Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 8, No. 3, November 2025, pp. 635 – 645
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Early Detection of Dengue Hemorrhagic Fever Using Patient Medical Data with Ensemble Learning Methods

[1*]Achmad Saleh, [2]Ridha Mukhtar, [3]Rusdah
[1,2,3]Master of Computer Science, Faculty of Information Technology, Budi Luhur University
Email: [1]2311601427@student.budiluhur.ac.id, [2]2311601443@student.budiluhur.ac.id, [3]rusdah@budiluhur.ac.id

Article Info
Article history:
Received Aug 03rd, 2025
Revised Sep 23th, 2025
Accepted Oct 15th, 2025

Keyword:
CRISP-DM
Dengue Hemorrhagic Fever
Early Detection
Medical Data
Random Forest

ABSTRACT
Dengue Hemorrhagic Fever (DHF) remains a major public health concern in Indonesia and worldwide, where delayed diagnosis increases the risk of severe complications and mortality. Conventional laboratory-based diagnostics are time-consuming and often less accessible in resource-limited healthcare settings. This study aims to develop an early detection model for DHF using only initial clinical symptoms and demographic data extracted from electronic medical records at RSUD Brigjend H. Hasan Basry Kandangan. A total of 649 patient records (352 DHF cases and 297 non-dengue) were analyzed using the CRISP-DM framework. Five ensemble learning algorithms, Random Forest, Bagging, AdaBoost, and Gradient Boosted Tree, were evaluated across 80:20, 70:30, and 60:40 data splits and validated using 5-fold and 10-fold cross-validation. Random Forest consistently delivered the best and most stable performance, achieving up to 90.00 % accuracy and 0.967 AUC in the 80:20 split and mean accuracies of 88.91 % (5-fold) and 88.29 % (10-fold) in cross-validation. Further hyperparameter tuning enhanced model stability and prevented overfitting. The findings confirm that initial clinical symptoms and demographic attributes can reliably identify DHF cases early, enabling faster and more affordable screening prior to laboratory confirmation. This machine learning based decision-support model has the potential to significantly improve early clinical management of dengue fever.

Corresponding Author:
Achmad Saleh,
Master of Computer Science, Faculty of Information Technology
Budi Luhur University
Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, DKI Jakarta 12260, Indonesia.
Email: 2311601427@student.budiluhur.ac.id
DOI: http://dx.doi.org/10.24014/ijaidm.v8i3.38088

1. INTRODUCTION

Dengue Hemorrhagic Fever (DHF) is one of the fastest-spreading infectious diseases in the world and poses a significant global health challenge. Over the past five decades, the incidence of DHF has increased by about 30-fold, expanding into new areas, including previously unaffected rural regions. The World Health Organization (WHO) estimates that around 50 million dengue infections occur every year, with more than 2.5 billion people living in endemic areas, and that 75 % of the global disease burden is concentrated in Southeast Asia and the Western Pacific. In addition to causing high morbidity and mortality, DHF places a substantial economic burden, with an average treatment cost of United States Dollar (USD) 514 for outpatient care and USD 1,491 for inpatient care, particularly in communities with limited healthcare access and environments favorable to the breeding of Aedes aegypti and Aedes albopictus mosquitoes [1].

The WHO identifies Indonesia as one of the countries with the highest dengue burden in Southeast Asia, reporting more than 100,000 cases annually and a Case Fatality Rate (CFR) of approximately 1%, which is significantly higher than that of Thailand, where the CFR has been reduced to below 0.2%. Compared with Latin America and the Caribbean, where the average CFR is about 1.2 %, Indonesia's

mortality rate is similar, but the frequency and magnitude of annual outbreaks tend to be greater, indicating more serious control challenges [1].

At the national level, DHF has been a serious public health problem since it was first reported in 1968. Cases continue to rise, with 143,000 cases reported in 2022 and an Incidence Rate (IR) of 52 per 100,000 population, far exceeding the national target of ≤10. Although the CFR has declined to below 1 % since 2008, the actual number of cases is estimated to be up to 50 times higher due to underreporting, rapid urbanization, climate change, and still-limited surveillance systems [2].

Numerous studies have demonstrated the growing application of machine learning for early prediction and diagnosis of dengue. One study combined five classification algorithms to compare the performance of decision tree–based and probabilistic methods, highlighting the importance of feature selection and data preprocessing, and reported that J48 achieved an accuracy of up to 96.58 % [3]. A Naïve Bayes approach utilized hematological data to distinguish DHF from typhoid fever, achieving 93.33 % accuracy and 97.62 % recall, and showing strength in handling overlapping symptoms [4].

Another study evaluated Support Vector Machine (SVM), Random Forest, Decision Tree, and K-Nearest Neighbor using patient symptom data and found that SVM achieved an accuracy of 87.76 %, underscoring the importance of algorithm selection for clinical symptom patterns [5]. The development of an Optimized Ensemble Classifier (OEC) that integrates CNN, ANN, and SVM introduced deep learning integration and feature optimization, and was reported to outperform individual methods [6]. Further research applied Backpropagation, Gaussian, and SVM to project long-term case trends and showed that Backpropagation yielded the lowest error (Mean Absolute Percentage Error (MAPE) 0.024), confirming its potential for long-term epidemic monitoring [7].

These findings demonstrate that machine learning can improve the accuracy and efficiency of early detection and monitoring of DHF, yet most studies still rely on laboratory data or limited datasets and rarely integrate multi-scenario validation or direct clinical implementation. This study provides several important novel contributions. First, the early detection model for DHF is developed solely from initial clinical symptoms and patient demographic factors, without laboratory test results, making it suitable for healthcare settings with limited resources.

Second, the study employs ensemble machine learning methods (Random Forest, Gradient Boosted Tree, AdaBoost, Bagging, and Random Tree) with comprehensive evaluation across multiple data-splitting scenarios and 5-fold and 10-fold cross-validation to ensure accuracy and model stability. Third, the research is conducted directly on electronic medical records of patients from the Regional Public Hospital (RSUD) Brigjend H. Hasan Basry Kandangan, reflecting real clinical conditions in Indonesia, thereby making the results more representative for the practical implementation of early dengue detection in healthcare services.

2. RESEARCH METHOD

This study followed the Cross-Industry Standard Process for Data Mining (CRISP-DM), which consists of six stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The study's workflow is illustrated in Figure 1.

2.1 Business Understanding

This stage aimed to identify the research problem and determine the appropriate data analysis approach. At RSUD Brigjend H. Hasan Basry Kandangan, 319 DHF cases were recorded in 2024, highlighting the urgency of early detection. Diagnosis delays often occur due to the reliance on laboratory tests, which require significant time and resources.

To address this issue, a literature review, direct observation, and interviews with medical staff were conducted. These efforts confirmed the potential of ensemble learning, particularly Random Forest, to handle overlapping clinical symptoms and provide accurate predictions. Ethical clearance was obtained from the Health Research Ethics Committee (KEPK), and permission was obtained from the hospital administration prior to data collection.

2.2 Data Understanding

Data were obtained from hospital medical records, including demographic (age, gender) and clinical attributes relevant to DHF detection. Attributes were selected in accordance with Ministry of Health guidelines and validated by medical staff to ensure clinical relevance. The focus was on features available at the first patient encounter, without laboratory dependency.

2.3 Data Preparation

Data preprocessing included three main steps:

1. Feature Selection – Relevant attributes were chosen based on literature, clinical validation, and information gain analysis. Features with low correlation or redundancy were excluded.
2. Data Transformation – Categorical variables were converted into numerical form. The attribute *fever* was excluded from training because of uniform values across all cases [8];[1], although it remains clinically important and is displayed in the prototype interface.
3. Data Splitting – The dataset was divided into training and testing sets with three ratios: 80:20, 70:30, and 60:40. Stratified sampling ensured balanced class distribution. To improve reliability, 5-fold and 10-fold cross-validation were applied.

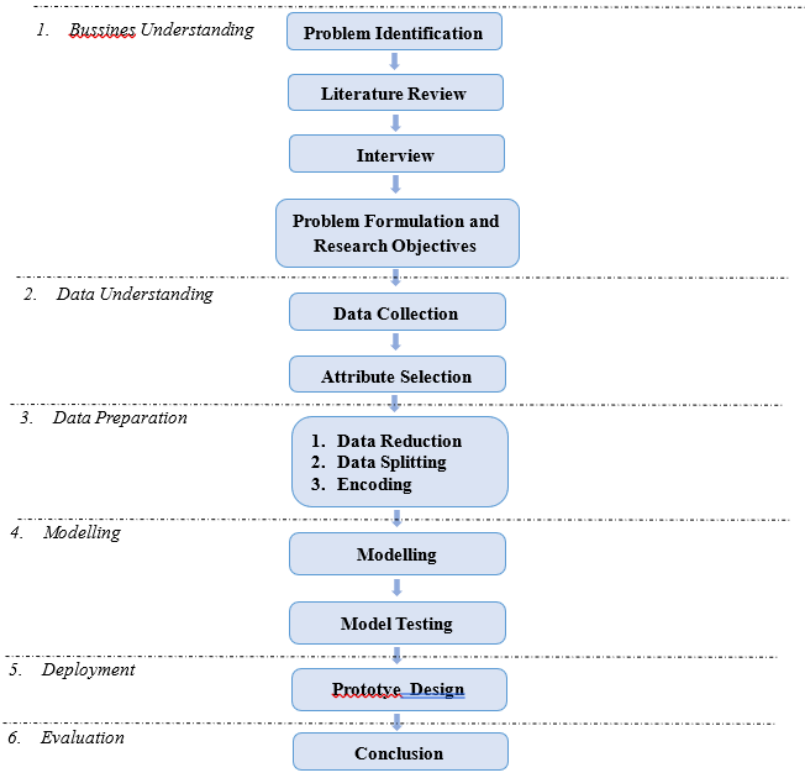

Figure 1. Research Methodology Framework Based on CRISP-DM

2.4 Modeling

This study applied five ensemble learning algorithms, Random Forest, Bagging, Gradient Boosted Tree, AdaBoost, and Random Tree, to build and evaluate predictive models for early dengue detection. Each algorithm is briefly described below, including its theoretical background, key mathematical formulation, and supporting references.

2.4.1 Random Forest

Random Forest is an ensemble learning algorithm for classification and regression. It constructs multiple decision trees on bootstrap samples of the training data and aggregates their outputs to improve predictive accuracy and reduce the risk of overfitting [9]. At each node, a random subset of features is selected, ensuring diversity among trees and yielding more stable combined predictions [10].

This algorithm can handle large, high-dimensional datasets, is robust to outliers and missing values, and provides feature-importance estimates by measuring the decrease in accuracy when feature values are permuted [11]. Such importance measures help identify key variables and support model interpretation [12].

The overall procedure of Random Forest can be described in three main stages, each contributing to the robustness and predictive power of the final model:

1. Bootstrap sampling from a dataset of size N, draw N samples with replacement to create a bootstrap sample for each tree.
2. Tree construction For each bootstrap sample, grow an unpruned decision tree. At each node, choose a random subset of features and determine the best split using criteria such as the Gini index or information gain [13].
3. Ensemble aggregation: Combine predictions from all trees. For classification, use majority voting; for regression, use the mean prediction.

The regression prediction for an input x is expressed as

$$\hat{f}(x) = \frac{1}{B}\sum_{b=1}^{B} f_b(x) \quad (1)$$

Where B is the number of trees and $f_b(x)$ is the prediction of the *b-th tree.*

For classification, the final output is the most frequently predicted class:

$$\hat{y} = mode\{f_b(x)\}_{b=1}^{B} \quad (2)$$

Because of its high accuracy, resistance to noise, and ability to explain variable importance, Random Forest is highly suitable for early detection of dengue fever in healthcare applications.

2.4.2 Bagging

Bagging (bootstrap aggregating) is an ensemble learning method designed to improve the stability and accuracy of predictive models by reducing variance. It trains multiple base learners on bootstrap samples random subsets of the training data drawn with replacement and then aggregates their predictions using majority voting for classification or averaging for regression [14].

The final prediction for an input x is expressed as

$$\widehat{f_{bag}}(x) = \frac{1}{B}\sum_{b=1}^{B} \widehat{f^{(b)}}(x) \quad (3)$$

Where B is the number of models and $\widehat{f^{(b)}}(x)$ is the prediction of the b-th model. For classification, the combined output is

$$\widehat{y_{bag}} = \text{mode}\left(\widehat{y^{(1)}}, \widehat{y^{(2)}}, \ldots, \widehat{y^{(B)}}\right) \quad (4)$$

Bagging operates through three principal stages:

1. Bootstrap sampling from a dataset of size N, draw N samples with replacement to form a bootstrap dataset for each base learner. This ensures diversity among training sets and lowers the correlation between models.
2. Model training: Build an independent base learner (e.g., decision tree) on each bootstrap sample without pruning.
3. Ensemble aggregation: Combine the outputs of all models by majority vote (classification) or mean (regression) to produce the final prediction.

From a statistical perspective, when B independent learners are aggregated, the prediction variance decreases according to

$$\text{Var}_{bag} = \frac{1}{B}\text{Var}(f) \quad (5)$$

yielding more stable and noise-resistant predictions [15].

Further developments such as evolutionary bagging, temporal bagging, and weighted bagging, enhance diversity or assign adaptive weights to base learners, thereby improving accuracy and generalization in dynamic or imbalanced data settings [16].

With its simple implementation, flexibility in base models, and proven ability to mitigate overfitting, Bagging remains a widely adopted and dependable ensemble technique in modern machine learning.

2.4.3 Gradient Boosting

Gradient Boosting is an ensemble learning algorithm that builds predictive models in a stage-wise fashion by iteratively minimizing a specified loss function. Each new weak learner corrects the residual errors of the previous model by following the negative gradient of the loss function with respect to the current predictions [17].

The initial model $F_0(x)$ is a constant that minimizes the overall loss. For a squared-error loss,

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} (y_i - \gamma)^2 \quad (6)$$

At each iteration m, the residual (negative gradient) is computed as

$$r_i^{(m)} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad (7)$$

and a weak learner $h_m(x)$ is fitted to these residuals. The model is updated by

$$F_m(x) = F_{m-1}(x) + \nu \cdot h_m(x) \quad (8)$$

where the learning rate ν (typically ≤0.1) serves as a regularization parameter to prevent overfitting.
After M iterations, the final prediction is

$$\hat{y}(x) = F_M(x) = \sum_{m=1}^{M} \nu \cdot h_m(x) \quad (9)$$

Gradient Boosting is essentially a functional gradient descent method, applicable to various differentiable loss functions, such as log-loss for classification or Huber loss for robust regression. Key variants include Stochastic Gradient Boosting, which uses random subsets of data at each iteration to reduce variance and improve generalization [18], and Regularized Gradient Boosting, which constrains tree complexity or applies L2 penalties to balance fit and model complexity [19].

With its stage-wise optimization, additive model structure, and flexibility for different loss functions, Gradient Boosting remains one of the most powerful and widely used ensemble methods in modern machine learning.

2.4.4 AdaBoost

Adaptive Boosting (AdaBoost) is an ensemble learning method that combines multiple weak learners typically decision stumps into a single strong classifier. The algorithm trains weak learners sequentially, where each successive model focuses on correcting the errors of the previous one [14].

Initially, all training samples are assigned equal weights $w_i^{(1)} = \frac{1}{N}$. At iteration t, a weak learner $h_t(x)$ is trained to minimize the weighted error.

$$\varepsilon_t = \sum_{i=1}^{N} w_i^{(t)} \cdot I(h_t(x_i) \neq y_i) \quad (10)$$

Where I is the indicator function and y_i is the true label of the sample i. if $\varepsilon_t < 0.5$, a learner's weight is calculated as

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \quad (11)$$

Sample weights are then updated to emphasize misclassified instances:

$$w_i^{(t+1)} = w_i^{(t)} \cdot \exp\left(-\alpha_t y_i h_t(x_i)\right) \quad (12)$$

and normalized so that $\sum_i w_i^{(t+1)} = 1$. This procedure ensures that difficult cases receive higher influence in subsequent iterations [20].
After T rounds, the final strong classifier is given by

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \quad (13)$$

The theoretical strength of AdaBoost lies in margin theory, which explains its ability to not only minimize training error but also enlarge the classification margin, thereby improving generalization and reducing overfitting even after many iterations [21].

Extensions such as AdaBoost.PCE and AdaBoost-CNN integrate boosting with regression or deep-learning models, while variants like Dynamic Clustering and Undersampling Boost (DYCUSBoost) and Particle Swarm Optimization - Adaptive Boosting (PSO-AdaBoost) address class imbalance and noisy data [22],[23]. With its adaptive weighting, mathematically grounded update rules, and flexibility to combine various base learners, AdaBoost remains one of the most influential and effective ensemble algorithms in modern machine learning.

2.5 Evaluation Confusion Matrix

In machine learning and data mining, model evaluation is crucial to assess how accurately a classifier predicts unseen data. A widely used tool is the confusion matrix, which compares predicted labels with actual labels to visualize classification performance [24].

A confusion matrix consists of four key components:

1. True Positive (TP) : correctly predicted positive samples.
2. True Negative (TN): correctly predicted negative samples.
3. False Positive (FP) : negative samples incorrectly predicted as positive (type I error).
4. False Negative (FN) : positive samples incorrectly predicted as negative (type II error).

It can be represented as

$$\begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \tag{14}$$

From this matrix, several quantitative metrics are derived:

1. Accuracy the proportion of correct predictions:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{15}$$

While intuitive, accuracy can be misleading when class imbalance exists [25].

2. Precision – the fraction of predicted positives that are true positives:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{16}$$

Precision is important when false positives are costly.

3. Recall (Sensitivity) – the fraction of actual positives correctly identified:

$$\text{Recall} = \frac{TP}{TP+FN} \tag{17}$$

High recall is crucial when minimizing false negatives is essential.

4. F1 Score the harmonic mean of precision and recall:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{18}$$

F1 is well-suited to imbalanced datasets.

5. Matthews Correlation Coefficient (MCC) is a balanced measure, even with class imbalance:
MCC ranges from −1 (inverse prediction) to +1 (perfect prediction).

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{19}$$

By applying these metrics, the performance of ensemble algorithms such as Random Forest, Bagging, Gradient Boosting, and AdaBoost can be objectively and comprehensively evaluated, ensuring the chosen model meets the practical requirements of specific applications.

2.6 Deployment

The best-performing model was deployed into a prototype web application using Python and Streamlit. The application enables healthcare practitioners to upload patient data, visualize model performance (confusion matrix, Receiver Operating Characteristic (ROC) curve, feature importance), and generate early detection predictions. This ensures that the study contributes not only theoretically but also practically in supporting medical decision-making.

3. RESULTS AND ANALYSIS

3.1 Dataset Characteristics

The dataset used in this study consists of 649 patient records collected from RSUD Brigjend H. Hasan Basry Kandangan. Among them, 352 patients (54.2%) were diagnosed with dengue fever (DBD) and 297 patients (45.8%) were classified as non-dengue. This relatively balanced distribution is important as it reduces potential bias during model training and evaluation. The class proportion is illustrated in Figure 2.

In addition to demographic attributes (gender and age), the dataset includes several early clinical features recorded during the patient's first medical examination. These attributes were validated by medical staff to ensure clinical relevance and consistency with national health guidelines. The complete list of attributes is presented in Table 1.

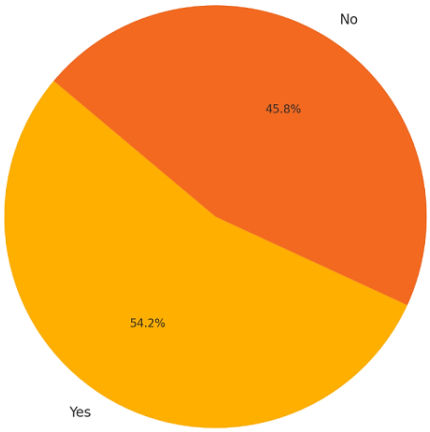

Figure 2. Distribution of Dengue and Non-Dengue Cases

Table 1. Selected Attributes

Category	Attribute	Description
Demographic	Gender	Male / Female
Demographic	Age	Patient's age
Clinical	Body Temprature	Temperature (°C)
Clinical	Fever	Yes / No
Clinical	Headache	Yes / No
Clinical	Muscle Pain	Yes / No
Clinical	Joint Pain	Yes / No
Clinical	Anorexia	Yes / No
Clinical	Nausea/Vomiting	Yes / No
Clinical	Abdominal Pain	Yes / No
Clinical	Skin Rash	Yes / No
Clinical	Mucosal Bleeding	Yes / No
Clinical	Lethargy	Yes / No
Label	Dengue Status	Dengue / Non-dengue

This dataset provides a comprehensive representation of demographic and early clinical information, making it highly relevant for developing an early detection model for dengue fever using clinical data that can be accessed more quickly compared to laboratory results.

3.2 Data Understanding

The data understanding phase aims to obtain a comprehensive overview of the dataset, including its structure, attributes, and the relationships between variables. This step is crucial to ensure that the dataset is representative, clinically relevant, and suitable for building an effective classification model.

The dataset consists of 12 predictor attributes and 1 target attribute (dengue status). The predictors include demographic features (gender, age) and clinical symptoms such as headache, muscle pain, joint pain, anorexia, nausea/vomiting, abdominal pain, skin rash, mucosal bleeding, and lethargy. All attributes were validated in collaboration with medical staff to ensure alignment with clinical practices.

To examine the relationships among variables, a correlation heatmap was generated, as presented in Figure 3. The visualization shows that most features exhibit low correlation with one another, indicating that they contribute complementary information for the model. This diversity in attributes is beneficial for machine learning, as it reduces redundancy and enhances classification performance.

The heatmap highlights that Mucosal Bleeding has a relatively strong correlation with the dengue status (0.68), suggesting that this attribute plays an important role in dengue classification. Additionally, Lethargy shows a moderate correlation with both dengue status (0.36) and mucosal bleeding (0.36), implying that patients with lethargy are more likely to experience mucosal bleeding and potentially dengue infection.

These findings confirm that the dataset contains a diverse set of attributes with varying levels of correlation, which provides a strong foundation for feature selection and the subsequent modeling phase.

3.3 Data Preparation

Before building the model, the dataset was preprocessed to ensure data quality and retain only informative attributes. One important finding was that the Fever attribute had a uniform value ("Yes") across all records. In machine learning, attributes without variation provide no information gain and cannot contribute to distinguishing between classes. Therefore, this attribute was excluded from the training process to prevent redundancy and reduce the risk of overfitting.

Nevertheless, fever remains a clinically significant symptom in the early diagnosis of dengue fever [1]. For this reason, it was preserved in the documentation and prototype interface for medical interpretation, even though it was not used during model training [8].

After preprocessing, the final dataset consisted of 12 predictor attributes and 1 target attribute. This refined dataset was then prepared for the modeling and evaluation phases described in the following sections.

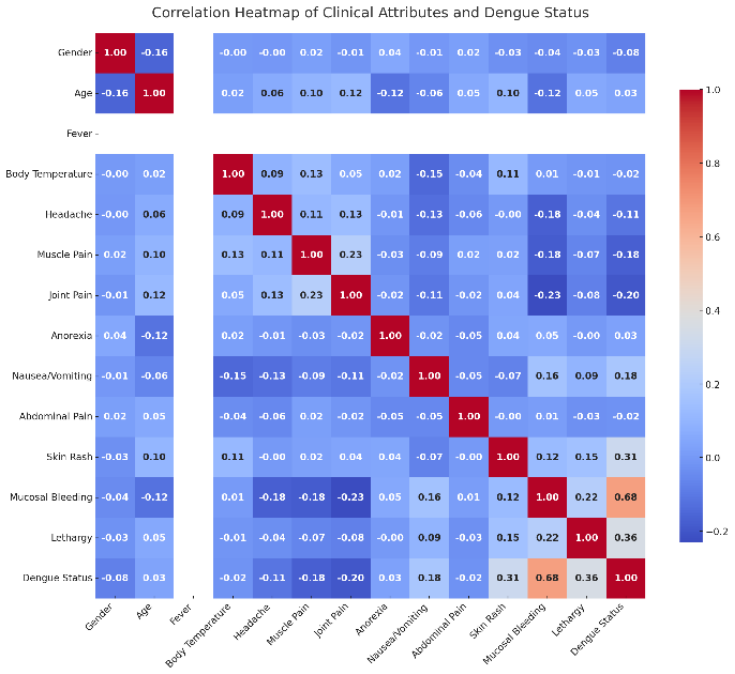

Figure 3. Correlation Heatmap of Clinical Attributes and Dengue Status

3.4 Modeling and Experimental Setup

This study implemented five machine learning algorithms: Random Forest, Random Tree, Bagging, Gradient Boosted Tree, and AdaBoost. These methods were selected because they are widely applied in medical classification problems, particularly for handling heterogeneous datasets with overlapping clinical features. Random Forest and Bagging are ensemble tree-based approaches known for robustness, while AdaBoost and Gradient Boosted Tree iteratively refine weak learners. Random Tree served as a baseline comparator.

The dataset was split into training and testing subsets using stratified random sampling with ratios of 80:20, 70:30, and 60:40, preserving class balance between dengue and non-dengue cases. To further evaluate robustness, k-fold cross-validation (k = 5 and k = 10) was applied.

The evaluation metrics included Accuracy, Precision, Recall (Sensitivity), Specificity, F1-score, and Area Under the Curve (AUC) ROC to ensure comprehensive performance assessment. The results of the experiments are summarized in Table 2. Random Forest consistently outperformed other algorithms across all split ratios. In the 80:20 split, it achieved the highest performance with 90.00% accuracy and 0.967 AUC. Bagging and AdaBoost also performed competitively, while Gradient Boosted Tree showed lower stability across different splits.

Table 2. Performance of Models Across Different Data Splits

Split Ratio	Random Forest (Acc/AUC)	Bagging (Acc/AUC)	AdaBoost (Acc/AUC)	Gradient Boosted Tree (Acc/AUC)
80:20	90.00% / 0.967	85.38% / 0.904	85.38% / 0.893	86.92% / 0.902
70:30	88.21% / 0.940	87.18% / 0.931	87.69% / 0.918	85.13% / 0.881
60:40	89.62% / 0.928	88.85% / 0.917	85.38% / 0.858	83.85% / 0.833

Since five ensemble algorithms were tested, a comprehensive performance comparison was first conducted. Table 3 summarizes the average accuracy and AUC obtained from 5-fold and 10-fold cross-validation for Random Forest, Gradient Boosted Tree, AdaBoost, Random Tree, and Bagging. Among these methods, Random Forest consistently achieved the highest and most balanced scores, indicating strong predictive power and stability. Gradient Boosted Tree and AdaBoost showed competitive accuracy but slightly lower AUC values, while Random Tree and Bagging performed less consistently across folds.

Because of this superior and stable performance, subsequent detailed analysis focused on Random Forest. Specifically, Random Forest reached an average accuracy of 88.91 % with an AUC of 0.939 in 5-fold validation and 88.29 % with an AUC of 0.947 in 10-fold validation, confirming its robustness and generalizability.

Table 3. Performance of Models Cross-Validation

Algorithm	Cross-Validation	Accuracy	AUC
Random Forest	5-fold	88.91%	0.939
	10-fold	88.29%	0.947
Bagging	5-fold	89.21%	0.913
	10-fold	87.98%	0.917
AdaBoost	5-fold	86.44%	0.904
	10-fold	87.36%	0.899
Gradient Boosted Tree	5-fold	86.90%	0.935
	10-fold	86.75%	0.938

Because Random Forest consistently provided the highest and most stable accuracy and AUC across all split ratios and cross-validation settings (Tables 2 and 3), it was selected for further optimization. Hyperparameter tuning was then performed to refine the model while preventing overfitting. The best configuration listed in Table 4 uses *n_estimators* = 100, *max_depth* = 5, and the *gini* criterion, achieving a mean cross-validation accuracy of 88.91 %.

Table 4. Best Random Forest Parameters

Parameter	Value
n_estimators	100
max_depth	5
criterion	gini
Mean CV Acc.	88.91%

The final evaluation of the Random Forest model was conducted using 10-fold stratified cross-validation. The model achieved an average cross-validation accuracy of 0.8891, while the final accuracy on the full training set reached 0.9122.

The classification report indicated balanced performance across both classes. For the non-dengue class, the model achieved a precision of 0.86, a recall of 0.96, and an F1-score of 0.91. For the dengue class, the precision was 0.97, the recall was 0.87, and the F1-score was 0.91. Overall accuracy across 649 patient records was 0.91, with macro- and weighted-average F1-scores also at 0.91. These results confirm that the Random Forest model provides a reliable and well-balanced classification, successfully identifying both dengue and non-dengue cases with high precision and recall.

3.5 Discussion

The superior performance of Random Forest in this study can be attributed to several technical advantages. Dengue clinical data typically show nonlinear relationships and overlapping symptoms with other febrile illnesses. By building multiple diverse decision trees through bootstrap sampling and random feature selection at each node, Random Forest effectively captures complex patterns while reducing overfitting. The algorithm can handle high-dimensional data without extensive feature selection, remains robust to missing values and outliers, and provides feature-importance analysis that highlights key predictors such as mucosal bleeding and lethargy. Together, these capabilities explain its stable and superior

performance across all data-split and cross-validation schemes, even when relying solely on early clinical and demographic indicators.

This work reinforces and extends previous research on machine learning–based dengue prediction. Decision tree methods such as J48 have achieved 96.58% accuracy using laboratory data [3], while Naïve Bayes distinguished dengue from typhoid fever with 93.33% accuracy and 97.62% recall using hematology inputs [4]. In contrast, our model relies only on initial clinical symptoms and demographic data, which are more accessible in primary care settings. On symptom-based datasets, SVM has been reported to reach 87.76% accuracy [5], whereas our Random Forest model surpassed this benchmark, demonstrating stronger robustness. Deep-learning hybrids such as the OEC, which combines Convolutional Neural Network (CNN), Artificial Neural Network (ANN), and SVM, have also shown high accuracy [6], but require significantly greater computational resources. By comparison, the optimized Random Forest in this study offers similar reliability with far lower computational cost, making it more suitable for real-time clinical deployment. At the population level, machine learning has also been applied to long-term forecasting, for example using backpropagation networks that achieved an MAPE of 0.024 for predicting dengue trends in Bali [7]. Our work complements such macro-level approaches by providing patient-level early detection, enabling immediate clinical action.

These findings carry important clinical and public health implications. By leveraging only early clinical and demographic information, the proposed model can function as a rapid, low-cost decision-support tool, particularly in primary health facilities with limited laboratory resources. Feature-importance analysis further improves clinical interpretability, helping healthcare workers prioritize high-risk patients. Other notable strengths include rigorous validation across multiple data-split and k-fold cross-validation schemes and the development of a web-based prototype that facilitates seamless integration into clinical workflows. Nevertheless, limitations remain. The dataset was drawn from a single hospital, so broader multi-center validation is needed to confirm generalizability. In addition, real-time integration with electronic health records and testing in diverse geographic settings are recommended for future work to enhance scalability and long-term reliability.

4. CONCLUSION

This study demonstrated that Random Forest is a powerful and reliable model for early detection of dengue fever using only initial clinical symptoms and demographic data. Among five ensemble learning algorithms tested, Random Forest, Bagging, AdaBoost, and Gradient Boosted Tree, Random Forest consistently achieved the highest and most stable accuracy and AUC across multiple train-test splits and cross-validation schemes. Careful hyperparameter optimization further improved its robustness, resulting in balanced precision, recall, and F1-scores for both dengue and non-dengue classes.

The model's ability to manage complex, nonlinear relationships, handle noisy or missing values, and identify the most influential features explains its superior performance. Clinically, these findings show that early, accurate dengue detection is possible without laboratory tests, supporting faster decision-making and treatment. The web-based prototype developed in this study offers practical integration into healthcare workflows.

While the dataset's single-hospital origin limits generalizability, the methodological rigor and strong results suggest broad potential for scaling and adaptation. Future research should expand validation to multi-center and real-time settings and explore integration with electronic health records to enhance clinical impact and long-term reliability.

REFERENCES

[1] WHO. Dengue Guidelines For Diagnosis, Treatment, Prevention And Control. Vol. 41, Psychiatric News. 2019. 29–29 p.
[2] Kemenkes RI. Membuka Lembaran Baru Untuk Hidup Sejahtera. Lap Tah 2022 Demam Berdarah Dengue. 2022;17–9.
[3] Patel KA, Sandhi MI. Dengue Disease Prediction Using Data Mining Techniques. 2020;8(9):1424–33.
[4] Handayani Y, Hakim AR. Classification of Naive Bayes Algorithm on Dengue Hemorrhagic Fever and Typhoid Fever Based on Hematology Results. J Appl Intell Syst. 2023;8(1):94–9.
[5] Rahman T, Rahman MM. Evaluation of Machine Learning Approaches for Prediction of Dengue Fever [Internet]. Vol. 141, Lecture Notes on Data Engineering and Communications Technologies. Springer Nature Singapore; 2023. 165–175 p. Available from: http://dx.doi.org/10.1007/978-981-19-3035-5_13
[6] Shaikh MSG, SureshKumar DB, Narang DG. Development of optimized ensemble classifier for dengue fever prediction and recommendation system. Biomed Signal Process Control [Internet]. 2023;85(March):104809. Available from: https://doi.org/10.1016/j.bspc.2023.104809
[7] Yudha Arya Dala IM, Gede Darma Putra IK, Wira Buana P. Forecasting Cases of Dengue Hemorrhagic Fever Using the Backpropagation, Gaussians and Support-Vector Machine Methods. J RESTI (Rekayasa Sist dan

Teknol Informasi). 2021;5(2):335–41.
[8] Ha J, Kambe M, Pe J. Data Mining [Internet]. Data Mining: Concepts and Techniques. Elsevier; 2012. 1–703 p. Available from: https://linkinghub.elsevier.com/retrieve/pii/C20090618195
[9] Biau G, Scornet E. A random forest guided tour. TEST [Internet]. 2016 Jun 19;25(2):197–227. Available from: http://link.springer.com/10.1007/s11749-016-0481-7
[10] Suci Amaliah, Nusrang M, Aswi A. Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. J Stat Its Appl Teach Res. 2022;4(3):121–7.
[11] Supriyadi R, Gata W, Maulidah N, Fauzi A. Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah. E-Bisnis J Ilm Ekon dan Bisnis. 2020;13(2):67–75.
[12] Amalia N, Asmunin. Optimasi Algoritma Random Forest dengan Hyperparameter Tuning Menggunakan GridSearchCV untuk Prediksi Nasabah Churn pada Industri Perbankan. Manaj Inf. 2024;16(1):1–9.
[13] Sagita A, Faqih A, Dwilestari G, Siswoyo B, Pratama D. Penerapan Metode Random Forest Dalam Menganalisis Sentimen Pengguna Aplikasi Capcut Di Google Play Store. JATI (Jurnal Mhs Tek Inform. 2024;7(6):3307–13.
[14] Huang F, Xie G, Xiao R. Research on Ensemble Learning. In: 2009 International Conference on Artificial Intelligence and Computational Intelligence. 2009. p. 249–52.
[15] Ngo G, Beard R, Chandra R. Evolutionary bagging for ensemble learning. Neurocomputing [Internet]. 2022;510:1–14. Available from: https://www.sciencedirect.com/science/article/pii/S0925231222010414
[16] Pham VT, Le Thi HA, Luu HPH, Damel P. DCA-Based Weighted Bagging: A New Ensemble Learning Approach BT - Intelligent Information and Database Systems. In: Nguyen NT, Boonsang S, Fujita H, Hnatkowska B, Hong TP, Pasupa K, et al., editors. Singapore: Springer Nature Singapore; 2023. p. 121–32.
[17] Biau G, Cadre B. Optimization by Gradient Boosting. Adv Contemp Stat Econom Festschrift Honor Christine Thomas-Agnan. 2021;23–44.
[18] Friedman JH. Stochastic gradient boosting. Comput Stat Data Anal [Internet]. 2002;38(4):367–78. Available from: https://www.sciencedirect.com/science/article/pii/S0167947301000652
[19] Cortes C, Mohri M, Storcheus D. Regularized gradient boosting. Adv Neural Inf Process Syst. 2019;32(NeurIPS).
[20] Cha Y. AdaBoost Algorithm with Classification Belief. J Appl Sci [Internet]. 2015; Available from: https://api.semanticscholar.org/CorpusID:123554062
[21] Sun B, Chen H, Wang J. An empirical margin explanation for the effectiveness of DECORATE ensemble learning algorithm. Knowledge-Based Syst [Internet]. 2015;78:1–12. Available from: https://www.sciencedirect.com/science/article/pii/S095070511500012X
[22] Li D, Liu Z, Armaghani DJ, Xiao P, Zhou J. Novel Ensemble Intelligence Methodologies for Rockburst Assessment in Complex and Variable Environments. Sci Rep. 2022;
[23] Li K, Zhou G, Zhai J, Li F, Shao M. Improved PSO_AdaBoost ensemble algorithm for imbalanced data. Sensors (Switzerland). 2019;19(6).
[24] Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics [Internet]. 2020 Dec 2;21(1):6. Available from: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-019-6413-7
[25] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag [Internet]. 2009;45(4):427–37. Available from: https://www.sciencedirect.com/science/article/pii/S0306457309000259

BIBLIOGRAPHY OF AUTHORS

Achmad Saleh is a Master of Computer Science student at Universitas Budi Luhur. As an emerging researcher from South Kalimantan, his primary interests lie in data mining and data science. His research focuses on applying data mining techniques to local problems in South Kalimantan.

Ridha Mukhtar is also a Master of Computer Science student at Universitas Budi Luhur. His research is focused on leveraging data science to improve the efficiency of information systems. He is interested in developing data-driven solutions for key sectors in South Kalimantan.

Rusdah, Dr. Rusdah, S.Kom., M.Kom is a lecturer at the Faculty of Computer Science, Universitas Budi Luhur. His research focuses on Data Mining, Machine Learning, Data Science, Decision Support Systems, and Information Systems. He has made significant contributions to these fields, as reflected in his scholarly publications.