may determine whether patients are more likely to die, targeting urgent help to them to ensure fewer individuals die

[5]–[7]. Therefore, to reduce the impact on the medical system and provide patients with the most effective care possible, it is crucial to exactly predict the prognosis of the disease and prioritize the treatment of patients who are in critical condition. Clinicians and health authorities have relied on computing and statistical model projections because of the unpredictable nature of their effects. [8], [9]. In solution to the abovementioned difficulties medical facilities worldwide are attempting to implement machine learning (ML) classifiers to achieve precise decision-making, hence reducing the inaccurate assessments of clinicians. [10], [11]. As a subfield of artificial intelligence (AI), ML provides the extraction of powerful predictive models from the mining of large raw databases [12]. ML algorithms may decrease inconsistency and complexity by utilizing evidence-based medicine for risk evaluation, assessment, prediction, and treatment strategies. They facilitate accurate medical decision-making and can be expected to enhance patients' condition and treatment quality [13], [14].

The objective of this study was to construct a model for predicting the risk of death for COVID-19 based on ML algorithms that utilized typical clinical, demographic, and laboratory data from patients. Specifically, we aimed to address the following questions. What are the most significant predictors that can be used to predict mortality in COVID-19 hospitalized patients? Additionally, we sought to determine the best ML method for constructing a model to predict mortality.

2. METHOD

2.1. Study population

This retrospective cohort study involved an overall of 4,314 COVID-19 patients who were registered to three referred hospitals in Yogyakarta, Indonesia, from January 2020 to December 2022. The hospitals studied were RS Akademik UGM, PKU Muhammadiyah Yogyakarta, and PKU Muhammadiyah Gamping General Hospital. The criteria for inclusion were: i) COVID-19 infection confirmed by reverse-transcriptase polymerase chain reaction (RT-PCR) assays on material gathered by a nasopharyngeal and oropharyngeal swab; ii) patients were hospitalized; and iii) they were aged over 18 years. We excluded hospitalized patients who had missing/incomplete data, were deceased while being admitted, or were sent to another hospital. From the patients' medical records, we gathered the following data: age, gender, cardiovascular risk factors (high blood pressure, diabetes, high cholesterol), primary comorbidities (e.g., chronic kidney disease), histories of coronary artery disease, chronic obstructive pulmonary disease, and peripheral vascular disease, as well as the laboratory results from their current treatment. The patients who died in the hospital and those who survived were evaluated independently. Figure 1 provides a flowchart of the study design.

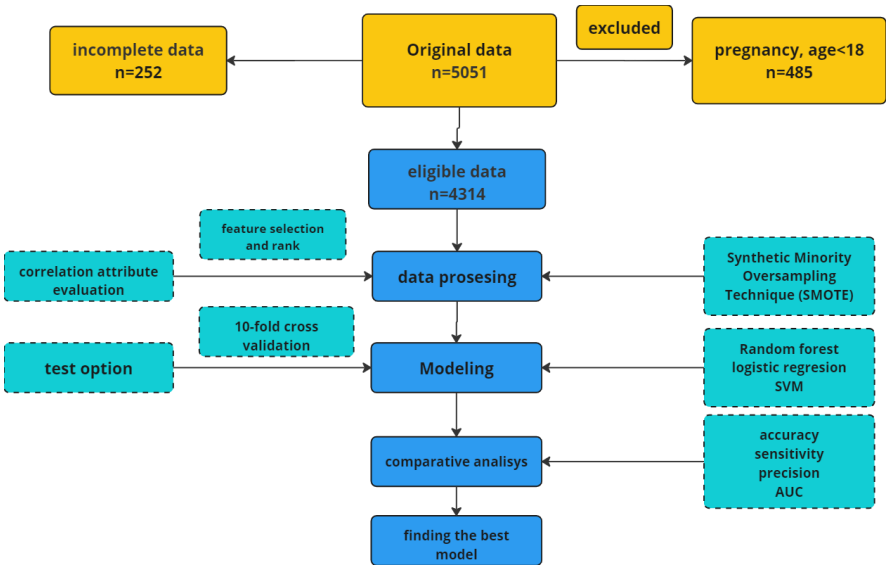

Figure 1. The study design flowchart

2.2. Machine learning algorithm

Four ML techniques-logistic regression (LR), support vector machines (SVMs), random forest (RF), and naïve Bayes (NB)-were used to construct our models using the open-source Waikato environment for knowledge analysis (Weka) software version 3.8.6 [15]. We selected these ML techniques because they are

commonly used and represent different approaches. LR and SVMs are statistical methods, while RF and NB are probability-based [16].

2.3. Cross-validation

Standard ten-fold cross-validation was used to test all the models. In this method, all the data is first randomly split into ten equal portions. Then, the models are trained on nine portions and tested on the tenth. This process is repeated ten times, each with a different tested portion. The evaluation scores are then made by summing the results [17]. This cross-validation method is commonly utilized in ML and data mining because it differs from the traditional split instance method. This method contributed to minimizing the difference in prediction errors, made better use of data for both training and validation, avoided any overlap or overfitting within the test and validation data, and protected the experimental theory by random data splitting [18].

2.4. Predictors

We selected 26 relevant predictors. The clinical variables included demographic data about the patient, such as age and gender. A description of chronic diseases such as diabetes, hypertension, heart and kidney disease, and cancer was recorded. Lymphocytes, leukocytes, thrombocytes, neutrophils, D-dimer, glucose, creatinine, and hemoglobin were evaluated as laboratory parameters.

2.5. Outcome variable

The outcome variable was 'deceased,' which accounted for mortality from COVID-19 while in the hospital. The variable had two possible values. These were 'Yes' if the patient was dead and 'No' if they survived.

2.6. Data balancing

The difficulties caused by unbalanced data-when the classes are not grouped equally-are among the main obstacles to using ML tools. In our dataset, there are disproportionately more samples associated with the survivor class (3,384 individuals) than the deceased class (930 individuals). Consequently, the trained models frequently produce biased results regarding the majority class, whereas the ML models are more likely to classify recent finding as belonging to the majority class. In this research, we used the synthetic minority over-sampling technique (SMOTE) from the imbalanced-learn toolkit to correct this class imbalance [19].

2.7. Feature selection

The selection of features is a method utilized frequently in prediction, structure identification, and classifier modelling to reduce the measurements and complexity of a dataset by eliminating unnecessary and unnecessary attributes. Several selections of attributes methods, including correlation attribute evaluation, information gain ratio attribute evaluation (GA), forward elimination, backward elimination, and one rule attribute evaluation, have been provided for determining appropriate attributes for predictive models [20]. In this research, the features were chosen using the correlation attribute evaluation method in the Weka v3.8.6 application.

2.8. Model evaluation

Constructing an effective ML model also requires model performance evaluation. Confusion matrix performance indicators were used to assess the predictive models as shown in Table 1. We used assessment indicators, such as accuracy, specificity, precision, sensitivity, and area under the curve (AUC) from the receiver operating characteristic (ROC) graph, to assess the performance of the predictive models. AUC value is the area under the ROC curve, with 1 representing the maximum (perfectly distinguish between all the positive and the negative class points). The ROC curve categorizes a patient's disease status as positive or negative, depending on test outcomes. It aims to identify the most effective cut-off value that provides the highest diagnostic performance [21]. Determining the cut-off value for ROC curve calculation involves identifying the point on the curve at which the sensitivity and specificity of the test exhibit equality [22], [23]. Various approaches are available for determining the location of this point, including an analytical method that relies on assessing the slope of the ROC curve [22], alternatively, it can be achieved through the minimization of the sum of absolute differences between the AUC and the corresponding sensitivity and specificity values [24]. Determining the best cut-off value involves more than just increasing sensitivity and specificity; it necessitates striking a suitable balance between these measures, considering other criteria. In situations where a disease exhibits high contagiousness or is linked to severe consequences, like the case of COVID-19, prioritizing sensitivity above specificity becomes essential [21]. We compared AUCs from four ML algorithm and identify the largest one as the best model. Lastly, all these evaluating criteria were analyzed based on how effectively they performed in Table 2 to find the best model for predicting COVID-19 mortality.

True positive (TP) is the number of instances that the algorithm accurately classifies as positive (deceased). False positive (FP) is the number of instances that are falsely classified as positive by the algorithm. False negative (FN) is the number of cases that are falsely classified as negative by the algorithm. True negative (TN) is the number of instances that the algorithm accurately classifies as negative (survived).

Table 1. Confusion matrix

Output	Predicted values	
	Deceased (+)	Survived (-)
Actual value		
Deceased (+)	TP	TN
Survived (-)	FP	FN

Table 2. The performance evaluation measures

Performance criteria	Formula
Accuracy	(TP+TN)/(TP+TN+FP+FN)
Precision	TP/(TP+FP)
Sensitivity/ Recall	TP/(TP+FN)
Specificity	TN/(TN+FP)

3. RESULTS AND DISCUSSION

3.1. Results

3.1.1. Feature identification and selection

Twenty-six features were examined as possible predictors of COVID-19 mortality risk. These features were classified into three different groups: demographics, comorbidities, and laboratory tests. Subsequently, using the correlation attribute evaluation method, the degree to which each factor contributed to the prediction of COVID-19 hospitalized mortality was evaluated as shown in Table 3. Myocardial infarction had the highest importance, whereas hypertension had the lowest importance in the feature selection set for predicting COVID-19 mortality. Their average correlation coefficients were 0.367±0.004 and 0.003±0.002, respectively. DM type 2: diabetes mellitus type 2; DM type 1: diabetes mellitus type 1; SpO_2: peripheral oxygen saturation; COPD: chronic obstructive pulmonary disease; CHF: congestive heart failure; AKD: acute kidney disease; CKD: chronic kidney disease.

Table 3. Features that have an impact on the ability to predict mortality in patients with COVID-19

No	Feature's name	Average degree of importance	No	Features name	Average degree of importance
1	Myocardial infarction	0.367 ± 0.004	14	DM type 1	0.076 ± 0.004
2	SpO_2	0.284 ± 0.005	15	Hemoglobin	0.064 ± 0.004
3	Neutrophil	0.263 ± 0.003	16	AKD	0.062 ± 0.005
4	D-dimer	0.229 ± 0.004	17	Asthma	0.061 ± 0.002
5	Creatinine	0.228 ± 0.003	18	Thrombocytes	0.061 ± 0.005
6	Respiratory failure	0.225 ± 0.005	19	Age	0.044 ± 0.004
7	Lymphocyte	0.191 ± 0.003	20	Cerebral Infraction	0.038 ± 0.004
8	Glucose	0.178 ± 0.004	21	Pneumonia	0.035 ± 0.005
9	Leukocytes	0.167 ± 0.006	22	DM type 2	0.022 ± 0.004
10	Blood pressure	0.119 ± 0.005	23	COPD	0.019 ± 0.004
11	Septic shock	0.102 ± 0.004	24	Anemia	0.016 ± 0.004
12	CKD	0.087 ± 0.003	25	CHF	0.014 ± 0.005
13	Gender	0.082 ± 0.004	26	Hypertension	0.003 ± 0.002

3.2. Patient selection

We obtained data from 4,314 hospitalized patients from three hospitals in Yogyakarta Province, Indonesia. In total, 2,356 patients (55.74%) were male and 1,958 (44.26%) were female, and the median age was 57.25 years (range 18–87). Of these, 3,384 (78.44%) recovered, and 930 (21.55%) died.

3.3. Data balancing

There are two created datasets listed in Table 4. The first is the original set of data, which contains 3,384 instances in the survivor class and 930 in the deceased class. Following the execution of the SMOTE algorithm, the number of minority class cases was increased by generating synthetic samples, and the resulting dataset was then stored in the second dataset. This data set includes 3,382 cases of the survivor class and 1,860

instances of the updated deceased class. We executed the SMOTE steps to construct a balanced dataset, which was then used to train and test the COVID-19 predictor.

Table 4. The SMOTE methods' results

Dataset number	Technique used	Numbers in the survivor class	Numbers in the deceased class
1	—	3,384	930
2	SMOTE	3,382	1,860

3.4. Developing and evaluating models

After choosing the best subset of features, we applied different ML methods to make a predictive model. Four ML algorithms (LR, RF, NB, and SVM) were presented to construct models that predict COVID-19 mortality. Then the performance of every model was assessed by its sensitivity, specificity, accuracy, precision, and AUC as shown in Table 5. The data presented in Table 5 demonstrates that the RF algorithms, which achieved a sensitivity of 84.1%, an accuracy of 84.15%, a precision of 84 %, and an AUC of 90.02%, demonstrated better results in predicting COVID-19 in-hospital mortality compared to the other ML algorithms. Figure 2 shows the results for the chosen ML algorithms. Figure 3 shows the comparison of ROC curves for the utilized algorithm.

Table 5. Performance evaluation of the selected ML mortality prediction algorithms for COVID-19

Algorithm	Sensitivity/Recall (%)	Accuracy (%)	Precision (%)	ROC (%)
LR	77.8	77.79	77.3	85.1
RF	84.1	84.15	84	90.02
SVM	77.9	77.85	77.4	73.9
NB	74.3	74.32	73.6	81.1

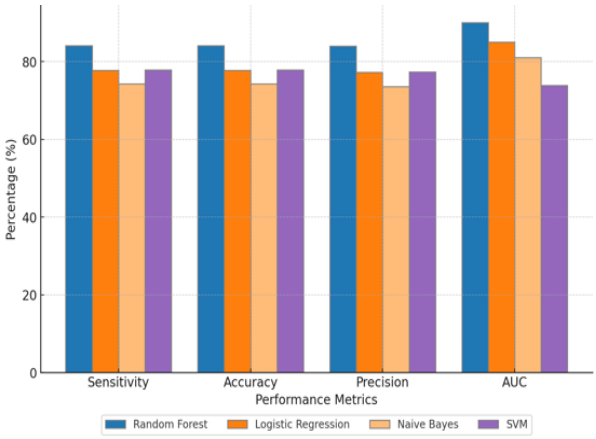

Figure 2. Visual comparisons of the capability of ML algorithms for COVID-19 mortality prediction

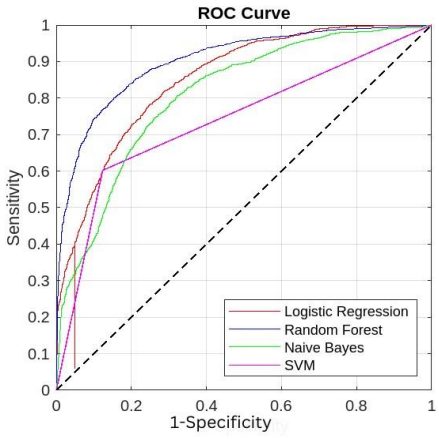

Figure 3. ROC curves for utilized ML algorithms

4. DISCUSSION

The objective of this study was to construct and evaluate ML models built using the most relevant factors in analyzing the mortality risk from COVID-19, which were found within an extensive examination of the existing research. Utilizing a dataset of confirmed COVID-19 hospitalized patients, the LR, RF, SVM, and NB algorithms were developed to achieve this objective. The findings demonstrated that the RF algorithm outperformed the other three ML techniques with an accuracy of 84.15%, a sensitivity of 84.1%, a precision of 84%, and an AUC of 90.02%. The ROC curve is a widely used technique for selecting the ideal cut-off point for a test and assessing the accuracy of diagnostic tests [25]. Our findings demonstrated that the RF model has a high accuracy prediction performance, as the AUC was above 90%. In contrast, the other models had a moderate accuracy prediction performance, with the AUC ranging from 70% to 90% [26].

ML methods have been evaluated for COVID-19 mortality prediction. Yadaw *et al.* [27] evaluated RF, LR, SVM, and XGBoost with a different dataset (n=3,841) to predict COVID-19 mortality. The XGBoost

model achieved the highest AUC at 91%. Gao *et al*. [14], analyzed data from 2,520 COVID-19 hospitalized patients, finding that the model with the best performance in terms of AUC (97.60%) was one constructed to predict mortality with a neural network, compared to another models constructed by LR, SVM, gradient boosting and a decision tree. According to Gao *et al.* [28], four ML models were tested using the data of 10,237 patients, finding that the best-performing technique was SVM, resulting in ROC of 96.3%.

The five most important variables in this study were myocardial infarction, SpO_2, neutrophil, D-dimer, and creatinine. Neutrophils are immune system cells that help the body fight off bacterial and viral infections [29]. However, their precise function in the immune system's defense against the COVID-19 virus is not clearly understood. In human COVID-19 studies, neutrophils have been found to get into the lungs, although this has not been found to be important in animal studies [30]. According to our findings, myocardial infarction was the most relevant in predicting death out of all 26 factors. A higher neutrophil count will result in increased platelet activity, lower fibrinolysis, and a general reduction in the endothelium's capacity to inhibit blood coagulation [31], [32]. Patients who experience a myocardial infarction have higher fibrinogen levels despite fibrinogen being a coagulant factor [33].

The COVID-19 patient population has been the subject of a number D-dimer studies, which have all found an association between higher D-dimer levels and an increased risk of mortality [34]–[36]. It was discovered that between 36% and 43% of the COVID-19 population had an abnormally high level of D-dimer [37]. When determining the severity of a thrombotic event, the D-dimer is a well-known laboratory measure that is frequently applied [38]. It is now generally accepted that people with COVID-19 have an elevated level of coagulation and that elevated D-dimer levels are a response to the prothrombotic process that occurs in advantage of the cardiovascular systems of these patients [39], [40].

Patients diagnosed with severe illness had serum creatinine and urea levels that were considerably higher than those defined as having mild or moderate illness. Higher than normal creatinine and urea levels in individuals with COVID-19 may indicate abnormal renal function. However, they may also indicate impaired glomerular filtration linked to heart failure [41]. An increased mortality risk is associated with respiratory symptoms such as respiratory failure and low oxygen levels (SpO_2<90%). There is an association between these respiratory symptoms and a higher mortality risk. Furthermore, this field of research has been explored in a population in the United States [42].

5. CONCLUSION

The RF model achieved the highest classification accuracy, sensitivity, precision, and AUC compared to the other three ML algorithms. The proposed model can be utilized effectively to predict the mortality risk of hospitalized COVID-19 patients and manage the use of limited hospital resources, especially in Yogyakarta, Indonesia. For future studies, the model can be optimized with additional important features (such as vaccines) and including larger and multicenter datasets. Moreover, the implementation of this predictive model is potential to help clinicians and hospitals predict COVID patients with a greater risk of death and effectively target more appropriate treatments. Limitations: first, radiological results, which may have been valuable as a predictor, were excluded from the data used for this analysis. Second, because the prognosis of patients may be substantially influenced by their treatments, we assumed that all of these patients had been given regular therapy. Third, the study did not consider whether or not patients had been vaccinated. Strength: this study is the first to document and utilized a large dataset of 4,314 COVID-19 cases in Yogyakarta, Indonesia to evaluate the actual distribution, giving an accurate estimation as the sample size increased.

ACKNOWLEDGEMENTS

The authors are grateful for Universitas Ahmad Dahlan in the form of a Penelitian Disertasi Doktor Program to Husnul Khuluq supervised by Dyah Aryani Perwitasari (Grant number 157/E5/PG.02.00.PT/2022).

REFERENCES

[1] K. Moulaei, M. Shanbehzadeh, Z. M. -Taghiabad, and H. K. -Arpanahi, "Comparing machine learning algorithms for predicting COVID-19 mortality," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, 2022, doi: 10.1186/s12911-021-01742-0.

[2] H. Shi *et al.*, "Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study," *The Lancet Infectious Diseases*, vol. 20, no. 4, pp. 425–434, 2020, doi: 10.1016/S1473-3099(20)30086-4.

[3] Y. Liu *et al.*, "A COVID-19 risk assessment decision support system for general practitioners: Design and development study," *Journal of Medical Internet Research*, vol. 22, no. 6, 2020, doi: 10.2196/19786.

[4] M. Z. Alom, M. M. S. Rahman, M. S. Nasrin, T. M. Taha, and V. K. Asari, "COVID_MTNet: COVID-19 detection with multi-task deep learning approaches," *arXiv-Electrical Engineering and Systems Science*, pp. 1-11, 2020, doi: 10.48550/arXiv.2004.03747.

[5] A. Bansal, R. P. Padappayil, C. Garg, A. Singal, M. Gupta, and A. Klein, "Utility of artificial intelligence amidst the COVID 19 pandemic: a review," *Journal of Medical Systems*, vol. 44, no. 9, 2020, doi: 10.1007/s10916-020-01617-3.

[6] Z. Zhao *et al.*, "Prediction model and risk scores of ICU admission and mortality in COVID-19," *PLoS ONE*, vol. 15, no. 7, 2020,

doi: 10.1371/journal.pone.0236618.
[7] H. Hu, N. Yao, and Y. Qiu, "Comparing rapid scoring systems in mortality prediction of critically ill patients with novel coronavirus disease," *Academic Emergency Medicine*, vol. 27, no. 6, pp. 461–468, 2020, doi: 10.1111/acem.13992.
[8] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *The BMJ*, vol. 369, 2020, doi: 10.1136/bmj.m1328.
[9] 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," *European Respiratory Journal*, vol. 56, no. 2, 2020, doi: 10.1183/13993003.01104-2020.
[10] L. Yan *et al.*, "An interpretable mortality prediction model for COVID-19 patients," *Nature Machine Intelligence*, vol. 2, no. 5, pp. 283–288, 2020, doi: 10.1038/s42256-020-0180-7.
[11] Z. Malki, E. S. Atlam, A. E. Hassanien, G. Dagnew, M. A. Elhosseini, and I. Gad, "Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches," *Chaos, Solitons and Fractals*, vol. 138, 2020, doi: 10.1016/j.chaos.2020.110137.
[12] M. Shanbehzadeh, R. Nopour, and H. Kazemi-Arpanahi, "Comparison of four data mining algorithms for predicting colorectal cancer risk," *Journal of Advances in Medical and Biomedical Research*, vol. 29, no. 133, pp. 100–108, 2021, doi: 10.30699/jambs.29.133.100.
[13] M. E. Shipe, S. A. Deppen, F. Farjah, and E. L. Grogan, "Developing prediction models for clinical use using logistic regression: An overview," *Journal of Thoracic Disease*, vol. 11, pp. S574–S584, 2019, doi: 10.21037/jtd.2019.01.25.
[14] Y. Gao *et al.*, "Machine learning based early warning system enables accurate mortality risk prediction for COVID-19," *Nature Communications*, vol. 11, no. 1, 2020, doi: 10.1038/s41467-020-18684-2.
[15] S. K. Joseph, "Weka about the tutorial copyright and disclaimer," *Tutorials Point*, pp. 1–13, 2019, [Online]. Available: https://store.tutorialspoint.com.
[16] R. J. Kate, R. M. Perez, D. Mazumdar, K. S. Pasupathy, and V. Nilakantan, "Prediction and detection models for acute kidney injury in hospitalized older adults," *BMC Medical Informatics and Decision Making*, vol. 16, no. 1, 2016, doi: 10.1186/s12911-016-0277-4.
[17] N. Japkowicz and M. Shah, "Evaluating learning algorithms: A classification perspective," *Evaluating Learning Algorithms: A Classification Perspective*, vol. 9780521196000, pp. 1–406, 2011, doi: 10.1017/CBO9780511921803.
[18] C. C. Wu *et al.*, "Prediction of fatty liver disease using machine learning algorithms," *Computer Methods and Programs in Biomedicine*, vol. 170, pp. 23–29, 2019, doi: 10.1016/j.cmpb.2018.12.032.
[19] M. Shanbehzadeh, A. Orooji, and H. K. -Arpanahi, "Comparing of data mining techniques for predicting in-hospital mortality among patients with covid-19," *Journal of Biostatistics and Epidemiology*, vol. 7, no. 2, pp. 154–169, 2021, doi: 10.18502/jbe.v7i2.6725.
[20] V. H. Nhu *et al.*, "Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms," *International Journal of Environmental Research and Public Health*, vol. 17, no. 8, 2020, doi: 10.3390/ijerph17082749.
[21] F. S. Nahm, "Receiver operating characteristic curve: overview and practical use for clinicians," *Korean Journal of Anesthesiology*, vol. 75, no. 1, pp. 25–36, 2022, doi: 10.4097/kja.21209.
[22] F. Habibzadeh, P. Habibzadeh, and M. Yadollahie, "On determining the most appropriate test cut-off value: The case of tests with continuous results," *Biochemia Medica*, vol. 26, no. 3, pp. 297–307, 2016, doi: 10.11613/BM.2016.034.
[23] K. H. -Tilaki, "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation," *Caspian Journal of Internal Medicine*, vol. 4, no. 2, pp. 627–635, 2013.
[24] I. Unal, "Defining an optimal cut-point value in ROC analysis: An alternative approach," *Computational and Mathematical Methods in Medicine*, vol. 2017, 2017, doi: 10.1155/2017/3762651.
[25] D. G. Altman and J. M. Bland, "Statistics notes: diagnostic tests 3: receiver operating characteristic plots," *BMJ*, vol. 309, no. 6948, p. 188, 1994, doi: 10.1136/bmj.309.6948.188.
[26] A. K. Akobeng, "Understanding diagnostic tests 3: Receiver operating characteristic curves," *Acta Paediatrica, International Journal of Paediatrics*, vol. 96, no. 5, pp. 644–647, 2007, doi: 10.1111/j.1651-2227.2006.00178.x.
[27] A. S. Yadaw, Y.-C. Li, S. Bose, R. Iyengar, S. Bunyavanich, and G. Pandey, "Clinical predictors of COVID-19 mortality," *medRxiv*, 2020, doi: 10.1101/2020.05.19.20103036.
[28] Y. Gao *et al.*, "Machine learning based early warning system enables accurate mortality risk prediction for COVID-19," *Nature Communications*, vol. 11, no. 1, 2020, doi: 10.1038/s41467-020-18684-2.
[29] C. An, H. Lim, D. W. Kim, J. H. Chang, Y. J. Choi, and S. W. Kim, "Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study," *Scientific Reports*, vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-75767-2.
[30] F. P. Veras *et al.*, "SARS-CoV-2-triggered neutrophil extracellular traps mediate COVID-19 pathology," *Journal of Experimental Medicine*, vol. 217, no. 12, 2020, doi: 10.1084/jem.20201129.
[31] B. Tomar, H. J. Anders, J. Desai, and S. R. Mulay, "Neutrophils and Neutrophil extracellular traps drive necroinflammation in COVID-19," *Cells*, vol. 9, no. 6, 2020, doi: 10.3390/cells9061383.
[32] B. Zhou, J. She, Y. Wang, and X. Ma, "Duration of viral shedding of discharged patients with severe COVID-19," *Clinical Infectious Diseases*, vol. 71, no. 16, pp. 2240–2242, 2020, doi: 10.1093/cid/ciaa451.
[33] T. Alnima, M. M. G. Mulder, B. C. T. V. Bussel, and H. T. Cate, "COVID-19 coagulopathy: from pathogenesis to treatment," *Acta Haematologica*, vol. 145, no. 3, pp. 282–296, 2022, doi: 10.1159/000522498.
[34] K. K. Bengtsen, L. Wilhelmsen, D. Elmfeldt, and G. Tibblin, "Blood coagulation and fibrinolysis in man after myocardial infarction compared with a representative population sample," *Atherosclerosis*, vol. 16, no. 1, pp. 83–88, Jul. 1972, doi: 10.1016/0021-9150(72)90010-X.
[35] B. Yu *et al.*, "Evaluation of variation in D-dimer levels among COVID-19 and bacterial pneumonia: a retrospective analysis," *Journal of Thrombosis and Thrombolysis*, vol. 50, no. 3, pp. 548–557, 2020, doi: 10.1007/s11239-020-02171-y.
[36] L. Zhang *et al.*, "D-dimer levels on admission to predict in-hospital mortality in patients with Covid-19," *Journal of Thrombosis and Haemostasis*, vol. 18, no. 6, pp. 1324–1329, 2020, doi: 10.1111/jth.14859.
[37] J. I. Weitz, J. C. Fredenburgh, and J. W. Eikelboom, "A test in context: D-Dimer," *Journal of the American College of Cardiology*, vol. 70, no. 19, pp. 2411–2420, 2017, doi: 10.1016/j.jacc.2017.09.024.
[38] J. Simes *et al.*, "D-dimer predicts long-term cause-specific mortality, cardiovascular events, and cancer in patients with stable coronary heart disease LIPID study," *Circulation*, vol. 138, no. 7, pp. 712–723, 2018, doi: 10.1161/CIRCULATIONAHA.117.029901.
[39] G. Lippi and M. Plebani, "Laboratory abnormalities in patients with COVID-2019 infection," *Clinical Chemistry and Laboratory Medicine*, vol. 58, no. 7, pp. 1131–1134, 2020, doi: 10.1515/cclm-2020-0198.

[40] M. I. Hayiroglu, T. Cinar, and A. I. Tekkesin, "Fibrinogen and D-dimer variances and anticoagulation recommendations in COVID-19: Current literature review," *Revista da Associacao Medica Brasileira*, vol. 66, no. 6, pp. 842–848, 2020, doi: 10.1590/1806-9282.66.6.842.
[41] K. W. Choi *et al.*, "Outcomes and prognostic factors in 267 patients with severe acute respiratory syndrome in Hong Kong," *Annals of Internal Medicine*, vol. 139, no. 9, 2003, doi: 10.7326/0003-4819-139-9-200311040-00005.
[42] G. Geri *et al.*, "Acute kidney injury in SARS-CoV2-related pneumonia ICU patients: a retrospective multicenter study," *Annals of Intensive Care*, vol. 11, no. 1, 2021, doi: 10.1186/s13613-021-00875-9.

BIOGRAPHIES OF AUTHORS

Husnul Khuluq is a doctor candidate in the Faculty of Pharmacy, University of Ahmad Dahlan, Yogyakarta, Indonesia. He obtained his professional degree from Universitas Gadjah Mada, Yogyakarta, Indonesia, and master degree from Faculty of Pharmacy, University of Ahmad Dahlan, Yogyakarta, Indonesia. He has worked as a lecturer in the Department of Pharmacy, Universitas Muhammadiyah Gombong, Kebumen, Indonesia. Currently, his study is focused on big data and artificial intelligence. He can be contacted at email: husnulkhuluq@unimugo.ac.id.

Prasandhya Astagiri Yusuf is an assistant professor in the Department of Medical Physiology and Biophysics, as well as the head of Medical Technology Cluster IMERI and head of Health Big Data Center IMERI, Faculty of Medicine, University of Indonesia. He obtained his bachelor's and master's degree from Institut Teknologi Bandung, Indonesia. He earned his Ph.D. degree in Auditory Neuroscience from the Hannover Medical School, Germany. Currently, his research is focused on neurotechnology, biomedical signal processing, health big data analyses, and medical technology. He can be contacted at email: prasandhya.a.yusuf@ui.ac.id.

Dyah Aryani Perwitasari is a Professor in the Faculty of Pharmacy, University of Ahmad Dahlan, Yogyakarta, Indonesia. She obtained her professional and Master's degree from Clinical Pharmacy at the Faculty of Pharmacy of Universitas Gadjah Mada, Yogyakarta, Indonesia. Since 2000, she has worked as a lecturer in the same Universitas Ahmad Dahlan. She earned his Ph.D. degree from the Leiden University Medical Center, Leiden, The Netherlands and Faculty of Medicine, Universitas Gadjah Mada, Yogyakarta, Indonesia. Currently, her study is focused on pharmacogenomics, patients' reported outcome, clinical pharmacy, personalized medicine and social-behavior-administrative pharmacy. She can be contacted at email: dyah.perwitasari@pharm.uad.ac.id.

Thang Nguyen is a Scientific Research and International Cooperation Office Can Tho University of Medicine and Pharmacy: Can Tho City, Vet Nham. He obtained his B.Pharm. (Pharmacology and Clinical Pharmacy), M.Sc. (Pharmacology and Clinical Pharmacy) degree from Ho Chi Minh City University of Medicine and Pharmacy: Ho Chi Minh City, VN. He earned his Ph.D. degree from the University of Groningen: Groningen, NL. He can be contacted at email: nthang@ctump.edu.vn.