Received: January 17, 2023. Revised: February 9, 2023.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

An Ensemble Feature Optimization for an Effective Heart Disease Prediction Model

Kavitha Chandrashekar[1]* **Anitha Tuluvanooru Narayanreddy[2]**

[1]*Department of Computer Science, Vijaya Vittala Institute of Technology, Bangalore, Karnataka, India*
[2]*Department of Computer Science and Engineering, Sir Mokshagundam Visvesvaraya Institute of Technology, Bangalore, Karnataka, India*
* Corresponding author's Email: ckavitac22@gmail.com

Abstract: The use of machine learning (ML) within medical field is on the rise, notably as a means to enhance both the speed and precision of diagnosis. Through evaluating large volumes of patient information, machine learning is able to provide disease prediction, giving both patients and doctors more control over their health. Predicting and preventing heart disease has become a major area of study in medical data processing as a result of the increased expense of therapy. Since there are so many factors that come into play, estimating one's heart disease risk manually is a challenging task. Moreover, there are very few methods which provide better accuracy for the prediction of the heart disease. Hence, by using openly accessible cleveland heart disease dataset, this research aims to design and evaluate several advanced technologies constructed employing machine leaning algorithms for diagnosing if an individual is going to get heart disease or not. In this paper, we propose an ensemble feature optimized (EFO) learning method which uses an enhanced extreme gradient boosting tree and feature level cross validation scheme for effective heart disease (EHD) prediction. The presented EFO prediction algorithm and other existing machine learning algorithms have been used for the prediction of the heart diseases. The performance of the existing algorithms (XGB-based, ensemble tree hyper optimization (ETHO), and MLP-PSO) and proposed EFO algorithm has been evaluated using the classification metrics. When compared with the XGB-based, ensemble tree hyper optimization (ETHO), and MLP-PSO algorithm, the EFO algorithm has attained an accuracy of 98.61%. The EFO algorithm provides the doctors to able to predict the heart disease more efficiently and effectively.

Keywords: Cross validation, Data imbalance, Ensemble learning, Feature-selection, Heart-disease, Machine-learning.

1. Introduction

Physical and mental health are crucial for having a good lifestyle. Nevertheless, millions of people around the world fall victim to chronic illness like cardiovascular diseases (CVD) [1, 2], that affects the blood vessels and heart, leading to disability or death due to the various reasons. It has been stated that in the upcoming years, CVD will become one of the leading cause of deaths in humans [3]. Heart failure is the end result of a cascade of health problems including high blood pressure, clots, high cholesterol and coronary disease. One of the leading causes of CVD is hypertension [4, 5]. Deaths from CVD were estimated at 7.4 million in 2012, whereas stroke took the lives of 6.7 million people [6]. According to world-health-organization (WHO), around 31% of all the diseases are caused due to the CVDs. This comes to nearly 17 million deaths annually. Patients may be cured and countless lives may be saved if CVD is detected and treated early.

There is still a lot of room for improvement in cardiologists' ability to detect and treat cardiac issues at their earliest stages. Conventional cardiovascular disease risk assessment models all make the assumption that all risk factor has a linear relationship to CVD outcome [7]. Models like these likely to reduce correlations too much, and that includes a number of risk indicators having non-

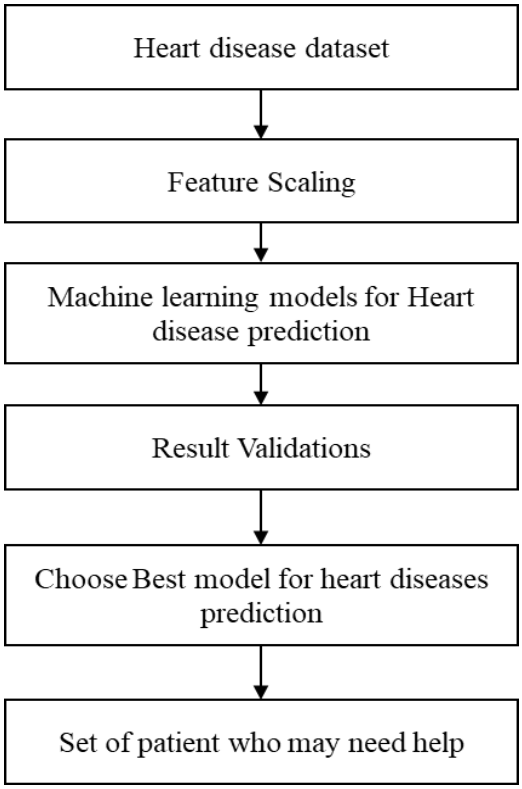

Figure. 1 General design of heart diseases prediction through machine learning models

linear relationships. More complex correlations among risk variables and results are needed to be uncovered, and a variety of factors need to be effectively included. For the purpose of assessing the prognosis of CVD, standard medical data and ML have not yet been employed in a large-scale investigation. The study's overall objective is to ascertain whether or not ML can improve heart disease risk predictive performance in population primary-care generally and to identify which ML method outcome has reasonably high efficiency. The general design of heart disease prediction through ML is shown in Fig. 1.

Recently, several CVD diagnosis algorithms using ML have already been presented [8]. Existing research is reviewed in order to define the issue and focus of each investigation. The use of ML [9] by a cardiologist aid in the early diagnosis and treatment of illness [10]. There have been many models which have been presented for prediction like support-vector-machine (SVM), K-nearest-neighbor (KNN), decision-trees (DT), artificial-neural-networks (ANN), extreme gradient boosting (XGBoost), and hybrid model [11, 12] combining ML and evolutionary computing algorithm each with its strengths and weaknesses. Recently, ensemble technique [12] combining several classifiers together have been presented for improving the accuracies of heart disease prediction. However, these models [13, 14] failed to establish which feature is important in predicting heart disease using standard cross validation algorithm; Further, existing heart disease prediction model [15, 16] exhibits very poor classification accuracy when data is imbalanced [17] in nature. This paper presents an ensemble feature optimization (EFO) for effective heart disease (EDH) predication model. The work introduces a novel feature learning optimization mechanism of standard XGBoost for minimizing false positive; further, introduce a new cross validation mechanism that identifies feature level ensemble addressing class imbalance issues. The significance of proposed EFO-based heart disease detection model is as follows:

- The presented heart disease prediction method uses an efficient ensemble gradient tree-based prediction model which provide better results when there exists imbalanced data.
- The ensemble gradient tree incorporates an enhanced cross validation method for studying the important features which impacts the performance of the prediction model.
- The presented method attains better performance in terms of sensitivity, specificity, accuracy and F1-Score when compared with the existing heart disease prediction models.

The organization of the paper is as follows, the literature survey has been given in section 2, the proposed method has bene explained in section 3 and on the basis of the proposed method, the results have been discussed in section 4. Finally, in section 5, the conclusion of the proposed work along with the future work has been given.

2. Literature survey

In this section, we present various machine learning model which have been presented for the detection and prediction of the heart diseases in the recent years. Multiple variables make it challenging to estimate one’s own risk of cardiovascular disease [13]. In [13], they have proposed a model for prediction of the heart disease, which uses the publicly accessible cleveland heart disease dataset to construct and evaluate various machine learning models. In this model they have used multi-layer perceptron (MLP) for training and a particle swarm optimization (PSO) method for heart disease identification. In this paper, they have compared the results with various other machine learning algorithms.

In [14], they modelled a PSO method and hybrid genetic algorithm (GA) built of random forest (RF),

GAPSO-RF. This method uses the only the optimal features which will help to increase the performance of the prediction of heart disease. In this model, in the initial stage the GAPSO-RF uses the multivariate statical evaluation model to select the most important features for the starting population. Further, the GA employs a discriminating mutation technique. At the final stage, the GAPSO-RF uses multivariate statistical analysis to pick the most important qualities for the seed population. Afterward, GA employs a discriminating mutation technique. Global search in GAPSO-RF is handled by a modified GA, while local search is handled by a PSO. Furthermore, PSO implemented the concept of rehabilitating people who were rejected during the selection process.

Using clinical data, in [15], they have proposed a machine learning model which makes early prediction for the CVD with the goal of reducing mortality rates. Many studies in the recent years have used various machine learning techniques for the detection of CVD or to determine the severity of the disease in a given patient. Although the existing research works showed promising results, but failed to optimize the machine learning techniques for better performance for the detection of DVD. In this model, they have used SMOTE (synthetic minority oversampling technique) for handling the data imbalance. Further, they have used six machine learning classification algorithms for the detection of the status of the patient. Also, a hyperparameter optimization (HPO) has been used for selection of best parameter for SMOTE and classification algorithms. For evaluation, they have used two datasets. The results show that their method attained good performance when compared with the existing models.

An assortment of fusion models for the diagnosis and severity prediction of CVDs are shown in [16]. Predictions of CVD have been made using a variety of ML methods, including ANN, SVM, logistic-regression (LR), decision trees (DT), random forests (RF), and AdaBoost. Multi-class classification class imbalance encouraged the implementation of a randomized over-sampler. Another technique based mostly on fusion of weighted scores was used to enhance performance of the classifier. The methods were initially trained. The results of two trained algorithms were blended using just a weighted rule. From these six ML algorithms, 3 fusion methods have been created. The performance indicator showed encouraging outcomes. Binary and multiclass classification problems have been tested with the presented method using varying test training ratios, with positive results for the fusion models. A 75% efficiency was obtained for multiclass-classification, whereas a 95% efficiency was found for binary-classification.

In [17] is focused with design of the cardiac disease diagnosis decision support system. This research makes advantage of a stream of data from the OpenML repository that contains 1,000,000 records of cardiac illness and 14 characteristics [18-20]. Following the application of pre-processing and feature selection techniques, machine learning methods such as RF, DT, gradient boosted trees (GBT), SVM, LR, MLP utilised to evaluate the multi-classification and binary classification on the data stream [21-23]. Binary classification achieved an accuracy of 94.8% whereas the multiclassification attained an accuracy of 88.2%. The results were satisfactory because MLP was used with the max abs scaler method. The GBT provided the correct result (accuracy of 95.8%) when contrasted with the other binary classifying methods [24, 25]. On the other hand, MLP excelled across a variety of categorization tasks. The accuracy of disease forecasts is affected by methods like

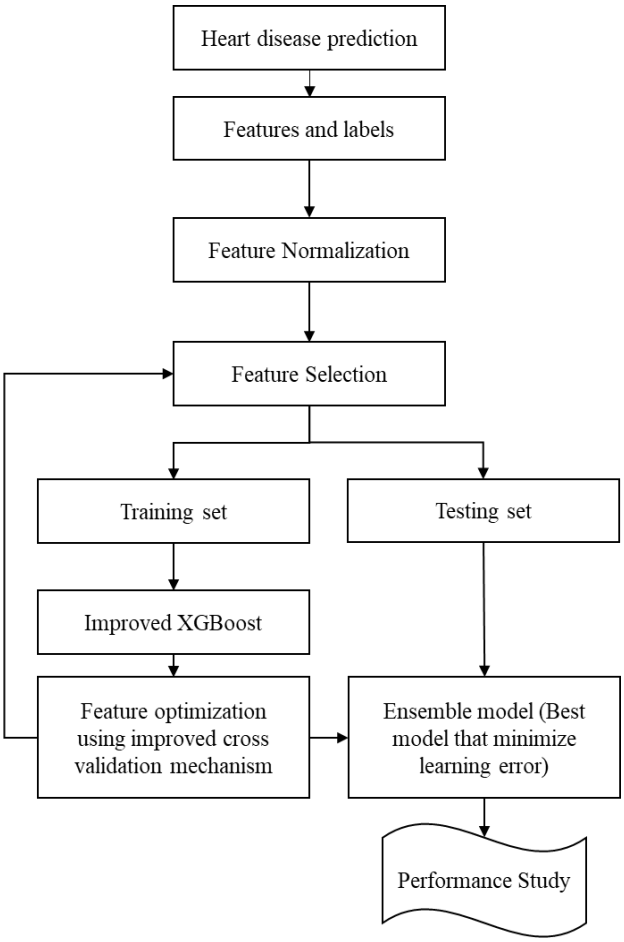

Figure. 2 Proposed Machine Learning model for heart disease prediction

oversampling and under sampling. Machine learning techniques, such as MLP and ensembles, can aid in the diagnosis of heart problems. Over-sampling and under-sampling are not viable options for this type of asymmetric data flow [26].

In [13, 14], they failed to establish which feature is important for the prediction of the heart disease using standard cross validation algorithm. The proposed model in [15-17] exhibited poor classification accuracy when the data is imbalanced in nature. Further, the proposed models in [21-25] mainly focussed on the binary classification and did not address the multi-classification issue. The proposed model of [26] did not address the issue of over-sampling and under-sampling data. Hence, from all the given problems there is a need for a model which addresses all the issues of the existing methods.

3. Methodology

The section introduces a methodology of ensemble feature optimization model for designing effective heart disease prediction model. In this proposed machine learning model for heart disease prediction, we have modified the process of feature selection during the XGBoost algorithm using the minimized objective function. The Fig. 2 shows the designed model.

3.1 XGBoost prediction algorithm

In improving the classification results, the XGBoost algorithm improves upon the gradient boosting technique [24] by combining multiple weak classifiers into a single robust one. The notation table has been given in Table 1.

For the prediction of the heart disease, consider the dataset of heart disease, $E = \{(y_j, z_j); j = 1 \dots o, y_j \in \mathcal{S}^n, z_j \in \mathcal{S}\}$, which contains various samples o of the data having n characteristics. Consider $\hat{z}_j$ which represents the prediction result of the model which can be represented using the following equation

$$\hat{z}_j = \sum_{l-1}^{L} g_l(y_j), \quad g_l \in G \tag{1}$$

Where, g_l represents the distinguishable regression tress and $g_l(y_j)$ represents the prediction result which has been accumulated from the $l-th$ tree using the $j-th$ sample. The distinguishable tree g_l and its attributes can be obtained using the minimization technique which is defined using the following equation

Table 1. Notation table

Notation	Meaning
$\hat{z}_j$	Prediction result of the model
g_l	Distinguishable regression tress
$g_l(y_j)$	Prediction result which has been accumulated from the $l-th$ tree using the $j-th$ sample
$\mathcal{O}$	Minimization variable
m	Training loss-operation
z_j	Actual result of the model
β	Penalizes the complexity of the prediction model
μ and δ	Regularization variable
U	Size of the leaf of the tree
x	Score of the various leaves of the tree
$\hat{z}_j^{(u)}$	Prediction result for the $j-th$ sample using the $u-th$ tree
g_u	Distinguishable tree
h_j	First-order gradient with respect to m
i_j	First-order gradient with respect to m
J_k	Sample set of the leaves of the gradient tree k
r	Fixed size of the gradient tree
x_k	Optimal-weights of the leaves j
$\mathcal{O}^s$	Optimal size of the tree
$CV(\sigma)$	K-fold cross validation
M	Heart disease training size
$P(\cdot)$	Loss function
$\hat{g}_\sigma^{-k(j)}(\cdot)$	Coefficient function
$\hat{\sigma}$	Selection feature

$$\mathcal{O} = \sum_{j=1}^{o} m(z_j, \hat{z}_j) + \sum_{l=1}^{L} \beta(g_l) \tag{2}$$

Where, m represents the training loss-operation which measures the variance between anticipated value $\hat{z}_j$ and the real value z_j. For avoiding the overfitting issue, the variable β is utilized which penalizes the complexity of the prediction model using the following equation

$$\beta(g_l) = \delta U + \frac{1}{2}\mu \|x\|^2 \tag{3}$$

Where, μ and δ is used for representing the regularization variable, U represent the size of the leaf of the tree, x represents the score of the various leaves of the tree. The construction of the ensemble tree is done using the summation technique. Consider, $\hat{z}_j^{(u)}$ which represents the prediction result for the $j-th$ sample using the $u-th$ tree, the distinguishable tree g_u should be considered and added to minimize the following function

$$\mathcal{O}^{(u)} = \sum_{j=1}^{o} m\left(z_j, \hat{z}_j^{(u-1)} + g_u(y_j)\right) + \beta(g_l) \quad (4)$$

After removing the constant variable using the second order Taylor expansion, the preceding equation becomes as follows.

$$\mathcal{O}^{(u)} = \sum_{j=1}^{o} \left[h_j g_j(y_j) + \frac{1}{2} i_j g_u(y_j)^2\right] + \beta(g_l) \quad (5)$$

Where, h_j is used to represent the first-order gradient with respect to m using the following equation.

$$h_j = \partial_{\hat{z}_j^{(u-1)}} m\left(z_j, \hat{z}_j^{(u-1)}\right) \quad (6)$$

Where, i_j is used to represent the first-order gradient with respect to m using the following equation.

$$i_j = \partial^2_{\hat{z}_j^{(u-1)}} m\left(z_j, \hat{z}_j^{(u-1)}\right) \quad (7)$$

Hence, the objective variable for the prediction can be obtained using the below equation

$$\mathcal{O}^{(u)} = \sum_{j=1}^{o} \left[h_j g_j(y_j) + \frac{1}{2} i_j g_u(y_j)^2\right] + \delta U + \frac{1}{2}\mu \sum_{k=1}^{U} x_k^2 \quad (8)$$

After simplification of the above equation, the given below equation is obtained

$$\mathcal{O}^{(u)} == \sum_{j=1}^{U} \left[\left(\sum_{j\in J_k} h_j\right) x_j \frac{1}{2}\left(\sum_{j\in J_k} i_j + \mu\right) x_k^2\right] + \delta U \quad (9)$$

Where, J_k represents the sample set of the leaves of the gradient tree k which is defined using the following equation

$$J_k = \{j | r(y_j = k)\} \quad (10)$$

Where, r represents the fixed size of the gradient tree, the optimal-weights x_k^* of the leaves j can be attained using the following equation

$$x_k = \frac{H_k}{I_k + \mu} \quad (11)$$

The optimal size of the tree can be attained using the following equation

$$\mathcal{O}^s = \frac{1}{2}\sum_{k-1}^{U} \frac{H_k^2}{I_k + \mu} + \delta U \quad (12)$$

Where, H_k is defined using the following equation

$$H_k = \sum_{j\in J_k} h_j \quad (13)$$

Similar to the H_k, I_k can be defined using the following equation

$$I_k = \sum_{j\in J_k} i_j \quad (14)$$

The optimal size of the tree $\mathcal{O}^*$ has the ability to define the structure of the tree. $\mathcal{O}^*$ also represent the quality of the tree r. The smaller the value of r, better the structure of the tree. Even though the XGBoost algorithm is an effective method for achieving higher prediction accuracy, but, when the data is imbalanced or due to improper feature selection, the accuracy for the prediction can degrade. In addressing research problem an ensemble feature optimization using XGBoost is presented in next sub-section.

3.2 Ensemble feature optimization

The traditional XGBoost algorithm the feature selection is done using K-fold cross validation presented in below equation

$$CV(\sigma) = \frac{1}{M}\sum_{k=1}^{K}\sum_{j\in G_{-k}} P\left(b_j, \hat{g}_\sigma^{-k(j)}(y_j, \sigma)\right) \quad (15)$$

However, the limitation of using above cross validation it fails to establish correlation among the features present within heard disease dataset; thus, result in impacting overall accuracies of prediction. In addressing this work introduces a modified cross validation as shown in below equation

$$CV(\sigma) = \frac{1}{SM}\sum_{s=1}^{S}\sum_{k=1}^{K}\sum_{j\in G_{-k}} P\left(b_j, \hat{g}_\sigma^{-k(j)}(y_j, \sigma)\right) \quad (16)$$

Where In Eq. (16), M represent heart disease training size, $P(\cdot)$ represent loss function, and $\hat{g}_\sigma^{-k(j)}(\cdot)$ coefficients function. Using above cross validation mechanism first the feature subset is obtained; later, the selected feature is used to construct the ensemble-based heart disease prediction by minimizing prediction error in iterative manner using below equation

$$\hat{\sigma} = \underset{\sigma\in\{\sigma_1,\ldots,\sigma_l\}}{\arg\min}\ CV_s(\sigma) \quad (17)$$

The proposed ensemble-based heart disease

prediction model is very efficient accurately prediction heart disease in comparison with existing heart disease prediction which is experimentally shown in experiment section.

4. Results and discussion

In this section the heart disease prediction using proposed EFO and other existing ML-based heart disease prediction method are studied [15, 17]. The UCI repository Cleveland dataset from [27, 28] is used for performance analysis. The selection of dataset is based on comparison paper [15, 17]. The machine learning model for performing heart disease prediction is implemented using python 3 framework. The dataset used in this research is the Cleveland-Heart-Disease dataset [19]. It is an imbalanced classification dataset consisting of 303 instances. The ROC performance metrics such as accuracy, sensitivity, specificity, precision, and F-measure are used for validating heart disease prediction model. The accuracy is computed as follows

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (18)$$

where TP defines true-positive, TN defines true-negative, FP defines false-positive, and FN defines false-negative. The sensitivity is computed as follows

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (19)$$

The specificity is computed as follows

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (20)$$

The precision is computed as follow

$$\text{Precision} = \frac{TP}{TP+FP} \quad (21)$$

The F-measure is computed as follows

$$\text{F} - \text{measure} = \frac{2\times Precision\times \text{Sensitivity}}{Precision\times \text{Sensitivity}} \quad (22)$$

4.1 Predictive model performance evaluation

This section studies the performance of different predictive models such as XGB-based [17], ensemble tree hyper optimization (ETHO) [15], MLP-PSO [13] and proposed EFO. The sensitivity performance in predicting heart disease is given in Fig. 3. The recall performance in predicting heart disease is given in Fig. 4. The accuracy performance

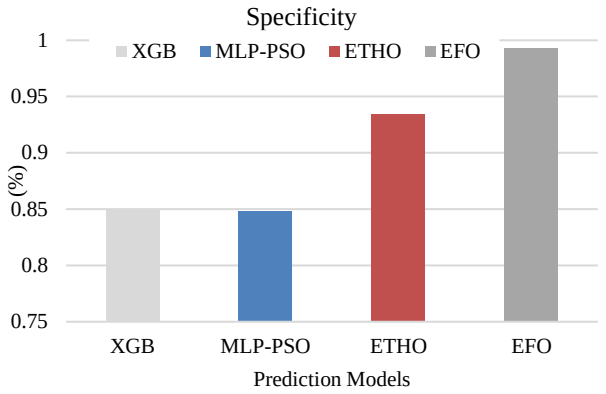

Figure. 3 Specificity performance

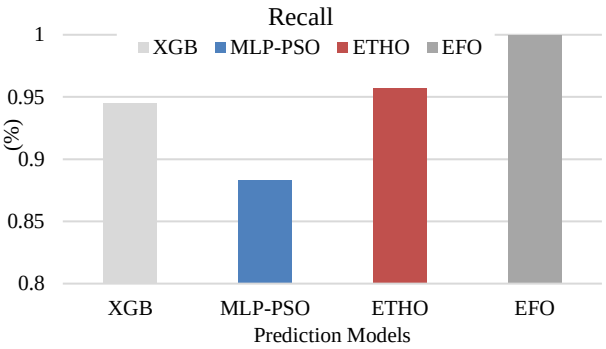

Figure. 4 Recall performance

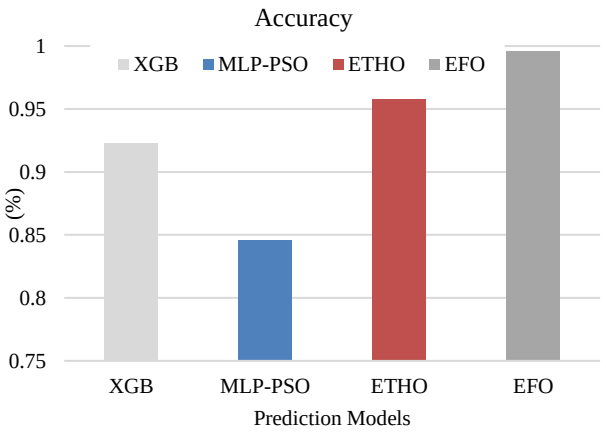

Figure. 5 Accuracy performance

in predicting heart disease is given in Fig. 5. The precision performance in predicting heart disease is given in Fig. 6. The F-measure performance in predicting heart disease is given in Fig. 7. Higher value in Figs. 3 to 7 indicates better predictive performance. The proposed EFO model achieves very good predictive performance in comparison with recent predictive model such as XGB-based, ETHO, MLP-PSO. The overall result shows the EFO-based predictive model is provides better predictive accuracy with minimal false positive.

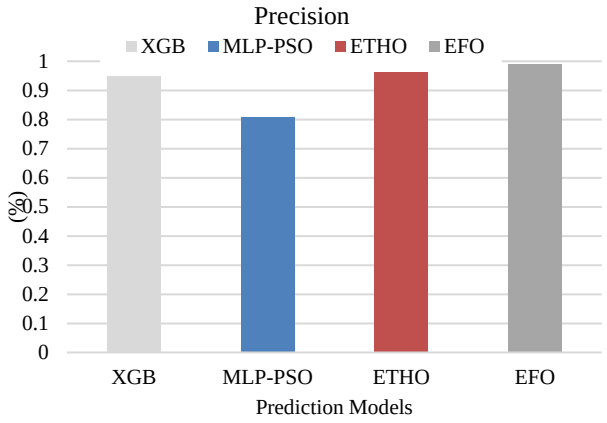

Figure. 6 Precision performance

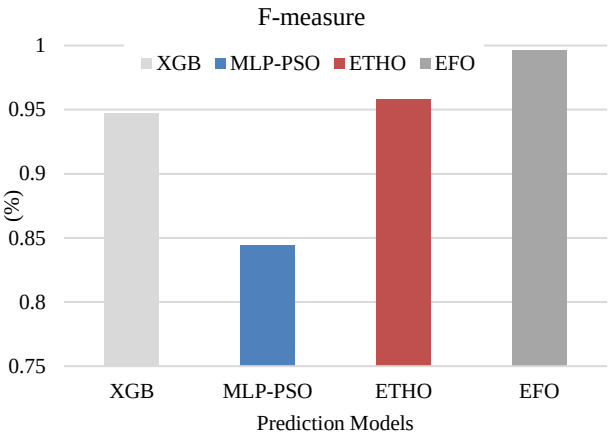

Figure. 7 F-measure performance

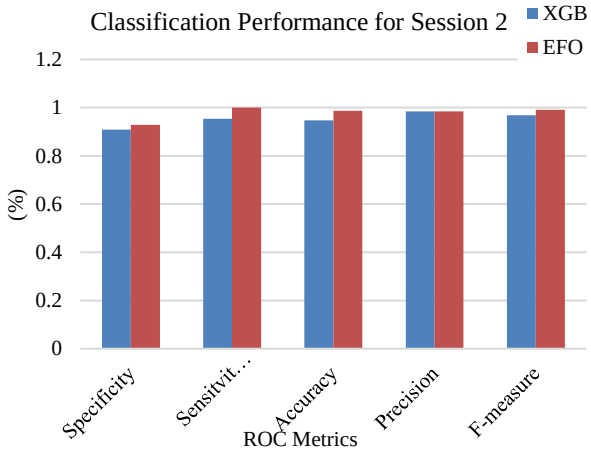

Figure. 8 Heart disease prediction for session 2

4.2 Heart disease prediction for different sessions

This section studies performance of standard XGB-based heart disease prediction with proposed improved version of XGB with ensemble feature optimization-based heart disease prediction. The work validated considering different sessions. In Fig. 8, only 20% data is considered for training and remaining 80% is considered for testing and heart disease prediction outcome is graphically shown. In Fig. 9, only 40% data is considered for training and remaining 60% is considered for testing and heart disease prediction outcome is graphically shown. In Fig. 10, only 60% data is considered for training and remaining 40% is considered for testing and heart

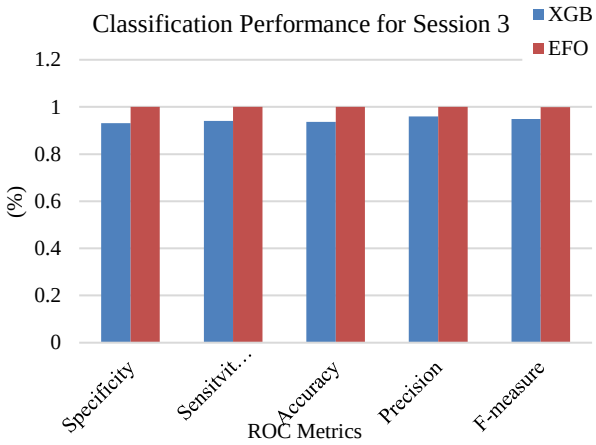

Figure. 9 Heart disease prediction model for session 3

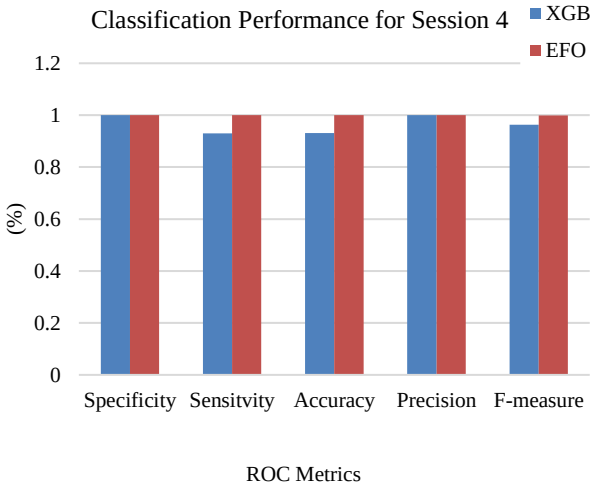

Figure. 10 Heart disease prediction for session 4.

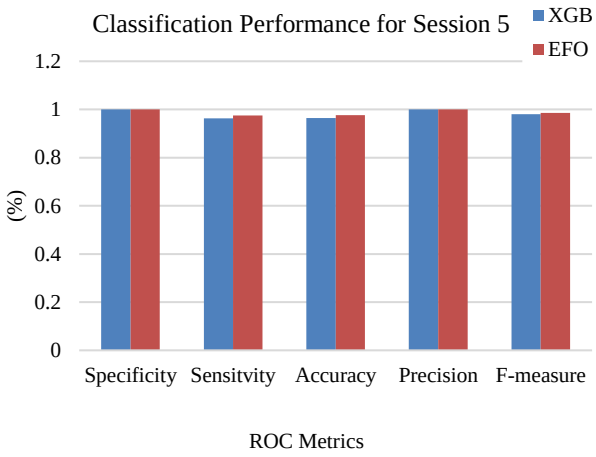

Figure. 11 Heart disease prediction for session 5

disease prediction outcome is graphically shown. In Fig. 11, only 80% data is considered for training and remaining 20% is considered for testing and heart disease prediction outcome is graphically shown. From Figs. 8 to 11 we can interpretate that proposed model work well for smaller and larger training data.

5. Conclusion

The most challenging task in the medical area is the prediction of cardiovascular disease (CVD) or heart disease by evaluating the session streams. In recent years, various machine learning algorithms have been proposed which show promising results for the prediction of heart diseases. Nevertheless, all the existing machine learning models have attained higher accuracy only for specific kinds of heart disease data and fail when a new dataset is introduced. To address this issue, recent work has proposed an ensemble-based machine learning prediction model for heart disease which selects the best model for performing the prediction task. But when there exists an imbalance in data the proposed ensemble-based machine learning prediction fails to attain higher accuracy for the prediction of heart disease. To address all these issues, in this research work, we propose an efficient ensemble machine learning technique that has been developed by modification of the XGBoost algorithm which provides better accuracy when the data is imbalanced. For attaining higher accuracy, this model uses a cross-validation technique which uses the effective feature ranking method for improving the accuracy of prediction by optimizing the prediction error. The model has been evaluated by conducting experimentation on heart disease session stream data. The results show that the EFO model provides better performance in terms of sensitivity, specificity, accuracy, and F-measure when compared with the existing machine learning model such as XGB-based, ETHO, MLP-PSO for predicting heart disease. For future work, the EFO model will be used for testing other datasets and to provide higher performance. Also, consider reducing the training error for multiclass classification.

Conflicts of interest

The authors declare no conflicts of interest.

Author contributions

Kavitha C came up with the original idea and contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript. Kavita C got help from T. N. Anitha to improve the manuscript.

References

[1] R. Selvi and I. Muthulakshmi, “An optimal artificial neural network based big data application for heart disease diagnosis and classification model”, *Journal of Ambient Intelligence Humanized Computing*, Vol. 12, No. 6, pp. 6129-6139, 2021.

[2] G. Bazoukis, S. Stavrakis, J. Zhou, S. Bollepalli, G. Tse, Q. Zhang, J. Singh, and A. Armoundas, “Machine learning versus conventional clinical methods in guiding management of heart failure patients: A systematic review”, *Heart Failure Review*, Vol. 26, No. 1, pp. 23-34, 2021.

[3] B. Patel and P. Sengupta, “Machine learning for predicting cardiac events: what does the future hold?”, *Expert Review of Cardiovascular Therapy*, Vol. 18, No. 2, pp. 77-84, 2020.

[4] M. Pal, S. Parija, G. Panda, K. Dhama, and R. Mohapatra, “Risk prediction of cardiovascular disease using machine learning classifiers”, *Open Medical Wars*, Vol. 17, No. 1, pp. 1100-1113, 2022.

[5] M. Niwariya, A. Rajput, and S. Jaloree, “A Nobel Hybrid Approach for CHDES Model for Predicting Heart Disease Patients”, In: *Proc. of 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, pp. 942-946, 2022.

[6] M. Tarawneh and O. Embarak, “Hybrid Approach for Heart Disease Prediction Using Data Mining Techniques”, *Advances in Internet, Data and Web Technologies*, pp. 447-454, 2019.

[7] A. Ullah, S. Khan, T. Alam, S. Osmani, and M. Sadie, “Heart disease classification using various heuristic algorithms”, *International Journal of Advances in Applied Sciences*, Vol. 56, pp. 158-167, 2022.

[8] K. Almustafa, “Prediction of heart disease and classifiers sensitivity analysis”, *BMC Bioinformatics*, Vol. 21, p. 278, 2020.

[9] A. Yazdani, K. Varathan, Y. Chiam, A. Malik, and W. Ahmad, “A novel approach for heart disease prediction using strength scores with significant predictors”, *BMC Medical Information Decision Making*, Vol. 21, No. 1, p. 194, 2021.

[10] H. Balaha, A. Shaban, and E. E. Gendy, “A multi-variate heart disease optimization and recognition framework”, *Neural Computing & Applications*, Vol. 34, pp. 15907–15944, 2022.

[11] H. Kibria and A. Matin, “An Efficient Machine

Learning-Based Decision-Level Fusion Model to Predict Cardiovascular Disease", *Intelligent Computing and Optimization, ICO 2020, Advances in Intelligent Systems and Computing*, Vol. 1324, pp. 1097-1110, 2021.

[12] F. Abdeldjouad, M. Brahami, and N. Matta, "A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques", *The Impact of Digital Technologies on Public Health in Developed and Developing Countries,. Lecture Notes in Computer Science*, Vol. 12157, pp. 299–306 2020.

[13] A. Bataineh and S. Manacek, "MLP-PSO Hybrid Algorithm for Heart Disease Prediction", *Journal of Personalized Medicine*, Vol. 12, No. 8, p. 1208, 2022.

[14] M. E. Shafiey, A. Hagag, and E. E. Dahshan, "A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest", *Multimedia Tools and Applications*, Vol. 81, pp. 18155–18179, 2022.

[15] A. Abdellatif, H. Abdellatef, J. Kanesan, C. Chow, J. Chuah, and H. Gheni, "An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods", *IEEE Access*, Vol. 10, pp. 79974-79985, 2022.

[16] H. Kibria and A. Matin, "The Severity Prediction of the Binary and Multi-Class Cardiovascular Disease−A Machine Learning-Based Fusion Approach", *Computational Biology and Chemistry*, Vol. 8, p. 107672, 2022.

[17] D. Hamid, S. Ullah, J. Iqbal, S. Hussain, C. Hassan, and F. Umar, "A Machine Learning in Binary and Multiclassification Results on Imbalanced Heart Disease Data Stream", *Journal of Sensors*, Vol. 2022, No. 8400622, pp. 1-13, 2022.

[18] A. Saboor, M. Usman, S. Ali, A. Samad, M. Abrar, and N. Ullah, "A method for improving prediction of human heart disease using machine learning algorithms", *Mobile Information Systems*, Vol. 2022, No. 1410169, pp. 1-9, 2022.

[19] K. Arumugam, M. Naved, P. Shinde, O. Chauca, A. Osorio, and T. Yanac, "Multiple disease prediction using machine learning algorithms", *Materials Today: Proceedings*, Vol. 67, No. 6, pp. 1-6, 2021.

[20] C. Gupta, A. Saha, N. Reddy, and U. D. Acharya, "Cardiac disease prediction using supervised machine learning techniques", *Journal of Physics: Conference Series*, Vol. 2161, No. 1, p. 012013, 2022.

[21] V. T. Truong, B. P. Nguyen, and T. H. Nguyen-Vo, "Application of machine learning in screening for congenital heart diseases using fetal echocardiography", *The International Journal of Cardiovascular Imaging*, Vol. 38, No. 5, pp. 1007–1015, 2022.

[22] A. S. Abdalrada, J. Abawajy, T. Quraishi, and S. Islam, "Machine learning models for prediction of co-occurrence of diabetes and cardiovascular diseases: a retrospective cohort study", *Journal of Diabetes & Metabolic Disorders*, Vol. 21, No. 1, pp. 251–261, 2022.

[23] B. P. Doppala, D. Bhattacharyya, M. Janarthanan, and N. Baik, "A reliable machine intelligence model for accurate identification of cardiovascular diseases using ensemble techniques", *Journal of Healthcare Engineering*, Vol. 2022, No. 2585235, pp. 1-13, 2022.

[24] A. Kondababu, V. Siddhartha, B. Kumar, and B. Penumutchi, "A comparative study on machine learning based heart disease prediction", *Materials Today: Proceedings*, Vol. 67, No. 6, pp. 1-10, 2021.

[25] I. Mienye, Y. Sun, and Z. Wang, "An improved ensemble learning approach for the prediction of heart disease risk", *Informatics in Medicine Unlocked*, Vol. 20, No. 1, p. 100402, 2020.

[26] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques", *IEEE Access*, Vol. 7, pp. 81542–81554, 2019.

[27] G. Magesh and P. Swarnalatha, "Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction", *Evolutionary Intelligence*, Vol. 14, No. 2, pp. 583-593, 2021.

[28] A. Janosi, W. Steinbrunn, M. Pfistere, and R. Detrano, "Heart Disease Data Set. The UCI KDD Archive", 1988. Available online: https://archive.ics.uci.edu/ml/datasets/heart+disease (accessed on 22 November 2022).