



ARTICLE

Modelling an Efficient Clinical Decision Support System for Heart Disease Prediction Using Learning and Optimization Approaches

Sridharan Kannan*

Department of Computer Science and Engineering, J K K Munirajah College of Technology, Erode, 638506, India

*Corresponding Author: Sridharan Kannan. Email: srijik77@gmail.com

Received: 04 August 2021 Accepted: 11 November 2021

ABSTRACT

With the worldwide analysis, heart disease is considered a significant threat and extensively increases the mortality rate. Thus, the investigators mitigate to predict the occurrence of heart disease in an earlier stage using the design of a better Clinical Decision Support System (CDSS). Generally, CDSS is used to predict the individuals' heart disease and periodically update the condition of the patients. This research proposes a novel heart disease prediction system with CDSS composed of a clustering model for noise removal to predict and eliminate outliers. Here, the Synthetic Over-sampling prediction model is integrated with the cluster concept to balance the training data and the Adaboost classifier model is used to predict heart disease. Then, the optimization is achieved using the Adam Optimizer (AO) model with the publicly available dataset known as the Stalog dataset. This flow is used to construct the model, and the evaluation is done with various prevailing approaches like Decision tree, Random Forest, Logistic Regression, Naive Bayes and so on. The statistical analysis is done with the Wilcoxon rank-sum method for extracting the p -value of the model. The observed results show that the proposed model outperforms the various existing approaches and attains efficient prediction accuracy. This model helps physicians make better decisions during complex conditions and diagnose the disease at an earlier stage. Thus, the earlier treatment process helps to eliminate the death rate. Here, simulation is done with MATLAB 2016b, and metrics like accuracy, precision-recall, F-measure, p -value, ROC are analyzed to show the significance of the model.

KEYWORDS

Heart disease; clinical decision support system; over-sampling; AdaBoost classifier; adam optimizer; Wilcoxon ranking model

1 Introduction

Globally, 42 million premature deaths are encountered annually due to non-communicable diseases reported by World Health Organization (WHO), i.e., roughly 72%. The total count of the non-communicable disease-based death reaches 52.1 million annually by 2030 [1–3] when these diseases are un-mitigated. Some general non-communicable conditions are hypertension and diabetes, about 46% of the total death rate. Type II diabetes occurs due to constant metabolic disorder that shows variation in blood glucose levels. It is generally a significant cause of the



human's ineffectiveness to adopt its own produced insulin. Likely, individuals are supposed to have diabetes with higher chances and risk of mortality and stroke [4]. However, constant monitoring of blood glucose levels can appropriately mitigate or prevent the complications of diabetes. In most developing countries, huge people have diabetes, and the diabetic rate will increase gradually from 84 to 229 million approximately by 2030 and burden the healthcare system significantly. Moreover, hypertension is a harmful condition where blood flows through blood vessels like arteries and veins at raised pressure. Globally, it is noted that one in three adults show increased blood pressure, and it is depicted as the root cause of mortality rate [5]. Roughly 640 million adults suffer from hypertension in developing countries and reach about 1 billion adults in 2025.

Various recent researches have adopted Machine Learning (ML) models as a decision-support system as the growing awareness of hypertension and diabetes risks for earlier prediction of the disease [6]. It produces higher accuracy in predicting hypertension and diabetes based on patients' present situation and therefore helps them take necessary actions earlier. Alam et al. [7] demonstrated ensemble model is a common ML approach that merges the outcomes from various classification models and proves that the anticipated model shows superior performance compared to various standard approaches [8]. Various other approaches have successfully adopted ensemble models to assist medical-based medical decision-based support systems and predict diabetes. In the ML field, various confronting issues arise, like imbalanced datasets and outlier data that cause the reduction of models' prediction accuracy [9,10]. Some approaches demonstrate adopting an isolated forest model to avoid outlier data and using the Synthetic Minority Over-sampling Technique (SMOTE) to balance the Not-A-Number (NaN) value and imbalanced data computation performance the ML approach is improved substantially.

However, none of the previous research works has integrated the characteristics of SMOTE with ML classifier and optimizer models to enhance the prediction accuracy of the model, specifically over hypertension and Type II diabetes datasets. Thus, this research concentrates on modelling and efficient decision-support systems using ML classification and optimization for predicting hypertension and Type II diabetes using individual risk factor data [11,12]. Here, Statlog UCI ML-based heart dataset is used for predicting (hypertension and Type-II diabetes) to measure the performance of the anticipated model. Some researchers concentrate on modelling IoT-based mobile applications for the healthcare system to offer superior proficiency over disease prediction and their suitable way for validating the present health condition via the mobile applications [13,14]. But, this work does not perform any form of mobile-based analysis to reduce the preliminary complexity [15]. The reason behind the avoidance of the mobile application is to eliminate the computational complexity and improve cost-efficiency. However, future research intends to include this idea to take necessary and immediate action to diminish and eliminate the individual risk during the unexpected worsening condition of health, i.e., higher BP and triggered diabetes level.

- The proposed decision support system model includes four major parts: 1) statlog dataset acquisition; 2) pre-processing with SMOTE; 3) classification with Adaboost classifier model; and 4) Adam Optimizer to attain a global solution. These methods adapt their way to predict the disease in the earlier stage to help the individual. The significant research contributions are given below.
- Generally, SMOTE technique is applied in the ML algorithm as the newly constructed instances are not the actual copies. Therefore, it softens the decision-support boundaries and helps the algorithm for hypothesis approximation more accurately. Also, it reduces the Mean Absolute Error (MAE) and gives better prediction accuracy.

- Here, the Adaboost Classifier model is adopted as it is a best-suited boosting algorithm. The AdaBoost model's significant benefits are reducing the generalization error, more straightforward implementation of the algorithm, and a broader range of classifier adaptability with no parameter adjustment. However, the classifier model needs to be provided with massive care as the algorithm is more sensitive to data outliers.
- Finally, the Adam Optimizer (AO) is used for this research as the model works well with the sparse gradients, noisy data, significant dataset parameters, and it needs lesser memory space. This optimizer model is well-known for its computational efficiency and more straightforward implementation.
- The simulation is done in the MATLAB 2018a environment. Various performance metrics like prediction accuracy, sensitivity, specificity, and execution time are evaluated to show the model's significance and improve prediction accuracy.

The proposed decision support system model includes four major parts: 1) statlog dataset acquisition; 2) pre-processing with SMOTE; 3) classification with Adaboost classifier model; and 4) Adam Optimizer to attain a global solution. These methods adapt their way to predict the disease in the earlier stage to help the individual. Generally, SMOTE technique is applied in the ML algorithm as the newly constructed instances are not the actual copies. Therefore, it softens the decision-support boundaries and helps the algorithm for hypothesis approximation more accurately. Also, it reduces the Mean Absolute Error (MAE) and gives better prediction accuracy. Here, the Adaboost Classifier model is adopted as it is a best-suited boosting algorithm. The AdaBoost model's significant benefits are reducing the generalization error, more straightforward implementation of the algorithm, and a broader range of classifier adaptability with no parameter adjustment. However, the classifier model needs to be provided with massive care as the algorithm is more sensitive to data outliers. Finally, the Adam Optimizer (AO) is used for this research as the model works well with the sparse gradients, noisy data, significant dataset parameters, and it needs lesser memory space. This optimizer model is well-known for its computational efficiency and more straightforward implementation. The simulation is done in the MATLAB 2018a environment. Various performance metrics like prediction accuracy, sensitivity, specificity, and execution time are evaluated to show the model's significance and improve prediction accuracy.

The work is structured as [Section 2](#) summarises various existing approaches used to predict the disease and its pros and cons. [Section 3](#) gives the ideology of the anticipated model that includes dataset acquisition, SMOTE technique, Adaboost classifier and Adam optimizer for enhancing the prediction rate. [Section 4](#) discusses the performance evaluation outcomes with the analysis of various metrics to balance the data. [Section 5](#) gives the summarization of the research work with research constraints and ideas for future research improvements.

2 Related Works

Generally, medical data is composed of various relevant features and categorized based on its usefulness, i.e., less practical and not useful features (redundant features) based on its progression and formulation of diverse practice measurements. The predictions of these attributes are essential for representing the proximity of the domain appropriately [16]. The labelling process is a non-trivial task as the specification of the prediction column is not influenced by the irrelevant features. The appropriate evaluation of maximal influencing elements affects the classification process, and the dataset size is reduced by eliminating redundant features. Some specific features are chosen to extract the sub-group of the most influencing features from all these clustered features. This section discusses two sub-sections of the recent research works. The first section discusses the

feature selection techniques and discusses heart disease prediction techniques. Usually, feature selection approaches are employed to predict heart disease risk, and the prediction models on the same dataset without feature selection approaches are laid out. Various datasets are involved in the heart disease prediction process like statlog, Hungarian, and Cleveland that are accessible in the repository of the UCI ML. Some investigators consider the Switzerland heart dataset from those available repositories and include it with the aforementioned datasets.

2.1 Reviews on Feature Selection Techniques

In this review section, some notable researches are performed for predicting the disease risk using the Cleveland heart dataset from the UCI ML repository. Avci [17] adopted an improved decision tree with four diverse features and attained 76% prediction accuracy. Also, the author uses the C4.5 tree model and attained 79% accuracy with five influencing features. Bakar et al. [18] used a Fuzzy-based decision support system (FDSS) with rule-based weighted fuzzy models from adaptive GN. The author retrieved eight significant features and attained 82% accuracy. Sujatha et al. [19] established the PI-sigma model and designed a multi-layer Pi-sigma neuron model for predicting heart disease. The proposed idea is provided for the k-fold validation method where 'k' is set as 5. The author selects four diverse attributes adjacent to the hyperplane of the SVM and Linear Discriminant Analysis approach and attains 89% accuracy. However, Taser et al. [20] noticed that adopting the DM voting technique and selecting nine diverse features are considered and merged with logistic regression and the naïve Bayes method. Also, the author adopts SVM, k-NN, NN and decision tree classifiers for the same dataset model.

With Tun et al. [21] perspective, two variants of modern algorithms are employed for feature selection like uni-variate and relief model are used for key feature sorting. The author uses SVM, DT, RF, MLP, and LR algorithms without feature selection. Here, hyper-parameters and cross-validation are tuned and validated to improve disease risk prediction accuracy. The author attains roughly 89% prediction accuracy with seven diverse features achieved using the univariate and LR feature selection approach. Also, he attained 90% accuracy with LR and the same number of features with the Relief-based feature selection approach. Rubino [22] adopted ensemble-based classification and integrate manifold classifier model with classifier models like Bayesian network, NB, MLP, and RF and later uses feature selection approach to improve the prediction accuracy. The outcomes are reformatory by adopting a majority voting model with MLP, RF, BN, and NB and attain the highest accuracy of about 86%. Heo et al. [23] discusses a modern ML intelligence framework and perform mixed data factor analysis for extracting and deriving features from the heart disease dataset for training the ML prediction model. The framework attains 92% and an AUC of 92.61% for ten various features and the proposed classifier model.

2.2 Reviews on Heart Disease Prediction Techniques

Singh et al. [24] performed arrhythmia data classification using the least-square twin SVM model, structural twin SVM model, structural least-square twin SVM model, and various structural data proofs and attain 88% prediction accuracy without performing feature selection over the online available Cleveland heart disease UCI ML dataset. Domingues et al. [25] anticipated a novel CDSS employing C5.0 and RF classifier model. Using C5.0, various rules are generated, and 91% prediction accuracy is attained with this system. Alfian et al. [26] used a bi-model Fuzzy-based analytic hierarchical process to legitimize the risky elements over the dataset using statistical data and provide the Bayesian formula's fuzzy rules. The anticipated model attains 87% accuracy. Yousefian et al. [27] used a clustering-based hybrid sampling technique and requires minimal sampling time. A confidence measurement-based approach is used to evaluate the testing set and

attain 78% prediction accuracy. Harliman et al. [28] anticipated a novel hybridization approach using ANN and DT model to achieve 78% and 79% prediction accuracy. Here, a 10-fold CV is used to substantiate the outcomes, and the experimental analysis is performed using WEKA. Also, the author noticed the functionality of the hybrid DT model, which outperforms the ANN algorithm during the prediction of heart ailment risk. Goel et al. [29] evaluated a novel nested ensemble model using a support vector classification approach for heart disease prediction. The proposed model relies on the conventional ML and ensemble classification approach. Prediction accuracy of 85% is attained with the sigmoid support vectors and conventional ML approaches. Calheiros et al. [30] employed a hybrid RF with a linearity model and achieve a prediction accuracy of 89% using the ML classifier model to predict heart disease risk. The author uses GA algorithm, re-sampling and class-balancer methods are used for balancing the dataset [31]. Then, sigmoid, linear, polynomial, and radial basis kernel functions are integrated with some base algorithms. However, the model fails to work effectually in heart disease prediction Freund et al. [32] commenced χ^2 statistical model for addressing the feature refinement and problem eradication issues during the heart disease prediction model, i.e., problems of over-fitting and under-fitting. Here, a novel DNN is used for the specific search purpose and attains a [33] prediction accuracy of 92%, respectively. Atiqur et al. [34] discussed performance analysis using PCA algorithm for heart disease prediction and attains a prediction accuracy of 91.11%. The author adopts various feature extraction algorithms to perform heart failure classification. Liaquat et al. [35] discussed a linear SVM model for heart disease prediction and attains a prediction accuracy of 92.22%. The model intends to eliminate irrelevant features by reducing the coefficients to zero. Liaquat et al. [36] anticipated an automated diagnostic system for heart disease prediction using the χ^2 model. The author acquires 93.33% prediction accuracy compared to various other approaches. Ali et al. [37] gave an approach based on mutually informed NN to optimize the generalization capabilities of DSS for heart disease prediction and attain a prediction accuracy of 93.33%. Table 1 depicts the comparison of prediction accuracy of various prevailing approaches using the statlog dataset.

Table 1: Comparison of various existing approaches using statlog dataset

References	Methodology	Accuracy (%)	AUC (%)	Total number of chosen features
[27]	Enhanced ID3	78.78	–	13
[28]	FILM	88.85	–	11
[29]	Voting method	88.42	–	9
[16]	Cuckoo search with k-NN	–	77.89	13
[16]	Discrete Cuckoo search with k-NN	–	77.58	13
[9]	MLP	85.08	–	13
[17]	SVM	83.60	–	13
[2]	J48	85.82	–	13
[8]	MRMD	86.98	–	13

(Continued)

Table 1: (continued)

References	Methodology	Accuracy (%)	AUC (%)	Total number of chosen features
[3]	SS-RF	79.69	–	13
[6]	PA	87.58	–	13
[17]	F-SVM	85.89	–	4
[3]	RF-relief attribute	87.95	84.3	13
[4]	NF-LDA	86.89	–	12
[17]	NFR + SVM	87.58	91.13	13
[3]	NFR + RF	88.56	90.83	8
[34]	PCA algorithm	91.11	–	–
[35]	Linear SVM model	92.22	–	–
[36]	Statistical χ^2 model	93.33	–	–
[37]	Mutually informed NN	93.33	–	–
[38]	DNN+PCA_- GWO	97.3	–	–
[39]	F-DA	95.59		

3 Methodology

This section discusses three phases of the proposed flow: data pre-processing, classification, and optimization. Here, the SMOTE technique pre-processes the missing data and handles the imbalanced data problem encountered in the statlog dataset. Moreover, the Adaboost-based classifier model is adopted to perform disease classification and evaluate the model's outcome. This work aims to classify the class labels and optimize the values to attain better results.

3.1 In Dataset Description

Starlog dataset is a heart disease dataset like another dataset available in the UCI Machine Learning (ML) repository but shows some slight difference in its form. Table 2 depicts the statlog dataset attributes.

Table 2: Starlog dataset attributes

S. No.	Attributes	Type	Information
1	Age	N	Patients' age
2	Sex	C	Gender
3	Chest Pain	C	1 \rightarrow normal; 2 \rightarrow atypical; 3 \rightarrow non – anginal $\frac{\text{strain}}{\text{pain}}$; 4 \rightarrow asymptomatic

(Continued)

Table 2: (continued)

S. No.	Attributes	Type	Information
4	Normal/Resting BP	N	120/80
5	Serum Cholesterol (mg/dl)	N	Cholesterol level
6	F. Blood Sugar	C	1 \rightarrow <i>yes</i> ; 0 \rightarrow <i>no</i>
7	Resting ECG	C	0 \rightarrow <i>normal/regular</i> ; 1 \rightarrow <i>abnormalST – T – wave</i> ; 2 \rightarrow <i>positivehypertrophy</i> <i>LSventricular</i>
8	Max. heart rate	N	72 normal
9	Exercise-induced angina	N	1 \rightarrow <i>yes</i> ; 0 \rightarrow <i>no</i>
10	Old peak (ST depression induced by exercise to rest)	N	–
11	Slope (peak exercise ST segment)	C	1 \rightarrow <i>up</i> ; 2 \rightarrow <i>flat</i> ; 3 \rightarrow <i>down</i> ;
12	Number of significant vessels coloured by fluoroscopy	N	0 to 3
13	Thal: 3 \rightarrow <i>normal</i> ; 6 \rightarrow <i>fixeddetect</i> ; 7 \rightarrow <i>reversibledefect</i>	C	0 \rightarrow <i>absent</i> ; 1, 2, 3, 4 \rightarrow <i>present</i>

Note: Here, ‘N’ is numeric, and ‘C’ is categorical.

3.2 SMOTE Technique

The provided statlog dataset comprises various positive and negative classes where positive is specified as 1, and negative class is defined as 0. However, there is an uneven distribution of classes over the dataset, and these uneven distributions are depicted as the foremost cause of the reducing prediction accuracy in the classifier model. The significant cause is that most ML algorithms do not drastically learn about the pattern for all the positive and negative classes due to the imbalanced dataset. However, negative classes are considered as the minority class with a smaller amount of instances. Therefore, the outcomes generated using the proposed classifier model for a class definition is often more inefficient. Various existing approaches do not consider this minority class as the major flaw in attaining classification outcomes. The key contribution of this pre-processing step is to deal with the imbalanced instances of the provided statlog dataset and handle this issue efficiently using SMOTE approach. Moreover, the outcomes for the minority and majority classes are separately documented to compute the class performance in generating overall classification outcomes.

The author describes that SMOTE is a prominent approach adopted for the classifier construction with the imbalance dataset. This imbalanced dataset is composed of uneven distribution of underlying output classes. It is most commonly used in classification issues of the imbalanced

dataset. The handling of uneven instances is a pre-processing technique and a reliable approach. With the origin of the SMOTE approach, there are diverse SMOTE variants deployed and proposed to improve the general SMOTE approach with efficient adaptability and reliability under various situations. It is depicted as the most innovative pre-processing approach in information mining and the ML domain. This approach aims to adopt interpolation with data of minor class instances, and therefore the numbers are improved. It is essential to attain classifier generalization and most extensively adopted to cater to the problems raised due to the imbalanced classification instances over the statlog dataset. Generally, the minority classes are oversamples with the production of various artificial models. The minority class-based feature space is adopted for the construction of these samples. Based on the classification and sampling requirements, the number of neighbours is selected. Lines are constructed with the minority class data points utilizing the neighbours and most efficiently used while handling the imbalanced datasets. It pretends to equalize the minority and majority class instances during the sample training process. A novel imblearn library is utilized for the SMOTE implementation to handle the imbalanced dataset.

Fig. 1 depicts the generic flow of the experimentation where the provided statlog dataset is processed to eradicate the null values. However, SMOTE approaches are provided to the statlog dataset in this research. Therefore, an equal amount of negative and positive samples are attained. However, the final phase provides diverse ML approaches over the given dataset, achieving the corresponding outcomes. Generally, k-fold CV is applied to the dataset to fulfil the reliability of the results. Fig. 1 depicts the workflow of the anticipated experiment.

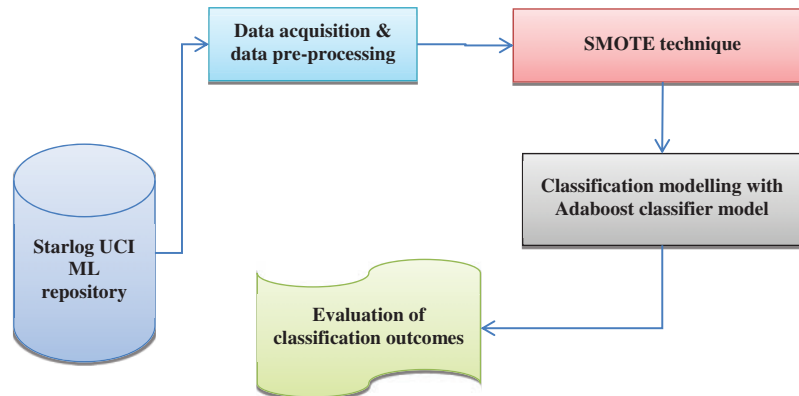


Figure 1: Workflow of the proposed experiment

3.3 Adaboost Classifier (AB)

Adaboosting helps in boosting the performance of the weak learner and improves the learning performance. Adaboosting is connected with the bagging and bootstrap model, and it is conceptually diverse from others. Generally, bagging is aggregation and bootstrapping, while bootstrap performs a specific sample-based statistical approach where the samples are randomly drawn with replacement. Here, three diverse methods are constructed with random sampling, and others are sampling with replacements. The general factors between these two approaches are depicted as a practical voting approach and the classifier model is more popular with the most satisfactory performance. It is experimentally proven that these model gains superior performance and are determined as the best classifier model. The classification process of this classifier is based on

the combination of learning outputs of the provided weak classifiers, and it is mathematically expressed as in Eq. (1):

$$H(x) = \text{sgn} \left(\sum_{n=1}^m \alpha_n H_n(x) \right) \quad (1)$$

Here, $H(x)$ is the output of the weak classifier, α_n is classifier weight. The preliminary idea behind this algorithm is to sequentially train all the classifiers and update weight vectors based on training set pattern accuracy among the present and previous classifiers. The weighted vectors are decreased or increased based on the decreasing or improving misclassified patterns among the iterations. The weight quality in increment and decrement $\Delta(x)$ is based on the error percentage among the present and previous patterns. This research adopts modest Adaptive boosting with lower generalization error compared to other forms of adaptive boosting algorithms. The algorithm for this classifier model is shown in Algorithm 1:

Algorithm 1: Modest Adaptive Boosting Algorithm

Input: $Z = \{z_1, z_2, \dots, z_N\}$ with $z_k = (x_k, y_k)$ where $k = 1, 2, \dots, N$ as training set, $M = \text{no.of classifiers}$.

Output: prediction accuracy based on $H(x)$

1. Initialize weighted values as $w_i \rightarrow \frac{1}{N}$ for all $i \in \{1, 2, \dots, N\}$ where w_i is determined as weighted value.

2. For all $m = 1 \rightarrow M; H_m \neq 0$;

3. **begin** the process

4. Train the classifiers $H_m(x)$ using w_i ;

5. Perform inverse distribution using Eq. (2):

$$\bar{w}_i = (1 - w_i); \text{normalize } \sum_i \bar{w}_i = 1; \quad (2)$$

6. Compute probability values using Eqs. (3)–(6):

$$P_m^{+1} = P_w(y = +1, H(x)); \quad (3)$$

$$\bar{P}_m^{+1} = P_w(y = +1, H(x)); \quad (4)$$

$$P_m^{-1} = P_w(y = -1, H(x)); \quad (5)$$

$$\bar{P}_m^{-1} = P_w(y = -1, H(x)); \quad (6)$$

7. Update classifier $H_m(x) = (P_m^{+1}(1 - P_m^{+1}) - P_m^{-1}(1 - P_m^{-1}))$;

8. Update the weighted value $w_i = w_i \exp(-y_i H_m(x_i))$ and normalize $\sum_i w_i = 1$;

9. **End** process;

10. Evaluate the classifier output using Eq. (7):

$$H(x) = \left(\text{sgn} \left(\sum_{m=1}^M H_m(x) \right) = \begin{cases} +1 & \text{for } \sum_{m=1}^M H_m(x) \geq 0 \\ -1 & \text{for } \sum_{m=1}^M H_m(x) < 0 \end{cases} \right) \quad (7)$$

3.4 Adam Optimizer (AO)

Adam Optimizer is used to handle the noisy problem and used to optimize the classifier model. Adam optimizer is known as a replacement optimization algorithm for training deep learning models. It merges the best properties of RMSprop and AdaGrad algorithms to enhance the optimization algorithm that handles sparse gradients over noisy problems. The main objective of integrating Adam Optimizer with AdaBoost is that Adaboost has some drawbacks, i.e., vulnerable to a constant noise (empirical evidence). Some weak classifiers turn to be weaker because of low margins and over-fitting issues. Thus, it helps to boost the weak classifier performance and works effectually over the large dataset and is computationally efficient. Adam is a well-known algorithm used in the Deep Learning (DL) field as it acquires better outcomes faster. Adam optimizer outcomes work well in real-time applications, and also it is one of the finest techniques among other stochastic optimization techniques. Adam optimization is an alternative to commonly used Stochastic Gradient Descent (SGD) for iteratively updating NN weights based on training data. The optimizer functionality of stochastic Gradient is different from conventional optimizer because it applies the learning rate of alpha for the overall weight updation. This updation is performed during the learning process. The learning rate is preserved for all network weights and adapted during the learning process separately. It integrates the benefit of two stochastic gradient descents as:

Adaptive Gradient: Learning rate is preserved for all attribute values with improved gradient descent performance.

Root Mean Square (RMS) propagation: The learning rate is maintained for all attributes; however, it is based on the average magnitudes of gradients for weight. The algorithm for Adam Optimizer is clearly described in Algorithm 2.

Algorithm 2: Adam Optimization Algorithm

Input: $\alpha, \beta_1 = 0.9, \beta_2 = 0.99$ are learning rate, hyperparameters, epochs, number of iterations, s = exponential average of squares of gradients, g - gradients at 'T' time, time step 'to = 0, ρ_1 and ρ_2 are exponential decay rates for moment 's' and 'r' respectively, θ is a parameter related to first and second moments.

Output: prediction accuracy

1. Initialize average weighted gradients, the weighted average square of past gradients, bias correction, bias, input multiplied with weights of neurons

2. Initialize first moment and second-moment vectors

3. Initialize timestep

4. After bias correction

$s \leftarrow \rho_1 s + (1 - \rho_1)g$ //first moment

$r \leftarrow \rho_2 r + (1 - \rho_2)g \odot g$ //second moment

5. Update parameters like learning rate and weighted average of gradients

$\hat{s} \leftarrow \frac{s}{1 - \rho_1}$ //bias correction second-order estimate

$\hat{r} \leftarrow \frac{r}{1 - \rho_2}$ //bias correction second-order estimate

$\Delta\theta = -\frac{\hat{s}}{\delta + \sqrt{\hat{r}}} \epsilon$ //operations applies element-wise $\epsilon = 10^{-8}$

6. Repeat Step 5 to Step 10 until no errors

$t \rightarrow$ no. of iterations

7. Specify epoch with no. of iterations

The optimized Adam output is expressed by Eq. (8):

$$\text{Output} = x.w + b \quad (8)$$

where 'x'-input parameters, i.e., input parameters needed for predicting brain tumour (tumour or non-tumour class). 'w'-weights that are varied for iterations. To improve the accuracy, 'b'-bias is added to nodes of NN where the learning rate is unique for all weights. Therefore, bias and weights are updated with the use of AO. It is utilized for updating the bias and weights of the network model used. Updated values are multiplied with given input parameters and classifier functions like ReLU, used for performing classification. The bias and weights are iteratively updated until the model acquires better classification accuracy.

4 Result and Discussion

The experimentation is done over the provided statlog dataset, and the associated outcomes are attained. The simulation is done in MATLAB 2018b environment and intends to free the results from biases. Classification needs to be achieved efficiently by balancing the instances with SMOTE technique. Pre-processing is performed over the raw data and later fed into the ML approaches, and the outcomes are evaluated and compared with various existing approaches. The major constraint with the evaluation of the ML approach is that the dataset needs to validate the same evaluation metrics like prediction accuracy, which leads to misleading outcomes due to the uneven distribution of the samples. The system weakness is measured with the uneven sample distribution. The predicted values are not measured as the true value with the missing samples. Most of the general approaches concentrate on the UCI ML dataset for accuracy evaluation. Standard evaluation metrics like accuracy, specificity, sensitivity, AUROC, F1-score and execution time compute the anticipated model. Generally, accuracy is measured as the ratio of proper predictions to the total amount of inputs. The confusion matrices are modelled with the evaluation of True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN), respectively.

Similarly, specificity and sensitivity are computed with $FP/(FP + TN)$ and $TP/(FN + TP)$, respectively. AUROC is another metric used for the evaluation of the classification accuracy of the provided model. Some special constraints are given for the evaluation of minor dataset samples. It is performed with the validation of accuracy measures. The experimentation is done with the Adaboost classifier model and SMOTE technique, outperforming the existing approaches. The methods include Linear Regression (LR), Random Forest (RF), Lightweight Gradient Boosting Machine approach (L-GBM), Support Vector Machine (SVM), Boosted Regression Tree and ensemble model, respectively. The evaluation is done with other datasets like Cleveland and the UCI Machine Learning repository (see Figs. 2–12). The accuracy of the AB-AO model is 90.15% without feature selection which is 3.9%, 9.17%, 5.43%, 9.17%, 10.28%, and 2.59% higher than LR, RF, L-GBM, SVM, BRT, and an ensemble model. The prediction accuracy after feature selection is 95.87% which is 3.33%, 8%, 88.98%, 86.76%, 8%, and 2.73% higher than other models. The sensitivity is 91% which is 17%, 14%, 20%, 17%, 6%, and 1% higher than other models. The specificity of AB-AO is 94% which is 24%, 26%, 23%, 22%, 5%, and 2% higher than other approaches. The F1-score is 20% which is 6%, 7%, 6%, 6%, 5% and 4% higher than other models. The Cleveland dataset consists of 13 features where LR selects 10 features, RF selects 9 features, L-GBM selects 11 features, SVM selects 7 features, BRT selects 12 features, and an ensemble model selects 8 features. The execution time (min) of all these models are 0.10, 0.10, 0.78, 6.80, 1.100, 0.06 and 0.04 for Linear Regression (LR), Random Forest (RF), Lightweight Gradient Boosting Machine approach (L-GBM), Support Vector Machine (SVM), Boosted Regression Tree,

ensemble model and AB-AO model, respectively. The AUROC of AB-AO is 96.78% which is 3.92%, 4.42%, 2.92%, 5.38%, 4.46% and 0.89% higher than other approaches.

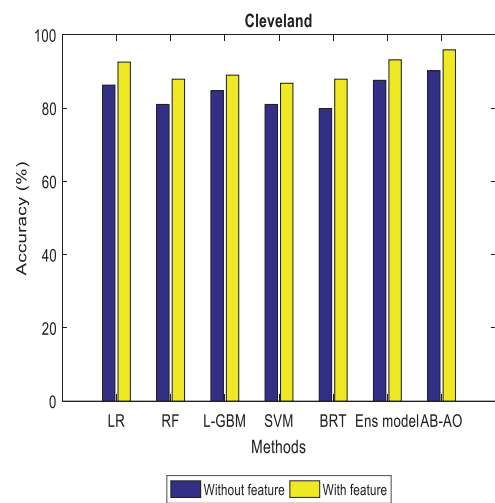


Figure 2: Accuracy computation in cleveland dataset

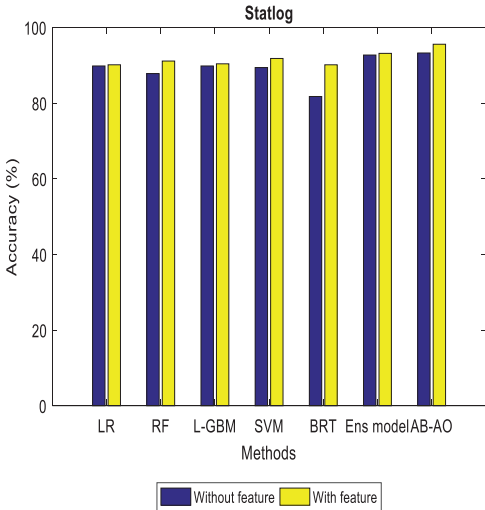


Figure 3: Accuracy computation in statlog dataset

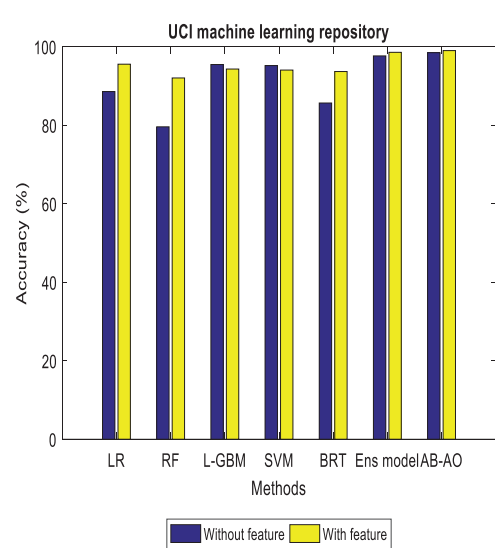


Figure 4: Accuracy computation in UCI ML dataset

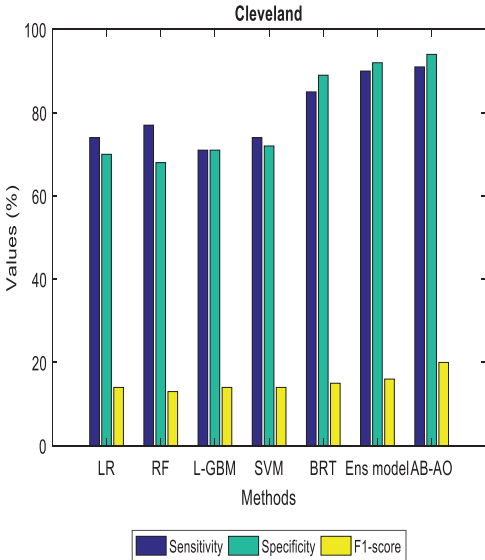


Figure 5: Sensitivity, specificity and F1-score computation in cleveland dataset

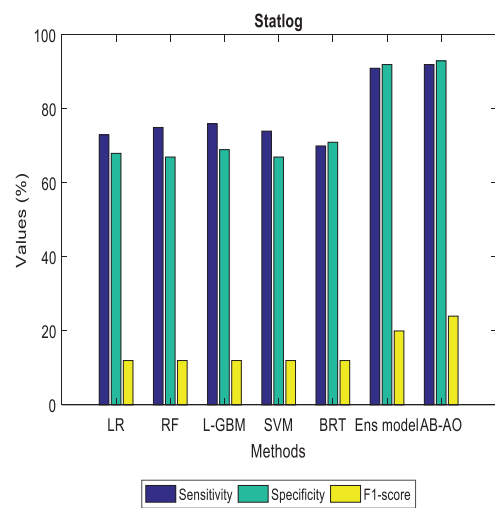


Figure 6: Sensitivity, specificity and F1-score computation in statlog dataset

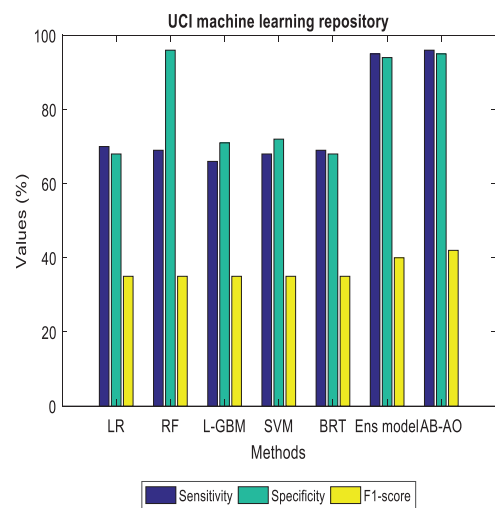


Figure 7: Sensitivity, specificity and F1-score analysis in UCI ML dataset

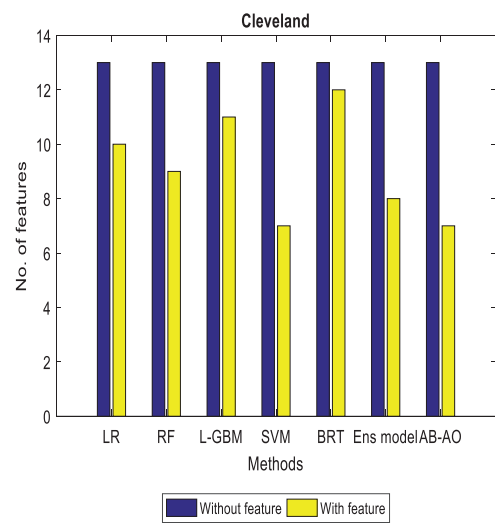


Figure 8: Feature specification in cleveland dataset

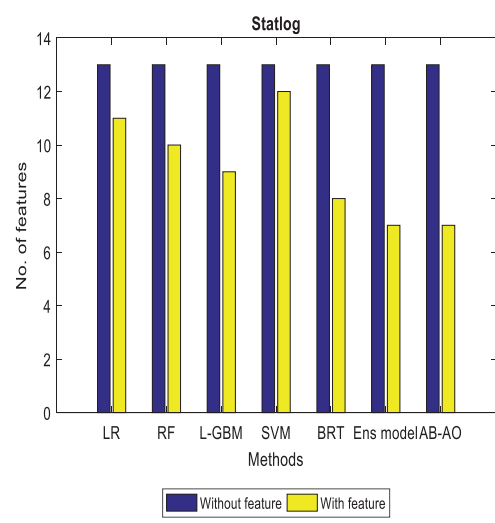


Figure 9: Feature specification in catalogue dataset

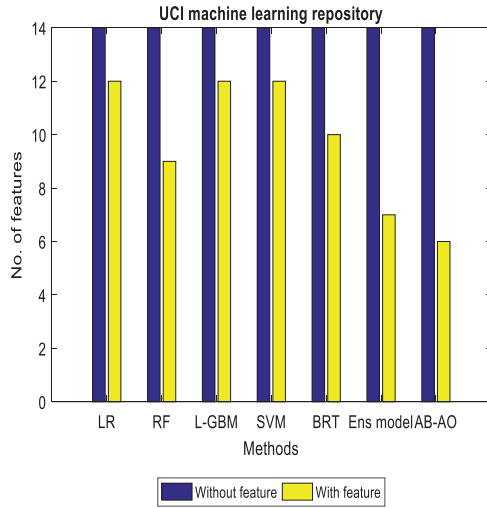


Figure 10: Feature specification in UCI ML dataset

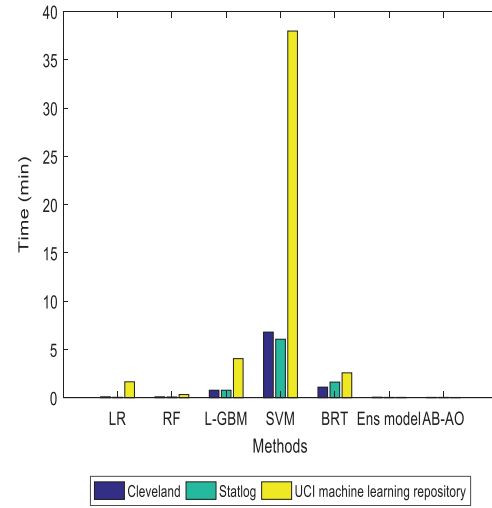


Figure 11: Execution time (min) computation

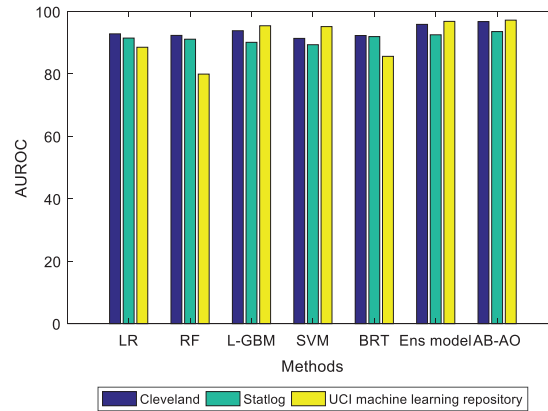


Figure 12: AUROC computation in all the three datasets

The performance metrics of the anticipated AB-AO model w.r.t. statlog dataset is evaluated. The accuracy without feature selection is 93.25% which is 3.43%, 5.45%, 3.43%, 3.86%, 11.5% and 0.54% higher than other approaches. The prediction accuracy of the AB-AO model with feature selection is 95.56% which is 5.43%, 4.44%, 5.17%, 3.76%, 5.43%, and 2.41% higher than other models. The sensitivity of AB-AO is 91% which is 18%, 16%, 15%, 17%, 21% and 1% higher than other models. The specificity of AB-AO is 92% which is 24%, 25%, 23%, 25%, 21%, and 1% higher than other models. The F1-score of AB-AO is 24% which is 12% higher than Linear Regression (LR), Random Forest (RF), Lightweight Gradient Boosting Machine approach (L-GBM), Support Vector Machine (SVM) and Boosted Regression Tree model and 4% lower than ensemble model. Here, the model considers 13 features where LR selects 11 features, RF chooses 10 features, L-GBM selects 9 features, SVM selects 12 features, BRT selects 8 features, ensemble and AB-AO selects least 7 features. The execution times of all these models are 0.05 min, 0.08 min, 0.78 min, 6.06 min, 1.62 min, and 0.04 min, respectively. The AUROC of AB-AO is 93.58% which is 2.08%, 2.44%, 3.44%, 4.2%, 1.61%, and 1.02%.

Now, the evaluation is done for the AB-AO model with all three datasets. The prediction accuracy without feature selection of AB-AO with statlog dataset is 93.25% which is 3.1% higher than Cleveland and 5.22% lower than UCI dataset. The prediction accuracy with feature selection is 95.56% which is 0.91% and 3.44% lower than another dataset. The sensitivity is 92% which is 2% higher for Cleveland and 4% lower than the UCI dataset. The specificity of AB-AO with statlog and Cleveland dataset is 92% which is 3% lower than the UCI dataset. The F1-score of statlog is 24% which is 4% higher than Cleveland and 18% lower than the UCI dataset. The Cleveland dataset and statlog dataset considers 13 features, while UCI considers 14 features. But, the average feature selection ratio of Cleveland is 8, statlog and UCI is 7 features. The average execution time for the Cleveland dataset and statlog dataset validation is 0.04 min, while for the UCI dataset, the execution time is 0.03 min, respectively. The AUROC value of statlog is 93.58% which is 3.2% and 3.67% lesser than other models. Table 3 depicts the comparison of various performance metrics.

Table 3: Comparison of various performance metrics

Dataset	ML algorithm	Accuracy (%)		Sensitivity	Specificity	F1-score	No. of features		Time (min)	AUROC
		Without feature	With feature				Without feature	With feature		
Cleveland	LR	86.25	92.54	0.74	0.70	0.14	13	10	0.10	92.86
	RF	80.98	87.87	0.77	0.68	0.13	13	9	0.10	92.36
	L-GBM	84.72	88.98	0.71	0.71	0.14	13	11	0.78	93.86
	SVM	80.98	86.76	0.74	0.72	0.14	13	7	6.80	91.40
	BRT	79.87	87.87	0.85	0.89	0.15	13	12	1.100	92.32
	Ensemble Model	87.56	93.14	0.90	0.92	0.16	13	8	0.06	95.89
	AB-AO	90.15	95.87	0.91	0.94	0.20	13	7	0.04	96.78
Catalogue	LR	89.82	90.13	0.73	0.68	0.12	13	11	0.05	91.50
	RF	87.80	91.12	0.75	0.67	0.12	13	10	0.08	91.14
	L-GBM	89.82	90.39	0.76	0.69	0.12	13	9	0.78	90.14
	SVM	89.39	91.80	0.74	0.67	0.12	13	12	6.06	89.38
	BRT	81.75	90.13	0.70	0.71	0.12	13	8	1.62	91.97
	Ensemble model	92.71	93.15	0.91	0.92	0.20	13	7	0.04	92.56
	AB-AO	93.25	95.56	0.92	0.93	0.24	13	7	0.04	93.58
UCI machine learning repository	LR	88.56	95.54	0.70	0.68	0.35	14	12	1.65	88.56
	RF	79.59	92.03	0.69	0.96	0.35	14	9	0.33	79.96
	L-GBM	95.45	94.3	0.66	0.71	0.35	14	12	4.05	95.45
	SVM	95.18	94.04	0.68	0.72	0.35	14	12	37.98	95.18
	BRT	85.65	93.67	0.69	0.68	0.35	14	10	2.58	85.65
	Ensemble model	97.65	98.56	0.95	0.94	0.40	14	7	0.04	96.85
	AB-AO	98.47	99	0.96	0.95	0.42	14	6	0.03	97.25

The performance metrics of the anticipated AB-AO model w.r.t. UCI ML dataset is evaluated. The accuracy without feature selection is 98.47% which is 9.91%, 18.88%, 3.02%, 3.29%, 12.82% and 0.82% higher than other approaches. The prediction accuracy of the AB-AO model with feature selection is 99% which is 3.46%, 6.97%, 4.7%, 4.96%, 5.33%, and 0.44% higher than other

models. The sensitivity of AB-AO is 96% which is 26%, 17%, 30%, 28%, 17% and 1% higher than other models. The specificity of AB-AO is 95% which is 27%, 1% lower than RF, 24%, 25%, 27%, and 1% higher than other models. The F1-score of AB-AO is 42% which is 7% higher than Linear Regression (LR), Random Forest (RF), Lightweight Gradient Boosting Machine approach (L-GBM), Support Vector Machine (SVM) and Boosted Regression Tree model and 2% higher than ensemble model. Here, the model considers 14 features where LR selects 12 features, RF chooses 9 features, L-GBM selects 12 features, SVM selects 12 features, BRT selects 10 features, ensemble selects 7, and AB-AO selects the least 6 features. The execution times of all these models are 1.65 min, 0.33 min, 4.05 min, 37.98 min, 2.58 min, 0.04 min and 0.03 min, respectively. The AUROC of AB-AO is 97.25% which is 8.69%, 17.29%, 1.8%, 2.07%, 11.6%, and 0.4%. The performance of the AB-AO over the statlog dataset is optimal as it eradicates uneven distribution of samples using SMOTE technique. However, the Cleveland and UCI ML repository approaches do not adopt SMOTE or other methods to handle these issues. However, it is noted that the performance metrics can be misleading results compared to the proposed model. Thus, it is proven that the anticipated model outperforms the existing approaches efficiently by adopting essential pre-processing steps.

5 Conclusion

This research work has provided a clinical decision support system for predicting heart diseases by evaluating the provided statlog UCI ML dataset. The anticipated Adaboost classifier and Adam optimizer work as a superior prediction approach to classify the imbalance dataset efficiently with the well-known SMOTE technique. Moreover, the execution time is considered a remarkable metrics with lesser computation time to show the reliability of the anticipated model. The classifier's performance entirely relies on parameter tuning with a '1' learning rate and a '30' estimator. The prediction accuracy of the AdaBoost classifier is 90.15% and 95.87% for without and with feature selection. The model's sensitivity is 91%, specificity is 94%, F1-score is 0.20, time for execution is 0.04 seconds, and 96.78% AUROC. The major constraint with this research work is the lesser number of available samples (statlog) for the evaluation process. However, the prediction accuracy achieved by this model is 95.87% and acts as a superior CDSS during the time of emergency. In the future, deep learning approaches are adopted to enhance the heart disease prediction rate, and statistical analysis with *p*-value computation is highly solicited to acquire higher validation accuracy. The statistical analysis will be given more importance in future research works.

Funding Statement: The author received no specific funding for this study.

Conflicts of Interest: The author declares that they have no conflicts of interest to report regarding the present study.

References

1. Groves, P., Kayyali, B., Knott, D., van Kuiken, S. et al. (2016). The big data revolution in healthcare: Accelerating value and innovation. *Science and Education*, 1(1), 1–22. DOI 10.12691/ajhr-4-1-1.
2. Paul, A. K., Shill, P. C., Rabin, M. R. I., Akhand, M. A. H. (2016). Genetic algorithm-based fuzzy decision support system for the diagnosis of heart disease. *International Conference on Informatics, Electronics and Vision*, pp. 145–150, Japan. DOI 10.1109/ICIEV.2016.7759984

3. Amin, M. S., Chiam, Y. K., Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Telematics Information*, 36(1), 82–93. DOI 10.1016/j.tele.2018.11.007.
4. Latha, C. B. C., Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked*, 16, 100203. DOI 10.1016/j.imu.2019.100203.
5. Shah, S. M. S., Batool, S., Khan, I., Ashraf, M. U., Abbas, S. H. et al. (2017). Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis. *Physica A: Statistical*, 482(1), 796–807. DOI 10.1016/j.physa.2017.04.113.
6. Kale, A. P., Sonavane, S. (2018). PF-FELM: A robust PCA feature selection for fuzzy extreme learning machine. *IEEE*, 12(6), 1303–1312. DOI 10.1109/JSTSP.2018.2873988.
7. Alam, M. Z., Rahman, M. S. (2019). A random forest-based predictor for medical data classification using feature ranking. *Informatics in Medicine Unlocked*, 15(5), 100180. DOI 10.1016/j.imu.2019.100180.
8. Shilaskar, S., Ghatol, A. (2018). Diagnosis system for imbalanced multiminority medical dataset. *Soft Computing*, 23(13), 4789–4799. DOI 10.1007/s00500-018-3133-x.
9. Mohan, S., Thirumalai, C., Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 7, 81542–81554. DOI 10.1109/ACCESS.2019.2923707.
10. Reddy, G. T., Reddy, M. P. K., Lakshmana, K., Rajput, D. S., Kaluri, R. et al. (2020). Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evolutionary Intelligence*, 13(2), 185–196. DOI 10.1007/s12065-019-00327-1.
11. Jensen, P. B., Jensen, L. J., Brunak, S. (2012). Mining electronic health records: Towards better research applications and clinical care. *Nature Reviews Genetics*, 13(6), 395–405. DOI 10.1038/nrg3208.
12. Li, Y., Bai, C., Reddy, C. K. (2016). A distributed ensemble approach for mining healthcare data under privacy constraints. *Information Sciences*, 330(2), 245–259. DOI 10.1016/j.ins.2015.10.011.
13. Khedr, A., Aghbari, Z. A. L., Kamel, I. (2018). Privacy-preserving decomposable mining association rules on distributed data. *International Journal of Engineering and Technology*, 7(3), 157–162. DOI 10.14419/ijet.v7i3.13.16343.
14. Nahar, N., Ara, F. (2018). Liver disease prediction by using different decision tree techniques. *International Journal of Data Mining Knowledge Management Process*, 8(2), 1–9. DOI 10.5121/ijdkp.2018.8201.
15. Hsu, C. H., Manogaran, G., Panchatcharam, P., Vivekanandan, S. (2018). A new approach for prediction of lung carcinoma using backpropagation neural network with decision tree classifiers. *IEEE 8th International Symposium on Cloud and Service Computing*, pp. 111–115. France. DOI 10.1109/SC2.2018.00023
16. Ramasamy, S., Nirmala, K. (2020). Disease prediction in data mining using association rule mining and keyword-based clustering algorithms. *International Journal of Computers and Applications*, 42(1), 1–8. DOI 10.1080/1206212X.2017.1396415.
17. Avci, E. (2009). A new intelligent diagnosis system for heart valve diseases by using the genetic-SVM classifier. *Expert Systems with Applications: An International Journal*, 36(7), 10618–10626. DOI 10.1016/j.eswa.2009.02.053.
18. Bakar, A. A., Kefli, Z., Abdullah, S., Sahani, M. (2011). Predictive models for dengue outbreaks using multiple rule-based classifiers. *International Conference on Electrical Engineering and Informatics*, pp. 1–6. Indonesia. DOI 10.1109/ICEEI.2011.6021830
19. Sujatha, R., Nithya, A. (2016). A survey of health care prediction using data mining. *International Journal of Innovative Research in Science, Engineering and Technology*, 5(8), 14538–14543. DOI 10.15680/IJIRSET.2016.0508032.
20. Taşer, P. Y., Birant, K. U., Birant, D. (2020). Multitask-based association rule mining. *Turkish Journal of Electrical Engineering and Computer Science*, 28(2), 933–955. DOI 10.3906/elk-1905-88.
21. Tun, N. N., Arunagirinathan, G., Munshi, S. K., Pappachan, J. M. (2017). Diabetes mellitus and stroke: A clinical update. *World Journal of Diabetes*, 8(6), 235–248. DOI 10.4239/wjd.v8.i6.235.
22. Rubino, F. (2008). Is type 2 diabetes an operable intestinal disease? A provocative yet reasonable hypothesis. *Diabetes Care*, 31(Supplement 2), S290–S296. DOI 10.2337/dc08-s271.

23. Heo, B. M., Ryu, K. H. (2018). Prediction of prehypertension and hypertension based on anthropometry, blood parameters, and spirometry. *International Journal of Environmental Research and Public Health*, 15(11), 2571. DOI 10.3390/ijerph15112571.
24. Singh, N., Singh, P., Bhagat, D. (2019). A rule extraction approach from support vector machines for diagnosing hypertension among people with diabetes. *Expert Systems with Applications*, 130(7), 188–205. DOI 10.1016/j.eswa.2019.04.029.
25. Domingues, R., Filippone, M., Michiardi, P., Zouaoui, J. (2018). A comparative evaluation of outlier detection algorithms: Experiments and analyses. *Pattern Recognition*, 74(1), 406–421. DOI 10.1016/j.patcog.2017.09.037.
26. Alfian, G., Syafrudin, M., Yoon, B., Rhee, J. (2019). False-positive RFID detection using classification models. *Applied Sciences*, 9(6), 1154. DOI org/10.3390/app9061154.
27. Yousefian-Jazi, A., Ryu, J. H., Yoon, S., Liu, J. J. (2014). Decision support in a machine vision system for monitoring of TFT-LCD glass substrates manufacturing. *Journal of Process Control*, 24(6), 1015–1023. DOI 10.1016/j.jprocont.2013.12.009.
28. Harliman, R., Uchida, K. (2018). Data- and algorithm-hybrid approach for imbalanced data problems in the deep neural network. *International Journal of Machine Learning and Computing*, 8(3), 208–213. DOI 10.18178/ijmlc.2018.8.3.689.
29. Goel, G., Maguire, L., Li, Y., McLoone, S. (2013). Evaluation of sampling methods for learning from imbalanced data. *International Conference on Intelligent Computing*, 7955, pp. 392–401. China. DOI 10.1007/978-3-642-39479-9_47.
30. Calheiros, R. N., Ramamohanarao, K., Buyya, R., Leckie, C., Versteeg, S. (2017). On the effectiveness of isolation-based anomaly detection in cloud data centres. *Concurrency Computation Practice and Experiences*, 29(18), e4169. DOI 10.1002/cpe.4169.
31. UCI Machine Learning Repository: Statlog (Heart) Data Set (2020). [https://archive.ics.uci.edu/ml/datasets/statlog+\(heart\)](https://archive.ics.uci.edu/ml/datasets/statlog+(heart)).
32. Freund, Y., Schapire, R. (1996). Experiments with a new boosting algorithm. *Proceedings of the Thirteenth International Conference*, pp. 148–156. France. DOI 10.5555/3091696.3091715
33. Freund, Y., Schapire, R. (1996). Game theory online prediction and boosting. *Proceedings of the Ninth Annual Conference on Computer Learning Theory*, pp. 325–332. Italy. DOI 10.1145/238061.238163.
34. Atiqur, R., Aurangzeb, K. (2020). Performance analysis of PCA, sparse PCA, kernel PCA and incremental PCA algorithm for heart failure prediction. *International Conference on Electrical, Communication and Computer Engineering*, pp. 1–15. Malaysia. DOI 10.1109/icecce49384.2020.9179199
35. Liaqat, A., Awais, N., Javed, A. K., Noorbakhsh, A. G., Xiong, X. (2019). An optimized stacked support vector machines based expert system for the effective prediction of heart failure. *IEEE Access*, 7, 54007–54014. DOI 10.1109/ACCESS.2019.2909969.
36. Liaqat, A., Atiqur, R., Aurangzeb, K., Mingyi, Z., Ashir, J. et al. (2019). An automated diagnostic system for heart disease prediction based on χ^2 statistical model and optimally configured deep neural network. *IEEE Access*, 7, 34938–34945. DOI 10.1109/ACCESS.2019.2904800.
37. Ali, L., Bukhari, S. A. C. (2020). An approach based on mutually informed neural networks to optimize the generalization capabilities of decision support systems developed for heart failure prediction. *Innovation and Research in Biomedical Engineering*, 42(5), 345–352. DOI 10.1016/j.irbm.2020.04.003.
38. Thippa, R. G., Neelu, K., Sweta, B., Saurabh, S., Praveen Kumar, R. M. et al. (2020). Deep neural networks to predict diabetic retinopathy. *Journal of Ambient Intelligence and Humanized Computing*, 57(13), 1–12. DOI 10.1007/s12652-020-01963-7.
39. Srinivas, K., Praveen Kumar, R., Gautam, S. (2020). Deep learning disease prediction model for use with intelligent robots. *Computers and Electrical Engineering*, 87(2), 1–13. DOI 10.1016/j.compeleceng.2020.106765.