

Machine Learning for Dengue Outbreak Prediction: A Performance Evaluation of Different Prominent Classifiers

Naiyar Iqbal and Mohammad Islam

Department of Computer Science and Information Technology

Maulana Azad National Urdu University, Hyderabad, Telangana, India

Email: naiyariqbal.rs@manuu.edu.in, islamcs1@gmail.com

Keywords: Dengue fever, machine learning, classification, ensemble classifier, clinical symptoms

Received: March 1, 2017

Dengue disease patients are increasing rapidly and actually dengue has recorded in every continent today according to the World Health Organization (WHO) record. By WHO report the number of dengue outbreak cases announced every year has expanded from 0.4 to 1.3 million during the period of 1996 to 2005 and then it has reached to 2.2 to 3.2 million during the year of 2010 to 2015 respectively. Consequently, it is fundamental to have a structure that can adequately perceive the pervasiveness of dengue outbreak in a large number of specimens momentarily. At this critical moment, the capability of seven prominent machine learning systems was assessed for the forecast of the dengue outbreak. These methods are evaluated by eight miscellaneous performance parameters. LogitBoost ensemble model is reported as the topmost classification accuracy of 92% with sensitivity and specificity of 90 and 94 % respectively.

Povzetek: Sedem algoritmov strojnega učenja je analiziranih na izbruhu mrzlice dengi in LogitBoost je dosegel najboljše rezultate.

1 Introduction

Dengue fever is the most well-known arboviral disease transmitted by female mosquitoes (*Aedes Aegypti*) in tropical and subtropical regions throughout the world [7]. Spanish word dengue is derived from dinga. Dengue fever also familiar as break-bone fever, break heart fever, and dandy fever. Dengue viral fever is originated by four concerned viruses known as DEN- (1 to 4). Now DEN-5 which is newly introduced in 2013 [13,3]. Dengue fever (DF), Dengue Hemorrhagic Fever (DHF), and Dengue Shock Syndrome (DSS) are the broad stages of dengue viral from normal to serious respectively [8,16].

According to WHO report the number of dengue outbreak cases announced every year has expanded from 0.4 to 1.3 million during the period of 1996 to 2005 and then it has reached to 2.2 to 3.2 million during the year of 2010 to 2015 respectively. Dengue outbreak is a champion among the most notable viral disease in human beings. Over 33% of the aggregate population of the world is under pitfall together with numerous urban communities of India. In due course, forecasting of dengue outbreak can protect the life of a human by alarming them to take appropriate treatment and care. Forecast of transmissible outbreaks like dengue disease is a challenging work and several prediction techniques are still in their early stages [10]. An Eco-bio-social framework for dengue vector breeding has been proposed by [2]. The researchers use six different Asian regions in their research work and as a gist, vector breeding and adult *Aedes aegypti* are determined by a complex interaction of the factor.

Souza et al, (2007) [19] shows the influence of dengue disease on liver activity. They found that liver damage is

more frequent in ladies. So, the liver test is more important that calculates the level of liver damage.

Machine learning is state of the art technology to embolden machines to perform without being explicitly customized to streamline performance standard use of case data or previous observations. Machine Learning model is used for the collection of precious information from the data by the normalized dataset. At this critical moment, the capability of many prominent machine learning systems was assessed for the forecast of the dengue outbreak. For the sake of this, seven machine learning algorithms have been used like LogitBoost, Logistic regression, Decision tree, Naive Bayes, Artificial neural network, Sequential minimal optimization, and k-nearest neighbor. Additionally, the ROC curve is also used for performance measurement. In table 4, we have shown the comparison among accuracy rate, sensitivity and specificity of the prominent classifier with two ensemble models i.e. Random forest [5] and LogitBoost.

2 Related Work

There are few other works concerned with the prediction of dengue outbreaks. Althouse et al. (2011) [1] applied three models, Linear regression (Step-down), Generalize Boosted regression and negative binomial Regression, as well as two another methods, logistic regression, and artificial neural network, are also applied for dengue disease prediction. They have performed their experiments for two regions Singapore and Bangkok. Authors found that the linear model is superior to other models; also support vector machine (SVM) performs

better than logistic regression in both regions. The selected linear model achieves a correlation of 0.86 and 0.93 between fitted and observed for Bangkok and Singapore region, respectively.

Brasier et al. (2012) [3] performed dengue disease prediction using CART and Random forest methods based on symptoms. They are performed 10 trails with 10-fold cross-validation that shows 84.0% (for DF) & 84.6% (for DHF) average accuracy result.

Support vector classification is used by Fathima et al, (2012) [6] for the prediction of arbovirus dengue. In their analysis, SVM gives 90.42% accuracy with 47.23% sensitivity and 97.59% specificity.

Fathima et al, (2015) [5] has done their experiment on dengue infection prognosis using random forest (one of the ensemble model) classifier on clinical parameters. As a result, they found 92% accuracy.

Ibrahim et al. (2005) [9] experiments dengue viral on 252 patients (4 DF & 248 DHF) using ANN with 9 input neurons and 5 hidden neurons on MATLAB simulator and their result showed 90% accuracy.

Rachata et al. (2010) [14] applied ANN using climate parameters like temperature, rainfall and relative humidity for dengue outbreak prediction. 85.92% accuracy is found in their experiment; also, they suggested using another feature selection method such as the hidden Markov model.

Decision Tree (C4.5) classifier is applied by Tanner et al. (2008) [21] on 1200 dengue samples (364 dengue positive & 836 dengue negative) consisting of five clinical parameters. Their experiments found 84.7% accuracy and 15.7% overall error rate and claims decision tree could be a useful classifier.

Additional review on related literature can be found in [10], which explores around thirty literature published between the year 1995 to 2013.

3 Methods & material

Data mining is an act of analyzing and extraction of substantial previous databases consider in mind that the end target is to the prediction of unknown information of a novel example from observed examples.

Data mining phases are as follow:

- Phase 1: Problem identification
- Phase 2: Formulation of the hypothesis
- Phase 3: Data collection
- Phase 4: Data Pre-process (scaling, encoding, and selecting features and outlier detection or removal)

- Phase 5: Model estimation
- Phase 6: Model interpret and draw conjecture

In this experiment, we use dengue disease dataset in CSV file format for the prediction on the WEKA data mining tool. This dataset consists of 75 samples with 36 samples without dengue disease (Negative) and 39 samples with dengue disease (Positive) [12,17,20]. The dataset is collected from test reports of different discharged patients. After that performs data pre-processing for smoothing some missing values using ReplaceMissingValues technique under filter option of WEKA tool. In this experiment, 8 distinct clinical attributes have been taken into account for the prediction of dengue diseases (Table 1).

Attribute Name	Data type	Range
1. Fever	Binary	No/Yes
2. Headache	Binary	No/Yes
3. Body ache	Binary	No/Yes
4. Abdominal pain	Binary	No/Yes
5. Vomiting	Binary	No/Yes
6. Haemoglobin	Numeric	12.0-17.5 (g/dL)
7. WBC	Numeric	4000-11000(/cumm)
8. Platelet	Numeric	1.5-4.5 (Lakh/mm ³)
Dengue	Binary	Negative/Positive

Table 1: Clinical attributes for dengue outbreak.

Dataset samples incorporated of the total of 75 samples with 8 clinical attributes for each sample. Samples with the absence of dengue outbreak were treated as a negative class, and samples with the presence of dengue outbreak were treated as positive samples for purpose of analysis. The correlation between

Eight attributes of negative and positive samples show the high correlation between the attributes of the two classes of samples as depicted in Figures 1 and 2. Figure 1 clearly shows that fever feature is positively correlated with all other parameters except a headache and platelet in samples without dengue outbreak (negative). Similarly, positively correlation between other parameters in negative class can be noticed in Figure 1.

Similarly, Figure 2 clearly demonstrates that the hemoglobin feature is positively correlated with all other parameters except a headache and platelet in samples with dengue outbreak (positive). Similarly, positively correlation between other parameters in positive class can be observed in Figure 2.

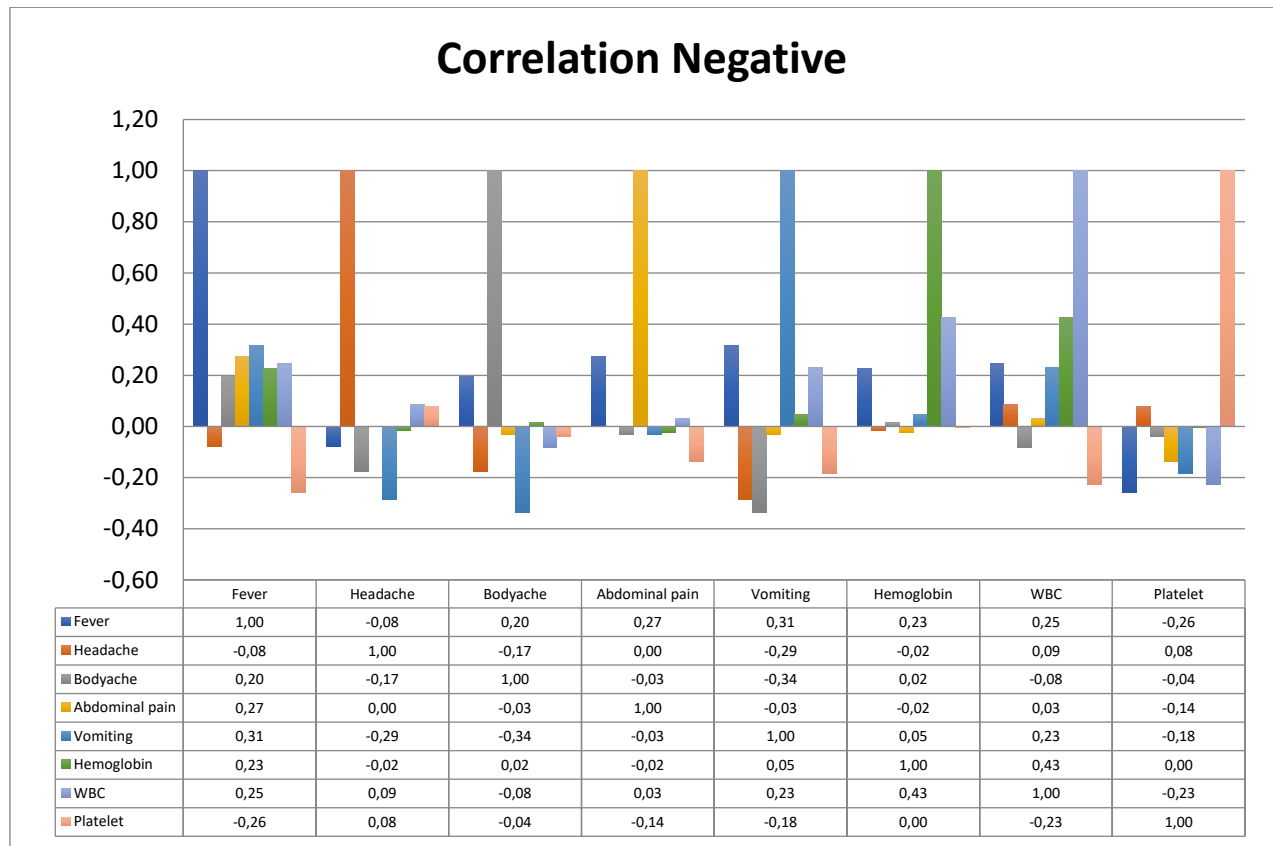


Figure 1: Linear correlation of negative cases.

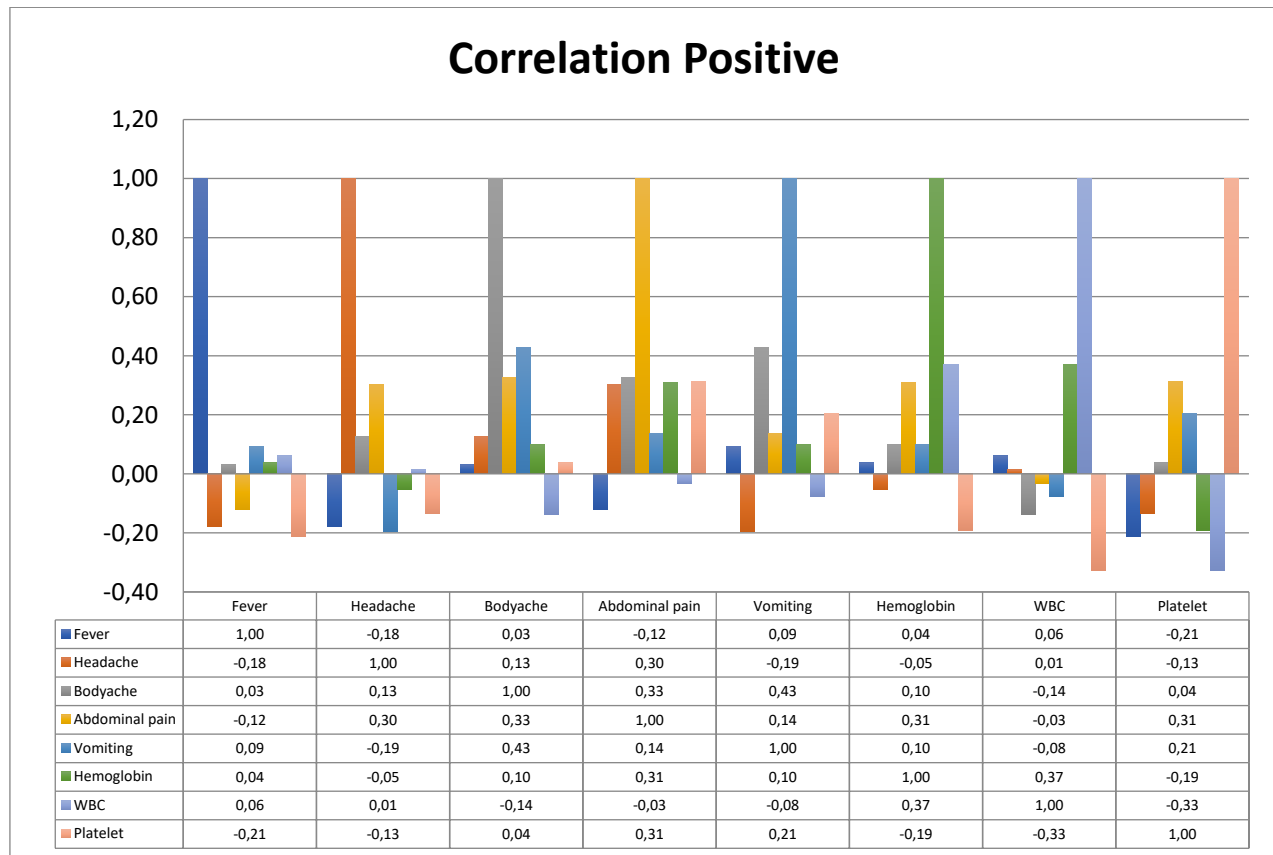


Figure 2: Linear correlation of positive cases.

4 Machine learning algorithms

4.1 K-nearest neighbour (kNN)

K-nearest Neighbour classifier is based on instance learning approach that is influenced by the lazy learning technique. Instance-based method, alternatively known as memory-based learning. In this approach, it matches novel problem instances with previously picked instances at training, which is stored in the memory. It is most fruitful for huge datasets with fewer features and provides global approximation and less time in training.

The k-NN method can be applied to both classification and regression. In both situations, the input composed of the k nearest training instances in feature space. The outcome is dependent on the application of k-NN is applied for classification or regression [10].

In k-NN classification, the result is a class belonging. The classification of entity is decided on the basis of a majority vote of their neighbor. In contrast k-NN regression, the outcome is the merit significance for the object. The significance is the means of the values of their kNN.

The k-NN model for continuous-valued objective functions that compute the average estimation of the k nearest neighbors. kNN is strong to noisy data by calculating the mean of k-nearest neighbors. The gap between neighbors can be overwhelmed by unnecessary features that lead to the curse of dimensionality. To defeat it, dimension stretch or elimination of the less significant features.

4.2 Support vector machine (SVM)

Support Vector Machine, also alternatively known as Support Vector Network introduced by Vladimir Vapnik, that is used for both classification and prediction. SVM is a machine learning method for binary classification problem, despite the fact that executions of multi-class SVMs exist to guide enter vectors to a multi-dimensional feature space. A straight decision environment is worked with exclusive competence guaranteeing high generalization capability of a machine learning strategy [6].

SVM depends on the statistical learning theory that there is an infinite line known as hyperplanes, isolating the two classes. SVM approach endeavoring to search the best one, that reduce the classification error on unknown data. SVM finds for the hyperplane with the biggest margin i.e. maximum marginal hyperplane (MMH).

The thought behind the SVM has been widely actualized in biology with some strategy for the limited situation where training data can be isolated error-free, additionally extending this outcome to non-separable training data. SVM is a deterministic approach that generates effective generalization properties. SVM has a strong mathematical function that uses kernel for complex learning.

Sequential minimal optimization (SMO) is a method for resolving quadratic programming issue which appears at training time of support vector machine [12,18].

A separating hyperplane can be calculated as:

$$H = W \cdot X + b = 0$$

Where, H hyperplane, W weight, X input vector, and b bias.

4.3 Artificial neural network (ANN)

The artificial neural network is powerful processing machine, that can be an algorithm or real hardware device that has the ability to recognize experience or contemplation knowledge represented through intermediary unit collectively features, and can make such learning knowledge available for usage.

The weighted sum of product $x_i w_{kj}$ (for $i=0$ to m) is usually denoted as net_k :

$$net_k = x_0 w_0 \sum_{i=1}^m x_i w_{kj}$$

Finally, an artificial neuron computes the output y_k as a certain function of net_k value:

$$y_k = f(net_k)$$

Where x and y are input and output signals respectively, w_{kj} synaptic weight, j synapse, and f is activation function [10].

4.4 Naive Bayes classifier

Bayesian learning is referred to as methods in probability and statistics. Bayes theorem illustrates the possibility of an event on the basis of conditions which may be respective to the event. It has a homological performance with chosen neural network classifiers and classification tree.

Every training sample can gradually increment or decrement the probability that a hypothesis is accurate means that previous knowledge could be associated accompanied by observed outcome. Naive Bayes is computability intractable and optimal decision making. Naive Bayes classifiers are applied for extraction of the appropriate grouping for a dataset wherever explicit elemental applications are conjoined [18].

The mathematical equation for Bayes theorem is stated as:

$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)}$$

Here X and Y represented as events, P(X) and P(Y) represents the ratios of X and Y without concern to each other. P(X|Y) is a conditional probability of observing occurrence X given that Y is correct. P(Y|X) is the ratio of observing occurrence Y specified that X is correct.

4.5 Decision tree

The decision tree is a hierarchical based prediction approach that sketches the observed attribute in the branches and the target value at their leaves. The predictions can be discrete values which is a classification decision tree or continuous values which is regression decision tree. The prominent algorithms have been developed e.g. ID3, C4.5, CART, CHAID and MARS for

decision tree prediction model. J48 decision tree [11] algorithm is a popular Java development under the C4.5 algorithm in WEKA tool that is applied as one of the experiments in this research.

Attribute selection measure by information gain is described as:

$$I(p, n) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

The entropy or requisite information required to the classification of objects in overall sub-trees is calculated as:

$$Entropy(A) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} I(p_i + n_i)$$

The encoded information that can be obtained by divaricating on A:

$$Gain(A) = I(p, n) - E(A)$$

Where A and I represent Attribute and Information gain respectively; p and n are an element of class P and N respectively.

4.6 Logistic regression classifier

Logistic regression is based on the regression technique in which the dependent variable is categorical. Logistic regression is a way to the prediction of a dichotomous result. Logistic regression can be binomial, ordinal and multinomial. In multinomial, the results can have more than two possible types.

Univariate logistic regression was applied for continuous covariates, whereas logistic regression techniques give odds proportion of interest, that is not easy to use as a diagnostic device because a computer would be required to compute dengue fever prediction. Consequently, we readjusted the two selected logistic regression technique that substituting continuous attributes with binary counterparts [4].

4.7 LogitBoost: an ensemble classifier

Various application of a data mining process demonstrated the legitimacy of mentioned No-Free-Lunch theorem [22]. According to No-Free-Lunch, a single learning model cannot be the best and most appropriate with the whole domain of application. Ensemble learning is an encouraging perspective strategy that combines weak learners to make a powerful model with a specific end goal to enhance the prediction model [15].

Ensemble model is a new way to the mixture of numerous prominent models for enhancement of the precision rate of a novel model for better prediction. It is a combination of k-learned models (M1, M2, M3...Mk) with the purpose of making an upgraded model M* [10], shown in figure 3.

In this research, LogitBoost algorithm has applied as an ensemble classifier for the prediction of dengue outbreak. LogitBoost follows the boosting approach as an ensemble. Boosting approach is most strong learning that is applied for both classification and regression analysis. Boosting approach first builds a weak classifier and test inputs are given starting weights and more often it begins with identical weighting. During iteration, the test inputs

are assigned with new weight value to center the systems that are not accurately classified with a newly learned classifier. At each progression of learning, increment weights of the input instance that are not accurately trained by the weak learner and reduction of weights of the input instance that are accurately trained by the weak learner. The ultimate classification model is built on a weighted vote of weak classifiers produced in the repetition.

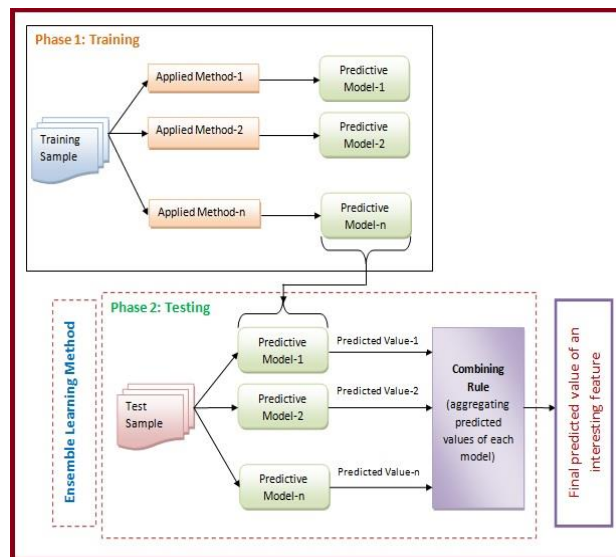


Figure 3: Ensemble model architecture [10].

In this comparative analysis, we found that LogitBoost performs better than another specific prominent classifier. LogitBoost ensemble model is reported as the topmost classification accuracy of 92% with sensitivity and specificity of 90 and 94 % respectively.

5 Classification performance metrics

In this research, seven supervised machine learning approaches were applied for the classification of dengue disease samples. Performance of the classification techniques was estimated on tenfold cross-validation. Eight quality parameters were taken into account for the assessment of classification models. Samples with the absence of dengue outbreak were treated as a negative class, and samples with the presence of dengue outbreak were treated as a positive class. Basic terminologies of confusion matrix as described here:

- **True Positive (TP)**- a number of records predicted as positive and it does have dengue outbreak.
- **True Negative (TN)**- a number of records predicted as negative and it doesn't have dengue outbreak.
- **False Positive (FP)**- a number of records predicted as positive but actually it doesn't have dengue outbreak. FP is also known as the *Type I Error*.
- **False Negative (FN)**- a number of records predicted as negative but actually it does have dengue outbreak. FN is also known as the *Type II Error*.

The quality measures on confusion matrix for binary classification are listed below as:

❖ **Classification Accuracy:**

The overall proportion of appropriately predicted samples to the total number of samples by the classifier model.

$$CA = (TP + TN)/(total\ sample)$$

❖ **True Positive Rate:**

The proportion of predicted positive sample to the total actually positive samples.

- Also known as *Sensitivity* or *Recall*
 $TPR = TP/(TP + FN)$

❖ **False Positive Rate:**

The proportion of predicted positive sample to the total actually negative samples.

$$FPR = FP/(FP + TN)$$

❖ **True Negative Rate:**

The proportion of predicted negative sample to the total actually negative samples.

- Also known as *Specificity*
 $TNR = TN/(TN + FP)$

❖ **Positive Predicted Value:**

The proportion of predicted positive sample to the total predicted positive samples.

- Also known as *Precision*
 $PPV = TP/(TP + FP)$

❖ **Negative Predictive Value:**

The proportion of predicted negative sample to the total predicted negative samples.

$$NPV = \frac{TN}{TN + FN}$$

Rate of Misclassification:

The proportion of overall incorrectly samples to the total number of samples. It can be also defined as the proportion of gross error (Type I Error and Type II Error) to the total number of samples

- $RMC=1-CA$
- Also known as "**Error Rate**"

$$RMC = \frac{Type\ I\ Error + Type\ II\ Error}{total\ sample}$$

❖ **F1 Score:** It is a weighted average of the recall and precision.

$$F1 = \frac{2TP}{2TP + FP + FN}$$

6 Results and discussion

The performance measurement of dengue outbreak prediction by seven machine learning algorithms is evaluated based on eight attributes as mentioned in the methods and materials section.

There was a total of 75 samples taken into account with 36 negative cases and 39 positive cases of dengue outbreak. Dengue dataset samples were divided in tenfold, each fold was used in testing and rest folds were applied as training throughout cross-validation.

Confusion matrix of prediction result is tabulated in Table 2 for LogitBoost, and other classifications like, Logistic regression, Decision tree, Naive Bayes, Artificial neural network, Sequential minimal optimization, and k-nearest neighbor are shown in figure 4.

Figure 4 depicts the predictions of these machine learning models. It is declared from the results that LogitBoost predicts the topmost number of true positives (number of records predicted as positive and it does have dengue outbreak) and it also predicts the topmost number of true negatives (number of records predicted as negative and it doesn't have dengue outbreak (Table 2; Figure 4).

Decision tree confusion matrix shows that it has the second highest true positives and Logistic regression predicts the second-highest true negatives (Figure 4).

Logistic regression confusion matrix shows that it has the third highest true positives and SMO confusion matrix predicts third highest true negatives (Figure 4).

Naive Bayes and ANN confusion matrix depicts that both are the fourth highest true positives and true negatives (figure 4).

SMO confusion matrix indicates that it has the fifth highest true positives and Decision tree predicts the fifth highest true negatives (Figure 4).

k-NN confusion matrix shows the worst performer in the sense of the lowest true positives and true negatives (Figure 4).

LogitBoost:

		Predicted Class		Total Actual
		Negative	Positive	
Actual Class	Negative	34 (89.47%)	2 (5.40%)	36
	Positive	4 (10.53%)	35 (94.59%)	39
Total Predicted		38	37	75

Table 2: Confusion matrix for LogitBoost algorithm.

Table 3 explains various classification chronicle measurements especially classification accuracy, specificity, sensitivity, precision, False Positive Rate, Negative predictive value, the rate of misclassification and F1 score.

Table 3 declared that LogitBoost outperformed over all other machine learning methods with the topmost classification accuracy of 92% while the second highest classification accuracy is achieved by Logistic regression of 85%. In addition, LogitBoost has found the highest sensitivity of 90% and Decision tree has got the second highest sensitivity of 87%. Logitboost also acquires topmost specificity of 94% and precision of 95% which declared that LogitBoost ensemble model is most appropriate for the prediction of patients with dengue outbreak (positive class).

Table 3 also shows other parameters like False Positive Rate, Negative predictive value, the rate of misclassification and F1 score of these machine learning methods. The table undoubtedly shows that LogitBoost

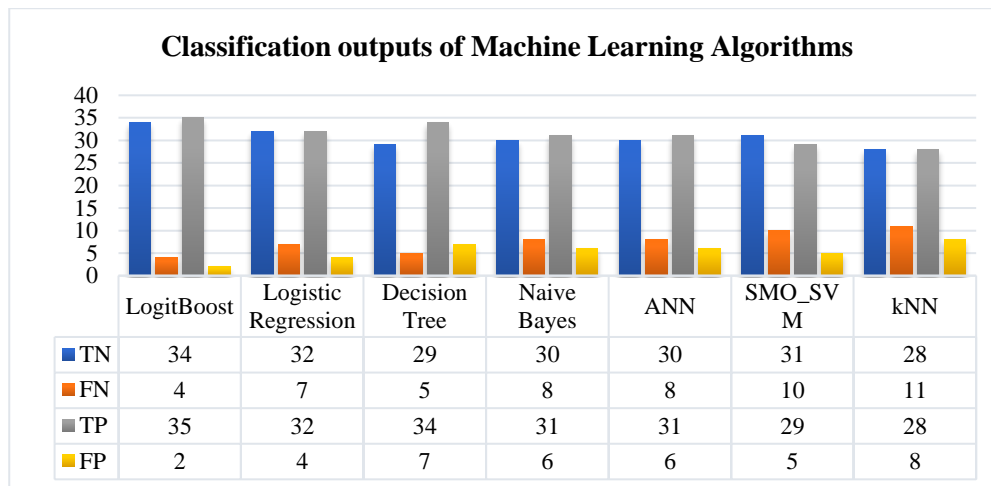


Figure 4: Classification output of machine learning algorithms

	CA	Sens.	Spec.	Prec.	FPR	NPV	RMC	F1
LogitBoost	0.92	0.90	0.94	0.95	0.06	0.89	0.08	0.92
Logistic Regression	0.85	0.82	0.89	0.89	0.11	0.82	0.15	0.85
Decision Tree	0.84	0.87	0.81	0.83	0.19	0.85	0.16	0.85
Naïve Bayes	0.81	0.79	0.83	0.84	0.17	0.79	0.19	0.82
ANN	0.81	0.79	0.83	0.84	0.17	0.79	0.19	0.82
SMO	0.80	0.74	0.86	0.85	0.14	0.76	0.20	0.79
kNN	0.75	0.72	0.78	0.78	0.22	0.72	0.25	0.75

Table 3: Classification performance metrics of machine learning algorithms.

MODEL	ACCURACY (%)	Sensitivity	Specificity	REFERENCE
Support Vector Machine	90.42%	47.23%	97.59%	[6]
Random Forest (Ensemble)	92%	94%	92%	[5]
Artificial Neural Network	90%	-	-	[9]
Artificial Neural Network	85.92%	-	-	[14]
Decision Tree (C4.5)	84.7%	78.2%	80.2%	[21]
Alternative Decision Tree	89%	89.2%	47.6%	[11]
LogitBoost (Ensemble)	92%	90%	94%	In this experiment

Table 4: Comparison of accuracy result of LogitBoost ensemble model among other experiments.

has the highest negative predictive value of 89% whereas it also defeats all other methods on the F1 score with 92%. LogitBoost also achieves the lowest FP rate of 6%, and also the lowest Rate of misclassification (8%).

6.1 ROC curve for performance evaluation

Receiver Operating Characteristic (ROC) curve is a generally employed diagrammatical representation which estimates the performance of the classification models over all feasible thresholds. ROC curve is generated by tracing the FPR on the x-axis with contrary to the TPR on the y-axis. ROC is impartial of both classes and important when the number of instances of both classes mutates at training. Range under ROC must be close to 1 for the best classifier.

Figure 5 enlighten that LogitBoost defeats all other methods in the prediction of negative dengue outbreak

case and Figure 6, LogitBoost beat other methods in the prediction of positive dengue outbreak case.

7 Limitation and future work

In this experimental work, we have used 8 clinical parameters with 75 dataset samples (36 dengue negative and 39 dengue positive samples) and performs classification tasks of data mining. After that, we applied seven prominent algorithms in which LogitBoost (one of the ensemble model) performs better than others. According to No-Free-Lunch [22], a single learning algorithm cannot be the best and at most appropriate with the whole domain of application. It may be the computing cost and processing time can increase due to ensemble model but subsequently, day by day the new technologies have come into existence like cloud computing services

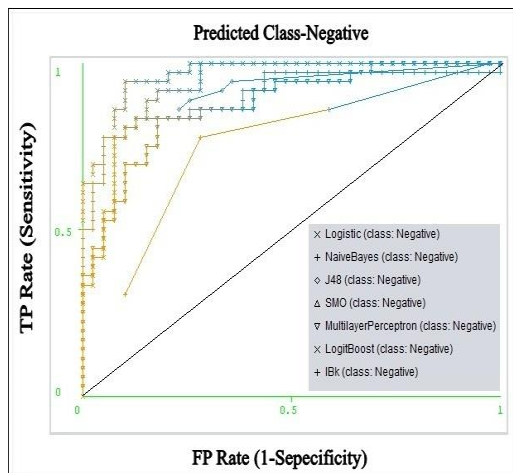


Figure 5: ROC for seven machine learning techniques tested for the negative case.

and distributed computing that reduced the computing cost and processing time.

In the future, one can use huge datasets with more related clinical parameters for their experiments and improvement of model accuracy as mention in the data classification section [10].

8 Conclusion

Dengue disease patients are increasing rapidly and actually, dengue has recorded in every continent today according to the World Health Organisation (WHO) record. Dengue outbreak prediction may save the life of people and can have valuable effectiveness on their diagnostic. This effort gives a work process established on machine learning techniques for the forecasting of the negative case or the positive case of dengue outbreak.

The prime focus of the research is toward prediction of dengue outbreak using WEKA tool. In this research article, seven prominent machine learning techniques have been applied and eight parameters are used for performance evaluation.

It has been concluded that LogitBoost ensemble model is the topmost performance classifier techniques that it has reached a classification accuracy of 92% with sensitivity and specificity of 90 and 94 % respectively and ROC area=0.967, and had the lowest error rate.

We have compared the accuracy rate of our analysis with other published results in Table 4. Based on our comparative analysis result using LogitBoost ensemble model as well as the Random forest classifier used by Fathima et al, (2015) [5] result concluded that ensemble model performs better than individual classifier (Table 4).

Furthermore, we are desirous to enhance the model accuracy with more related expressed and sensitive clinical features on a huge amount of dataset in future and as well as we are also interested to develop a web-based tool that helps doctors to take a decision with more accurate dengue outbreak.

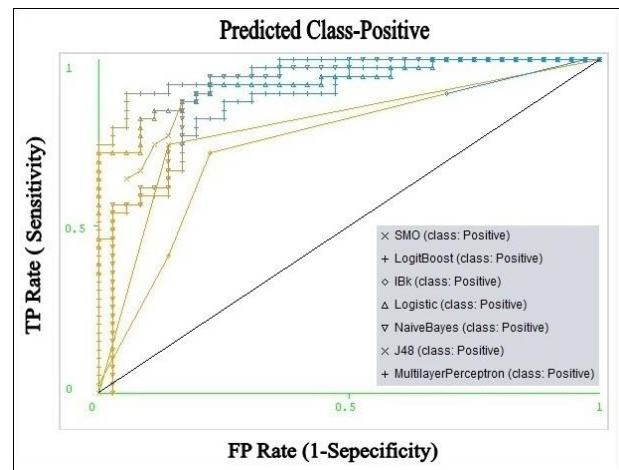


Figure 6: ROC for seven machine learning techniques tested for the positive case.

List of abbreviations

DEN	: Dengue
DF	: Dengue Fever
DHF	: Dengue Haemorrhage Fever
DSS	: Dengue Shock Syndrome
CSV	: Comma Separated Values
WBC	: White Blood Count
ANN	: Artificial Neural Network
SVM	: Support Vector Machine
SMO	: Sequential Minimal Optimization
ADT	: Alternating Decision Tree
NB	: Naive Bayes
RF	: Random Forest
MNB	: Modified Naive Bayes
MFNN	: Multilayer Feedforward Neural Network
ROC	: Receiver Operative Characteristics

9 References

- [1] Althouse, B. M., Ng, Y. Y., & Cummings, D. A. (2011). Prediction of dengue incidence using search query surveillance. *PLoS Negl Trop Dis*, 5(8), e1258. <https://doi.org/10.1371/journal.pntd.0001258>
- [2] Arunachalam, N., Tana, S., Espino, F., Kittayapong, P., Abeyewickrem, W., Wai, K. T., ... & Petzold, M. (2010). Eco-bio-social determinants of dengue vector breeding: a multicountry study in urban and periurban Asia. *Bulletin of the World Health Organization*, 88(3), 173-184. <https://doi.org/10.2471/BLT.09.067892>
- [3] Brasier, A. R., Ju, H., Garcia, J., Spratt, H. M., Victor, S. S., Forshey, B. M., ... & Rocha, C. (2012). A three-component biomarker panel for prediction of dengue hemorrhagic fever. *The American journal of tropical medicine and hygiene*, 86(2), 341-348. <https://doi.org/10.4269/ajtmh.2012.11-0469>
- [4] Chadwick, D., Arch, B., Wilder-Smith, A., & Paton, N. (2006). Distinguishing dengue fever from other infections on the basis of simple clinical and laboratory features: application of logistic regression analysis. *Journal of Clinical Virology*, 35(2), 147-

153.
<https://doi.org/10.1016/j.jcv.2005.06.002>
- [5] Fathima, A. S., & Manimeglai, D. (2015). Analysis of Significant Factors for Dengue Infection Prognosis Using the Random Forest Classifier. *Analysis*, 6(2).
<https://doi.org/10.14569/IJACSA.2015.060235>
- [6] Fathima, A., & Manimegalai, D. (2012). Predictive analysis for the arbovirus-dengue using svm classification. *International Journal of Engineering and Technology*, 2(3), 521-7.
- [7] Gibbons, R. V., & Vaughn, D. W. (2002). Dengue: an escalating problem. *BMJ: British Medical Journal*, 324(7353), 1563.
<https://doi.org/10.1136/bmj.324.7353.1563>
- [8] Horstick, O., Farrar, J., Lum, L., Martinez, E., San Martin, J. L., Ehrenberg, J., ... & Kroeger, A. (2012). Reviewing the development, evidence base, and application of the revised dengue case classification. *Pathogens and global health*, 106(2), 94-101.
<https://doi.org/10.1179/2047773212Y.0000000017>
- [9] Ibrahim, F., Taib, M. N., Abas, W. A. B. W., Guan, C. C., & Sulaiman, S. (2005). A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN). *Computer methods and programs in biomedicine*, 79(3), 273-281.
<https://doi.org/10.1016/j.cmpb.2005.04.002>
- [10] Iqbal, N. and Islam, M. (2017). Machine learning for Dengue outbreak prediction: An outlook, *International Journal of Advanced Research in Computer Science*, 8(1):93-102.
- [11] Kumar, M. N. (2013). Alternating decision trees for early diagnosis of dengue fever. *arXiv preprint arXiv:1305.7331*.
- [12] Nandini, V., Sriranjitha, R., & Yazhini, T. P (2016). Dengue detection and prediction System using data mining with Frequency analysis. *Computer Science & Information Technology*, [DOI: 10.5121/csit.2016.60906].
<https://doi.org/10.5121/csit.2016.60906>
- [13] Online Available
[\[https://en.wikipedia.org/wiki/Dengue_fever\]](https://en.wikipedia.org/wiki/Dengue_fever)
- [14] Rachata, N., Charoenkwan, P., Yooyativong, T., Chamnongthai, K., Lursinsap, C., & Higuchi, K. (2008, October). Automatic prediction system of dengue haemorrhagic-fever outbreak risk by using entropy and artificial neural network. In *Communications and Information Technologies, 2008. ISCIT 2008. International Symposium on* (pp. 210-214). IEEE.
<https://doi.org/10.1109/ISCIT.2008.4700184>
- [15] Raza, K. (2019). Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule. In *U-Healthcare Monitoring Systems* (pp. 179-196). Academic Press.
<https://doi.org/10.1016/B978-0-12-815370-3.00008-6>
- [16] Santamaria, R., Martinez, E., Kratochwill, S., Soria, C., Tan, L. H., Nunez, A., ... & Castelobranco, I. (2009). Comparison and critical appraisal of dengue clinical guidelines and their use in Asia and Latin America. *International health*, 1(2), 133-140.
<https://doi.org/10.1016/j.inhe.2009.08.006>
- [17] Shakil, K. A., Anis, S., & Alam, M. (2015). Dengue disease prediction using weka data mining tool. *arXiv preprint arXiv:1502.05167*.
- [18] Shaukat, K., Masood, N., Mehreen, S., & Azmeen, U. (2015). Dengue Fever Prediction: A Data Mining Problem. *Journal of Data Mining in Genomics & Proteomics*, 2015.
<https://doi.org/10.4172/2153-0602.1000181>
- [19] Souza, L. J. D., Nogueira, R. M. R., Soares, L. C., Soares, C. E. C., Ribas, B. F., Alves, F. P., ... & Pessanha, F. E. B. (2007). The impact of dengue on liver function as evaluated by aminotransferase levels. *Brazilian Journal of Infectious Diseases*, 11(4), 407-410.
<https://doi.org/10.1590/S1413-86702007000400007>
- [20] Stany Leena Princy, S., & Muruganandam, A. (2016). An Implementation of Dengue Fever Disease Spread Using Informatica Tool with Special Reference to Dharmapuri District. *International Journal of Innovative Research in Computer and Communication Engineering*, 4(9). [DOI: 10.15680/IJIRCCE.2016.0409031].
- [21] Tanner, L., Schreiber, M., Low, J. G., Ong, A., Tolfvenstam, T., Lai, Y. L., ... & Simmons, C. P. (2008). Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. *PLoS Negl Trop Dis*, 2(3), e196.
<https://doi.org/10.1371/journal.pntd.0000196>
- [22] Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1), 67-82.
<https://doi.org/10.1109/4235.585893>

