

<sup>1</sup>Kiruthiga G  
<sup>2</sup>Shakkeera L  
<sup>3\*</sup>Vinodkumar  
 Jacob  
<sup>4</sup>Anita Venaik  
<sup>5</sup>Asha A  
<sup>6</sup>Dhiyanesh B

## Enhanced Detection of Chronic Heart Disease: A Novel Approach With Extreme Gradient Boosting- Logistic Regression Algorithm



**Abstract:** - In recent years, biomedical healthcare applications for managing and treating chronic heart diseases have increased. Chronic heart disease is a serious and sometimes deadly ailment that affects millions of individuals globally. Therefore, timely and effective detection of heart disease is crucial in medicine, particularly in cardiology. These complications can lead to reduced quality of life, increased medical costs, and increased risk of death. Obesity, smoking, and high blood pressure are some of the risk factors that worsen chronic heart disease. However, challenges occur in accurately and sensitively classifying heart disease in performance evaluation. To resolve this problem, we propose an Extreme Gradient Boosting-Logistic Regression (EGBLR) algorithm to accurately and quickly predict chronic heart disease. Furthermore, pre-processing can be performed using the Normalized Minimum-Maximum Scaling Feature (NMMSF) technique to estimate missing or duplicate values. Next, the Inter-Quartile Range Optimization (IQRO) technique can analyze heart disease data to identify and remove outliers. After that, the Relief Feature Vector (RFV) algorithm can be used to find each feature weight. Finally, the proposed EGBLR method can be employed to detect normal or abnormal chronic heart diseases and increase accuracy in diagnosis. The evaluation of the proposed techniques can be predicted using performance matrices such as recall, sensitivity score, precision, false rate, and precision. The accuracy of the proposed EGBLR detection approach exceeds that of previous techniques. Moreover, the proposed method for diagnosing chronic heart disease can be easily adapted to biomedical healthcare applications.

**Keywords:** Chronic heart disease, Logical Gradient Boosting, Logistic Regression, Normalized Scaling Feature, Optimization, Feature Vector

### I. INTRODUCTION

The healthcare sector can gather extensive data, enabling experts to derive valuable insights from hidden information. Moreover, advanced biomedical applications allow healthcare providers to make data-driven decisions that result in more relevant and accurate outcomes. Similarly, the diagnosis of heart disease relies on the patient's symptoms and physical assessment. However, with data mining and the storage of clinical information, there is a significant opportunity to improve patient management within the healthcare sector. Healthcare providers can improve their capacity to identify and treat numerous disorders by combining technology and clinical diagnostics, thereby enhancing patient data analytics and overall healthcare quality [1-2].

Moreover, it is crucial to acknowledge the significance of Electronic Medical Records (EMRs) in the healthcare industry. The identification of heart disease has seen significant progress with the advent of wearable sensors that enable the continuous monitoring of patient's internal and external physiological parameters. By integrating the data from these sensors with the information extracted from EMRs, healthcare providers can effectively identify subtle signs of potential heart issues that may have otherwise gone unnoticed. Furthermore, using brainy systems to analyse this aggregated data can advance heart disease prediction as it becomes clinically apparent. Similarly, improving patient results and healthcare quality through technology can enhance the efficiency of the healthcare sector, highlighting their significance [3].

When heart disease progresses, it may lead to a heart attack, which poses a serious risk to the patient's life. Therefore, early identification and prompt therapy are critical. Timely identification of heart issues and regular monitoring by medical professionals can lower mortality rates. Nevertheless, accurately diagnosing heart disease in every instance is challenging, necessitating round-the-clock patient care from doctors for a deeper understanding, more time, and specialized knowledge. Heart attack, heart failure, and arrhythmia are serious consequences that can arise from inadequate blood flow to the heart muscles. In some circumstances, angioplasty or surgery may be necessary to increase blood flow to the heart [4-5].

<sup>1</sup>Professor/CSE, Professor/IT, Karpagam College of Engineering, Coimbatore, Tamil Nadu, India

<sup>2</sup>Associate Professor/CSE, Presidency University, Bengaluru, Karnataka, India

<sup>3\*</sup>Professor/ECE, M.A College of Engineering, Kothamangalam, Kerala, India

<sup>4</sup>Professor/IT, Amity Business School, Amity University, Noida, India.

<sup>5</sup>Professor/ECE, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India.

<sup>6</sup>Associate Professor/CSE, Dr. N.G.P. Institute of Technology, Coimbatore, Tamil Nadu, India

<sup>1</sup>kiruthiga.g@kce.ac.in, <sup>2</sup>shakkeera.l@presidencyuniversity.in, <sup>3\*</sup>vkj@mace.ac.in, <sup>4</sup>avenaik@amity.edu, <sup>5</sup>ngash78@gmail.com,

<sup>6</sup>dhiyanu87@gmail.com

\*Corresponding author: vkj@mace.ac.in

Copyright © JES 2024 on-line: journal.esrgroups.org

This paper introduces the EGBLR algorithm to provide accurate and quick predictions of chronic heart disease. The NMMSF method can be implemented to estimate missing or duplicate values during the data pre-processing stage. IQRO techniques are then employed to analyze cardiac data and remove outliers. Then, the RFV algorithm can be used to find the weight of each feature in the dataset. The proposed EGBLR method based on the ML technique can diagnose chronic heart diseases and improve their accuracy. This proposed approach for diagnosing chronic heart disease can be easily integrated into biomedical healthcare applications.

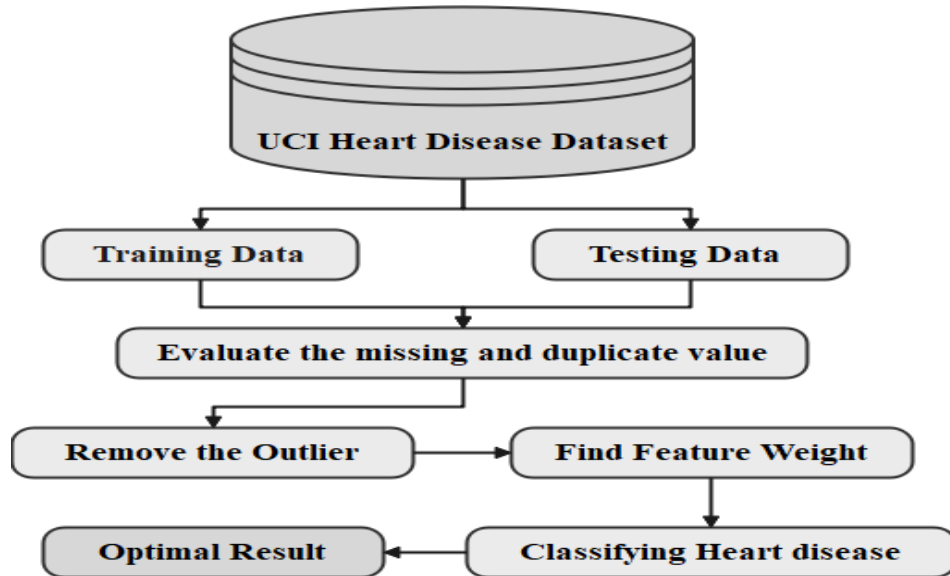


Fig. 1: Diagnosing Chronic Heart Disease of Architectural Diagram

As shown in Figure 1, the data collected from the Kaggle dataset can remove outliers, assess missing and duplicate data, and predict the accuracy of chronic heart disease diagnosis based on the architectural graph, which is the contribution of this section. The feature weights are updated and tabulated by the Heart classification system.

## II. LITERATURE SURVEY

Electronic Health Records (EHRs) are integrated and analyzed to develop suitable treatment plans and safer patient processes for cardiovascular patients. Moreover, the suggested association rule mining technique precisely forecasts heart disease using heart disease data. Nevertheless, several obstacles could be overcome by the broad acceptance of centralized EHRs [6]. Thus, identifying heart disease is an extremely complex medical undertaking. Furthermore, using various data mining approaches to assess the likelihood of cardiac disease might help classify individuals into different risk groups [7]. They offer accurate, measurable precision medicine to improve health and prevent chronic disease. Similarly, smartphone app integration enables real-time monitoring of lifestyle and environmental parameters [8]. Therefore, the OCI-DBN method may compare the trial findings with other cutting-edge methods to enhance the system's performance [9]. Similarly, Support Vector Machine (SVM) algorithms can achieve high coronary heart disease prediction accuracy by processing datasets with essential features [10]. An efficient and accurate cardiac diagnosis system can be observed utilizing Machine Learning (ML) techniques with implementation metrics to evaluate the classifier's performance [11]. Furthermore, a classification system for valvular heart disease could be established by implementing a Convolutional Neural Network (CNN) approach utilizing opacity charts and depth dream images for local and global analysis [12]. By utilizing DBSCAN approaches for outlier identification and removal, as well as XGBoost, to establish the feasibility of accurately forecasting the development of heart disease, the threshold is enhanced [13]. Using the Cleveland cardiovascular dataset, they produce a Recurrent Fuzzy Neural Network (RFNN) technique for identifying heart disease. The dataset comprised 297 patient records, with 252 allocated for training and 45 for testing [14]. A Deep Neural Network (DNN) algorithm and an embedded feature selection method can be utilized to predict heart disease. Nevertheless, the abundance of data during analysis complicates the predictions of the proposed method [15].

Table 1: Healthcare based on Chronic Heart Disease

Author/Ref	Dataset	Learning Model	Concert Achieved	Limitations
A. Noor et al [16]	Imbalance dataset	Majority Weighted Minority Oversampling Technique (MWMOTE)	90%	Early identification is critical in recognising cardiac disease, which contributes significantly to worldwide death rates.
K. Raza et al [17]	Framingham Heart Study dataset	K Nearest Neighbors (KNN)	92.37%	Diagnosing heart disease is a considerable challenge.
Muhammad, Y et al [18]	Cleveland heart disease dataset	Fast Correlation-Based Filter (FCBF)	90%	Prompt diagnosis of patients is crucial to prevent further damage.
V. Chaurasia et al [19]	coronary illness dataset	Multilayer Perceptron (MLP)	92.03%	Heart failure can occur if the disease is not adequately treated.
D. Hassan et al [20]	Heart Disease dataset	Artificial Neural Networks (ANNs)	91.79%	Early detection is extremely important for effective prevention and treatment of heart disease.

Table 1 completely summarises the dataset detection, learning model, predicted results, and constraints in chronic heart disease within the health sector. This cardiovascular data is crucial for understanding the methodologies used in analyzing and forecasting health-related outcomes in training.

Table 2: Heart Disease Forecast Using the Feature Selection Method

Author	Year	Feature Selection	Classifier	Performance achieved
T. Ullah [21]	2024	LASSO	KNNBM	92%
Kaushalya Dissanayake [22]	2021	Recursive Feature Elimination (RFE)	Logistic Regression (LR)	88.52%
S. M. Saqlain [23]	2019	RFE	Random Forest Ensemble Method (RFEM)	93.08%
A.K.Gárate-Escamila [24]	2020	Mean Fisher Score-Based Feature Selection Algorithm (MFFSA)	SVM	81.19%
R.Sivaranjani [25]	2019	Maximal Redundancy Maximal Relevancy (MRMR)	KNN, SVM	83%

Table 2 summarises the performance characteristics of a previously described classifier that uses feature selection methodologies to predict the development of heart disease. Furthermore, the data offers a thorough analysis of the outcomes, evaluating how well the classifier correctly identified those at risk of heart disease.

Similarly, using the DNN algorithm to predict cardiovascular disease in critical areas can enhance the overall quality of cardiovascular disease classification [26]. They also offer medical information that can be extracted using the Natural Language Processing (NLP) technique. Similarly, phrase embedding can enhance the connection between diagnosis and medication treatments [27]. Decision trees (DT) and fuzzy rule-based techniques can create predictive models for diagnosing cardiac disease. Furthermore, employing multiple methods to diagnose heart disease patients leads to significant accuracy [28]. To classify the features efficiently, the GF-DE method is adopted for variable selection.

Further, using the proposed classification model, a neural network can quickly adjust multiple hyperparameters and determine suitable weights [29]. The dataset classifier employs Stochastic Gradient Boosting (SGB) for loss optimization with performance evaluation. However, traditional methods have proven ineffective in enhancing the precision of heart disease classification [30].

### III. PROPOSED METHODOLOGY

The suggested EGBLR algorithm will increase the accuracy and performance of chronic heart disease prediction. Additionally, the NMMSF method evaluates missing or duplicate values during data pre-processing. The IQRO technique can then place and remove outliers in the heart disease data. Similarly, the RFV algorithm can be used to find the weights of each dataset feature, optimizing the overall analysis process. The results suggest that the new EGBLR detection approach exceeds the accuracy of prior strategies, giving it a superior way to forecast chronic heart disease. The EGBLR algorithm significantly improves prognosis in heart disease and provides a comprehensive and efficient method to improve diagnostic accuracy.

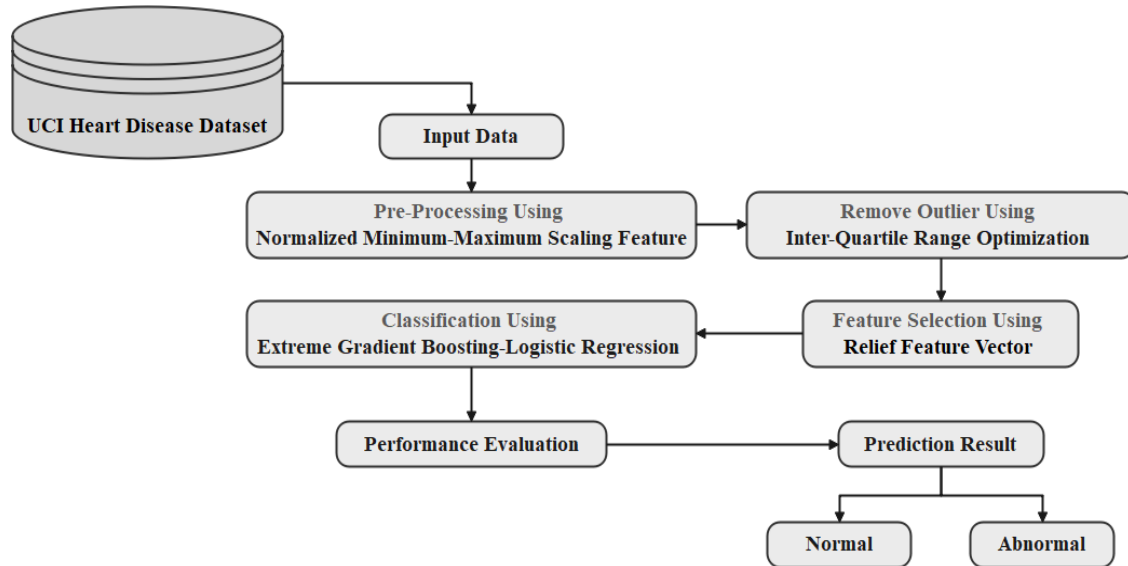


Fig. 2: The Proposed EGBLR Method Architecture Diagram

The findings in Figure 2 demonstrate that the diagram implementation of the EGBLR method surpasses the accuracy of conventional techniques in predicting chronic heart disease. This indicates that the proposed method is highly effective in improving cardiac disease prognosis. Enhancing diagnostic accuracy, the EGBLR method offers a valuable approach to identifying and managing heart conditions. Its ability to provide more precise predictions can lead to improved patient outcomes and better treatment strategies.

#### A. Data Collection

The heart disease dataset obtained from Kaggle would provide significant value in this section. By utilizing data from a dataset like Kaggle, their assessment can greatly enhance precision and dependability. The dataset consists of 76 attributes, covering both predicted and actual data. Within these attributes, the crucial "target" field denotes if the patient has heart disease, with 0 meaning no disease and 1 indicating its presence. In addition, the heart disease dataset can be easily accessed via the following URL: <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>.

Table 3: Dataset Feature Collection

<b>Id</b>	<b>Age</b>	<b>sex</b>	<b>cp</b>	<b>chol</b>	<b>restecg</b>	<b>trestbps</b>	<b>thalch</b>
1	63	Male	typical angina	233	lv hypertrophy	145	150
2	67	Male	asymptomatic	286	lv hypertrophy	160	108
3	67	Male	asymptomatic	229	lv hypertrophy	120	129
4	37	Male	non-anginal	250	normal	130	187
5	41	Female	atypical angina	204	lv hypertrophy	130	172
6	56	Male	atypical angina	236	normal	120	178
7	62	Female	asymptomatic	268	lv hypertrophy	140	160

The Kaggle Heart Disease dataset is the driving force behind the valuable collection of dataset attributes in Table 3. Estimate accuracy and dependability may be significantly increased using data from websites like Kaggle.

B. *Normalized Minimum-Maximum Scaling Feature (NMMSF)*

The normalized minimum-maximum scaling feature method can identify missing or duplicate values during data pre-processing. Furthermore, the NMMSF approach may pre-process data and identify missing values. Similarly, the performance of the analysis can be enhanced by filling in the missing values with constants or the dataset's average value. Datasets with increasing dimension values represent age, treetops, chol, and old peak features. Moreover, the NMMSF method analyses standard scalar features by converting eigenvalues to standard deviations. Furthermore, the NMMSF approach may be used for all features to normalize them, estimate significant eigenvalues, and find the values of the feature dataset by using the min-max range for normalization.

Equation 1, the dataset's missing values and redundant data are evaluated with the mean value. Let us assume the S-mean value and T-total data are in the dataset and represent the data row.

$$S(g) = \frac{\sum_{a=0}^t g^a}{t} \tag{1}$$

Two main processes are involved in feature scaling, namely normalization and standardization. Standardization is performed by subtracting the mean from the variance in the distribution and dividing it by the standard deviation. Calculate the normalization values as shown in Equation 2. Where  $g'$  -participation value,  $g'$  -standardized value,  $\mu$  -mean value,  $\sigma$  -standard deviation.

$$g' = \frac{g - \mu}{\sigma} \tag{2}$$

Calculate the standard deviation of the standardized mean value as presented in equation 3.

$$\mu = \frac{\sum_{a=0}^t a}{t} \tag{3}$$

Calculate the dataset's total number of columns. The age, trestbps, chol, and oldpeak features available from the datasets have the most significant dimension values. Estimate the standard scalar to convert the eigenvalues into a uniform scaling factor, as illustrated in Equation 4.

$$\sigma = \sqrt{\frac{1}{J} \sum_{a=1}^t (g_a - \mu)^2} \tag{4}$$

After scaling the large eigenvalues, normalize them using the min-max scale. This process confines the eigenvalues to intervals between the minimum and maximum values. Equation (4) is employed for the minimum scaling normalization. Let's assume  $G_{ma} - G_{mi}$  -minimum and maximum value.

$$g' = \frac{g - G_{mi}}{G_{ma} - G_{mi}} \tag{5}$$

The dataset's lowest and maximum values for each feature are used to normalize all features, and an estimate may be made for the missing value. This normalization process can analyze the precise values of different scaling features.

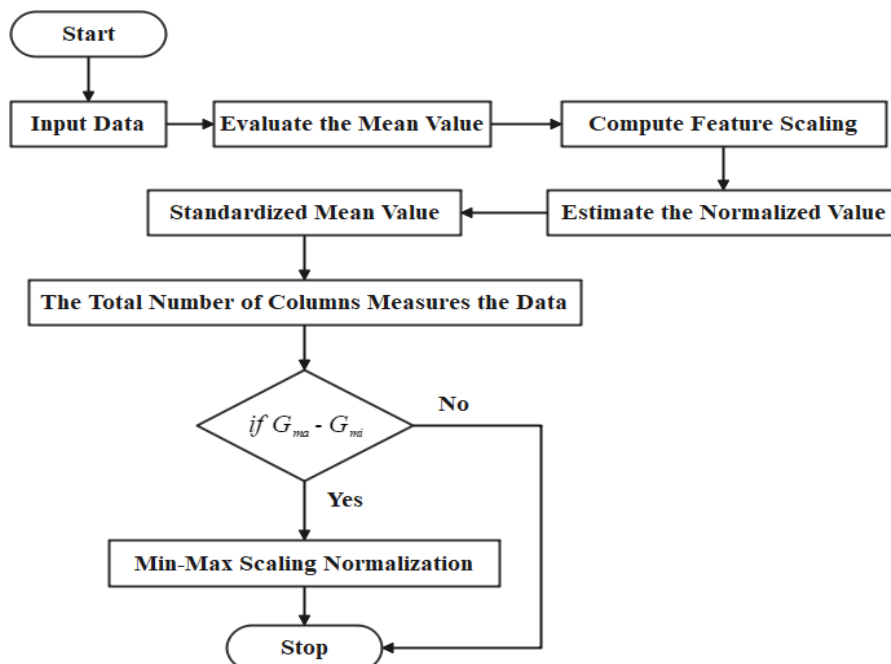


Fig. 3: Normalization Minimum-Maximum Scaling Feature Flowchart Diagram

The minimum-maximum scaling feature normalization method can detect missing values in a dataset and normalize the features using minimum-maximum scaling, as shown in figure 3.

C. *Inter-Quartile Range Optimization (IQRO)*

Cardiac data can be effectively analyzed using temporal threshold optimization techniques to detect and accurately remove outliers in cardiac disease datasets. The IQRO method quantifies chronic heart disease prevalence data by calculating the difference between the optimization of the third and first data quarter ranges. The IQRO method detects and removes outliers in heart disease prediction models. In addition, the lower and upper bounds can be estimated by optimizing the mean difference between the third and first intermediate bounds. Likewise, the IQRO technique utilizes chol and trestbps columns, which are generally assumed to pursue a normal distribution.

Furthermore, using a box plot, both columns exhibit anomalies outside the range of expected values. Data standardization transforms the original data into a customarily distributed format without varying the design of the actual data distribution. Eliminating outliers using the IQRO approach can make models for predicting chronic heart disease more accurate and reliable.

When evaluating heart disease data, it can remove outliers in the dataset and predict the difference between the third and first interim ranges. The lower and upper limits can be computed using equations 6 to 8 provided.  $A^k$  –Inter range,  $V$ -quartile,  $j$ -bound,  $V_1$  –first quartile,  $V_3$  –third quartile.

$$A^k = V_3 - V_1 \tag{6}$$

$$J_\ell = V_1 - 1.5 * \mathcal{A}_v(\mathcal{K}) \tag{7}$$

$$B_u = V_3 + 1.5 * A_v(\ell) \tag{8}$$

As shown in equation 9, remove outliers in the calculation range. Let’s assume  $J_L$  –lower bound,  $J_p$  –upper bound, and  $\ell_R$  –outside range.

$$\ell_{\hat{k}} = J_\ell \sim J_p \tag{9}$$

As shown in equations 10 and 11, clean the data and remove outliers. Where  $F$ -data,  $\sigma^\ell$  –outlier,  $p_j$  –upper bound,  $L_j$  –lower bound,  $S$ -mean,  $t$ -standard deviation.

$$F = \begin{cases} \sigma^\ell > p_j \\ \sigma^\ell < L_j \end{cases} \tag{10}$$

$$O^\ell - S > |3 * m| \tag{11}$$

The accuracy of chronic heart disease prediction models can be enhanced by translating raw data into a more broad distribution pattern and modifying the distribution.

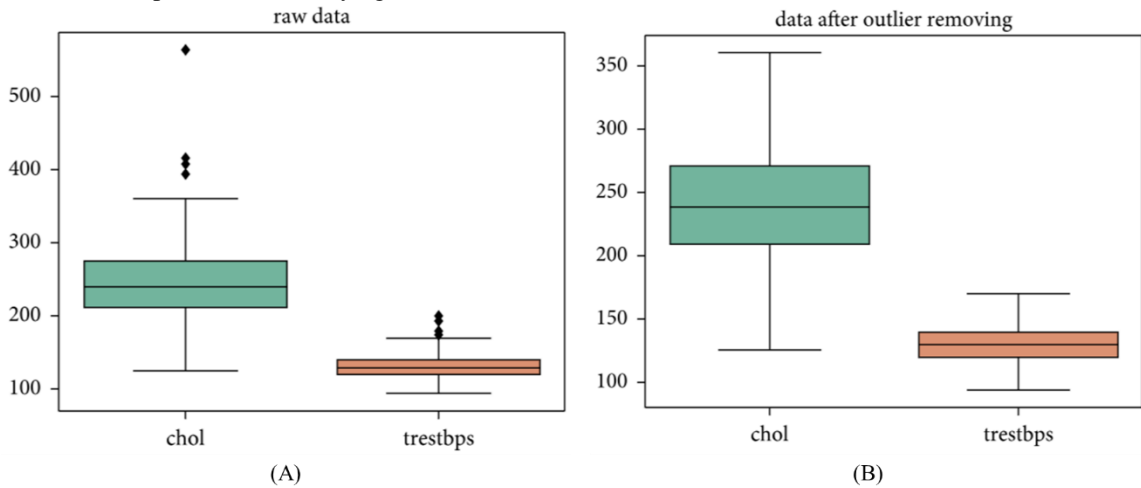


Fig. 4: Boxplot with Outlier Removal Using IQRO Before and After

Figure 4 shows (A) the raw data and (B) the outlier removal boxplot of the chronic heart disease dataset before and after outlier removal.

D. *Relief Feature Vector (RFV)*

The relief feature vector algorithm can find weights assigned to individual features in a dataset. The RFV algorithm in data analysis processes can significantly contribute to the optimization of feature weighting and

ultimately enhance the overall effectiveness of the analysis. The relief method applies weights to each dataset's characteristics and automatically updates them. The heart disease dataset can be utilized to identify features with high-weight values. Likewise, relief can be iteratively assessed by selecting variations with random training samples. Moreover, the target model and weights can be updated for each heart disease data feature vector.

Algorithm 1: RFV

Input: Outlier removal  $O^{\ell}$

Output: Update each feature weight  $Z$

Begin

Evaluate the total number of training samples  $\leftarrow t$

Calculate the feature dimension  $\leftarrow F$

If the feature weight set  $z[I] \leftarrow 0.0$

For each  $R \leftarrow 1$  to  $S$  do,

Select the target samples randomly  $\leftarrow K_R$

Find the nearest hits-and-miss  $\leftarrow Y_S$

For each  $I \leftarrow 1$  to  $a$  do,

    Compute the weight of each feature vector

$$z[I] \leftarrow z[I] - d_{if}(I, K_R, Y) / s + d_{if}(J, K_R, S) / s \tag{12}$$

End for each

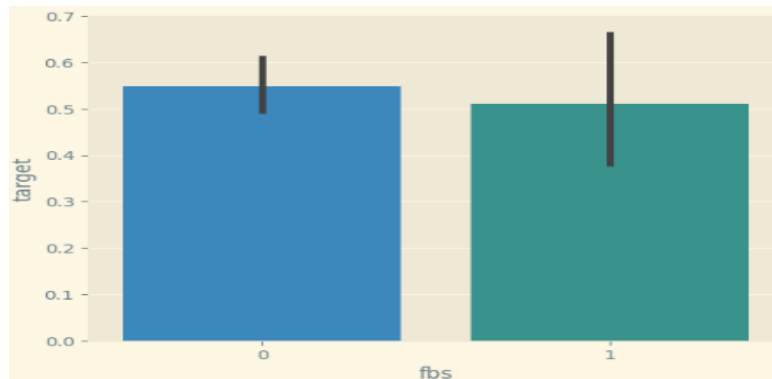
End for each

End if

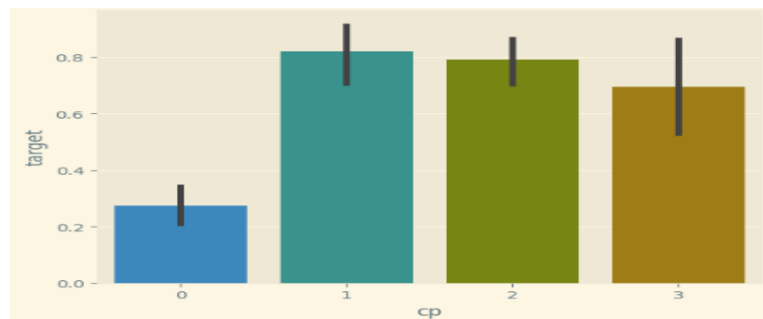
Return  $\leftarrow$  Update feature weight  $Z$

End

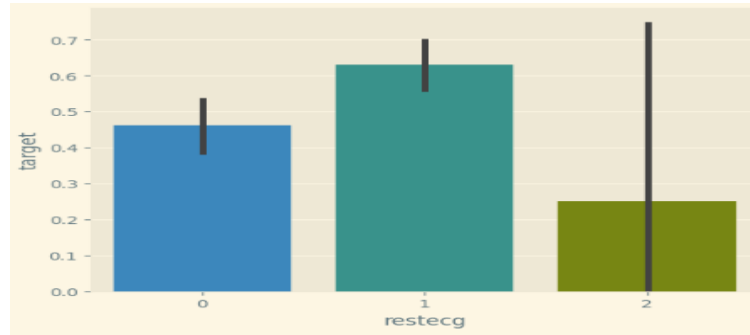
Also, the accuracy of data related to heart disease data can be improved by analyzing and updating the target model and weights of each feature vector. Let's assume  $Z$ -Weight,  $F$ -hit value,  $S$ -miss value,  $Z[I]$  -feature weight set,  $d_{if}$  -difference sample,  $K_R$  -target sample,  $t$ -number of training sample,  $f$ -feature dimension.



(A) FBS- Fasting Blood Sugar



(B) CP- Chest Pain



(C) RESTECG-Resting Electrocardiographic

Fig. 5: Feature Selection Target Sample Data

Figure 5 illustrates that feature weights from the heart diseases dataset, such as CP, thal, exchange, and restecg, may be updated for the target sample.

*E. Extreme Gradient Boosting-Logistic Regression (EGBLR)*

The EGBLR algorithm can predict normal or abnormal chronic heart disease and improve accuracy. Besides, a gradient-boosting classifier utilizing the EGBLR algorithm may be employed to forecast normal or abnormal chronic heart disease patients and enhance performance accuracy. It may be retrained using the heart disease dataset to infer covariance relationships between leaf nodes and row characteristics. Thus, it is a logistic gradient-boosting technique to enhance tree classifiers' efficiency. Furthermore, it frequently employs improved logistic regression to minimize logistic loss. Moreover, the EGBLR technique typically indicates a bias towards zero, which can lead to overfitting. Therefore, assigning new weights to the EGBLR method can enhance the accuracy of cardiac data and categorize normal and abnormal patients.

Algorithm 2: EGBLR

Input: Update Feature Weight Z

Output: Heart disease prediction  $\mathcal{O}_p$

Begin

Calculate the number of heart disease patients for each tree weight in the base classifier

$$d(i) = \sum J_R(i) \tag{13}$$

For each  $R = 1, 2, \dots, S$

Calculate the gradient number of iterations

$$X_R = \frac{\delta \mathcal{L}(h, d)}{\delta_d} \tag{14}$$

Evaluate the structure by choosing the split with the highest gain.

$$I = \frac{1}{2} \left[ \frac{x_\ell^{(2)}}{y_\ell} + \frac{x_R^{(2)}}{y_R} + \frac{x^2}{y} \right] \tag{15}$$

Calculate the weight of the leaves.

$$Z^* = \frac{x^2}{y} \tag{16}$$

For each weight vector  $z = \frac{1}{T} \quad a = 1, 2, \dots, T$

If fit function  $D_{(g)} = 0$

If Probability  $u(g_a) = 1/2$

Compute the weight for each iteration

$$Z_a = u(G_a) / (1 - u(g_a)) \tag{17}$$

End if

End if

End for each

End for each

Determination of base classifier data

$$J_g = \sum N (b - 1) z^a \tag{18}$$

Update the function to calculate the least square weight

$$\text{Heart disease } \mathcal{O}_p = [D_{(g)} = \sum [D_{\mathcal{R}}(g)]] \tag{19}$$

Return  $\leftarrow \mathcal{O}_p$

End

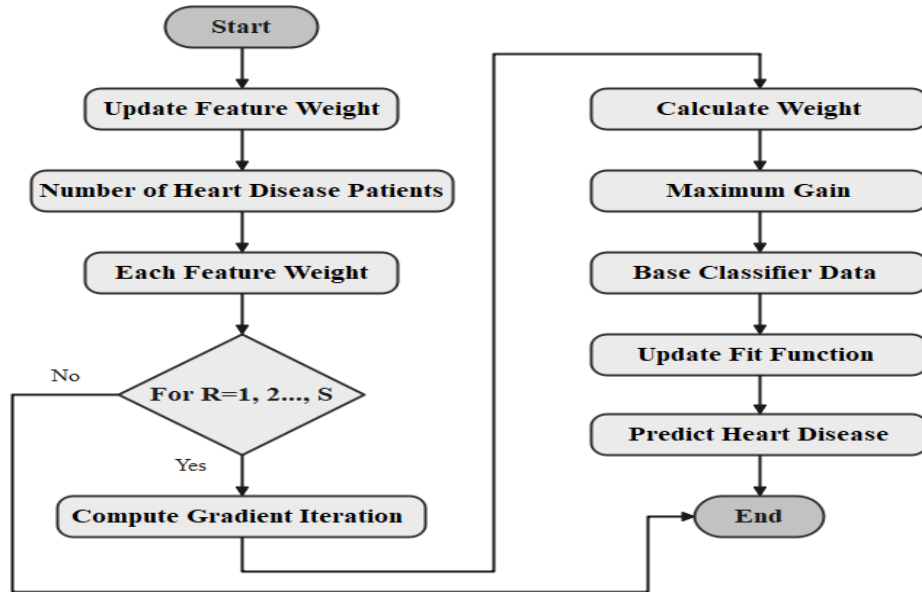


Fig. 6: Extreme Gradient Boosting-Logistic Regression Based on flowchart Diagram

This model calculates tree weights using the EGBLR method in Algorithm 2 to predict and classify heart disease patients at each iteration. Let's assume  $Z$ -weight,  $R$ -number of iteration,  $D(i)$  –number of heart disease patients,  $Y_R(g)$  –represent base classifier each tree,  $x_R$  –gradient iteration,  $L$  –loss function,  $g$ -gradient,  $D(g)$  –fit function,  $S$ -number of base classifier heart disease patients,  $X$ -gain,  $\frac{x^2}{y}$  –leaf weight,  $N$ -tree,  $G$ -feature vector,  $T$ -data,  $O_p$  –heart disease output,  $u(g_a)$  –eqch iteration weight.

Figure 6 illustrates the flowchart utilizing the EGPLR method to predict chronic heart disease patients to enhance the model's accuracy. Further, the squared weights of the base classifier data were computed for the update function.

#### IV. RESULT AND DISCUSSION

The EGBLR algorithm was recently introduced and implemented using the heart disease dataset in Python. An optimal classifier can be identified by improving existing classifiers for predicting heart disease. Indicators like recall, precision, accuracy, specificity, and false positives are crucial when assessing classifier performance. The accuracy of the classifier can be measured in predicting heart disease. Using a confusion matrix, these metrics can also be computed using true negatives, false positives, and false positives. It is evident from the numbers in this matrix that the classifier correctly identifies instances of heart disease. Furthermore, the EGPLR algorithm demonstrated superior results in heart disease prediction.

Table 4. Simulation parameter

Parameter	Value
Dataset name	Heart Disease Predictions
Number of records	921
Language	Python
Tool	Jupyter
Training data	723
Testing data	198

As shown in table 4, the simulation parameters can be tested and trained from the data obtained from the heart disease dataset using the Python language-based Jupyter Notebook.

Table 5: Performance Matrices

Performance	Matrix
Specificity	$\frac{t^n}{t^n + f^p}$

Recall (rec)	$\frac{t^p}{t^p + f^n}$
False Score	$\frac{2 * pre * rec}{pre + rec}$
Precision (pre)	$\frac{t^p}{t^p + f^p}$
Accuracy	$\frac{t^p + t^n}{t^p + t^n + f^p + f^n}$

Table 5 describes the proposed method’s accuracy in predicting and identifying heart diseases based on performance metrics derived from testing data. Further, the performance matrix is represented as  $t^p$  –true positive,  $f^p$  –false positive,  $t^n$  –true negative, and  $f^n$  –false negative.

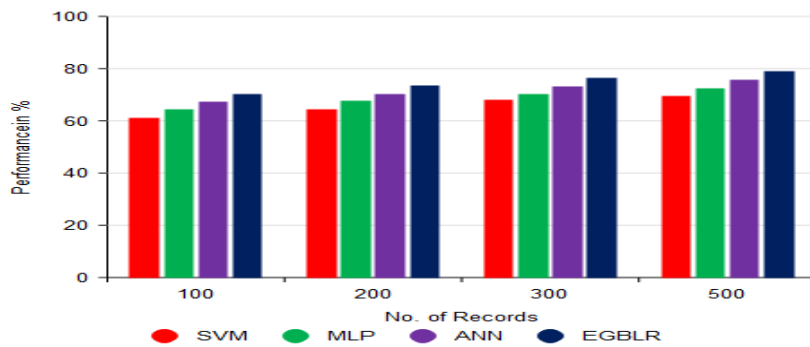


Fig 7: Evaluation of Specificity

As shown in Figure 7, the suggested EGBLR technique based on specificity analysis may properly predict heart disease. Furthermore, compared to approaches such as SVM, ANN, and MLP obtained in the literature, the suggested method has a lesser accuracy of 69.46% in specific analysis. Furthermore, the specific analytical value of the proposed method in predicting heart disease increased to 79.18%.

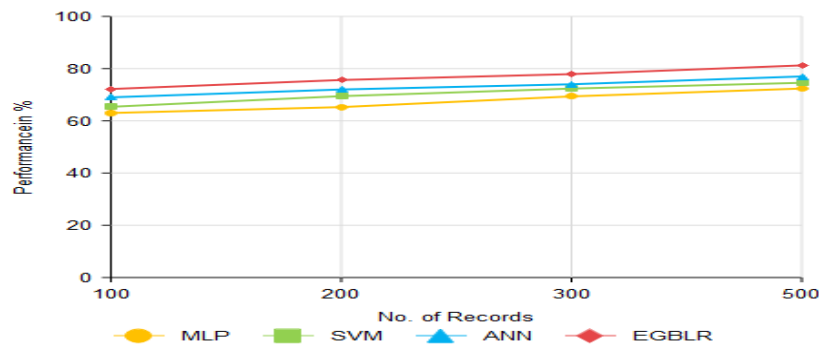


Fig 8: Analysis of Recall

In Figure 8, the efficacy of predicting heart disease can be notably enhanced using the proposed EGBLR method, especially regarding recall performance. Furthermore, the comparative analysis with established techniques such as SVM, ANN, and MLP from existing literature, where the proposed method demonstrates a lower accuracy rate of 72.36% in recall analysis. However, it is imperative to highlight that the recall performance metric of the proposed method in predicting heart disease has shown a significant improvement, reaching 81.23%.

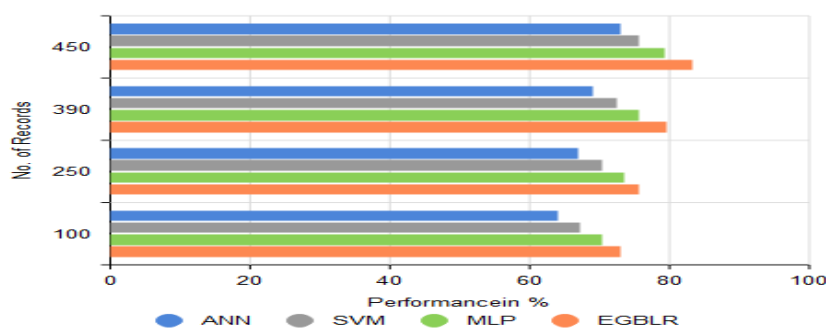


Fig 9: Analysis of Precision

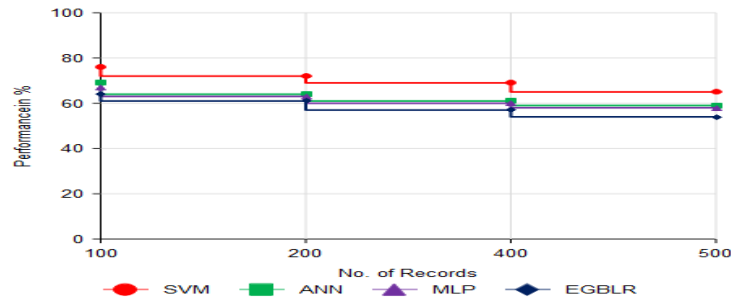


Fig. 10: Analysis of False Rate

Figure 9 demonstrates that the suggested EGBLR approach may greatly enhance heart disease prediction performance, especially precision performance. Nevertheless, it indicates that the accuracy of the proposed method has improved by 83.14% in predicting heart disease. Furthermore, in the comparative analysis with the existing methods such as MLP, SVM, and ANN in the current literature, the precision performance of the proposed method is 73% low.

The proposed EGBLR method can predict false rates using datasets obtained from Kaggle, significantly improving the performance of heart disease forecasting, as illustrated in Figure 10. Furthermore, using the proposed method to predict heart disease, the accuracy of single methods of MLP, ANN, and SVM analysed in the literature increased to 65%. Likewise, the false rate analysis of the proposed systematic EGBLR method is as low as 54%.

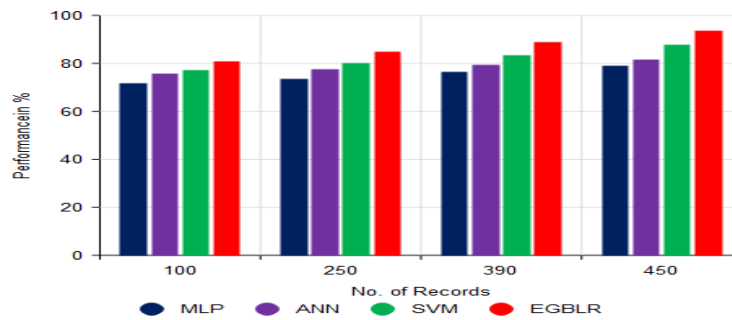


Fig. 11: Analysis of Accuracy

As shown in Figure 11, the proposed EGBLR method demonstrates a significant advancement in predicting accuracy performance based on datasets sourced from Kaggle, particularly in enhancing the efficacy of forecasting heart disease. By utilizing this innovative approach for heart disease prediction, it becomes evident that the accuracy levels achieved surpass those of traditional methods such as MLP, ANN, and SVM, as reported in existing literature, where the accuracy stands at a modest 79.18%. In contrast, the systematic application of the EGBLR method reveals a remarkable accuracy rate of 93.64%, underscoring its superiority in predictive analytics for heart disease. This substantial improvement in accuracy showcases the EGBLR method’s potential and highlights its practical relevance in enhancing healthcare outcomes through more precise and reliable disease forecasting techniques.

## V. CONCLUSION

This study provides an effective chronic heart disease prediction based on optimum feature selection utilising RFV and EGBLR classification. These procedures perform outlier removal and data classification with pre-processing and feature selection to effectively predict heart disease. First, the NMMSF approach is used in the data pre-processing step to evaluate missing values. After the pre-processing, an outlier removal phase is performed using the IQRO algorithm. Similarly, the RFV method selects from all features and finds each feature’s weight. Finally, the EGBLR method based on the proposed ML technique can detect normal or abnormal chronic heart diseases and improve analytical accuracy. The application of the EGBLR method in the healthcare field shows an impressive accuracy of 93.64%, which highlights its superiority in the predictive analysis of chronic heart disease. Besides, the offered EGBLR technique effectively improves the outcomes of biomedical health applications through more accurate heart disease prediction systems.

## REFERENCE

- [1] Singh P, Singh S, Pandi-Jain GS. Effective heart disease prediction system using data mining techniques. *Int J Nanomedicine*. 2018 Mar 15;13(T-NANO 2014 Abstracts):121-124.
- [2] Hossain, M.I., Maruf, M.H., Khan, M.A.R. et al. Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison. *Iran J Comput Sci* 6, 397–417 (2023).
- [3] Farman Ali, Shaker El-Sappagh, "A Smart Healthcare Monitoring System for Heart Disease Prediction Based on Ensemble Deep Learning and Feature Fusion", June 2020 *Information Fusion* 63.
- [4] Shadman Nashif, Md. Rakib Raihan, Md. Rasedul Islam, Mohammad Hasan Imam, "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System," *World Journal of Engineering and Technology* > Vol.6 No.4, November 2018.
- [5] Ahmad, A.A.; Polat, H. Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm. *Diagnostics* 2023, 13, 2392.
- [6] A. M. Khedr, Z. A. Aghbari, A. A. Ali and M. Eljamil, "An Efficient Association Rule Mining from Distributed Medical Databases for Predicting Heart Diseases," *IEEE Access*, vol. 9, pp. 15320-15333, 2021.
- [7] W. Zhang, W. Cheng, K. Fujiwara, R. Evans, and C. Zhu, "Predictive Modeling for Hospital Readmissions for Patients with Heart Disease: An updated review from 2012-2023," *IEEE Journal of Biomedical and Health Informatics*.
- [8] T. Wu et al., "A Precision Health Service for Chronic Diseases: Development and Cohort Study Using Wearable Device, Machine Learning, and Deep Learning," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, pp. 1-14, 2022, Art no. 2700414.
- [9] S. A. Ali et al., "An Optimally Configured and Improved Deep Belief Network (OCI-DBN) Approach for Heart Disease Prediction Based on Ruzzo–Tomba and Stacked Genetic Algorithm," *IEEE Access*, vol. 8, pp. 65947-65958, 2020.
- [10] A. Kumar, K. U. Singh and M. Kumar, "A Clinical Data Analysis Based Diagnostic Systems for Heart Disease Prediction Using Ensemble Method," *Big Data Mining and Analytics*, vol. 6, no. 4, pp. 513-525, December 2023.
- [11] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," *IEEE Access*, vol. 8, pp. 107562-107582, 2020.
- [12] A. Bhardwaj, S. Singh and D. Joshi, "Explainable Deep Convolutional Neural Network for Valvular Heart Diseases Classification Using PCG Signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-15, 2023, Art no. 2514215.
- [13] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System," *IEEE Access*, vol. 8, pp. 133034-133050, 2020.
- [14] Kaan Uyar, Ahmet İlhan, "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks," *Procedia Computer Science*, Volume 120, 2017, Pages 588-593, ISSN 1877-0509.
- [15] D. Q. Zhang, Y. Y. Chen, Y. X. Chen, S. Y. Ye, W. Y. Cai, J. X. Jiang, et al., "Heart disease prediction based on the embedded feature selection method and deep neural network," *Journal of Healthcare Engineering*, vol. 2021, pp. 6260022, 2021.
- [16] A. Noor, N. Javaid, N. Alrajeh, B. Mansoor, A. Khaqan and S. H. Bouk, "Heart Disease Prediction Using Stacking Model with Balancing Techniques and Dimensionality Reduction," *IEEE Access*, vol. 11, pp. 116026-116045, 2023.
- [17] K. Raza, *Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule, in U-Healthcare Monitoring Systems*. (Academic Press, Cambridge, 2019), pp.179–196
- [18] Muhammad, Y., Tahir, M., Hayat, M. et al. Early and accurate heart disease detection and diagnosis using an intelligent computational model. *Sci Rep* 10, 19747 (2020).
- [19] V. Chaurasia and A. Chaurasia, "Novel method of characterization of heart disease prediction using sequential feature selection-based ensemble technique," *Biomed. Mater. Devices*, Jan. 2023.
- [20] D. Hassan, H. I. Hussein, and M. M. Hassan, "Heart disease prediction based on pre-trained deep neural networks combined with principal component analysis," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023.
- [21] T. Ullah et al., "Machine Learning-Based Cardiovascular Disease Detection Using Optimal Feature Selection," *IEEE Access*, vol. 12, pp. 16431-16446, 2024.
- [22] Kaushalya Dissanayake, Md Gapar Md Johar, "Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms," *Applied Computational Intelligence and Soft Computing*, vol. 2021, Article ID 5581806, 17 pages, 2021.

- [23] S. M. Saqlain, M. Sher, F. A. Shah, I. Khan, M. U. Ashraf, M. Awais, et al., "Fisher score and Matthew's correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines," *Knowl. Inf. Syst.*, vol. 58, no. 1, pp. 139-167, Jan. 2019.
- [24] A. K. Gárate-Escamila, A. H. El Hassani and E. Andrès, "Classification models for heart disease prediction using feature selection and PCA," *Informat. Med. Unlocked*, vol. 19, Jan. 2020.
- [25] R. Sivaranjani, V. S. Naresh and N. V. Murthy, "4 coronary heart disease prediction using genetic algorithm-based decision tree", *Intell. Decis. Support Syst. Appl. Signal Process.* vol. 4, pp. 71, Oct. 2019.
- [26] Mahesh Parmar, "Heart Diseases Prediction using Deep Learning Neural Network Model," *International Journal of Innovative Technology and Exploring Engineering* 9(3):2244-2248, 2020.
- [27] Houssein, E.H., Mohamed, R.E. & Ali, A.A. Heart disease risk factors detection from electronic health records using advanced NLP and deep learning techniques. *Sci Rep* 13, 7173 (2023).
- [28] Pathak, A.K., Arul Valan, J. (2020). A Predictive Model for Heart Disease Diagnosis Using Fuzzy Logic and Decision Tree. In: Elçi, A., Sa, P., Modi, C., Olague, G., Sahoo, M., Bakshi, S. (eds) *Smart Computing Paradigms: New Progresses and Challenges*. *Advances in Intelligent Systems and Computing*, vol 767. Springer, Singapore.
- [29] D. Deepika & N. Balaji (2022) Effective heart disease prediction with Grey-wolf with Firefly algorithm-differential evolution (GF-DE) for feature selection and weighted ANN classification, *Computer Methods in Biomechanics and Biomedical Engineering*, 25:12, 1409-1427.
- [30] Jawalkar, A.P., Swetcha, P., Manasvi, N. et al. Early prediction of heart disease with data analysis using supervised learning with stochastic gradient boosting. *J. Eng. Appl. Sci.* 70, 122 (2023).