An Efficient Method for Heart Disease Prediction Using Hybrid Classifier Model in Machine Learning

Karthikeyan G[1], Komarasamy G[2], Daniel Madan Raja S[3]

[1]*Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore*
[2]*Department of Computer Science and Engineering, GITAM School of Technology,*
GITAM Deemed to Be University, Bengaluru
[3]*Department of Information Technology, Bannari Amman Institute of Technology, Sathyamangalam*

Heart disease is the foremost and leading cause of major death rate all over the world today. Cardiovascular disease prediction is considered to be an extremely challenging factor in the clinical decision support system (CDSS). Recently, Machine learning (ML) has shown its significance effectively for making predictions and assistance for a huge amount of data generated by the healthcare industries. Similarly, ML approaches have been adopted in various research developments in the E-healthcare monitoring field. Many existing investigations have given a glimpse for disease prediction using ML approaches. Here, a novel method is known as the Hybrid Linear stacking model for feature selection and Xgboost algorithm for heart disease classification (HLS-Xgboost). This model initially predicts the significant or most influencing features of the disease and is classified using the Xgboost algorithm for enhancing prediction accuracy. This model is initiated with various feature subset combinations and diverse classification approaches. This model produces an improved performance level with 96% accuracy using the HLS-Xgboost model. The proposed model gives a better trade-off with any prediction accuracy of 96% when compared to other approaches.

Key words – cardio-vascular disease, clinical decision support system, machine learning, e-healthcare, stacking model, Xgboost classifier

1. Introduction

Recently, it is extremely complex to predict the occurrence of various contributory risk factors like abnormal pulse rate, diabetes, high cholesterol, high blood pressure, and other factors [1]. Various approaches in neural networks and data mining have been adopted to predict the severity of cardiovascular disease among the human community. The severity is classified based on diverse approaches such as Decision Tree (DT), k-Nearest Neighbour (k-NN), Naive Bayes (NB), and Genetic Algorithm (GA) [2]. Disease severity is extremely intricate and therefore needs to be treated more cautiously. The disease does not show any symptoms but leads to sudden death for sometimes [3]. The wider research perspectives of data mining and medical science are adopted for predicting diverse metabolic syndromes [4]. Mining with disease classification plays an essential role in heart disease prediction and data analysis.

It is observed that DT is also widely used for computing disease prediction accuracy for the events associated with heart disease [5]. There are diverse approaches that are provided for knowledge abstraction with well-known approaches of DM for heart disease prediction. In this investigation, various observations have been performed to generate the prediction model with distinct approaches; however, related to two or more approaches. These integrated methods are generally known as hybrid methods. In some cases, neural networks are considered (NN) for measuring heart rate time series [6]. This approach is widely used by various clinical records for measuring factors like second-degree block, normal sinus rhythm, atrial flutter, right and left bundle branch block, atrial fibrillation, premature ventricular contraction, and sinus bradycardia to predict the present condition of patients which is related to heart disease [7]. The benchmark dataset is adopted to perform classification with a training rate of 70% and a testing rate of 30% which is used for disease classification.

Various researchers have concentrated on Computer-Aided Decision Support System (CADSS) in both the research and medical field. In various prevailing work, the data used by the DM approaches over healthcare industries have been assumed to take lesser time for disease prediction with more appropriate outcomes [8]. Some existing methods concentrate on heart disease prediction with the Genetic Algorithm (GA). GA adopts effectual association rules for tournament selection, mutation, and crossover outcomes in a newer form of the fitness function. By experimental validation, some benchmark and well-known datasets are gathered from the UCI machine learning repository. This approach gives some prominent outcomes and proves to be efficient when compared to other supervised learning approaches. Similarly, some powerful algorithms known as Particle Swarm Optimization (PSO) are also studied to generate rules for disease prediction [8]. The generated rules are adopted randomly with encoding approaches which outcomes in overall accuracy enhancements. The prediction of heart disease relies on some necessary symptoms like age, pulse rate, sex, and many other symptoms. Generally, ML approaches with NN were adopted which outcomes in more appropriate and reliable results.

Generally, NN is considered one of the most reliable tools for predicting diseases like heart and brain. The anticipated model considers 13 different attributes for predicting heart disease. The outcomes demonstrate that the enhancement level of model performance is compared with prevailing approaches to show better results [9]. The artery stenting is also considered a more appropriate treatment process in the medical field. It model determines the occurrence of the adverse consequences of cardiovascular events during disease prediction for the elder people. The evaluation process is considered to be extremely significant. The results are evaluated using Artificial Neural Networks (ANN) which gives superior performance over heart disease prediction. The initiation of NN with the combination of some posterior probabilities also predicts the multiple predecessor approaches. This model has attained an accuracy of 89% which gives stronger outcomes when compared to prevailing works. For performing the experimentation, the standard Cleveland dataset is utilized to enhance heart disease performance as shown in [9].

There are also some extensive developments in ML approaches over the data from the Internet of Things. The ML approaches are adopted over network traffic data to validate the prediction accuracy of IoT devices towards the network. The author in [10], gathered and labelled the network traffic data from various IoT devices through smartphones and personal computers. With supervised learning approaches, a multi-stage Meta-classifier is trained for extracting the outcomes. In the initial phase, the classifiers have to classify the data from IoT devices. In the successive stage, the IoT devices are related to certain classes of IoT devices. Similarly, deep learning gives a promising solution for information extraction from sensor data which is deployed over a complex environment. With the multi-layer structure, DL is also a more appropriate way for edge computing.

In this work, a novel hybridization approach known as the linear stacking model is merged with the Xgboost algorithm (HLS-Xgboost). The ultimate research objective is to enhance prediction accuracy. Various investigationsare performed that outcomes in restriction with feature selection for various algorithm adoption. On contrary, the HLS-boost method utilizes all features without any constraint towards the selection process. Finally, the Xgboost classifier is used to classify the class labels. Here, various experiments have been carried out to perform heart disease prediction with the anticipated hybrid approach which is strongly able to identify heart disease in contrast to other diseases.

The remainder of the work is organized as: section 2 gives the extensive analysis of the existing research work on heart disease prediction; section 3 is the research methodology known as hybrid linear stacking model with Xgboost algorithm (HLS-Xgboost). Section 4 is the numerical results and discussions. Section 5 is a conclusion with future research directions.

2. EXISTING WORK

Various investigations and researchers have reported heart disease prediction-based development model using ML approaches with the target to offer enhanced performance. There are two diverse publicly available datasets termed Cleveland and Statlog dataset which is extensively utilized for evaluating the diagnostic performance among diverse researchers. With the Statlog dataset, an extensive heart disease CDSS based on rough-set and chaos algorithm for attribute reduction which is anticipated by Long et al., [11]. The former model is adopted for several attributes while the latter model is utilized to categorize the disease. The anticipated model is compared to other prevailing approaches like ANN, SVM, and NB respectively. The model attains better performance when compared to prevailing models with accuracy, specificity, and sensitivity. BPNN and rough set-based attribute integration are anticipated by the author in [12]. With chosen attributes, the model attains give better accuracy up to 90% respectively. Verma et al., [13] evaluated ML approaches like k-NN, LR, SVM, NB, ANN, and DT respectively with various performance metrics. The outcomes revealed that the performance of LR is superior to other approaches by attaining 86%, 90%, 82%, and 86% for accuracy, sensitivity, specificity, and precision respectively. Haq et al., [14] evaluated by predicting the significant attributes and adopts ML approaches like k-NN, LR, SVM, NB, ANN, and DT along with a hybrid model. The experimented outcome reveals that the hybrid approach with chosen attributes attains superior accuracy.

The conventional Cleveland dataset is extensively adopted by investigators to produce various predictive models. Saqlain et al., [15], modelled a hybrid prediction approach based on MLP, k-means clustering, particle swarm optimization, correlation feature subset, and so on. The outcome gives an accuracy of 91% respectively. Gupta et al., [16], designed a comparative analysis over the hybrid model on diverse feature selection approaches like minimal redundancy and maximal relevancy, relief approach, selection approach, and least absolute shrinkage, along with ML approaches like k-NN, LR, SVM, NB, ANN, and DT. The experimentation has revealed feature reduction influences the performance of the given model. Various investigations have been

conducted by integrating the LR-based ML approaches and the Relief-based feature selection algorithm which gives higher accuracy when compared to other hybrid algorithms. Latha et al., [17], anticipates an approach based on SVM classification and fisher score feature selection model. The features are chosen based on superior fisher score mean value. Then, the SVM model is applied to choose the feature subset to compute and learn MCC via the validation process. The research work shows that the hybridization of SVM and fisher score to produce specificity, sensitivity, and accuracy up to 88%, 73%, and 82% respectively. Latha et al., [17], anticipated a hybrid model with MLP, NB, RF, and BN respectively. The anticipated model attains 85% accuracy respectively.

Ali et al., [18], anticipated a two-stacked SVM model to enhance the prediction process. The anticipated SVM model is adopted to eliminate the non-relevant feature and the second model is used to identify heart disease. The outcomes show that this model attains superior performance to other models. Mohan et al., [19], initiated hybrid RF with a linear model to improve the model functionality. The author predicted that the anticipated model attains better specificity, F-measure, sensitivity, precision, and accuracy of up to 88%, 90%, 92%, 90%, and 83% respectively. Recently, Liu et al., [20], modelled an intelligent ML framework comprising of various analysing factors with mixed data and RF-based MLA approaches. The anticipated model is utilized to predict the appropriate features and identify the disease. The experimental analysis shows that the anticipated model outperforms other approaches by attaining better sensitivity, specificity, and accuracy of 93%, 89%, and 96% respectively.

The approaches mentioned above have not been adopted with feature attribute selection and balancing data during testing/training the data to enhance the accuracy of the prediction model, specifically, in the case of heart disease datasets. Therefore, this study concentrates on predicting the appropriate feature selection model, and a classifier is utilized to learn and to generate a prediction model. The anticipated model will give better performance than the prevailing approaches. At last, this work pretends to model an efficient CDSS model for E-healthcare monitoring that assists the clinicians to predict heart disease of patients according topatients' condition. Therefore, the earlier treatment process is carried out to prevent future risk.

3. PROPOSED WORK

This section elaborates the methodology of the proposed HLS-Xgboost model which is performed in two stages: HLS based feature selection and Xgboost based classification. Finally, the metrics are evaluated to analyze the significance of HLS-Xgboost over other models. Figure 1 shows the block diagram of the HLS-Xgboost classifier model.
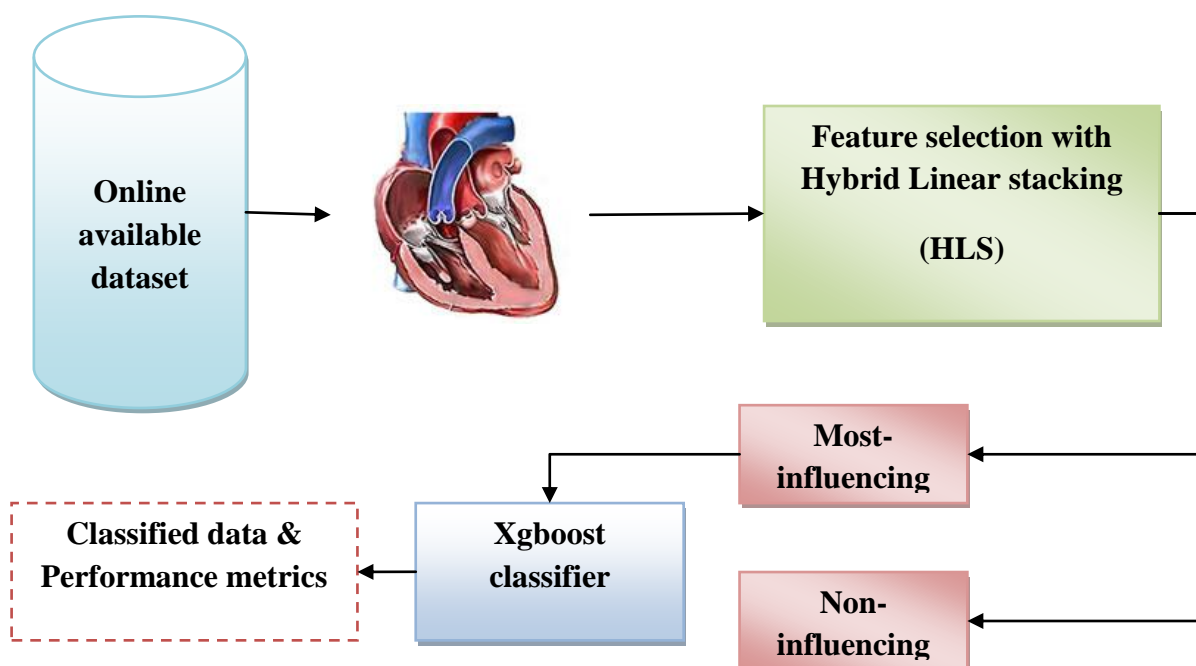

Fig 1. Block diagram of proposed HLS-Xgboost classifier

a. Dataset description

Here, the data have been collected from the online available UCI machine learning repository. The data is pre-processed once it is collected from various patients' records. It comprises of 303 patients' record, while 6 records show missing values. It is removed from the dataset while the others are maintained, i.e. 297 records. The binary classification and multi-class variables are initiated for the dataset attributes. These variables are adopted to validate the occurrence/non-occurrence of heart disease. The values are set as 1 or 0. The value '0' specifies that there are no disease symptoms. Here, 161 shows the absence of disease and 137 shows the presence of disease respectively. Table 1 depicts the dataset descriptions.

Table 1 UCI dataset description

Attributes	Descriptions	Type
Age	Patients' age	Nominal
Sex	Patients gender	Nominal
CP	Chest pain type 1. Typical and atypical angina, 3. Non-anginal and 4. asymptomatic	Numerical
Tresbps	Blood pressure level	Numerical
Chol	Serum cholesterol	Nominal
FBS	Blood sugar level	Nominal
Resting	Electrocardiogram results	Numerical
Thali	Maximal rate of heart	Nominal
Exang	Angina due to exercise	Numerical
OldPeak	Exercise due to depression	Nominal
Slope	Segment during peak exercise	Numerical
Ca	Fluoroscopy colored vessels	Nominal
Thal	Heart status	Nominal
Num	Values for heart disease diagnosis	Nominal

b. Linear stacking-based feature selection

From Table 1, 14 features are observed. Among the 13 dataset attributes, two attributes, i.e. sex and age are adopted to predict the information related to the patients. The remaining attributes are related to the clinical records for learning and predicting the severity of the disease. Assume, χ is the input, and $g_1, g_2, ..., g_L$ specifies the prediction function with $g_i: \chi \rightarrow R, \forall_i$.similarly,$f_1, f_2, ..., f_M$ is collection of the meta-feature function used for selecting the features. The meta-features are used to map the data-points $x \in \chi$ to corresponding $f_i(x) \epsilon R$. the linear stacking is used for predicting the function which is expressed as in Equation(1):

$$b(x) = \sum_i w_i g_i(x), \forall_x \in \chi \tag{1}$$

Here, the weight of learned model w_i is a constant value of R. The feature-based linear stacking is merged with weighted function, which is a linear function expressed as in Equation (2):

$$w_i(x) = \sum_j v_{ij} f_j; \ \ \forall_x \in \chi \tag{2}$$

Equation (2) is to express the learning weight. The above Equation is re-written as in Equation (3):

$$b(x) = \sum_{i,j} v_{ij} f_j(x) g_i(x), \ \ \forall_x \in \chi \tag{3}$$

The linear stacking-based optimization is given as in Equation (4):

$$\min_v \sum_{x\epsilon\chi} \sum_{i,j} \left(v_{ij} f_j(x) g_i(x) - y(x) \right)^2 \tag{4}$$

Here, $y(x)$ is target prediction and $'x'$ and χ is the subset used for stacking parameters training. Thus, the Equation() is linear with parameters v_{ij} and uses linear regression to compute the feature parameters. The

independent regression variables are the product of meta-feature functions $f_j(x), g_i(x)$ where the model predictor evaluates all $x \in \chi$.

The following algorithm explains the feature selection based on linear stacking

Algorithm 1: Feature selection based on linear stacking
Input: UCI machine learning dataset- features with the target class
Output: Feature parameters based on linear stacking
1. For all features do
2. For all samples do
3. Execute linear stacking function as in Equation (1);
4. End for
5. Identify the weighted features from the meta-feature functions as in Equation (2);
6. End for
7. Obtain the total number of feature products using linear regression
8. Partition the features based on most influencing and non-influencing
9. End process

c. Xgboost classifier

After balancing the imbalanced data over the dataset, the classifier model is adopted for learning and generating successive iterative values. Here, the Xgboost classifier model is used for predicting the occurrence/non-occurrence of heart disease. It is a supervised learning approach that is adopted for regression and classification. This is considered as an improved algorithm based on gradient boosting implementation with modification towards column sampling, loss function, and regularization. It is a technique to analyse the error residuals with a score to attain final prediction outcomes. It helps to reduce the loss score minimization when the model is constructed. The objective function is used to evaluate the model performance which includes two factors: regularization and training loss. The former term penalizes the model complexity and eliminates the overfitting issues. The loss function and regularization are expressed as in Equation (5)& (6):

$$L(\emptyset) = \sum_i l\left(\widehat{y_i}, y_i\right) + \sum_k \Omega\left(f_k\right) \tag{5}$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \left|\left|w\right|\right|^2 \tag{6}$$

Here, $'l'$ is a convex loss function that evaluates the difference between the target and prediction model. The regularization term Ω penalizes model complexity and the number of leaves over the tree which is specified as $'T'$. Moreover, f_k specifies the leaf weight and independent tree structure. Finally, γ is related to threshold and pre-pruning during the optimization to restrict the tree growth and λ is adopted to smooth the learning rate to evade the over-fitting problem.

The Xgboost model is implemented by considering the library. The data outliers are eliminating the training datasets using the Xgboost classifier. The linear stacking model is adopted to balance the dataset and generate the appropriate feature selection. This classifier model is adopted for learning the training dataset and classifies the state of heart disease appropriately. Here, various metrics are evaluated to analyse the functionality of the anticipated HLS-Xgboost classifier with the prevailing approaches.

Algorithm 2: Xgboost classifier
1. Initialize the target prediction function from the linear stacking model
2. For $i = 1$ to n do
3. Compute the loss function and regularization using a gradient model
4. Fit the model using Equation (5) to avoid over-fitting issues
5. Choose the gradient descent step size
6. Update the estimated function
7. End for
8. Obtain the final regression function model
9. Classify the labels for better prediction

This model needs to ensure the functionality of the anticipated model by executing the diagnose of subjects based on the present scenario. The measures like True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) respectively. TP and TN depict the total patients appropriately categorized as positive

(occurrence) and negative (non-occurrence) of heart disease. FP and FN depict the number of patients inappropriately categorized as positive (occurrence) and negative (non-occurrence) of heart disease. The classifier model undergoes cross-validation to analyse the iterative outcomes of training data. The statistical value gives $a\ p-value$ lesser than 1. Therefore, the anticipated model gives a significant comparison with other models as in Table 2.

Table 2 Statistical value comparison

Statistical value	$p-value$
HLS-Xgboost Vs NB	0.0001
HLS-Xgboost Vs LR	0.0002
HLS-Xgboost Vs MLP	0.0003
HLS-Xgboost Vs SVM	0.0
HLS-Xgboost Vs DT	0.0
HLS-Xgboost Vs RF	0.001

4. Results and discussion

This section discusses the numerical values attained by performing experimentation with the proposed HLS-Xgboost model. The simulation is carried out in a MATLAB environment. Here, metrics like accuracy, sensitivity, F-measure, precision, Matthew'scorrelation coefficients, False Negative rate, False Positive Rate, and True Negative Rate. The evaluation is done existing approaches like Naive Bayes (NB) classifier, Linear Regression (LR), Multi-Layer perceptron (MLP), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), and Density base Spatial clustering model respectively. The performance metrics are measured using below given Equation (7) to Equation (14):

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{7}$$

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

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

$$F - measure = \frac{2 * precision}{p + r} \tag{10}$$

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{11}$$

$$FPR = \frac{FP}{FP + TN} \tag{12}$$

$$FNR = \frac{FN}{FN + TP} \tag{13}$$

$$TNR = \frac{TN}{TN + FP} \tag{14}$$

FPR is specified as the miss rate which is the probability of positive prediction results missed during the test. Similarly, TNR (specificity) is the probability with actual negative will be tested negative.

Table 3: Overall comparison of HLS-Xgboost with existing approaches

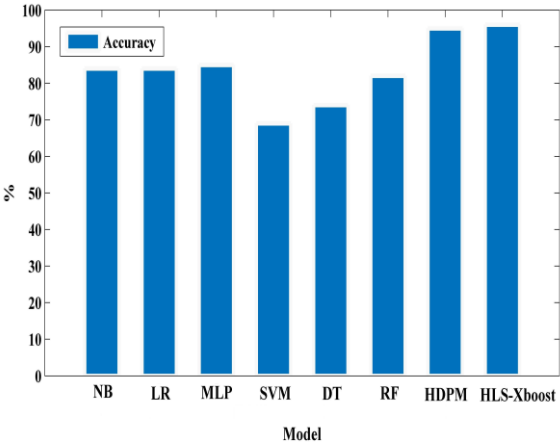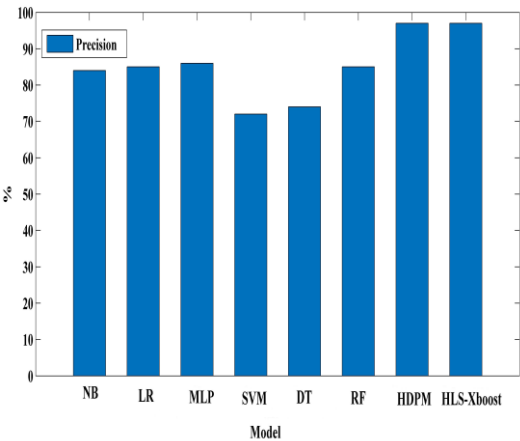| Model	Accuracy	Precision	Sensitivity	F-measure	MCC	FPR	FNR	TNR
NB	84%	84%	80%	81%	68%	12%	20%	87%
LR	83%	85%	80%	82%	70%	12%	19%	88%
MLP	85%	86%	81%	82%	67%	11%	18%	88%
SVM	69%	72%	50%	59%	39%	15%	49%	84%
DT	74%	74%	70%	71%	50%	3.33%	29%	76%
RF	82%	85%	75%	79%	68%	12%	23%	88%
HDPM	95%	97%	94%	94%	92%	4%	3.32%	95%
HLS-Xgboost	96%	97%	95%	95%	93%	4.53%	3.10%	96%

Fig 1: Accuracy comparison
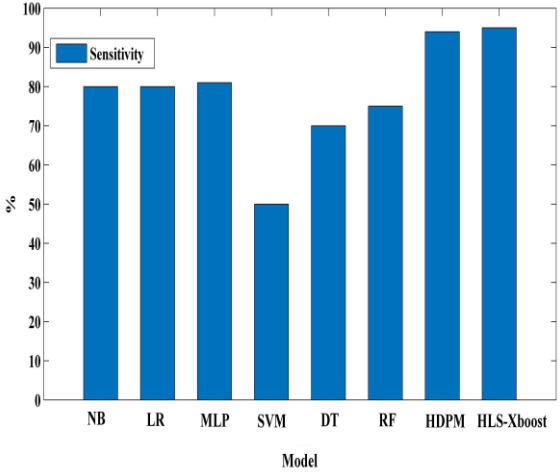

Fig 2: Precision comparison
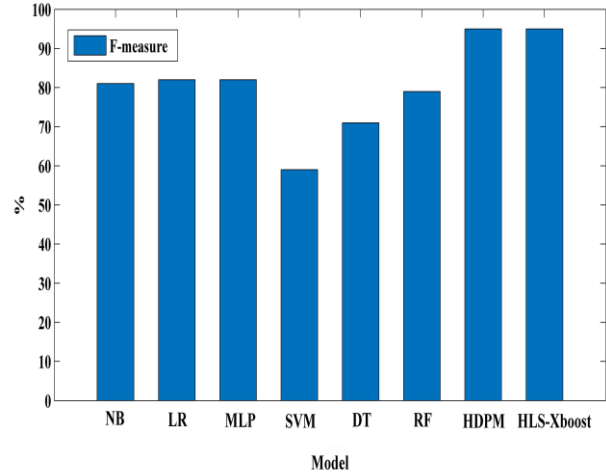

Fig 3: Sensitivity comparison
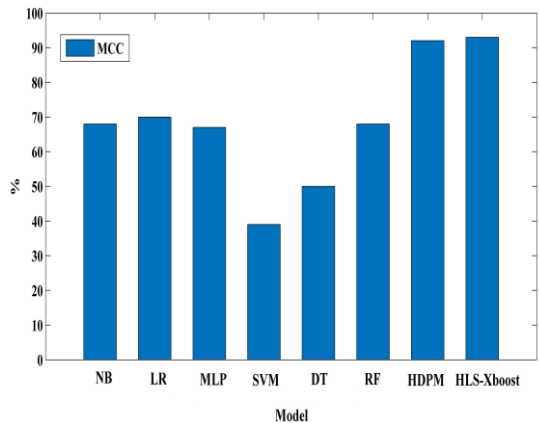

Fig 4: F-measure comparison
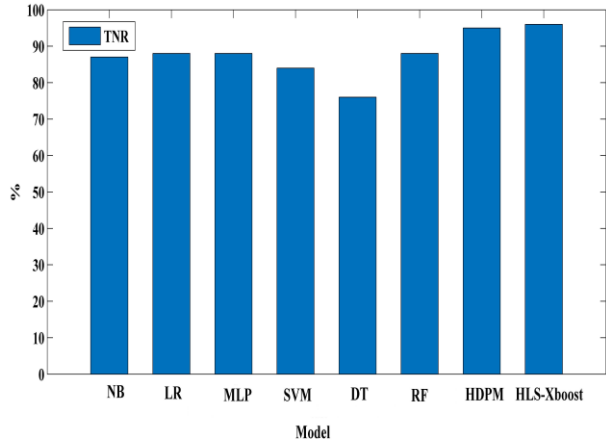

Fig 5: MCC comparison
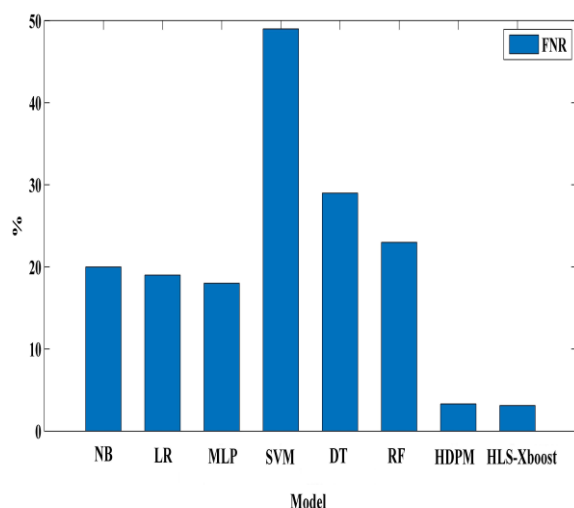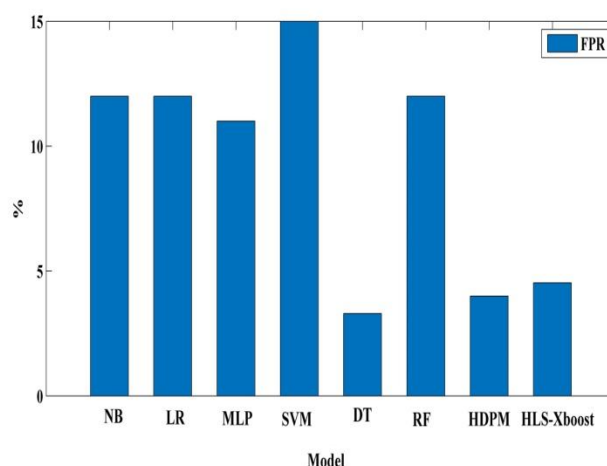

Fig 6: TNR comparison
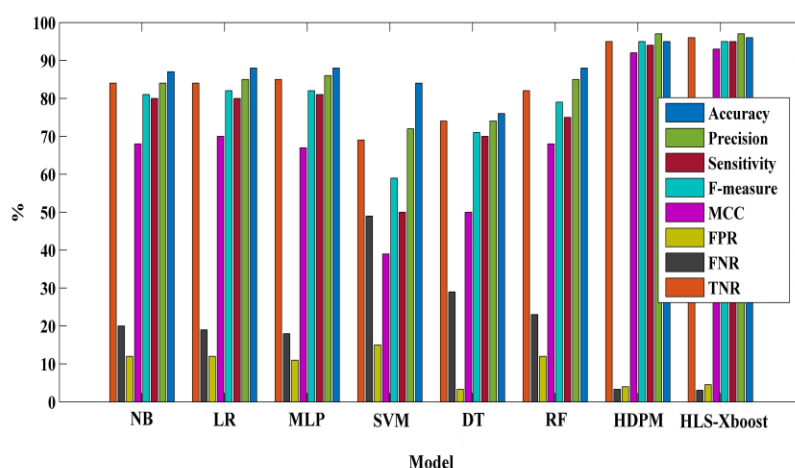
Fig 7: FNR comparison

Fig 8: FPR comparison

Fig 9: Overall performance metrics comparison

Figure 2 to Figure 9 shows the comparison of metrics like accuracy, sensitivity, precision, TNR, FPR, TNR, MCC, F-measure, and the overall comparison of HLS-Xgboost with other models. From the observations, it is known that the HLS-Xgboost model outperforms other approaches and gives better performance. The accuracy is 96% higher than other models and shows better trade-off in contrast to other approaches. Thereby, the prediction of heart disease by HLS-Xgboost helps the physicians to take an appropriate decision during the time of critical condition and acts as a better CDSS.

5. Conclusion

This investigation gives a new hybrid model for heart disease prediction. The anticipated HLS-Xgboost model is used for predicting the optimal feature subset and the classifier algorithm for predicting the disease. The proposed HLS-Xgboost classifier helps to address various problems like over-fitting issues while dealing with healthcare data from the available dataset. This HLS-Xgboost model is compared with traditional approaches like LR, NB, MLP, DT, SVM, RF, and HDPM models respectively. An extensive analysis is carried out by evaluating the results of the proposed and existing approaches, i.e. Accuracy, precision, sensitivity, F-measure, MCC, FNR, FPR, and TNR respectively. The comprehensive analysis demonstrates that the HLS-Xgboost classifier model outperforms the prevailing approaches and prevailing ML approaches. The comparison of various feature selection approaches with HLS-Xgboost and the classification model enhances the prediction performance. From the experimental analysis, the HLS-Xgboost model enhances the prediction rate with an accuracy of 96% and assists the physicians to formulate an effectual decision during the complication period.

REFERENCES

[1] Tada, H., Melander, O., Louie, J. Z., Catanese, J. J., Rowland, C. M., Devlin, J. J., ... &Shiffman, D. (2015). Risk prediction by geneticrisk scores for coronary heart disease is independent of self-reportedfamily history. European heart journal, 37(6), 561-567.

[2] Narasimhan, B., &Malathi, A. (2019). Improved Fuzzy ArtificialNeural Network (IFANN) Classifier for Coronary Artery HeartDisease Prediction in Diabetes Patients. Indian Journal of AppliedResearch, 9(04).

[3] Wu, C. S. M., Badshah, M., &Bhagwat, V. HeartDisease Prediction Using Data Mining Techniques. In Proceedingsof the 2019 2nd International Conference on Data Science andInformation Technology (pp. 7-11). ACM, 2017.

[4] Baskar, S., Dhulipala, V.R.S., Shakeel, P.M., Sridhar, K. P., Kumar,R. Hybrid fuzzy-based spearman rank correlation for cranial nervepalsy detection in the IoT environment. Health Technology. (2019).https://doi.org/10.1007/s12553-019-00294-8

[5] Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M.V., Corrado, G. S., ... & Webster, D. R. (2018). Prediction ofcardiovascular risk factors from retinal fundus photographs via deeplearning. Nature Biomedical Engineering, 2(3), 158.

[6] Chen, J., Valehi, A., &Razi, A. (2019). Smart Heart Monitoring:Early Prediction of Heart Problems Through Predictive Analysis ofECG Signals. IEEE Access, 7, 120831-120839.

[7] Singh, P., Singh, S., &Pandi-Jain, G. S. (2018). Effective heartdisease prediction system using data miningtechniques. International journal of nanomedicine, 13(T-NANO2014 Abstracts), 121.

[8] Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T.(2017). Artificial intelligence in precision cardiovascular
medicine. Journal of the American College of Cardiology, 69(21),2657-2664.26. Page, R. L., Joglar

[9] Lin, K.-C. Hsu, K. R. Johnson, M. Luby, and Y. C. Fann, ``Applyingdensity-based outlier identi_cations using multiple datasets for validationof stroke clinical outcomes,'' Int. J. Med. Inform., vol. 132, Dec. 2019

[10] Ismail, A. K. K. Chun, and M. I. S. Razak, ``Ef_cient herd Outlier detection in livestock monitoring system based on density_Basedspatial clustering,'' IEEE Access, vol. 7, pp. 175062_175070, 2019

[11] N. C. Long, P. Meesad, and H. Unger, ``A highly accurate _re_y basedalgorithm for heart disease prediction,'' Expert Syst. Appl., vol. 42, no. 21,pp. 8221_8231, Nov. 2015

[12] K. B. Nahato, K. N. Harichandran, and K. Arputharaj, ``Knowledge miningfrom clinical datasets using rough sets and backpropagation neural network,''Comput. Math. Methods Med., vol. 2015, pp. 1_13, Mar. 2015

[13] L. Verma, S. Srivastava, and P. C. Negi, ``A hybrid data mining modelto predict coronary artery disease cases using non-invasive clinical data,''J. Med. Syst., vol. 40, no. 7, p. 178, Jul. 2016

[14] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, ``A hybrid intelligentsystem framework for the prediction of heart disease using machinelearning algorithms,'' Mobile Inf. Syst., vol. 2018, pp. 1_21, Dec. 2018.

[15] S. M. Saqlain, M. Sher, F. A. Shah, I. Khan, M. U. Ashraf, M. Awais,and A. Ghani, ``Fisher score and Matthews correlation coef_cient-basedfeature subset selection for heart disease diagnosis using support vectormachines,'' Knowl. Inf. Syst., vol. 58, no. 1, pp. 139_167, Jan. 2019

[16] A. Gupta, R. Kumar, H. S. Arora, and B. Raman, ``MIFH: Amachine intelligence framework for heart disease diagnosis,'' IEEEAccess, vol. 8, pp. 14659_14674, 2020

[17] C. B. C. Latha and S. C. Jeeva, ``Improving the accuracy of predictionof heart disease risk based on ensemble classi_cation techniques,''Inform. Med. Unlocked, vol. 16, Jan. 2019

[18] L. Ali, A. Niamat, J. A. Khan, N. A. Golilarz, X. Xingzhong, A. Noor,R. Nour, and S. A. C. Bukhari, ``An optimized stacked support vectormachines based expert system for the effective prediction of heartfailure,'' IEEE Access, vol. 7, pp. 54007_54014, 2019

[19] S. Mohan, C. Thirumalai, and G. Srivastava, ``Effective heart diseaseprediction using hybrid machine learning techniques,'' IEEE Access, vol. 7,pp. 81542_81554, 2019.

[20] X. Liu, Q. Yang, and L. He, ``A novel DBSCAN with entropy and probabilityfor mixed data,'' Cluster Comput., vol. 20, no. 2, pp. 1313_1323,Jun. 2017