Leveraging a Random Forest Classifier and SVMSMOTE for an Early-stage Dengue Prediction

Yulianti Paula Bria
Department of Computer Science, Universitas Katolik Widya Mandira, Indonesia
yulianti.bria@unwira.ac.id (corresponding author)

Paskalis Andrianus Nani
Department of Computer Science, Universitas Katolik Widya Mandira, Indonesia
paskalisnani@unwira.ac.id

Yovinia Carmeneja Hoar Siki
Department of Computer Science, Universitas Katolik Widya Mandira, Indonesia
yoviniacarmeneja@unwira.ac.id

Natalia Magdalena Rafu Mamulak
Department of Computer Science, Universitas Katolik Widya Mandira, Indonesia
mamulaknatalia@unwira.ac.id

Emiliana Metan Meolbatak
Department of Computer Science, Universitas Katolik Widya Mandira, Indonesia
emilianameol@unwira.ac.id

Robertus Dole Guntur
Department of Mathematics, Universitas Nusa Cendana, Indonesia
robertus_guntur@staf.undana.ac.id

Received: 1 March 2025 | Revised: 28 March 2025 | Accepted: 7 April 2025

 DOI: https://doi.org/10.48084/etasr.10762

ABSTRACT

Dengue is a significant global health issue, and in its severe form, it can be fatal. Accurate early-stage prediction tools are crucial in resource-limited areas to prevent dengue's progression to a severe state. This study aims to develop machine learning classifiers for dengue to assist medical personnel in differentiating the latter from other diseases, thereby helping with its earlier prognosis. Early-stage dengue classifiers were developed using medical records collected from two hospitals in East Nusa Tenggara Province, Indonesia. Eight machine learning techniques were leveraged to develop the classifiers, including Logistic Regression (LR), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), K-nearest Neighbors (KNN), Naïve Bayes (NB), Classification and Regression Tree (CART), Random Forest (RF), and extreme Gradient Boosting (XGBoost). To address the imbalance in the dataset, we utilized the SVM Synthetic Minority Oversampling Technique (SVMSMOTE) was utilized. The dataset was finalized through the expertise of 15 medical doctors and insights gathered from four Indonesian digital health platforms. The key findings of this study are: i) important features for early-stage dengue prediction include fever, duration of fever, headache, arthralgia, myalgia, nausea, shivering, loss of appetite, bitter mouth, temperature, and age, ii) machine learning techniques, including RF, NB, KNN, and XGBoost, were found to be suitable for dengue prediction, and iii) RF combined with SVMSMOTE, outperformed other techniques, achieving an accuracy of 94.99% and an F1-score of 85.65% for early-stage dengue prediction.

Keywords-dengue fever; dengue classifier; dengue prediction; random forest

I. INTRODUCTION

Dengue is a life-threatening disease transmitted by female mosquitoes, particularly Aedes aegypti. It has become a significant global health burden, claiming many lives each year. The World Health Organization (WHO) estimates that between 100 and 400 million people are infected annually, with nearly half of the global population being at risk, especially in tropical and subtropical regions [1]. Dengue is also a major health concern in Indonesia. As of July 1, 2024, Indonesia has reported 149,866 confirmed dengue cases, roughly three times the number of the cases reported during the same period in 2023. These cases have been recorded across 465 districts in all 38 provinces, leading to 884 deaths [2].

Early-stage diagnosis of dengue is challenging due to its symptoms being similar to those of other diseases, such as malaria, typhoid fever, and COVID-19. It is important to note that delays in treating dengue patients can lead to severe conditions like Dengue Hemorrhagic Fever (DHF) and Dengue Shock Syndrome (DSS), which can be fatal [1, 3]. In the context of rural Indonesia, where resources are often limited, reliable tools for early-stage dengue diagnosis are crucial to prevent the progression to these severe states. Although the WHO [1] states that there is no specific cure for dengue or severe dengue, early diagnosis can significantly reduce the fatality rates associated with severe cases.

In recent years, machine learning models for dengue prediction have gained attention. Most studies have focused on predicting dengue outbreaks [4-12], clustering dengue incidence areas [13-14], and identifying characteristics linked to the spread of the disease [15]. However, few studies have explored using symptoms and risk factors for early-stage dengue prediction. For instance, authors in [16] conducted a dengue prediction study based on selected features but used environmental factors, like humidity and temperature, rather than medical records. Another study [17] focused on early-stage dengue prediction using clinical data but did not exclude severe symptoms of dengue fever.

The current study aims to fill this gap by providing insights into the symptoms and risk factors that are significant for early-stage dengue prediction, based on medical records. This approach will help raise awareness among the Indonesian population to seek medical advice early. Additionally, this study aims to identify suitable classifiers for early-stage dengue prediction, develop an intelligent recommendation system, and find the best classifier for this purpose.

II. METHOD

The dataset for this study was acquired from medical records collected from two hospitals in East Nusa Tenggara Province, Indonesia: Kewapante Hospital in Sikka and Soe Hospital in South Central Timor. A total of 561 medical records were collected, comprising patients diagnosed with either dengue or non-dengue diseases, such as malaria, COVID-19, and typhoid fever. Of these, 88 records corresponded to dengue patients and 473 to non-dengue patients. Feature selection was performed based on a previous study, [18], which identified 15 significant symptoms and one significant risk factor for dengue prediction. Additionally, important symptoms and risk factors for the clinical diagnosis of dengue were gathered through interviews with fifteen medical doctors and from four Indonesian digital health platforms [19–22].

The features collected from medical experts and digital platforms were utilized to eliminate noisy data, resulting in a final dataset containing 439 medical records. The dataset was then split into training and test sets using stratified 10-fold cross-validation, ensuring the preservation of the class distribution [16–17]. Due to the imbalance between dengue and non-dengue records, SVMSMOTE was applied to handle data imbalance.

Subsequently, eight machine learning classifiers were developed, including LR, LDA, SVM, KNN, NB, CART, RF, and XGBoost. The classifiers were evaluated using two performance metrics: accuracy and F1-score. Hyperparameter tuning was performed for all classifiers to optimize their performance. Finally, the evaluation results were used to identify the machine learning techniques most suitable for early-stage dengue prediction and to determine the best-performing classifier.

A. Data Collection – Characteristics of Medical Records

Medical records were collected from patients diagnosed with either dengue fever or other diseases with similar symptoms (including typhoid fever, malaria, COVID-19, dyspepsia, and pneumonia) during the years 2017–2023. The data were manually recorded using an Excel spreadsheet. Exclusion criteria were applied to refine the dataset. First, to reduce the noise caused by the symptom overlap between dengue and other diseases, medical records with fewer than three reported symptoms were excluded. Second, patients diagnosed with asymptomatic dengue were excluded, as this study focuses on symptom- and risk-factor-based dengue classification.

This refined dataset represents a reduction from the 561 medical records used in [18], resulting in a total of 439 medical records: 79 records of dengue cases and 360 records of non-dengue cases. The findings from [18] identified 15 significant symptoms—including fever, fever duration, headache, muscle and joint pain (arthralgia/myalgia), nausea, abdominal pain, shivering, malaise, loss of appetite, shortness of breath, rash, bleeding nose, bitter mouth, and body temperature—along with one significant risk factor, age, for dengue prediction. However, as this study specifically targets early-stage dengue prediction, symptoms that typically manifest in the later stages of dengue (abdominal pain, malaise, persistent vomiting, shortness of breath or rapid breathing, rash, and bleeding nose) were excluded [1].

Therefore, this study utilized 10 input features, labeled S_1–S_9 and F, and one target feature, Diagnosis, which is represented as a binary variable (1 for dengue and 0 for non-dengue diseases). Seven of the input features (S_1, S_3–S_8) are binary variables, while the remaining three features (S_2, S_9, and F) are numeric. Table I presents the normalized feature values of the dengue dataset used in this study.

TABLE I. THE NORMALIZED VALUES FOR THE FEATURES OF THE DENGUE DATASET

Notation	Feature	Values
Symptoms		
S_1	Fever	1-yes, 0-no
S_2	Duration of fever (days)	Mean: 2.17, SD: 3.90
S_3	Headache	1-yes, 0-no
S_4	Arthralgia/myalgia	1-yes, 0-no
S_5	Nausea	1-yes, 0-no
S_6	Shivering	1-yes, 0-no
S_7	Loss of appetite	1-yes, 0-no
S_8	Bitter mouth	1-yes, 0-no
S_9	Temperature (°C)	Mean: 37.00, SD: 0.99
Risk factor		
F	Age (years)	Mean: 33.80, SD: 24.11

B. Data Collection – Interviews with 15 Doctors

Structured interviews were conducted with fifteen Indonesian medical doctors to gather knowledge related to important symptoms and risk factors contributing to the clinical diagnosis of dengue. This information was critical for finalizing the dengue dataset, particularly during the outlier removal process. All interview questions were delivered in Bahasa Indonesia and subsequently translated into English for analysis.

All fifteen medical doctors identified fever as the most critical symptom for diagnosing clinical dengue, followed by rash and bleeding nose (each cited by 14 doctors), and fever duration along with abdominal pain (each cited by 13 doctors). The collected responses regarding symptoms were then categorized based on their severity using the WHO standards [1]. A summary of the interviews is provided in the Appendix.

C. Data Collection - Gathering Data from Digital Health Platforms in Indonesia

This study utilized four Indonesian digital health platforms—Alodokter.com, Klikdokter.com, Halodoc.com, and Ayosehat.kemkes.go.id—to gather additional information on symptoms and risk factors associated with dengue fever. Combined with the results from the structured interviews with fifteen medical doctors, this acquired knowledge was used to finalize the dengue dataset during the data preprocessing phase, particularly for outlier elimination. Table II presents the symptoms and risk factors collected from the four Indonesian digital health platforms.

D. Machine Learning Techniques

Dengue classifiers were developed using eight popular machine learning techniques in dengue prediction including SVM, LR, KNN, NB, RF, XGBoost, LDA, and CART [4, 16-18, 23].

E. Performance Metrics

To evaluate the early-stage dengue classifiers, two performance metrics were utilized: accuracy and F1-score. The F1-score calculation also required the computation of precision and recall.

TABLE II. COLLECTED SYMPTOMS AND RISK FACTORS FROM THE INDONESIAN DIGITAL HEALTH PLATFORMS

Symptoms and risk factors	[19]	[20]	[21]	[22]
Fever and temperature	V (39-40 °C)	V (39 °C)	V (39-40 °C)	V (39-40 °C)
Fever duration	3 days	2-7 days	2-7 days	2-7 days
Headache	V	V	V	V
Loss of appetite	V	V	V	V
Nausea	V	V	V	V
Shivering	-	V	-	V
Arthralgia	V	V	V	V
Myalgia	V	V	V	V
Orbital pain	V	V	V	V
Age	-	V (<15 yo)	-	V (<15 yo)
Visiting/living in endemic areas	-	V	V	V
Dengue infection history	-	V	V	V
Sore throat	-	V	-	-

yo: years old

F. Data Preprocessing

Data preprocessing was carried out by removing outliers and duplicate records to improve the performance of the machine learning classifiers. The final dataset used in this study consisted of 439 records, comprising 79 dengue cases and 360 non-dengue cases. For feature normalization, a min-max scaler was applied, rescaling the feature values to a range between 0 and 1 to optimize the performance of the machine learning classifiers. Additionally, repeated stratified k-fold was applied with ten folds [16-17], and a pipeline was employed to prevent data leakage from the training set into the testing set.

G. Imbalance Data Handling

The difference between the 79 dengue cases and the 360 non-dengue cases in this dataset created a class imbalance. To improve the performance of the machine learning classifiers [24], several commonly employed techniques for handling imbalanced datasets were compared, including Random Oversampling (ROS), Synthetic Minority Oversampling Technique (SMOTE), Borderline-SMOTE (BLSMOTE), SVMSMOTE, and Adaptive Synthetic Sampling (ADASYN) [25–26], to determine the most appropriate method for this study.

III. RESULTS AND DISCUSSION

A. Applying Imbalance Methods

Figure 1 shows the comparison of the performance of different imbalance handling methods. The experimental results revealed that SVMSMOTE yielded the highest performance with a geometric mean score of 0.915 and a Standard Deviation (SD) of 0.056. It was followed by ADASYN (0.908 ± 0.073), SMOTE (0.902 ± 0.081), ROS (0.901 ± 0.070), and BLSMOTE (0.886 ± 0.090). Based on the best-performing method, SVMSMOTE was chosen for handling the imbalanced dengue dataset in this study.

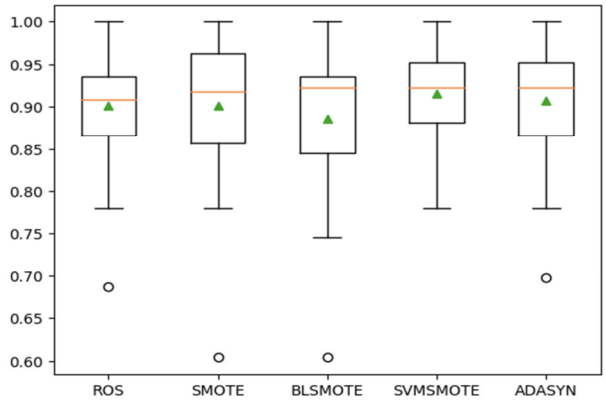

Fig. 1. Comparison of imbalance methods.

B. Developing and Improving the Classifiers' Performance

The present work successfully developed eight classifiers by combining SVMSMOTE with LR, LDA, KNN, CART, XGBoost, RF, NB, and SVM. To further enhance the performance of the classifiers, the hyperparameters of each dengue classifier were tuned with the values detailed in Table III. A performance comparison of the eight machine learning classifiers—before data preprocessing, after applying SVMSMOTE, and after hyperparameter tuning—is provided in Table IV.

TABLE III. HYPERPARAMETERS USED FOR THE EIGHT CLASSIFIERS

Classifier	Hyperparameter	Value
LR	solver	lbfgs
LDA	solver	lsqr
KNN	n_neighbors	5
CART	max_depth min_samples_split min_samples_leaf	10 2 2
XGBoost	scale_pos_weight	99
RF	n_estimator	100
NB	var_smoothing	0.000001
SVM	gamma C	auto 5

TABLE IV. PERFORMANCE COMPARISON OF EIGHT DENGUE CLASSIFIERS

Classifier	Accuracy (%)			F1-score (%)		
	BP	S	Tn	BP	S	Tn
LR	85.218	92.484	92.484	19.222	74.609	74.609
LDA	84.449	93.177	92.484	19.444	80.175	74.609
KNN	**89.301**	89.524	93.177	**66.606**	70.422	80.175
CART	84.449	92.257	89.524	46.621	79.066	70.422
XGBoost	84.962	92.722	93.166	45.794	78.910	78.937
RF	88.006	**93.631**	**94.989**	52.530	**83.122**	**85.654**
NB	74.468	89.308	94.975	47.025	75.770	85.648
SVM	87.519	82.008	89.535	51.466	68.119	75.986

BP: Before data preprocessing, S: After SVMSMOTE, Tn: Tuning

Figure 2 displays the comparison of the F1-score of the eight machine learning classifiers pre- and post-hyperparameter tuning.

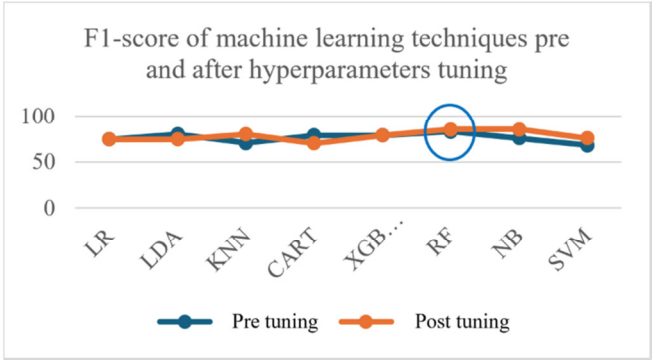

Fig. 2. The performance comparison of F1-score.

C. Discussion

Based on the results in Table IV, before data preprocessing, the KNN classifier displayed the best performance, achieving an accuracy score of 89.301% and an F1-score of 66.606%. After implementing SVMSMOTE, both before and after hyperparameter tuning, the RF classifier yielded the highest performance, with an accuracy score of 93.631% and an F1-score of 83.122%, which further improved to an accuracy score of 94.989% and an F1-score of 85.654%, respectively. Post-tuning, the NB classifier performed very similarly to RF, achieving an accuracy of 94.975% and an F1-score of 85.648%. Furthermore, the third and fourth best performing classifiers were KNN and XGBoost. Based on Figure 2, RF demonstrated the most stable performance, with almost identical accuracy and F1-score values before and after tuning. Interestingly, the NB classifier experienced significant improvement after hyperparameter tuning, with accuracy rising from 89.308% to 94.975% and F1-score increasing from 75.770% to 85.648%.

Further experiments were conducted to evaluate how well the RF classifier can predict 28 randomly selected dengue cases from the dataset. The results indicated that RF achieved perfect accuracy in this task. These experimental results can be found in Appendix. The RF classifier has proven to be highly effective in dengue prediction [27-28] and in predicting other diseases, such as diabetes, when integrated with IoT [29], malaria [30], and COVID-19 [31].

The experimental results in Table IV also suggest that the combination of nine symptoms and one risk factor is sufficient for early-stage dengue prediction. These features include fever, fever duration, headache, arthralgia/myalgia, nausea, shivering, loss of appetite, bitter mouth, temperature, and age. This study successfully predicted early-stage dengue by incorporating significant symptoms and a risk factor while excluding severe symptoms, which are often included in other studies [17]. This study extends the work in [18], where feature selection methods were used to identify important features, including those related to severe symptoms.

The insights gained from this study regarding significant features for dengue prediction can raise public awareness about the importance of early detection. It highlights the symptoms and risk factors that should not be overlooked, encouraging individuals to seek medical advice promptly to prevent the

progression to severe dengue, such as DHF or DSS, which can be life-threatening [3, 32-33].

Further improvements to the study could include adding features identified by medical doctors and Indonesian digital health platforms, such as orbital pain and regional endemic factors, which were not available in the medical records. The findings from this research could contribute to the development of an intelligent expert system for dengue diagnosis, helping users decide whether to seek medical advice based on the input symptoms.

D. Ethical Statement

This study obtained ethical clearance from the Human Ethics Committee of Widya Mandira Catholic University (reference number: 001/WM.H9/LPPM/SKKEP/X/2023). Written consent was obtained from the 15 medical doctors who were interviewed. Permission to collect the medical records was granted by the Departments of Permission Affairs in East Nusa Tenggara Province, as well as in the Sikka and South-Central Timor Districts. Additionally, the directors of Kewapante Hospital and Soe Hospital provided their approval for the medical record collection.

IV. CONCLUSION

In this study, 8 different classifiers were implemented: Logistic Regression (LR), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), K-nearest Neighbors (KNN), Naïve Bayes (NB), Classification and Regression Tree (CART), Random Forest (RF), and extreme Gradient Boosting (XGBoost), using a dataset of 439 medical records (79 dengue cases and 360 non-dengue cases). The dataset's imbalance was addressed using the SVM Synthetic Minority Oversampling Technique (SVMSMOTE). This work aimed to identify suitable classifiers for early-stage dengue prediction using nine significant and early symptoms (fever, fever duration, headache, arthralgia, myalgia, nausea, shivering, loss of appetite, bitter mouth, temperature) and only one risk factor (age).

The four best-performing classifiers in terms of accuracy and F1-score for dengue prediction were, in order of performance, RF, NB, KNN, and XGBoost. This study determined that the combination of the nine symptoms and the one risk factor was sufficient for early-stage dengue prediction. Unlike other studies that incorporated severe symptoms based on clinical data, this study focused exclusively on early-stage symptoms. This approach is crucial for raising awareness and encouraging individuals to seek medical advice upon experiencing these symptoms, thereby preventing progression to severe dengue, which can be life-threatening.

Overall, the RF, NB, KNN and XGBoost classifiers in conjunction with SVMSMOTE, can be used to develop intelligent expert systems for dengue diagnosis, even when working with imbalanced datasets.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Widya Mandira Catholic University for providing the research grant for this study. They also thank the Directors of Kewapante and Soe Hospitals. Finally, the authors deeply appreciate the contributions of the fifteen medical doctors who participated in the interviews, as well as the medical records staff at both hospitals for their valuable assistance.

APPENDIX

THE FIFTEEN MEDICAL DOCTORS' RESPONSES ABOUT SYMPTOMS AND RISK FACTORS FOR CLINICAL DENGUE DIAGNOSIS

Symptoms and risk factor	**Code [1]**	**D_1**	**D_2**	**D_3**	**D_4**	**D_5**	**D_6**	**D_7**	**D_8**	**D_9**	**D_{10}**	**D_{11}**	**D_{12}**	**D_{13}**	**D_{14}**	**D_{15}**	**Total Y**
Fever (S_1)	NS	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	15
Fever duration (S_2)	NS	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	-	Y	Y	13
Headache (S_3)	NS	Y	Y	-	-	Y	Y	Y	Y	Y	Y	-	-	-	-	Y	9
Arthralgia/joint pain and Myalgia/muscle pain (S_4)	NS	Y	Y	-	-	Y	Y	Y	Y	Y	Y	-	-	Y	-	-	9
Nausea (S_5)	NS	Y	-	-	-	-	-	-	-	-	-	Y	-	-	-	-	2
Shivering (S_6)	NS	Y	Y	-	-	Y	Y	Y	Y	Y	Y	-	-	-	-	-	8
Loss of appetite (S_7)	NS	Y	Y	-	-	Y	Y	Y	Y	Y	Y	-	-	-	-	-	8
Bitter mouth (S_8)	NS	Y	Y	-	-	Y	Y	Y	Y	Y	Y	-	-	-	-	-	8
High Temperature (S_9)	NS	Y	Y	Y	-	Y	Y	Y	Y	Y	Y	-	-	Y	-	Y	11
Vomiting	S	Y	-	-	Y	-	-	-	-	-	-	Y	-	-	-	-	3
Abdominal pain	S	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	-	13
Body pain	NS	Y	Y	-	-	Y	Y	Y	Y	Y	Y	-	-	-	-	-	8
Heartburn	NS	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	1
Dizziness	NS	Y	Y	-	-	Y	Y	Y	Y	Y	Y	-	-	Y	-	-	8
Malaise	S	Y	Y	-	-	Y	Y	Y	Y	Y	Y		Y			Y	10
Shortness of breath	S	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	1
Rash	S	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	Y	Y	14
Bleeding nose	S	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	Y	Y	Y	14
Seizure	S	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Orbital pain	NS	-	Y	-	Y	Y	Y	Y	Y	Y	Y	-	-	-	-	-	8
Loss of consciousness	S	Y	Y	-	Y	Y	Y	Y	Y	Y	Y	-	Y	Y	-	-	11
Age (F)	R	-	Y	-	-	Y	Y	Y	Y	Y	Y	Y	-	-	Y	-	9
Gender	R	-	Y	-	-	Y	Y	Y	Y	Y	Y	-	-	-	Y	-	8
Endemic area	R	-	Y	-	Y	Y	Y	Y	Y	Y	Y	Y	-	Y	-	Y	11

D: medical doctor; Y: Yes; NS: Non-severe; S: Severe; R: Risk factor

THE PREDICTION RESULTS OF THE RF CLASSIFIER WITH SVMSMOTE ON DENGUE CASES IN THE DATASET

Input	Prediction	Input	Prediction
[3,38.9,1,5,0,0,0,0,1,0]	>Predicted=1 (expected 1)	[16,36.8,1,4,0,1,0,0,1,0]	>Predicted=1 (expected 1)
[3.6,37,1,3,0,1,0,0,0,0]	>Predicted=1 (expected 1)	[16,37.4,1,4,0,1,1,0,1,0]	>Predicted=1 (expected 1)
[4,40.6,1,3,0,0,0,0,1,0]	>Predicted=1 (expected 1)	[16,38,1,4,0,1,0,0,1,0]	>Predicted=1 (expected 1)
[5,39.2,1,7,0,1,0,0,1,0]	>Predicted=1 (expected 1)	[16,37.5,1,1,0,1,0,0,1,0]	>Predicted=1 (expected 1)
[5,38.9,1,2,0,0,1,0,1,0]	>Predicted=1 (expected 1)	[16,38,1,4,0,1,1,0,0,0]	>Predicted=1 (expected 1)
[6,36.2,1,3,0,1,0,0,1,0]	>Predicted=1 (expected 1)	[17,38,1,3,0,1,1,0,1,0]	>Predicted=1 (expected 1)
[6,37.1,1,4,0,0,0,0,1,0]	>Predicted=1 (expected 1)	[17,37.6,1,8,0,0,0,0,1,0]	>Predicted=1 (expected 1)
[6,36.7,1,3,0,1,1,0,1,0]	>Predicted=1 (expected 1)	[17,39.5,1,3,0,1,1,0,1,0]	>Predicted=1 (expected 1)
[6,38.4,1,4,0,1,0,0,1,0]	>Predicted=1 (expected 1)	[18,37.6,1,3,0,1,0,0,1,0]	>Predicted=1 (expected 1)
[7,38.9,1,5,1,0,0,0,0,0]	>Predicted=1 (expected 1)	[19,37,1,3,0,1,0,0,0,0]	>Predicted=1 (expected 1)
[8,37.5,1,4,1,0,0,0,0,0]	>Predicted=1 (expected 1)	[19,37.9,1,4,0,1,1,0,0,0]	>Predicted=1 (expected 1)
[8,36.7,1,1,0,0,0,0,1,0]	>Predicted=1 (expected 1)	[19,37.5,1,2,0,1,1,0,1,0]	>Predicted=1 (expected 1)
[8,36.5,1,5,0,0,0,0,1,0]	>Predicted=1 (expected 1)	[22,39.2,1,3,0,1,1,0,1,0]	>Predicted=1 (expected 1)
[8,37.1,1,6,0,1,0,0,1,0]	>Predicted=1 (expected 1)	[22,36.5,1,3,0,1,1,0,0,0]	>Predicted=1 (expected 1)

REFERENCES

[1] World Health Organization, "Dengue and severe dengue." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue#:~:text=The%20incidence%20of%20dengue%20has,dengue%20cases%20are%20under-reported.

[2] World Health Organization, "Indonesia takes decisive, pioneering action to strengthen multisource collaborative surveillance for dengue," [Online]. Available: https://www.who.int/indonesia/news/detail/30-07-2024-indonesia-takes-decisive--pioneering-action-to-strengthen-multisource-collaborative-surveillance-for-dengue.

[3] D. H. Somasetia, T. T. Malahayati, F. M. Andriyani, D. Setiabudi, and H. M. Nataprawira, "A fatal course of multiple inflammatory syndrome in children coinfection with dengue. A case report from Indonesia," *IDCases*, vol. 22, 2020, Art. no. e01002, https://doi.org/10.1016/j.idcr.2020.e01002.

[4] R. A. K. Tazkia, V. Narita, and A. S. Nugroho, "Dengue outbreak prediction for GIS based Early Warning System," in *2015 International Conference on Science in Information Technology (ICSITech)*, Yogyakarta, Oct. 2015, pp. 121–125, https://doi.org/10.1109/ICSITech.2015.7407789.

[5] M. Nabilah, R. Tyasnurita, F. Mahananto, W. Anggraeni, R. A. Vinarti, and A. Muklason, "Forecasting the number of dengue fever based on weather conditions using ensemble forecasting method," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 1, Mar. 2023, Art. no. 496, https://doi.org/10.11591/ijai.v12.i1.pp496-504.

[6] M. F. F. Mardianto, S. H. Kartiko, and H. Utami, "The Fourier series estimator to predict the number of dengue and malaria sufferers in Indonesia," in *International Conference on Mathematics, Computational Sciences and Statistics 2020*, Surabaya, Indonesia, 2021, Art. no. 060002, https://doi.org/10.1063/5.0042115.

[7] A. L. Ramadona, Y. Tozan, J. Wallin, L. Lazuardi, A. Utarini, and J. Rocklöv, "Predicting the dengue cluster outbreak dynamics in Yogyakarta, Indonesia: a modelling study," *The Lancet Regional Health - Southeast Asia*, vol. 15, Aug. 2023, Art. no. 100209, https://doi.org/10.1016/j.lansea.2023.100209.

[8] T. H. F. Harumy, H. Y. Chan, and G. C. Sodhy, "Prediction for Dengue Fever in Indonesia Using Neural Network and Regression Method," *Journal of Physics: Conference Series*, vol. 1566, no. 1, Jun. 2020, Art. no. 012019, https://doi.org/10.1088/1742-6596/1566/1/012019.

[9] M. Mistawati, Y. Yasnani, and H. Lestari, "Forecasting prevalence of dengue hemorrhagic fever using ARIMA model in Sulawesi Tenggara Province, Indonesia," *Public Health of Indonesia*, vol. 7, no. 2, pp. 75–86, Jun. 2021, https://doi.org/10.36685/phi.v7i2.411.

[10] W. Anggraeni *et al.*, "Modified Regression Approach for Predicting Number of Dengue Fever Incidents in Malang Indonesia," *Procedia Computer Science*, vol. 124, pp. 142–150, 2017, https://doi.org/10.1016/j.procs.2017.12.140.

[11] M. A. Majeed, H. Z. M. Shafri, Z. Zulkafli, and A. Wayayok, "A Deep Learning Approach for Dengue Fever Prediction in Malaysia Using LSTM with Spatial Attention," *International Journal of Environmental Research and Public Health*, vol. 20, no. 5, Feb. 2023, Art. no. 4130, https://doi.org/10.3390/ijerph20054130.

[12] S. K. Dey *et al.*, "Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: A machine learning approach," *PLOS ONE*, vol. 17, no. 7, Jul. 2022, Art. no. e0270933, https://doi.org/10.1371/journal.pone.0270933.

[13] Mamenun, Y. Koesmaryono, A. Sopaheluwakan, R. Hidayati, B. D. Dasanto, and R. Aryati, "Spatiotemporal Characterization of Dengue Incidence and Its Correlation to Climate Parameters in Indonesia," *Insects*, vol. 15, no. 5, May 2024, Art. no. 366, https://doi.org/10.3390/insects15050366.

[14] A. Aswi, S. Cramb, E. Duncan, W. Hu, G. White, and K. Mengersen, "Climate variability and dengue fever in Makassar, Indonesia: Bayesian spatio-temporal modelling," *Spatial and Spatio-temporal Epidemiology*, vol. 33, Jun. 2020, Art. no. 100335, https://doi.org/10.1016/j.sste.2020.100335.

[15] R. Gangula, L. Thirupathi, R. Parupati, K. Sreeveda, and S. Gattoju, "Ensemble machine learning based prediction of dengue disease with performance and accuracy elevation patterns," *Materials Today: Proceedings*, vol. 80, pp. 3458–3463, 2023, https://doi.org/10.1016/j.matpr.2021.07.270.

[16] G. Gupta *et al.*, "DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms," *Diagnostics*, vol. 13, no. 6, Mar. 2023, Art. no. 1093, https://doi.org/10.3390/diagnostics13061093.

[17] B. Abdualgalil, S. Abraham, and W. M. Ismael, "Early Diagnosis for Dengue Disease Prediction Using Efficient Machine Learning Techniques Based on Clinical Data," *Journal of Robotics and Control (JRC)*, vol. 3, no. 3, pp. 257–268, May 2022, https://doi.org/10.18196/jrc.v3i3.14387.

[18] Y. P. Bria, "Determining Important Features for Dengue Diagnosis using Feature Selection Methods," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 47–59, Jan. 2024, https://doi.org/10.47738/jads.v6i1.445.

[19] Alodokter Ministry of Health of the Republic of Indonesia, "Demam Dengue." [Online]. Available: https://www.alodokter.com/demam-berdarah.

[20] Halodoc, "Demam Berdarah." [Online]. Available: https://www.halodoc.com/kesehatan/demam-berdarah.

[21] Klikdokter, "Demam Berdarah Dengue." [Online]. Available: https://www.klikdokter.com/penyakit/masalah-infeksi/demam-berdarah-dengue.

[22] Ministry of Health of Indonesia, "Demam Berdarah Dengue." [Online]. Available: https://ayosehat.kemkes.go.id/topik/demam-berdarah-dengue.

[23] V. Ramasamy, S. Vadivel, S. Kothandapani, J. Mahilraj, P. Sivaram, and B. Sharma, "An Optimal Feature Selection with Neural Network-Based Classification Model for Dengue Fever Prediction," in *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India, Mar. 2023, pp. 1–5, https://doi.org/10.1109/ISCON57294.2023.10112011.

[24] S. Matharaarachchi, M. Domaratzki, and S. Muthukumarana, "Minimizing features while maintaining performance in data classification problems," *PeerJ Computer Science*, vol. 8, Sep. 2022, Art. no. e1081, https://doi.org/10.7717/peerj-cs.1081.

[25] Y. P. Bria, C.-H. Yeh, and S. Bedingfield, "Significant symptoms and nonsymptom-related factors for malaria diagnosis in endemic regions of Indonesia," *International Journal of Infectious Diseases*, vol. 103, pp. 194–200, Feb. 2021, https://doi.org/10.1016/j.ijid.2020.11.177.

[26] J. Brownlee, *Imbalanced classification with Python: choose better metrics, balance skewed classes, and apply cost-sensitive learning*, v1.2, 2020.

[27] P. Silitonga, B. E. Dewi, A. Bustamam, and H. S. Al-Ash, "Evaluation of Dengue Model Performances Developed Using Artificial Neural Network and Random Forest Classifiers," *Procedia Computer Science*, vol. 179, pp. 135–143, 2021, https://doi.org/10.1016/j.procs. 2020.12.018.

[28] C. Y. Santos *et al.*, "A machine learning model to assess potential misdiagnosed dengue hospitalization," *Heliyon*, vol. 9, no. 6, Jun. 2023, Art. no. e16634, https://doi.org/10.1016/j.heliyon.2023.e16634.

[29] R. Rastogi, M. Bansal, N. Kumar, S. Singla, P. Singla, and R. A. Jaswal, "Effective Diabetes Prediction using an IoT-based Integrated Ensemble Machine Learning Framework," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 20064–20070, Feb. 2025, https://doi.org/10.48084/etasr.8869.

[30] A. D'Abramo *et al.*, "A machine learning approach for early identification of patients with severe imported malaria," *Malaria Journal*, vol. 23, no. 1, Feb. 2024, Art. no. 46, https://doi.org/10.1186/s12936-024-04869-3.

[31] F. Özen, "Random forest regression for prediction of Covid-19 daily cases and deaths in Turkey," *Heliyon*, vol. 10, no. 4, Feb. 2024, Art. no. e25746, https://doi.org/10.1016/j.heliyon.2024.e25746.

[32] S. Haryanto *et al.*, "Clinical features and virological confirmation of perinatal dengue infection in Jambi, Indonesia: A case report," *International Journal of Infectious Diseases*, vol. 86, pp. 197–200, Sep. 2019, https://doi.org/10.1016/j.ijid.2019.07.019.

[33] R. T. Sasmono *et al.*, "Molecular epidemiology of dengue in North Kalimantan, a province with the highest incidence rates in Indonesia in 2019," *Infection, Genetics and Evolution*, vol. 95, Nov. 2021, Art. no. 105036, https://doi.org/10.1016/j.meegid.2021.105036.