Pamukkale Univ Muh Bilim Derg, 32(3), 551-561, 2026

Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi

Pamukkale University Journal of Engineering Sciences

# Improved arbovirus suspected case analysis via ensemble methods with parameter tuning: Insights from SISA dataset

## Parametre ayarlaması ile geliştirilmiş topluluk yöntemleri kullanarak arbovirüs şüpheli vaka analizi: SISA veri kümesinden çıkarımlar

*Alican Doğan*[1*]

[1]Department of Management Information Systems, Faculty of Applied Sciences, Bandırma Onyedi Eylul University, Bandırma, Türkiye.

alicandogan@bandirma.edu.tr

Received/Geliş Tarihi: 30.01.2025 Revision/Düzeltme Tarihi: 26.09.2025 doi: 10.65206/pajes.24040

Accepted/Kabul Tarihi: 01.10.2025 Research Article/Araştırma Makalesi

**Abstract**

*Hospital admission necessity of a patient who is under care for the possibility of arbovirus infection is a critical decision for healthcare practitioners. Medical staff may experience stress when making this decision due to the potential risks it poses to the broader community. Current capacities for diagnosis can be confusing. For this reason, data mining approaches have been proven to be highly effective in the diagnosis of diseases as well as in many other fields. As many research studies suggest, they can also be used to decide whether a patient with arbovirus infection should be hospitalized or not. For this purpose, this study uses Severity Index for Suspected Arbovirus (SISA) dataset and implements various machine learning classification techniques with the aim of binary classification to detect the hospitalization status of a specific patient. Several neural networks, single classifiers, and ensemble supervised learning methods are selected as classifiers during the experiments. The best classification accuracy value is obtained by Random Forest (RF) model with 0.9908. This model has been shown to outperform many data mining techniques previously applied in prominent studies. This improved result leads to additional experiments with a different number of estimators when implementing RF. The outcome also improves the maximum classification performance up to 0.9926 using 25 estimators. The study reveals the effectiveness of ensemble models, especially bagging and boosting approaches, for Arbovirus suspected case analysis.*

**Keywords:** Machine learning, Arboviral infection, Random forest, Hospitalization.

**Öz**

*Arbovirüs enfeksiyonu şüphesiyle gözlem altında tutulan bir hastanın hastaneye yatış gerekliliği, sağlık profesyonelleri için kritik bir karardır. Tıbbi personel, bu kararın sağlıklı bireyler üzerindeki potansiyel riskleri nedeniyle baskı altında olabilir. Mevcut teşhis olanakları zaman zaman kafa karıştırıcı olabilir. Bu nedenle, veri madenciliği yaklaşımlarının hastalıkların teşhisinde ve birçok farklı alanda oldukça etkili olduğu kanıtlanmıştır. Yapılan araştırmalar, veri madenciliği yöntemlerinin arbovirüs enfeksiyonu taşıyan bir hastanın hastaneye yatırılıp yatırılmaması kararında da kullanılabileceğini göstermektedir. Bu amaç doğrultusunda, bu çalışma Şüpheli Arbovirüs Vakaları için Şiddet İndeksi (SISA) veri kümesini kullanarak, bir hastanın hastaneye yatış durumunu belirlemek için ikili sınıflandırma gerçekleştiren çeşitli makine öğrenmesi tekniklerini uygulamaktadır. Deneylerde sınıflandırıcı olarak çeşitli yapay sinir ağları, tekil sınıflandırıcılar ve topluluk destekli öğrenme yöntemleri kullanılmıştır. En yüksek sınıflandırma doğruluğu, %99,08 ile Rastgele Orman (Random Forest-RF) modeli tarafından elde edilmiştir. Bu modelin, literatürdeki önemli araştırmalarda uygulanan birçok veri madenciliği tekniğinden daha etkili olduğu kanıtlanmıştır. Bu olumlu sonuçlar, RF modelinin farklı sayıda tahmin edici (estimators) ile ek deneyler yapılmasını teşvik etmiştir. Çalışma sonucunda, en yüksek sınıflandırma performansı 25 tahmin edici kullanılarak 0.9926'ya yükseltilmiştir. Elde edilen bulgular, arbovirüs şüpheli vaka analizinde özellikle torbalama (bagging) ve güçlendirme (boosting) yaklaşımlarına dayalı topluluk modellerinin etkinliğini ortaya koymaktadır.*

**Anahtar kelimeler:** Makine öğrenimi, Arbovirüs enfeksiyonu, Rastgele orman, Hastane yatış durumu.

## 1 Introduction

In recent years, the number of visits to hospital emergency departments (ED) has risen substantially. Making clinical decisions in the context of arboviral infections is especially complex in settings with limited resources. This situation causes emergency department overcrowding, increasing morbidity and mortality and disrupting clinical workflows. Severe crowding in EDs has resulted in increased mortality due to delays in care and heightened risks for critically ill patients. Infrastructural constraints create a demanding clinical environment, particularly as healthcare providers must decide whether hospitalization is necessary for patients suspected of having arboviral infections. Managing a surge of cases within a short timeframe has significantly strained the functioning of healthcare systems, making their operation increasingly challenging. All these circumstances necessitate the implementation of an effective and efficient triage system to prioritize care and optimize resource utilization [1],[2].

Triage is a common method used to determine which patients should be hospitalized. Triage improves decision performance by healthcare providers, adjusts the timing, and queue position of patients. This decision should be taken before starting the treatment process. Efficient use of the resources of a clinic depends on this decision. The right decision leads to a decrease in treatment expenditure. Triage, while conceptually straightforward, becomes challenging due to factors such as time constraints, limited patient information, diverse medical conditions, and a heavy dependence on clinicians' intuition. Current triage systems often struggle to assess critical illnesses because too few physicians are available for too many patients, limiting the time spent on each case. Innovative computational

*Corresponding author/Yazışılan Yazar

algorithms present a promising solution to support decision-making in optimizing triage processes [3]-[5].

Machine learning combines principles of statistics and computer science to analyze and derive insights from extensive datasets effectively. Unlike traditional statistical methods like regression models, machine learning approaches impose fewer assumptions about data distributions and variable relationships, allowing for greater flexibility in handling complex and non-linear patterns. They are not similar to statistical learning modelling. Machine learning approach provides a wide range of assumptions about the data distribution and is capable of discovering interesting relationships between features. An increase in the number of features and instances generally improves prediction performance for machine learning models [6].

This study focuses on creating a machine learning (ML) application to predict whether patients should be hospitalized during the triage phase. The goal is to enhance the decision-making process by providing physicians with more data in a short time, improving accuracy, and optimizing resource use. This study focuses on patients with suspected arboviral infections, which are common in tropical areas and cause fever, fatigue, and joint pain. These infections, transmitted by mosquitoes and flies, include diseases like malaria, river blindness, and yellow fever, which account for millions of deaths globally. For instance, malaria caused an estimated 435,000 deaths in 2017.

Emerging arboviruses, such as the Zika virus, dengue virus (DENV), chikungunya virus (CHIKV), and Zika virus (ZIKV), present significant public health challenges, particularly under changing climatic conditions. While most infected patients recover with mild symptoms, a small percentage develop severe conditions that can result in death. The study addresses the critical need for improved tools to manage the clinical challenges associated with these infections effectively [7],[8].

In the rapidly evolving landscape of healthcare, the integration of advanced computational techniques has become essential for optimizing patient outcomes and resource allocation. Among these techniques, data mining and ML have emerged as pivotal tools in the early detection and prediction of a patient's hospitalization status. These methods use large healthcare datasets to reveal patterns that traditional statistics often miss [9]-[11].

Data mining approaches are strategies to extract valuable knowledge from vast and unorganized data. Data mining involves extracting meaningful information from large datasets, often by identifying hidden patterns and relationships. In the context of healthcare, this process can analyze electronic health records (EHRs), clinical test results, and demographic information to predict hospitalization risks. For instance, factors such as age, comorbidities, medication adherence, and previous hospital visits can be correlated to anticipate a patient's likelihood of requiring hospital care. By automating the analysis of these complex variables, data mining not only enhances diagnostic accuracy but also aids in prioritizing at-risk patients for preventive interventions [12]-[16].

ML uses data mining insights to build predictive models that improve over time. These models use algorithms to learn from historical data, enabling them to forecast hospitalization outcomes with remarkable precision. Techniques such as supervised learning, unsupervised learning, and deep learning can identify subtle trends in patient data that may indicate worsening health conditions. For example, supervised learning algorithms can classify patients based on hospitalization risk levels by analyzing labeled datasets with known outcomes. Similarly, unsupervised learning can cluster patients into groups with similar risk factors, aiding in personalized care planning. Deep learning methods, particularly neural networks, have demonstrated efficacy in processing unstructured data such as medical imaging or physician notes, further expanding the scope of hospitalization prediction [17].

The application of data mining and machine learning in hospitalization prediction offers several advantages, which are early detection and prevention, resource optimization, cost reduction, and improved patient outcomes. Identifying high-risk patients allows healthcare providers to intervene early, potentially preventing hospital admissions. Hospitals can allocate beds, staff, and medical equipment more efficiently by predicting patient inflows. By reducing unnecessary admissions and optimizing care pathways, these technologies help decrease healthcare costs for both providers and patients. Finally, proactive management of health risks enhances the overall quality of care, leading to better health outcomes [18],[19].

Despite their transformative potential, the adoption of data mining and ML in healthcare is not without challenges. Issues such as data privacy, algorithm bias, and the need for high-quality datasets remain significant hurdles. Additionally, integrating these technologies into existing healthcare workflows requires collaboration between data scientists, clinicians, and policymakers.

Future research should focus on developing interpretable ML models that provide actionable insights to healthcare professionals. Moreover, ensuring the ethical use of patient data and addressing disparities in algorithmic predictions will be critical for maximizing the impact of these technologies.

## 2 Related work

Several important studies in the literature have used the SISA dataset and made predictions about the hospitalization status of a patient with arbovirus infection using machine learning methods. Lee et al. conducted a retrospective case-control study involving 117 patients diagnosed with chikungunya infection, confirmed via reverse transcription-polymerase chain reaction (RT-PCR). These cases were identified during the August 2008 outbreak and involved individuals who were hospitalized at Tan Tock Seng Hospital in Singapore, the designated national outbreak response center. Predictive tools were developed using classification and regression trees (CART) to differentiate between dengue fever (DF), dengue hemorrhagic fever (DHF), and chikungunya at the time of presentation. These tools aim to support clinical decision-making by identifying patterns and key features associated with each condition, enhancing diagnostic accuracy in the early stages of illness [20].

Sippy et al. analyzed data retrospectively from a prospective arbovirus surveillance study conducted in Machala, Ecuador. The study spanned November 2013 to September 2017 and included participants aged six months and older who were recruited from clinical sites of the Ecuadorian Ministry of Health (MoH). This analysis aimed to investigate patterns and outcomes related to arboviral infections in the region. In the Surveillance for Arboviral Infections in Southern Ecuador

(SISA) study, several machine learning algorithms were employed to predict outcomes using only symptom and demographic data. Among these methods, generalized boosting models, elastic net, neural networks, and logistic regression demonstrated strong performance on the test set, achieving accuracies ranging from 89.8% to 96.2%. The Cohen's kappa values varied from 0.00 to 0.77, while the AUC ranged between 0.50 and 0.91. The generalized boosting model (GBM) emerged as the top performer, achieving the highest AUC of 0.91 on the test dataset. It also ranked as the second-best algorithm when evaluated on the training set, indicating its robustness and predictive strength in this context [3].

Huang et al. developed a machine learning (ML) model utilizing an artificial neural network (ANN) to predict dengue outcomes, leveraging patient demographic data and laboratory test results. Upon evaluating its prognostic performance, our ANN-based ML model outperformed established models in predicting severe dengue cases. This indicates its potential as a superior tool for early and accurate prognosis, aiding in clinical decision-making and resource allocation [21].

Gorur et al. classified individuals as either hospitalized or outpatient by applying shallow machine learning algorithms to the SISA and SISAL (Severity Index for Suspected Arbovirus with Laboratory) datasets. The algorithms included Feed Forward Neural Network (FFNN), Probabilistic Neural Network (PNN), and Decision Tree (DecT) with three splitting criteria. The classification results demonstrated significant improvements in performance, achieving an area under the curve (AUC) score of 0.973 and accuracy rates up to 98.73% with FFNN. These outcomes surpassed the performance metrics reported in prior studies related to machine learning and arbovirus prediction, highlighting the effectiveness of our approach [22].

Fathima et al. compared the performance of Support Vector Machine and Naïve Bayes for the same dataset. They differentiated dengue from other febrile illnesses in primary care and forecasted the likelihood of severe arboviral disease within populations, and obtained 90.43% accuracy at most. Likely, Akhtar et al. introduced a dynamic neural network model for real-time prediction of the geographic spread of outbreaks. They forecasted the geographic spread of Zika across the Americas with an average accuracy exceeding 85%, even for prediction horizons extending up to 12 weeks. Additionally, Salim et al. analyzed to determine the most effective machine learning model for predicting dengue outbreaks of five districts in Selangor, Malaysia, with the highest dengue fever incidence from 2013 to 2017. They obtained the best prediction performance with 70.12% accuracy using Support Vector Machine [23]-[25].

Apart from SISA dataset, there are also different studies using deep learning models with the aim of arbovirus infection detection and hospital admission prediction. For instance, Neto et al. provided an overview of current research on automated classification of arboviral diseases, aimed at supporting clinical diagnosis through Machine Learning (ML) and Deep Learning (DL) approaches. Additionally, Neto et al. evaluated different methods for converting tabular data into images, identify the most effective approach, optimize a CNN using random search, and compare its performance against an optimized machine learning algorithm, XGBoost. Ozer et al. intended to inform clinicians about the effectiveness of machine learning with transfer learning for diagnosis, along with a comprehensive comparison of classification performances on hospitalization status datasets, particularly in resource-limited settings where arboviral infection data are often scarce. Similarly, Sciannameo et al. employed a deep learning approach to forecast the spatio-temporal spread of COVID-19, predicting new cases and hospital admissions in Reggio Emilia, Northern Italy [26]-[29].

## 3 Material and method

### 3.1 Dataset description

The dataset, part of an ongoing surveillance study in Machala, Ecuador, includes comprehensive data for individuals with suspected arboviral infections. The information consists of demographic details, past medical histories, symptom records, and laboratory results. Outputs from the analysis focus on determining individuals' hospitalization status and recognizing gender distinctions. The dataset is composed of 543 instances and 28 attributes. These attributes provide information on symptoms, demographics, and medical history of patients. For example, Table 1 shows binary symptom values for some attributes. The value 1 indicates existence and the value 0 indicates absence of the related symptom. Table 2 reveals the data distribution of SISA dataset with respect to hospitalization status.

Table 1. A snapshot of five instances of the dataset with values of some symptomatic attributes.

| Symp Fever 7Days | Symp Head | Symp Nausea | Symp Muscle | Symp Rash | Symp Bleed |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 1 | 0 | 0 |
| 1 | 1 | 0 | 1 | 0 | 0 |

Table 2. Data distribution for all classes in SISA Dataset.

| Classes of Data | # of instances | % |
|---|---|---|
| Hospitalized | 59 | 11.05 |
| Outpatient | 475 | 88.95 |
| Total | 534 | 100 |

Several preprocessing steps were applied to ensure data consistency and model compatibility. First, missing or empty values were replaced with zeros, as the machine learning models employed cannot process null entries. Second, comma-based decimal separators were converted to periods to standardize numeric formats across the dataset. All numerical features were then cast from string to float. Additionally, categorical variables such as gender were encoded into integer format (e.g., female = 0, male = 1). These preprocessing steps were essential for enabling effective model training and ensuring reproducible results.

In medical datasets, class imbalance is a common challenge, particularly when the positive class represents a rare but clinically significant condition. In the SISA dataset used in this study, only approximately 11% of the instances correspond to hospitalized patients, while the remaining 89% represent outpatient cases. This substantial imbalance can mislead evaluation metrics such as overall accuracy, as a model may achieve high accuracy simply by predicting the majority class.

To mitigate this issue, we employed several strategies. First, evaluation metrics that are more informative in imbalanced scenarios, such as precision, recall, and F1-score, were

emphasized alongside accuracy. These metrics provide a more nuanced understanding of the model's ability to correctly identify true positives (hospitalized cases) without being biased toward the majority class.

Second, we used stratified k-fold cross-validation to maintain the original class distribution across all training and testing folds. This ensures that each fold reflects the inherent class proportions in the dataset, allowing for more reliable and generalizable performance evaluation.

Overall, these measures help ensure that the models are not only accurate in general, but also effective in identifying the minority class, which is critical for real-world clinical applications where under-detection of hospitalized cases could have serious consequences.

### 3.2 Ensemble learning and random forest method

Ensemble methods in machine learning improve predictive accuracy and robustness by combining multiple models to reach a collective decision. These methods use strategies such as stacking, bagging, boosting, and majority voting. Each technique has unique characteristics. The mathematical model of general ensemble learning is given in Equation (1), where $M$ is the number of models and F is the aggregation function:

$$\hat{y}_{ensemble} = F(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_M) \quad (1)$$

Stacking integrates different models by training a meta-model to synthesize their outputs, offering flexibility but requiring higher computational resources. In stacking, a meta-model $g$ learns to combine the predictions of multiple base models as shown in Equation (2):

$$\hat{y}_{ensemble} = g(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_M) \quad (2)$$

Bagging (e.g., Random Forest) reduces variance by averaging predictions from independently trained models, enhancing stability. It is the short form of bootstrap aggregation. The mathematical model of bagging for regression is given in Equation (3) and for classification is given in Equation (4), which is equal to majority voting.

$$\hat{y}_{ensemble} = \frac{1}{M}\sum_{i=1}^{M} \hat{y}_i \quad (3)$$

$$\hat{y}_{ensemble} = mode(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_M) \quad (4)$$

Boosting (e.g., Gradient Boosting) focuses on correcting errors by training models sequentially, improving accuracy, but at the cost of sensitivity to overfitting. The prediction is given in Equation (5).

$$\hat{y}_{ensemble} = \sum_{i=1}^{M} \hat{y}_i a_i \quad (5)$$

Majority Voting uses a straightforward approach of selecting the most common output, making it efficient but sometimes less nuanced in decision-making.

One of the most popular ensemble learning classifiers is Random Forest algorithm. It is a widely used classification method offering numerous advantages. Known for its robustness compared to other algorithms, it is less susceptible to overfitting. This resilience stems from its approach of averaging predictions across multiple decision trees, which reduces the impact of any single tree that may overfit the data. Additionally, random forest is highly flexible, capable of handling various input data types, including categorical and numerical values, as well as accommodating missing data. Furthermore, it excels at identifying outliers, as they are often isolated within their own trees.

Random Forest (RF) classifiers are also capable of maintaining accuracy even when a substantial portion of the data is missing. Their ability to estimate feature importance and model complex interactions between features with minimal preprocessing makes them especially suitable for tasks such as spam detection. There are $N$ decision trees in RF model. When they are represented as $T$, the final prediction is obtained as given in Equation (6).

$$\hat{y}_{RF} = majorityvote(T_1(x), T_2(x), \dots, T_M(x)) \quad (6)$$

Random Forest (RF) consists of a hierarchy of base classifiers organized in a tree topology. Text data, often characterized by high dimensionality, typically contains numerous irrelevant attributes, while only a limited number of key features are informative for the classifier model. The RF method utilizes a straightforward and predetermined probability to identify and select the most relevant features.

The approach, introduced by Breiman, generates multiple decision trees by randomly sampling subspaces of features and mapping them to subsets of the data. The process begins with constructing individual RF trees, followed by iterative development to enhance the model. The architecture of RF, illustrating its hierarchical structure and feature selection mechanism, is shown in Figure 1.

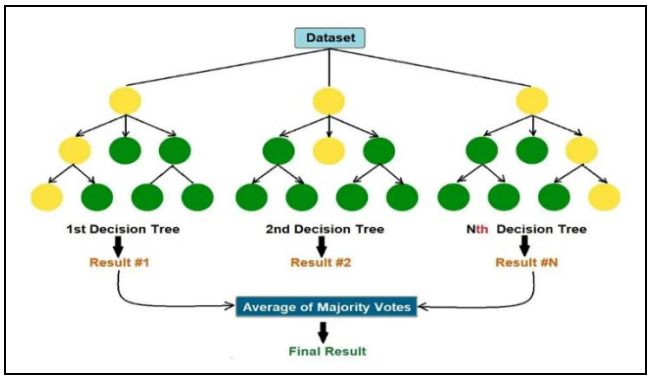


Figure 1. Random forest architecture.

### 3.3 Classification performance evaluation metrics

Evaluating the performance of classification models is critical in determining their effectiveness for specific tasks. Various metrics are used to measure the performance of these models, taking into account their ability to handle imbalanced data, trade-offs between precision and recall, and overall predictive capability. Below, the most commonly used evaluation metrics are discussed along with their formulas.

Accuracy measures the proportion of correctly classified instances among all instances as shown in Equation 7.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (7)$$

where TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) denote correctly classified positive instances, correctly classified negative instances,

incorrectly classified negative instances as positive, and incorrectly classified positive instances as negative, respectively.

Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive as shown in Equation 8.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Recall measures the proportion of correctly predicted positive instances out of all actual positive instances as shown in Equation 9.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

Specificity measures the proportion of correctly predicted negative instances out of all actual negative instances as shown in Equation 10.

$$Specificity = \frac{TN}{FP + TN} \tag{10}$$

The F1-score is the harmonic mean of precision and recall, providing a balance between them as shown in Equation 11.

$$F1 - Score = 2\left(\frac{Precision \times Recall}{Precision + Recall}\right) \tag{11}$$

The ROC curve plots the True Positive Rate (Recall) against the False Positive Rate (1-Specificity) at various threshold settings. The AUC measures the area under this curve. The AUC provides an aggregated measure of performance across all classification thresholds as shown in Equation 12.

$$AUC = \int_0^1 TPR(x)dx \tag{12}$$

K-fold cross-validation is a resampling procedure to evaluate model performance by splitting the dataset into k subsets (folds). The model is trained on k−1 folds and tested on the remaining fold. This process is repeated k times, with each fold used as a test set exactly once.

## 4 Experimental studies

This section outlines the prediction accuracies, recall, and f-score values attained through the implemented methods. The SISA dataset was used as input for the implementation of various machine learning methods. In the first part of the experiments, the dataset is divided into two parts randomly. Those parts are called training set and test set. Training set is composed of 80% of the whole dataset. In contrast, the remaining 20% of the dataset becomes test set.

All model training and evaluation procedures were performed on a standard laptop computer running Windows 11, equipped with an Intel Core i7-1165G7 CPU @ 2.80 GHz and 16 GB of RAM. No GPU acceleration was used during experimentation. Training time for the Random Forest model with 25 estimators was approximately 3.2 seconds using 10-fold cross-validation. XGBoost, CatBoost, and LightGBM models were also trained under 5 seconds each. These results demonstrate that the proposed approach is computationally lightweight and can be executed efficiently on modest hardware, making it suitable for real-world applications in clinical environments with limited computational resources.

At this stage, 22 different machine learning methods have been applied in order to classify arbovirus as hospitalized or outpatient. These methods are support vector machine (SVM), decision trees (DT), random forest (RF), logistic regression (LR), Naïve Bayes (NB), Ridge Classifier, Stochastic Gradient Descent (SGD), Restricted Boltzmann Machine (RBM), Multi-layer Perceptron (MLP), Gaussian Mixture, Logistic Regression (LR), Perceptron, Passive Aggressive Classifier (PA), Linear Support Vector Machine (LinearSVC), K-Nearest Neighborhood (KNN), Nearest Centroid, Gaussian Process, Gaussian Naïve Bayes, Bernoulli Naïve Bayes, Decision Trees (DT), Gradient Boosting (GB), AdaBoost, XGBoost, CatBoost, and LightGBM classifiers. RF obtained 0.9908 accuracy values. Table 3 reveals the classification performance results with accuracy, precision, recall, and F-score values. Figure 2 visualizes this comparison with a bar chart.

Table 3. Classification performance results of the mentioned machine learning algorithms applied to SISA dataset divided into two parts: 80% training and 20% testing sets.

| | Accuracy | F1 Score | Precision | Recall |
|---|---|---|---|---|
| RF | 0.9908 | 0.9677 | 1.000 | 0.9333 |
| RBM | 0.8624 | 0.0000 | 1.000 | 0.000 |
| MLP | 0.8716 | 0.1250 | 1.000 | 0.0667 |
| Gaussian Mixture | 0.8807 | 0.6977 | 0.5357 | 1.000 |
| LR | 0.9358 | 0.7742 | 0.7500 | 0.800 |
| Ridge Classifier | 0.9817 | 0.9275 | 0.8824 | 1.000 |
| SGD | 0.9083 | 0.5455 | 0.8571 | 0.40 |
| Perceptron | 0.8807 | 0.6977 | 0.5357 | 1.000 |
| PA | 0.8889 | 0.6250 | 0.5882 | 0.6667 |
| SVM | 0.8624 | 0.000 | 1.000 | 0.000 |
| LinearSVM | 0.9882 | 0.9655 | 0.9375 | 1.000 |
| KNN | 0.8899 | 0.6250 | 0.5882 | 0.6667 |
| Nearest Centroid | 0.8257 | 0.4571 | 0.4000 | 0.5333 |
| Gaussian Process | 0.8624 | 0.000 | 1.000 | 0.000 |
| Gaussian NB | 0.9174 | 0.7692 | 0.6250 | 1.000 |
| Bernoulli NB | 0.9174 | 0.7692 | 0.6250 | 1.000 |
| DT | 0.9817 | 0.9333 | 0.9333 | 0.9333 |
| Gradient Boosting | 0.9358 | 0.6957 | 1.000 | 0.5333 |
| AdaBoost | 0.9577 | 0.9442 | 1.000 | 1.000 |
| XGBoost | 0.9713 | 0.9267 | 0.9329 | 0.9240 |
| CatBoost | 0.9734 | 0.9269 | 0.9197 | 0.9381 |
| LightGBM | 0.9771 | 0.9411 | 0.9483 | 0.9381 |

As can be seen in Table 3, RF has the highest accuracy and F-Score values while RBM, MLP, SVM, Gaussian Process, GB, and AdaBoost obtain the best precision values. Nearest Centroid acquires the far worse performance values compared to other methods.

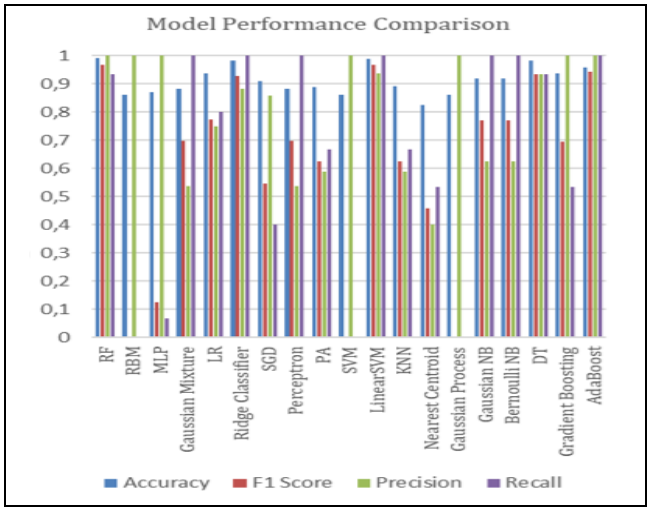


Figure 2. Bar chart comparing tested methods applied to SISA dataset divided into two parts: 80% training and 20% testing.

In the second part of the experiment, all of these mentioned five models have been implemented, but the dataset is split into two parts using the 10-folds cross validation technique. It means that the testing part is selected randomly as 10% of the dataset. In each of ten iterations, a different 10% of the data becomes testing set and the remaining instances become training set. Classification is performed in each iteration and finally, the average results are recorded as classification performance values of the applied classifier. As a result, Table 4 shows the outcomes of the same classifiers for 10-folds cross validation process. It can be observed that there is no significant difference in the results compared to Table 3.

As shown in Table 3 and Table 4, the most successfull classifiers appear to be Random Forest, Ridge, and Linear SVM. Their prediction about the status of patients in test sets when 10-folds cross validation process was implemented are presented in the confusion matrices given in Figure 3, Figure 4, and Figure 5.

Standart deviation values are also revealed in Table 5 for each evaluation metric result when 10-folds cross validation is performed. Overall, models such as Random Forest, Ridge Classifier, Linear-SVM, Decision Tree, and AdaBoost exhibited very low variability across folds, indicating high consistency and robustness. In contrast, models like RBM, SVM, and Gaussian Process showed either extreme values or zero variance in certain metrics, suggesting unreliable or skewed performance, often due to predicting only one class. These findings highlight the importance of including stability metrics, in addition to average scores, when evaluating machine learning models for clinical applications.

Table 4. Classification performance results of machine the mentioned learning algorithms applied to apple plant leaf dataset divided into two parts using 10-folds cross validation.

| Model | Accuracy | F1 Score | Precision | Recall |
|---|---|---|---|---|
| RF | 0.9908 | 0.9510 | 0.9633 | 0.9437 |
| RBM | 0.8914 | 0.000 | 1.000 | 0.000 |
| MLP | 0.9098 | 0.5476 | 0.7196 | 0.6443 |
| Gaussian Mixture | 0.8079 | 0.5833 | 0.4408 | 0.9000 |
| LR | 0.9337 | 0.7009 | 0.6974 | 0.7488 |
| Ridge Classifier | 0.9741 | 0.8806 | 0.8652 | 0.9238 |
| SGD | 0.8842 | 0.3407 | 0.6507 | 0.4476 |
| Perceptron | 0.7376 | 0.3594 | 0.3704 | 0.60 |
| PA | 0.8931 | 0.4201 | 0.6715 | 0.5444 |
| SVM | 0.8914 | 0.000 | 1.000 | 0.000 |
| LinearSVM | 0.9797 | 0.9169 | 0.9198 | 0.9238 |
| KNN | 0.9153 | 0.6206 | 0.6330 | 0.6726 |
| Nearest Centroid | 0.8656 | 0.5029 | 0.4189 | 0.6480 |
| Gaussian Process | 0.8914 | 0.000 | 1.000 | 0.000 |
| Gaussian NB | 0.9115 | 0.7029 | 0.5480 | 1.000 |
| Bernoulli NB | 0.9189 | 0.7189 | 0.5712 | 1.0000 |
| DT | 0.9742 | 0.8980 | 0.9149 | 0.8885 |
| Gradient Boosting | 0.9577 | 0.7622 | 0.9417 | 0.7095 |
| AdaBoost | 0.9889 | 0.9442 | 1.000 | 0.9571 |
| XGBoost | 0.9752 | 0.9187 | 0.9300 | 0.9214 |
| CatBoost | 0.9752 | 0.9367 | 0.9348 | 0.9437 |
| LightGBM | 0.9726 | 0.9644 | 0.9657 | 0.9690 |

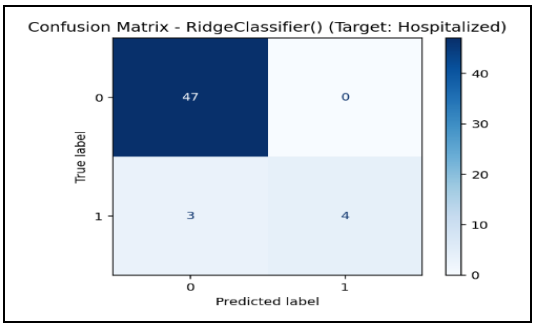


Figure 3. Confusion matrix of ridge classifier.

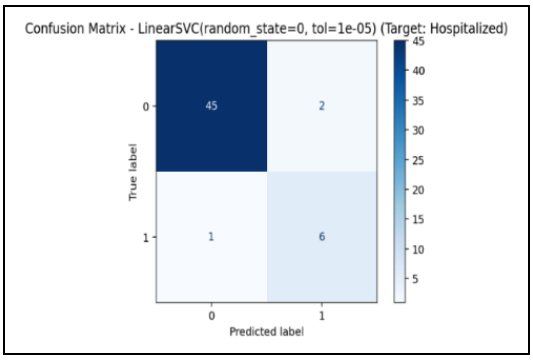


Figure 4. Confusion matrix of Linear-SVM classifier.

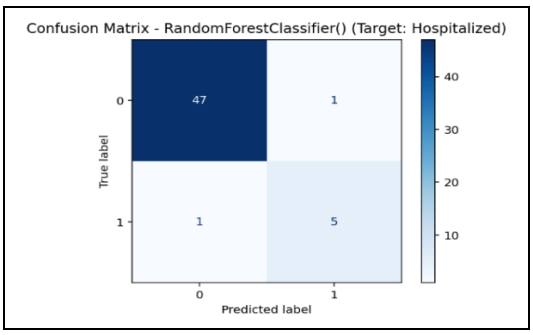


Figure 5. Confusion matrix of RF classifier.

Table 5. Standart deviation values of each evaluation for 10-folds cross validation.

| Model | St. Dev Accuracy | St. Dev F1 Score | St. Dev Precision | St. Dev Recall |
|---|---|---|---|---|
| RF | 0.0051 | 0.0124 | 0.0000 | 0.0205 |
| RBM | 0.0143 | 0.0000 | 0.0000 | 0.0000 |
| MLP | 0.0128 | 0.0365 | 0.0000 | 0.0111 |
| Gaussian Mixture | 0.0134 | 0.0221 | 0.0194 | 0.0000 |
| LR | 0.0083 | 0.0195 | 0.0222 | 0.0200 |
| Ridge Classifier | 0.0059 | 0.0131 | 0.0183 | 0.0000 |
| SGD | 0.0112 | 0.0342 | 0.0211 | 0.0234 |
| Perceptron | 0.0140 | 0.0209 | 0.0178 | 0.0000 |
| PA | 0.0123 | 0.0251 | 0.0203 | 0.0186 |
| SVM | 0.0131 | 0.0000 | 1.0000 | 0.0000 |
| LinearSVM | 0.0044 | 0.0118 | 0.0152 | 0.0000 |
| KNN | 0.0108 | 0.0217 | 0.0193 | 0.0165 |
| Nearest Centroid | 0.0149 | 0.0301 | 0.0242 | 0.0228 |
| Gaussian Process | 0.0131 | 0.0000 | 1.0000 | 0.0000 |
| Gaussian NB | 0.0092 | 0.0184 | 0.0175 | 0.0000 |
| Bernoulli NB | 0.0089 | 0.0179 | 0.0180 | 0.0000 |
| DT | 0.0058 | 0.0127 | 0.0127 | 0.0127 |
| Gradient Boosting | 0.0084 | 0.0231 | 0.0000 | 0.0182 |
| AdaBoost | 0.0067 | 0.0110 | 0.0000 | 0.0000 |
| XGBoost | 0.0055 | 0.0117 | 0.0130 | 0.0098 |
| CatBoost | 0.0062 | 0.0123 | 0.0142 | 0.0105 |
| LightGBM | 0.0051 | 0.0108 | 0.0125 | 0.0091 |

To further evaluate the classification performance of the selected models, ROC curves were generated for Ridge Classifier, Linear-SVM, and RF. The corresponding ROC curves, are presented in Figure 6, Figure 7, and Figure 8, respectively. A higher AUC indicates a stronger discriminatory capability of the model.

The initial experiments revealed that the Random Forest (RF) algorithm outperformed other models in terms of classification performance, achieving the highest accuracy and robustness across multiple evaluation metrics. Given its superior performance, additional experiments were conducted to optimize the RF model through hyperparameter tuning.

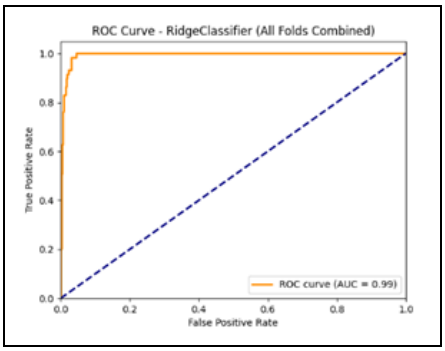


Figure 6. ROC curve of Ridge classifier.

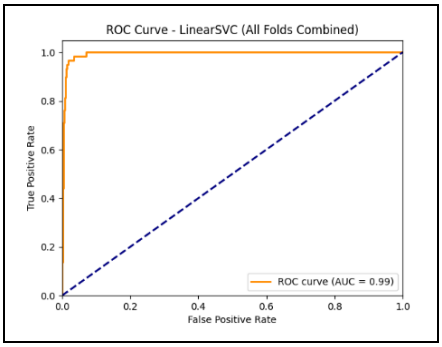


Figure 7. ROC curve of Linear-SVM classifier.

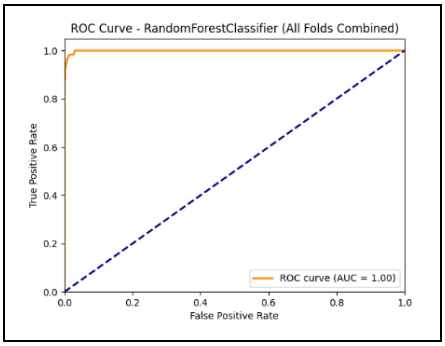


Figure 8. ROC curve of RF classifier.

One of the most critical hyperparameters in RF is the number of estimators, which determines the number of decision trees in the ensemble. To investigate its impact on classification performance, we systematically varied the number of estimators from 1 to 100. For each configuration, the model was trained using k-fold cross-validation (k=10) to ensure generalizability and mitigate overfitting. The results were analyzed to determine the optimal number of estimators that maximized predictive performance while maintaining computational efficiency. The findings from this parameter tuning experiment provide valuable insights into the optimal configuration of RF for medical triage applications, reinforcing the role of ensemble learning in healthcare decision-making.

We evaluated RF performance across varying numbers of estimators ranging from 1 to 100 to better understand the influence of hyperparameters in this model. The results revealed a clear trend where increasing the number of trees improved accuracy, particularly up to 25 estimators, after which the performance plateaued. This analysis was conducted separately for male and female patients, and the results are summarized in Table 6.

Table 6. Classification accuracy values for different number of estimators in RF for each gender.

| Number of Estimators | Accuracy for Male Patients | Accuracy for Female Patients |
|---|---|---|
| 1 | 0.9609 | 0.9584 |
| 2 | 0.9445 | 0.942 |
| 3 | 0.9718 | 0.9693 |
| 5 | 0.9737 | 0.9712 |
| 10 | 0.9866 | 0.9841 |
| 15 | 0.9903 | 0.9878 |
| 20 | 0.9847 | 0.9822 |
| **25** | **0.9958** | **0.9933** |
| 30 | 0.9921 | 0.9896 |
| 40 | 0.9903 | 0.9878 |
| 50 | 0.9865 | 0.984 |
| 60 | 0.994 | 0.9915 |
| 70 | 0.9884 | 0.9859 |
| 80 | 0.9921 | 0.9896 |
| 90 | 0.994 | 0.9915 |
| 100 | 0.994 | 0.9915 |

The data presented in Figure 9 illustrates the classification accuracy of RF model as a function of the number of estimators. As the number of estimators increases from 1 to 100, we observe a general improvement in accuracy.

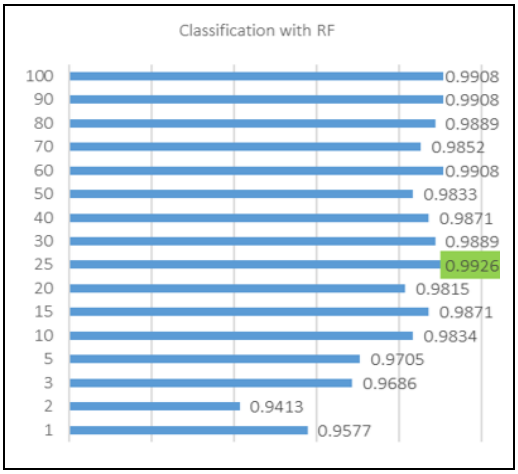


Figure 9. Classification accuracy values for different number of estimators in RF applied to SISA dataset with 10-folds cross validation.

The model achieves its highest classification accuracy of 0.9926 at 25 estimators, which is a notable peak compared to the other values. However, after reaching this maximum at 25 estimators, the accuracy stabilizes, with values ranging from 0.9833 to 0.9908 for estimator counts of 50 or higher. This suggests that after a certain threshold (around 25 estimators), adding more estimators does not substantially improve the model's performance, though there is slight variation.

The initial increase in accuracy with the addition of more estimators is consistent with the typical behavior of Random Forests, where more estimators tend to reduce variance and improve model stability. However, the plateau observed beyond 25 estimators indicates that the model might be reaching a point of diminishing returns. This finding suggests that while increasing the number of estimators can improve the performance, an optimal number exists, beyond which additional estimators may provide minimal additional benefits, while also increasing computational cost. A Random Forest model with 25 estimators seems to offer the best balance between classification accuracy and computational efficiency in this dataset.

Additionally, gender-based experiments were also implemented in these experimental studies. Among the 543 total records, 299 (55.1%) correspond to female patients and 244 (44.9%) to male patients. This slight predominance of female subjects has been taken into account during subgroup analyses to ensure balanced evaluation across genders. The gender distribution is illustrated in Figure 10.

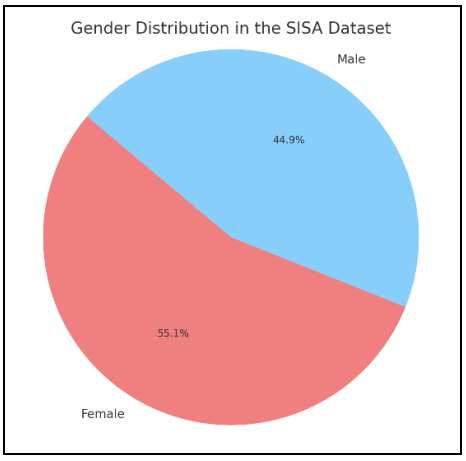


Figure 10. Gender distribution.

We evaluated the effect of varying the number of estimators in the RF classifier separately for male and female patients. As shown in Table 6, both subgroups demonstrated a similar upward trend in accuracy as the number of estimators increased. While initial accuracy levels were relatively close (e.g., 0.9609 for males and 0.9584 for females at 1 estimator), the performance steadily improved with more trees. Peak performance was observed around 25 estimators, with accuracies of 0.9958 for males and 0.9933 for females, after which marginal fluctuations occurred. These results suggest that the model generalizes well across genders and benefits equally from ensemble depth, with only minimal differences in predictive power between male and female subgroups.

Selection of different number of estimators affects highly the classification performance as can be seen in the experiments, but the implementation of parameter tuning of RF method for another dataset is required in order to generalize these effects. Therefore, same approach was applied to SISAL dataset which is related to hospitalization status of patients from the same source of data. Figure 11 demonstrates the classification accuracies of RF with different number of estimators for SISAL dataset.

SISAL dataset contains similar features to SISA but incorporates laboratory test results. We evaluated the Random Forest classifier with varying numbers of estimators (from 1 to 100) and recorded the classification accuracy for each configuration. As shown in Figure 9, model performance improved notably with an increasing number of estimators, reaching a peak accuracy of 0.9700 at 40 estimators. Beyond this point, the accuracy plateaued or slightly declined, indicating diminishing returns. These findings demonstrate that the ensemble method maintains robust performance on a related but distinct dataset, thereby supporting the model's generalizability across different data contexts.

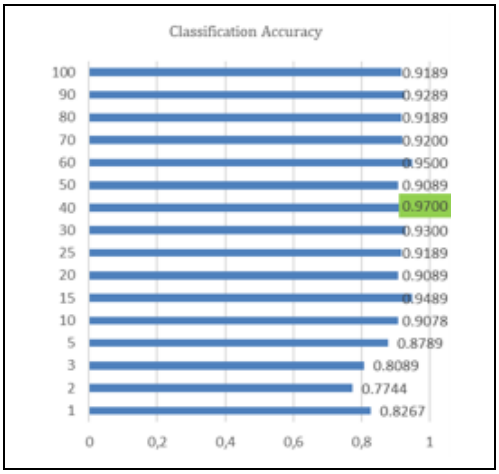


Figure 11. Classification accuracy values for different number of estimators in RF applied to SISAL dataset with 10-folds cross validation.

## 5 Discussion

This study highlights the potential of machine learning (ML) approaches to enhance clinical decision-making during the triage process, particularly for patients with suspected arboviral infections. By employing the SISA dataset, our research underscores the importance of advanced computational techniques in addressing the resource and decision-making challenges faced by emergency departments (EDs).

The results of the experimental studies demonstrate the effectiveness of ML models, particularly Random Forest (RF), in accurately predicting hospitalization status with high precision, recall, and F1-score values. The RF model consistently outperformed other algorithms, achieving robust classification metrics, which is consistent with findings in prior studies [22]. This performance is attributed to the ensemble nature of RF, which combines multiple decision trees to reduce overfitting and enhance generalizability.

This study confirms the suitability of ML methods such as RF, Gradient Boosting, and neural networks for handling complex, non-linear relationships in healthcare data. Notably, the strong predictive power of these models, with accuracy rates exceeding 90%, aligns with previous research utilizing the same dataset, such as Sippy et al.'s GBM study and Gorur et al.'s FFNN results [3, 22]. These findings suggest that integrating ML models into triage workflows can significantly optimize resource allocation, improve patient outcomes, and reduce ED overcrowding.

Several recent studies have explored the SISA dataset using deep learning or ensemble-based methods. For instance, Sippy et al. [2] employed neural networks alongside other algorithms and reported test accuracies ranging from 0.8980 to 0.9620, with the best-performing model (GBM) achieving an AUC of

0.91. Similarly, Gorur et al. [1] applied feedforward neural networks and reported up to 0.9873 accuracy and an AUC of 0.973. Huang et al. [21] also reported improved prognosis in dengue prediction using an ANN-based model. While these studies confirm the utility of deep models, our findings demonstrate that a properly tuned RF model with 25 estimators can outperform or match these results, achieving a peak accuracy of 0.9926 on the same dataset. This suggests that traditional ensemble models, when carefully optimized, can offer competitive performance with lower computational complexity and greater interpretability. Such models are particularly advantageous in clinical settings, where transparency and efficiency are often critical considerations.

Although the classification models achieved notably high accuracy values, up to 99% in some cases, such results must be interpreted with caution. One possible explanation is that the SISA dataset contains features that strongly differentiate between hospitalized and non-hospitalized cases, making certain patterns easier to learn. However, we also recognize the potential risk of overfitting, especially given the class imbalance, where only approximately 11% of instances belong to the positive class. To mitigate this, we applied stratified 10-fold cross-validation and emphasized performance metrics such as F1-score, precision, and recall rather than relying solely on overall accuracy. Furthermore, we report standard deviations across folds to assess stability, and we acknowledge that external validation with independent datasets will be necessary to fully evaluate generalizability.

The integration of machine learning-based arboviral case prediction models into clinical workflows offers significant potential for improving patient triage and resource allocation, particularly in outbreak settings. Given its strong performance and interpretability, the Random Forest model used in this study could be embedded within a clinical decision support system (CDSS) to assist frontline healthcare workers in identifying patients at higher risk of hospitalization. Such a system could support early intervention decisions even in resource-limited environments where laboratory testing is not immediately available. For real-world implementation, it is essential that the system be designed with user-friendly interfaces, require minimal training, and provide timely outputs that align with clinical workflows. These practical considerations are vital for ensuring adoption and effectiveness in public health settings.

## 6 Conclusion

In the United States alone, the cost of managing dengue illness is estimated to be around $9 billion. Arboviral infections impose significant financial and healthcare burdens on countries, and numerous studies have been conducted on their diagnosis. However, research using machine learning to predict the severity or hospitalization status of arboviral infections remains limited. Machine learning algorithms can assist clinicians in resource-constrained settings by providing accurate classification models to predict whether patients require hospitalization or outpatient care.

This study compared various data mining algorithms using the SISA dataset, which includes symptom and demographic information. The models achieved notable results, including an accuracy score of 0.9926 with the RF algorithm, reaching 0.9677 F1-Score in 10-folds cross validation. Additionally, other ensemble methods, such as Adaboost and Gradient Boosting, achieved accuracy scores ranging from 0.95 to 0.99. DT also performed well, with accuracy scores of 0.9817 to 0.9742 among rule-based methods. These findings could contribute to the development of more reliable and gender-sensitive arbovirus control programs.

The proposed models demonstrated very high accuracy on the SISA dataset, but these results should be interpreted with caution. Dataset-specific patterns and class imbalance may have contributed to the performance metrics. While validation on the SISAL dataset provided some support for generalizability, external validation on independent and more diverse patient populations will be essential to establish the robustness and clinical utility of the approach. Future work should therefore prioritize multi-center validation studies to ensure the model's reliability in real-world settings.

Even though our study primarily focused on tuning the number of estimators for RF model, more advanced hyperparameter optimization techniques such as Grid Search, Random Search, or Bayesian Optimization could be employed to further enhance model performance. Incorporating such methods represents a promising avenue for future research, especially in clinical prediction settings where subtle parameter adjustments may yield significant gains in reliability and generalizability.

A gender-based analysis was conducted using the RF model with varying numbers of estimators to assess potential variability in model performance across patient subgroups. Results indicated consistent accuracy across male and female patients, suggesting that the model performs robustly regardless of gender. However, subgroup-specific patterns such as age-related or comorbidity-driven variations could influence model outcomes in broader clinical settings. Also, we observed that class imbalance may contribute to a higher risk of false positives or false negatives. To mitigate such biases, future work may incorporate techniques such as class weighting, data augmentation (e.g., SMOTE), and calibration-based ensembling strategies to enhance generalizability and fairness.

Deploying ML models in real-world healthcare environments poses several practical challenges. First, integration into existing hospital information systems requires compatibility with current EHR infrastructures. In addition, clinical decision support tools must align with fast-paced triage workflows, especially in emergency departments where time-sensitive decisions are critical. Data quality and heterogeneity also present barriers, particularly in low-resource settings where incomplete or inconsistent entries may affect model reliability. Moreover, for such systems to be adopted by healthcare professionals, they must provide transparent outputs, require minimal training, and comply with ethical and regulatory standards concerning data privacy and patient safety.

The modeling framework can be generalized to other infectious or acute illnesses with similar clinical presentations. For instance, diseases such as influenza, leptospirosis, or bacterial sepsis often present with overlapping symptoms and could benefit from early hospitalization prediction models based on demographic and symptomatic features. Furthermore, adapting the model to different patient groups such as pediatric or geriatric populations could improve personalized care in diverse healthcare settings. Future studies should explore these extensions to assess the adaptability and robustness of ensemble-based classification models across varying clinical contexts.

## 7 Author contribution statement

Alican Doğan, confirms the sole responsibility for the conception of the study, presented results and manuscript preparation. In the conducted study, Alican Doğan contributed to the formation of the idea, design process, and literature review, the evaluation of the obtained results, procurement of materials used, and analysis of the results, and the proofreading and content review of the article.

## 8 Ethics committee approval and conflict of interest statement

"The authors declare no conflicts of interest. Ethical committee approval was not required for this article". "The authors declare that there is no need to obtain permission from the ethics committee for the review paper prepared".

## References

[1] Görür K, Çetin O, Özer Z, Temurtaş F. "Hospitalization status and gender recognition over the arboviral medical records using shallow and RNN-based deep models". *Results in Engineering*, 18, 101109, 2023.

[2] Sippy R, Farrell DF, Lichtenstein DA, Nightingale R, Harris MA, Toth J, Hantztidiamantis P, Usher N, Cueva Aponte C, Barzallo Aguilar J, Puthumana A, Lupone CD, Endy T, Ryan SJ, Stewart Ibarra AM. "Severity Index for Suspected Arbovirus (SISA): Machine learning for accurate prediction of hospitalization in subjects suspected of arboviral infection". *PLoS Neglected Tropical Diseases*, 14(2), e0007969, 2020.

[3] Fite J, Baldet T, Ludwig A, Manguin S, Saegerman C, Simard F, Quenel P. "A one health approach for integrated vector management monitoring and evaluation". *One Health*, 20, 100954, 2025.

[4] Gutiérrez-López R, Ruiz-López MJ, Ledesma J, Magallanes S, Nieto C, Ruiz S, Sanchez-Pena C, Ameyugo U, Camacho J, Varona S, Cuesta I, Jado-Garcia I, Sanchez-Seco M, Figuerola J, Azquez A. "First isolation of the Sindbis virus in mosquitoes from southwestern Spain reveals a new recent introduction from Africa". *One Health*, 20, 100947, 2025.

[5] Nenaah GE, Mahfouz ME, Almadiy AA, Albogami BZ, Alasmari SM, Gazzy AA, Fadl AE. "Fabrication, characterization, and mosquitocidal activity of CuO nanoparticles against Aedes aegypti L. (Diptera: Culicidae), the main vector of dengue arbovirus, with environmental risk assessment". *BioNanoScience*, 15(1), 93, 2025.

[6] Panmei K, Syed AH, Okafor O, Mammen S, Abraham A. "Performance evaluation of TaqMan™ Arbovirus Triplex Kit (ZIKV/DENV/CHIKV) for detection and differentiation of dengue and chikungunya viral RNA in serum samples of symptomatic patients". *Journal of Virological Methods*, 332, 115072, 2025.

[7] Xiao P, Hao Y, Yuan Y, Ma W, Li Y, Zhang H, Li N. "Emerging West African genotype chikungunya virus in mosquito virome". *Virulence*, 16(1), 2444686, 2025.

[8] Saihar A, Yaseen AR, Suleman M, Parveen R, Bashir H. "From bytes to bites: In-silico creation of a novel multi-epitope vaccine against Murray Valley Encephalitis Virus". *Microbial Pathogenesis*, 198, 107171, 2025.

[9] Hefti E, Xie Y, Engelen K. "Machine learning model better identifies patients for pharmacist intervention to reduce hospitalization risk in a large outpatient population." *Journal of Medical Artificial Intelligence*, 8, 11, 2025.

[10] Rosman L, Lampert R, Wang K, Gehi AK, Dziura J, Salmoirago-Blotcher E, Brandt C, Sears SF, Burg M. "Machine learning-based prediction of death and hospitalization in patients with implantable cardioverter defibrillators". *JACC-Journal of the American College of Cardiology*, 85(1), 42-55, 2025.

[11] Chang CY, Chen CC, Tsai ML, Hsieh MJ, Chen TH, Chen SW, Chang SH, Chu PH, Hsieh IC, Wen MS, Chen DY. "Predicting mortality and hospitalization in heart failure with preserved ejection fraction by using machine learning". *JACC-Asia*, 4(12), 956-968, 2024.

[12] Larkin JW, Lama S, Chaudhuri S, Willetts J, Winter AC, Jiao Y, Stauss-Grabo M, Usvyat LA, Hymes JL, Maddux FW, Wheeler DC, Stenvinkel P, Floege J, Zimbelman J. "Prediction of gastrointestinal bleeding hospitalization risk in hemodialysis using machine learning". *BMC Nephrology*, 25(1), 366, 2024.

[13] Albrecht S, Broderick D, Dost K, Cheung I, Nghiem N, Wu M, Zhu J, Poonawala-Lohani N, Jamison S, Rasanathan D, Huang S, Trenholme A, Stanley A, Lawrence S, Marsh S, Castelino L, Paynter J, Turner N, McIntyre P, Riddle P, Grant C, Dobbie G, Wicker JS. "Forecasting severe respiratory disease hospitalizations using machine learning algorithms". *BMC Medical Informatics and Decision Making*, 24(1), 293, 2024.

[14] Fan S, Abulizi A, You Y, Huang C, Yimit Y, Li Q, Zou X, Nijiati M. "Predicting hospitalization costs for pulmonary tuberculosis patients based on machine learning". *BMC Infectious Diseases*, 24(1), 875, 2024.

[15] Peacock J, Stanelle EJ, Johnson LC, Hylek EM, Kanwar R, Lakkireddy DR, Mittal S, Passman RS, Russo AM, Soderlund D, Hills MT, Piccini JP. "Using atrial fibrillation burden trends and machine learning to predict near-term risk of cardiovascular hospitalization". *Circulation-Arrhythmia and Electrophysiology*, 17(11), e012991, 2024.

[16] Wu WT, Kor CT, Chou MC, Hsieh HM, Huang WC, Huang WL, Lin SY, Chen MR, Lin TH. "Prediction model of in-hospital cardiac arrest using machine learning in the early phase of hospitalization". *Kaohsiung Journal of Medical Sciences*, 40(11), 1029-1035, 2024.

[17] Abdul-Samad K, Ma S, Austin DE, Chong A, Wang CX, Wang X, Austin PC, Ross HJ, Wang B, Lee DS. "Comparison of machine learning and conventional statistical modeling for predicting readmission following acute heart failure hospitalization". *American Heart Journal*, 277, 93-103, 2024.

[18] Davoudi A, Chae S, Evans L, Sridharan S, Song J, Bowles KH, McDonald MV, Topaz M. "Fairness gaps in machine learning models for hospitalization and emergency department visit risk prediction in home healthcare patients with heart failure". *International Journal of Medical Informatics*, 191, 105534, 2024.

[19] Windle N, Alam A, Patel H, Street JM, Lathwood M, Farrington T, Maruthappu M. "Predicting hospitalization risk among home care residents in the United Kingdom: Development and validation of a machine learning-based predictive model". *Home Health Care Management and Practice*, 36(4), 253-261, 2024.

[20] Lee VJ, Chow A, Zheng X, Carrasco LR, Cook AR, Lye DC, Ng LC, Leo YS. "Simple clinical and laboratory predictors of chikungunya versus dengue infections in adults". *PLoS Neglected Tropical Diseases*, 6(9), e1786, 2012.

[21] Huang SW, Tsai HP, Hung SJ, Ko WC, Wang JR. "Assessing the risk of dengue severity using demographic information and laboratory test results with machine learning". *PLoS Neglected Tropical Diseases*, 14(12), e0008960, 2020.

[22] Görür K, Başçıl MS, Bozkurt MR, Temurtaş F. "Classification of thyroid data using decision trees, kNN, and SVM methods". *International Artificial Intelligence and Data Processing Symposium (IDAP'16),* Malatya, Turkey, 130-134, 17-18 September 2016.

[23] Fathima SA, Hundewale N. "Comparative analysis of machine learning techniques for classification of arbovirus." *IEEE-EMBS International Conference on Biomedical and Health Informatics,* Hong Kong, China, 376-379, 05-07 January 2012.

[24] Akhtar M, Kraemer MU, Gardner LM. "A dynamic neural network model for predicting risk of Zika in real time". *BMC Medicine*, 17(1), 171, 2019.

[25] Salim NAM, Wah YB, Reeves C, Smith M, Yaacob WFW, Mudin RN, Dapari RNFF, Haque U. "Prediction of dengue outbreak in Selangor, Malaysia using machine learning techniques". *Scientific Reports*, 11(1), 939, 2021.

[26] Neto LM, da Silva Neto SR, Endo PT. "A comparative analysis of converters of tabular data into image for the classification of Arboviruses using Convolutional Neural Networks". *PLoS ONE*, 18(12), e0295598, 2023.

[27] Neto SRDS, Oliveira TT, Teixeira IV, Oliveira SBAD, Sampaio VS, Lynn T, Endo PT. "Machine learning and deep learning techniques to support clinical diagnosis of arboviral diseases: A systematic review". *PLoS Neglected Tropical Diseases*, 16(1), e0010061, 2022.

[28] Ozer I, Cetin O, Gorur K, Temurtas F. "Improved machine learning performances with transfer learning to predicting need for hospitalization in arboviral infections against the small dataset". *Neural Computing and Applications*, 33(21), 14975-14989, 2021.

[29] Sciannameo V, Goffi A, Maffeis G, Gianfreda R, Jahier Pagliari D, Filippini T, Mancuso P, Giorgi-Rossi P, Alberto Dal Zovo L, Corbari A, Vinceti M, Berchialla P. "A deep learning approach for Spatio-Temporal forecasting of new cases and new hospital admissions of COVID-19 spread in Reggio Emilia, Northern Italy". *Journal of Biomedical Informatics*, 132, 104132, 2022.