An Enhanced Random Forest Model for Detecting Effects on Organs after Recovering from Dengue

Avijit Kumar Chaudhuri[1], Arkadip Ray[2], Prof. Dilip K. Banerjee[3], Dr. Anirban Das[4]

[1]Research Scholar, Department of Computer Application, SEACOM SKILLS UNIVERSITY, Kendradangal, Bolpur, Birbhum, 731236, West Bengal, India

[2]Department of Information Technology, Government College of Engineering and Ceramic Technology, Kolkata, West Bengal, 700010, India

[3]Professor, Department of Computer Application, SEACOM SKILLS UNIVERSITY, Kendradangal, Bolpur, Birbhum, 731236, West Bengal, India

[4]University of Engineering & Management, Kolkata, West Bengal, India

Abstract : One of the most important issues in healthcare and machine learning research is determining the likelihood of organ failure in dengue fever (DF). Researchers have demonstrated the effectiveness of artificial intelligence and machine-learning models in a variety of practical classification problems. The current study focuses on the development of a prediction model for organ failure in dengue. This paper proposes an Enhanced random forest (ERF) model that employs an ensemble of classification methods to achieve this goal. The proposed ERF classifier is tested on a dengue fever dataset collected from dengue patients from all over the West Bengal state in India from 2016 to 2019, from several hospitals, and by interacting with people previously infected with DF individually using online and offline questionnaire methods. The proposed classifier is also compared to some cutting-edge machine-learning classifiers, including random forest, naive Bayes, support vector machine with radial basis function kernel, and decision tree. To assess the strength of the proposed ERF classifier, various performance metrics such as accuracy, sensitivity, specificity, receiver operating characteristic, area under the curve, and some statistical tests such as kappa statistics were used to test the classifiers. To test the credibility of the classification models in dealing with unbalanced data, various splits of training and testing data – namely, 50–50 percent, 66–34 percent, 80–20 percent, and 10-fold cross-validation – were used in this study. The output results were also compared to previous research on the same dataset, where the proposed classifiers were found to be the best across all performance dimensions.

Keywords: Enhanced Random Forest (ERF), Dengue Fever (DF), Organ Failure, Ensemble of Classification Methods, Kappa Statistic, ROC-AUC

1. INTRODUCTION

Human well-being is facing significant challenges as instances of various virus-related diseases are increasing. In these situations, focused research and development are a pressing need. Given the availability of broad datasets, data mining is a recommended approach to gain insights. The occurrence of multiple organ dysfunctions is due to the development of mortality in acute dengue infections. Severe dengue symptoms vary, and consequently, the precise morbidity and mortality in terms of organ failure are not well studied in the Indian environment. From 2019 to 2021, the authors conducted a prospective multicenter retrospective analysis in some territory care and intensive care units in Kolkata. The aim was to calculate the likelihood of organ failure in serious dengue infection [11]. The availability of pathological test data offers the potential to gain insights with a higher accuracy level. Some authors have used data-mining approaches in the study of dengue but have concluded that these approaches have varied accuracy in prediction. There is very little evidence in studies on the application of data mining to predict the effects of dengue on several organs.

Bagging and boosting are two ways to improve the classification. Bagging performs a bootstrap to obtain object samples to train a classifier in dealing with individual samples [3]. Although the classifier models used in bagging are sensitive to minor data changes, bootstrap sampling appears to lead to low-diversity ensembles compared to other

methods of creating an ensemble. Bagging needs larger sizes of ensembles. This issue can be overcome by using the random forest (RF) method [4]. RF method also uses bootstrapping to create samples, but it differs in how it builds individual decision trees. RF makes use of an F algorithm that leads to the implementation of higher diversity without reducing the accuracy of every classifier. The individual classifiers are built parallel to each other, rather than being dependent on one another. However, this creation does not suggest that the outputs are independent.

Boosting, on the other hand, helps to enhance the efficiency of a weak classifier. It refers to a family of methods, with AdaBoost being the most popular member of the family. On average, AdaBoost turns out to be better when compared to bagging and RF. A previous study has shown that bagging and RF have their application niches [12]. All classifiers behave identically in the presence of large group sizes [1]. Thus, the challenge lies in the case of small ensemble sizes.

AdaBoost has a greater ability to create diversity, but at the same time, it can generate incorrect classifiers because it forces the classifiers to concentrate on hard objects and ignores the rest of the data. Research has demonstrated that it is challenging to get classifiers both to create high levels of diversity and to achieve greater accuracy [17]. In this paper, we propose the use of the Enhanced random forest (ERF) classifier to address these issues – that is, to ensure both diversity and accuracy.

This paper attempts to answer the following research questions: Which is the best Data Mining Technique (DMT) for the prediction of dysfunction of multi-organs in dengue? We consider the most popular approaches and explore their ensemble to arrive at the highest levels of consistency, sensitivity, and specificity. Previous authors have emphasized the reduction of variables to improve prediction. However, this approach leads to a loss of information. Thus, in this paper, we establish a framework that proposes the application of data-mining approaches, the measurement of consistency using kappa statistics, and suggested improvement of the specificity and sensitivity parameters using an ensemble learning approach, without any variable reduction.

In this research, we propose an ERF approach to construct an ensemble of classifiers. This classifier utilizes n number of RF classifiers (instead of the decision tree (DT) default classifier) with the adoptive boosting technique by changing its defined base classifier to RF and changing its weight iteratively. This proposed classifier will integrate the n individual decisions and generate a robust classifier with 84% accuracy. Thus, the framework in this paper contributes to the well-being of humankind by enabling higher levels of prediction of multi-organ failure in dengue.

2. RELEVANT LITERATURE

Diagnostic test evaluation is a challenge in the current healthcare industry, not just for proving the existence of illness but also for ruling out disease in healthy subjects. The mainstream approach to diagnostic test assessment uses sensitivity and specificity as indicators of test accuracy in addition to gold standard status of diagnostic tests with dichotomous outcomes (positive/negative test results) [15]. When the test findings are reported in ordinal scale (e.g., 5 ordinal scale: "definitely regular", "probably regular", "uncertain", "probably irregular", "definitely irregular") or on continuous scale, the sensitivity and specificity may be measured over all possible threshold values. As a result, sensitivity and specificity vary depending on the threshold, and sensitivity is inversely related to specificity [2,8,13]. The plot of sensitivity versus 1-specificity is then known as the receiver operating characteristic (ROC) curve, and the field under the curve (AUC) has been considered as an important indicator of accuracy for meaningful interpretations [9]. This curve is critical in assessing a test's ability to discriminate the true state of a subject, determining the best cut off values, and comparing two alternative diagnostic tasks as each task is done on the same subject [9,10]. According to a Pubmed search, this study has been commonly used in clinical epidemiology for evaluating the diagnostic ability of biomarkers (e.g., serum markers) and imaging studies in the differentiation of diseased from healthy subjects [5,6,7,14,16,18]. This mathematical approach is often used in medical trials to measure the likelihood of some negative outcome based on the patient's risk profile. This paper discusses the benefits of the ROC curve, accuracy experiments that use the ROC curve and their predictive behaviours, as well as bias and confusing issues in ROC analysis.

3. METHODOLOGY

3.1 Dataset

The authors collected data of dengue patients from the year 2016 to 2019 of all over the West Bengal state in India, from several hospitals and also by interacting with the people previously infected with DF individually by using online and offline questionnaire methods. The authors gather information from a population of 268 people are involved in this study with 131 females and 137 males. The dataset for this study is extracted to predict whether a patient is likely to face organ failure based on the following attribute information (as depicted in Table 1).

Table. 1. Description of the Dataset

Sl. No.	Feature	Description	Range of Values
1	Age	Age at exam time in years	Continuous
2	Gender	Male or Female	0 = Female, 1 = Male
3	Symptoms	Symptoms before detecting Dengue	Fever, Vomiting, Headache, Body Pain, Urine output etc.
4	Temperature Level	Range of temperature during the disease	
5	Fever Stayed in days		Continuous
6	Symptoms after detecting Dengue		Fever, Fatty liver, Prostatomegaly, Vomiting etc.
7	Plate late Counting		Continuous
8	Average BP	Blood pressure	mmHg
9	Bleeding episode present or not		0 = no bleeding episodes; 1 = bleeding episodes present;
10	Symptoms after recovery of Dengue		Red eyes, Headache, Loss of Appetite, Vomiting etc
11	Effect on other organ after recovery		0 = Organ not effected, 1 = Organ effected

3.2 Algorithm for ERF

Input: Sequence of M examples,

S = { $(x_1, y_1)\ldots\ldots(x_m, y_m)$ } where, $x_i \in X$ with labels $y_i \in Y = \{\omega_1 + \omega_2 + \omega_3 + \ldots\ldots + \omega_c\}$, where ω_i is the total number of classes and number of iterations for learning = t.

Initialization: Distribution, $D_i^t = \dfrac{1}{M}$, i = 1, 2, 3, …, M

Neighbour(s): K = n, 1≤n≤M

For t = 1 to t, perform the following –

Step 1: Select the subset for training the data S_{Set}, peaked from the distribution D_j.

Step 2: Train the base classifier with S_{Set} and obtain the hypothesis h_t, where $h_t : X \rightarrow Y$.

Step 3: Compute the error of h_t.

$$h_t . \varepsilon_t = \sum_i^M D_i^t e_i^t \qquad \text{where} \quad e_i^t \begin{cases} (h_t(x_i) = y_i) = 1 \\ (h_t(x_i) \neq y_i) = 0 \end{cases}$$

Step 4: If $\varepsilon_t > 0.5$, then set t = t -1 and exit from the loop.

Step 5: Set weight, $\beta_t = \dfrac{\varepsilon_t}{1 - \varepsilon_t}$.

$$D^t : D_i^{(t-1)} = \frac{(D_i^t)}{Z_t} \times \Phi_t$$

Update the distribution,

where $\Phi_t = e_i^t \begin{cases} (h_t(x_i) = y_i) = \beta_t \\ (h_t(x_i) \neq y_i) = 1 \end{cases}$ and $z_t = \sum_i^M D_i^t$ is a constant for normalization, so that D_j^{-1} becomes a proper distribution.

Output: Given an unlabeled instance x, select the class that has the maximum total vote as the optimum classification.

$$h_f(x) = \arg\underset{y=Y}{\arg\max} \cdot \sum_{t=1}^r \log \frac{1}{\beta_t} \times v, \text{ where } V = \begin{cases} (h_t(x) = \omega_j) = 1 \\ (h_t(x) \neq \omega_j) = 0 \end{cases}.$$

3.3 Assessment of Performance of Machine Learning Algorithms

In this paper, statistical metrics are applied to evaluate the classification performance of machine-learning algorithms. The metrics include (1) Accuracy, (2) Kappa statistic for each model and (3) receiver operating characteristic (ROC) curve and area under the curve (AUC) values.

3.3.1 Accuracy

For the binary classification, evaluation metrics include accuracy, sensitivity, and specificity. The metrics are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3}$$

where TP: true positive - patient with dengue correctly classified; TN: true negative - patient without dengue correctly classified; FP: false positive - patient with dengue incorrectly classified; FN: false negative - patient without dengue incorrectly classified.

In other words, the term accuracy measures the rate of correctly classified instances, sensitivity is the rate of correctly classified instances with dengue, and specificity is the rate of correctly classified instances without dengue.

3.3.2 Kappa Statistic for Each Model

Cohen's kappa statistic enables accuracy checks of classification. The kappa score gives the measure of the accuracy of classification in the range [-1,1]. This statistic compares the observed (between the coders) and predicted (the probability that the coders agree by chance) agreement [19]. A value of -1 indicates a sharp disparity between observed and predicted outcomes, while a score of 1 indicates the opposite. The zero value indicates that observed and predicted outcomes are equal.

3.3.3 ROC Curve and AUC Values

In a binary classifier, AUC is a performance metric. It measures the degree to which the curve is up in the northwest corner by contrasting the ROC curve with the area below the curve. A score of 0.5 is regarded as better than a random guess. A value close to 0.9 indicates a good model, and a score equal to 1 would indicate an excellent model. The basic theory and rationale for ROC analysis of ranking and continuous diagnostic test outcomes versus a gold standard [7,16]

are presented in this article. For disease classification from healthy subjects, derived accuracy indexes, especially the area under the curve (AUC), have a meaningful interpretation. The methods for estimating AUC and evaluating it in single diagnostic tests as well as comparative studies, the benefit of using a ROC curve to assess the best cut-off values, and bias and confounding issues have all been addressed.

4. RESULTS AND DISCUSSION

We developed and simulated a proposed model by using the Python programming language. In this model, we performed a comparative study between five state-of-the-art machine-learning algorithms, namely LR, RF, NB, SVM, and DT, and the proposed model. Among these five popular machine-learning techniques, some show better accuracy, whereas the performances of others are inferior. To boost the accuracy and performance of the weak classifier, we used advanced ensemble machine learning and proposed an ensemble meta-algorithmic technique (as shown in Fig. 1).

The machine-learning techniques applied on the collected dataset showed 84% accuracy.

As shown in Table 2 different approaches yielded different levels of accuracy, with LR and NB recording an accuracy of 52% while the proposed ERF exhibited 84% accuracy.

A confusion matrix presents the statistics of real and projected classifications achieved from the analysis of different classification systems. The performance of all such systems is generally assessed by using the data generated in this matrix. Table 3 shows the results generated from confusion matrices by using different machine-learning algorithms. The performance of our proposed model, along with the performances of other methods, was evaluated based on sensitivity, specificity, and accuracy tests, which use the true positive (TP), true negative (TN), false negative (FN), and false positive (FP) terms.

The results of sensitivity and specificity in Table 3 demonstrate the potential of our proposed model in the classification of two classes. The comparison of our proposed model with other widely used independent classification techniques is shown in Table 3. It is clear from the comparative results that our proposed classification technique along with RF has the highest accuracy, sensitivity and specificity values (accuracy = 84%, sensitivity = 0.84, specificity = 0.83) for the identification of the probability of effecting organs in dengue.
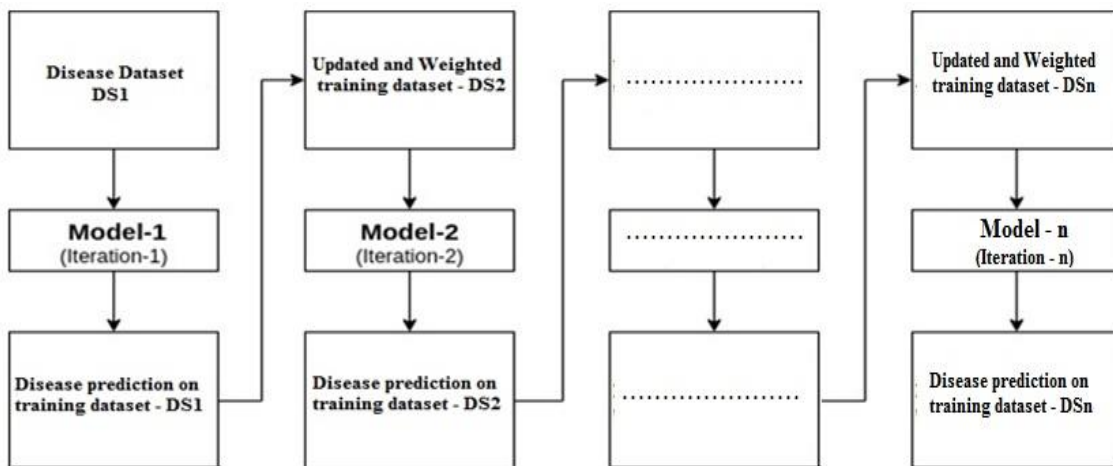

Fig. 1. Working of The Proposed Classifier

We found from different analyses in our study that the classification accuracy of the LR was 57%, with 56% sensitivity and 63% specificity. The RF achieved an accuracy of classification of 82%, with 87% sensitivity and 88% specificity. The accuracy of classification of the SVM was 57%, with 62% sensitivity and 64% specificity. However, the best performance of the six classifiers evaluated was that of our proposed classifier, which achieved 84% accuracy in classification, with 84% sensitivity and 83% specificity. Table 2 and Table 3 show the complete set of results.

Our proposed classifier's analytical findings are positive. An important result is the improved specificity and sensitivity to predict using our proposed classifier. The LR, NB and SVM classifiers produced very low sensitivities and specificities. Our proposed classifier removed that drawback, producing 84% accuracy results for sensitivity and 83% accuracy for specificity, which would mean that fewer patients would need to be tested for organ failure due to higher

specificity. At the same time, a higher sensitivity value would also save money and shorten the waiting times of the genuinely ill patients, which would be critical to saving lives.

The ROC charts for these experiments with individual machine-learning techniques are depicted in Table 4. In this table, six ROC charts are drawn in different parts for a 10-fold cross-validation shown in blue. Experimental results show that, in terms of cross-validation accuracy, along with RF, our proposed classifier outperformed all the other previously used methods discussed in the literature review. With the proposed model, the generated AUC value reaches 0.89.

Comparing the performances of different machine-learning classifiers might generate an ambiguous result if the comparison has been based only on accuracy-based metrics. The Cohen's Kappa Statistic (CKS) value is used to help produce error-free comparative efficiency of different classifiers. The cost of error must be considered in such evaluations. In this respect, the CKS is an excellent measure for inspecting classifications that may be due to chance. Usually, the CKS takes a value between -1 and +1. As the classifier's calculated kappa value approaches 1, its performance is assumed to be more realistic than "by chance." Therefore, the CKS value is a suggested metric for measurement purposes in the performance analysis of classifiers [19]. This kappa value is calculated by using Equation 4.

$$CKS = \frac{(pa - pac)}{(1 - pac)} \qquad (4)$$

where, pa represents total agreement probability and pac represents probability "by chance."

The results of the CKS analysis of the five popular machine-learning techniques and our proposed model are shown in Table 5. These results clearly demonstrate that our proposed model performed much better than other classifiers (value = 67.9).

Table. 2. Comparison of Accuracy

Training -Testing Partition	Accuracy					
	LR	RF	ERF	NB	SVM	DT
50-50	0.57	0.80	0.84	0.52	0.57	0.83
66-34	0.53	0.79	0.82	0.59	0.53	0.75
80-20	0.52	0.80	0.70	0.59	0.54	0.74
10-fold Cross Validation	0.53	0.82	0.84	0.59	0.56	0.84

Table. 3. Comparison of Sensitivity and Specificity

Training - Testing Split	LR		RF		ERF		NB		SVM		DT	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
50-50	0.52	0.60	0.87	0.76	0.84	0.83	0.46	0.56	0.52	0.62	0.84	0.82
66-34	0.56	0.61	0.81	0.79	0.80	0.84	0.50	0.61	0.55	0.63	0.68	0.80
80-20	0.36	0.63	0.74	0.83	0.60	0.82	0.33	0.63	0.62	0.64	0.64	0.81
10-fold Cross Validation	0.56	0.47	0.82	0.88	0.84	0.83	0.58	0.60	0.56	0.53	0.84	0.83

Table. 4. Comparison of ROC Curve and AUC Values
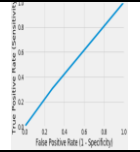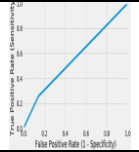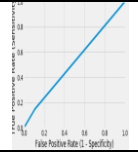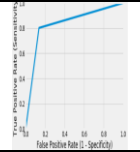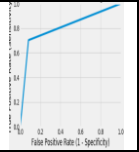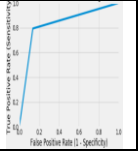
Training-Testing Partition	LR	NB	SVM	DT	RF	ERF
10-fold Cross Validation						
AUC	52.25	57.26	50.7	83.6	90	89

Table. 5. Kappa Statistic for Each Model

Training - Testing Split	Kappa Statistic					
	LR	RF	ERF	NB	SVM	DT
50-50	11.36	58.27	**66.6**	02.81	13.3	65.02
66-34	52.39	56.21	**61.74**	49.04	07.57	49.04
80-20	−1.15	55.87	**40.17**	−2.06	01.75	45.53
10-fold Cross Validation	02.77	63.14	**67.9**	11.78	04.38	66.92

CONCLUSION

A comparison of the performance of the popular machine-learning models with that of our proposed model requires benchmarking. Our proposed classifier was compared to assess whether the proposed model is the best and whether it improves performance and classification accuracy. Accuracy is determined by the number of feature selection techniques and results produced by the other models in most of the research papers. In selecting the features to use, our proposed classifier had no restrictions. The best results were obtained by considering all features available in the dataset in this model. The comparison of performances of different models with our proposed method is shown in Table 2, Table 3, Table 4 and Table 5. For our proposed classifier, all 16 attributes were selected and the classification was made on the dataset. The results obtained by using our proposed classifier with all features from the dataset proves that it is efficient in accurately predicting the probability of organ failure in dengue patients compared with other known existing machine-learning algorithms.

So, the main contribution of this research paper is not only the development of an ensemble learning model but also the restructuring of the default structure of the boosting algorithm by changing the base estimator. Our proposed classifier also has important features, such as its ability to determine the best possible ratio on the training dataset versus the testing dataset to simultaneously find the optimum combination of both sets based on the defined ratio, as well as its experimentation to find an accurate rule using the ERF. Test results show that our proposed ensemble learning classification model is effective in improving performance metrics and classification accuracy compared to its foundation learner and other independent learners specified in the literature.

REFERENCES

1. Banfield, R. E., Hall, L. O., Bowyer, K. W., Bhadoria, D., Kegelmeyer, W. P., & Eschrich, S. (2004, June). A comparison of ensemble creation techniques. In International Workshop on Multiple Classifier Systems (pp. 223-232). Springer, Berlin, Heidelberg.

2. Ben-Shakhar, G., Lieblich, I., & Bar-Hillel, M. (1982). An evaluation of polygraphers' judgments: A review from a decision theoretic perspective. Journal of Applied Psychology, 67(6), 701.

3. Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.

4. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

5. Darmon, M., Vincent, F., Dellamonica, J., Schortgen, F., Gonzalez, F., Das, V., ... & Schlemmer, B. (2011). Diagnostic performance of fractional excretion of urea in the evaluation of critically ill patients with acute kidney injury: a multicenter cohort study. Critical care, 15(4), 1-8.

6. Daubin, C., Quentin, C., Allouche, S., Etard, O., Gaillard, C., Seguin, A., ... & Du Cheyron, D. (2011). Serum neuron-specific enolase as predictor of outcome in comatose cardiac-arrest survivors: a prospective cohort study. BMC cardiovascular disorders, 11(1), 1-13.

7. Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian journal of internal medicine, 4(2), 627.

8. Hanley, J. A. (1989). Receiver operating characteristic (ROC) methodology: the state of the art. Crit Rev Diagn Imaging, 29(3), 307-335.

9. Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1), 29-36.

10. Kumar, R., & Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. Indian pediatrics, 48(4), 277-287.

11. Jog, S., Prayag, S., Rajhans, P., Zirpe, K., Dixit, S., Pillai, L., ... & Patel, D. (2015). Dengue infection with multiorgan dysfunction:-sofa score, arterial lactate and serum albumin levels are predictors of outcome. Intensive Care Medicine Experimental, 3(1), 1-2.

12. Li, J., Heap, A. D., Potter, A., & Daniell, J. J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. Environmental Modelling & Software, 26(12), 1647-1659.

13. Linnet, K. (1987). Comparison of quantitative diagnostic tests: type I error, power, and sample size. Statistics in Medicine, 6(2), 147-158.

14. Reddy, S., Dutta, S., & Narang, A. (2008). Evaluation of lactate dehydrogenase, creatine kinase and hepatic enzymes for the retrospective diagnosis of perinatal asphyxia among sick neonates. Indian pediatrics, 45(2), 144.

15. Swets, J. A. (1979). ROC analysis applied to the evaluation of medical imaging techniques. Investigative radiology, 14(2), 109-121.

16. Vandewiele, G., Dehaene, I., Kovács, G., Sterckx, L., Janssens, O., Ongenae, F., ... & Demeester, T. (2021). Overly optimistic prediction results on imbalanced data: a case study of flaws and benefits when applying over-sampling. Artificial Intelligence in Medicine, 111, 101987.

17. Xiao, H., Xiao, Z., & Wang, Y. (2016). Ensemble classification based on supervised clustering for credit scoring. Applied Soft Computing, 43, 73-86.

18. Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation, 115(5), 654-657.

19. Ben-David, A. (2008). Comparison of classification accuracy using Cohen's Weighted Kappa. Expert Systems with Applications, 34(2), 825-832.

20. Kowshik B; Savitha V; Nimosh madhav M; Karpagam G; Sangeetha K. "Plant Disease Detection Using Deep Learning". International Research Journal on Advanced Science Hub, 3, Special Issue ICARD-2021 3S, 2021, 30-33. doi: 10.47392/irjash.2021.057

21. Salini Suresh; Suneetha V; Niharika Sinha; Sabyasachi Prusty; Sriranga H.A. "Machine Learning: An Intuitive Approach In Healthcare". International Research Journal on Advanced Science Hub, 2, 7, 2020, 67-74. doi: 10.47392/irjash.2020.67

22. Kowshik B; Savitha V; Nimosh madhav M; Karpagam G; Sangeetha K. "Plant Disease Detection Using Deep Learning". International Research Journal on Advanced Science Hub, 3, Special Issue ICARD-2021 3S, 2021, 30-33. doi: 10.47392/irjash.2021.057

23. Maneesha M; Savitha V; Jeevika S; Nithiskumar G; Sangeetha K. "Deep Learning Approach For Intelligent Intrusion Detection System". International Research Journal on Advanced Science Hub, 3, Special Issue ICARD-2021 3S, 2021, 45-48. doi: 10.47392/irjash.2021.061

24. Sona Solanki; Asha D Solanki. "Review of Deployment of Machine Learning in Blockchain Methodology". International Research Journal on Advanced Science Hub, 2, 9, 2020, 14-20. doi: 10.47392/irjash.2020.141

25. Mohd. Akbar; Prasadu Peddi; Balachandrudu K E. "Inauguration in Development for Data Deduplication Under Neural Network Circumstances". International Research Journal *on Advanced Science Hub*, 2, 6, 2020, 154-156. doi: 10.47392/irjash.2020.55