Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

Employing Machine Learning with Discriminative Function for Predict Heart Diseases

Firas Abdulrahman Yousif[a], Rasha Raad Al-Mola[b], Muthanna Subhi Sulaiman[c]

[a]*Department of Computer Science, University of Al-Hamdaniya, Mosul, Iraq. Email: firasabdulrahman@uohamdaniya.edu.iq*

[b]*Department of Computer Science, University of Al-Hamdaniya, Mosul, Iraq. Email: rasharaad@uohamdaniya.edu.iq*

[c]*Department of Statistics and Informatics, University of Mosul, Mosul, Iraq. Email: muthanna.sulaiman@uomosul.edu.iq*

ARTICLE INFO

Article history:
Received: 22 /06/2025
Rrevised form: 03/07/2025
Accepted : 28/07/2025
Available online: 30/09/2025

Keywords:

Heart disease

Machine learning

Discriminant Function

Prediction

ABSTRACT

Heart disease is also one of the leading causes of death and therefore there is need to have a readily accurate way of diagnosing it. Based on this study, a strong basis of machine learning based on predicting heart disease has been proposed with the combination of ensemble learning algorithms (XGBoost, Random Forest, Gradient Boosting, and Extra Trees) and classic discriminant analysis based (Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA)). The models were tested with two test benchmark datasets following rigorous pre-processing and feature engineering. The experimental findings show that XGBoost performed best in the first dataset with the accuracy of 98.54, perfect precision of 100 and the F1 score of 0.985. The best results by Random Forest on the second dataset, compared to the rest were 94.96 and the F1 score of 0.955. Comparatively, LDA achieved 82.44 and 87.39, accuracy rates out of the first and the second dataset respectively, whereas, QDA did not do as well in 54.15 and 44.96 respective accuracy levels. The results provided clearly indicate that compared to alternative discriminant models, machine learning models considerably outperformed the discriminant ones, which also shows that the XGBoost model would be the most suitable option in classifying the data. The outcomes confirm the usefulness of machine learning as a reliable and accurate diagnostic aid in detecting preliminary heart diseases in the clinical setting.

MSC..

https://doi.org/10.29304/jqcsm.2025.17.32379

∗Corresponding author: Firas Abdulrahman Yousif

Email addresses: firasabdulrahman@uohamdaniya.edu.iq

Communicated by ‘sub etitor’

1. Introduction

During the last decade, heart disease remains to be top of the list for most fatal diseases worldwide. Cardiovascular diseases are responsible for about 17.9 million deaths each year, and more than 80% of these related to coronary heart disease and cerebrovascular analysis [1]. Ninety percent of the deaths are in low and middle-income countries [2]. A number of factors are responsible for developing heart diseases including lifestyle choices people make due to their personal as well professional life, genetic factor and behavioral risks such as smoking, excessive alcoholism or caffeine in-take, stress and physical inactivity. Physical conditions including obesity, hypertension, a history of heart disease or high cholesterol are also important risks.

Scientists have been working on better diagnostic options for HD, in part because many of the current methods are inaccurate, or take too long to process and return a diagnosis [3]. In region without medical technology and expertise are unable to diagnosis, treat for heart disease becomes very difficult [4]. Diagnostics and therapy should be efficient to save lives given the probable number of deaths. In 2023, the global population is approaching 8 billion, with around 620 million people experiencing cardiac problems worldwide. About 60 million new heart disease cases are identified annually [5] [6]. According to the British Heart Foundation, 1 in 3 people die from heart disease in 2021, or 20.5 million per 1.5 seconds [7] [8].

The traditional method of HD diagnosis is by looking at the history and then a standard physical examination assessing specific symptoms, but these methods are usually subjective with poor specificity or require complicated tools which none-the-less can be inaccurate.

Technology and data, in a world of artificial intelligence [9] have allowed progress to be made through the integration with machine learning (ML) features capable to improve detection capabilities on an automatic way. LDA and QDA are among the techniques applied for predicting heart disease before it occurs [10]. Such detection systems can decrease misdiagnosis by significantly helping medical professionals. In addition to this, these systems enable the fast and accurate assessment of medical data.

Healthcare is one of the most affected sectors where ML has played a key role in Heart disease management. In terms of patient-specific data, risk prediction is heavily dependent on its functioning. ML models can detect diseases based on patterns it sees in medical data. ML algorithms are also being used to integrate imaging data from MRI and CT scan. These algorithms can be implemented within clinical decision support systems and are used to personalize the care of individual patients. An additional example is the use of ML to improve wearable technology, which can monitor a patient's vital signs around-the-clock and identify any anomalies with such range in detail that significantly de- creases hospital readmissions. Furthermore, ML algorithms help track possible compounds to treat heart diseases in the field of drug discovery [11]. Basic ML methods for data collection, feature selection, and preprocessing prepare the initial inputs used in ML algorithms. It is crucial to choose the right form of algorithms such as supervised, ensemble learning efficiently [12]. The efficiency of these algorithms is then evaluating using evaluation metrics. Many of these algorithms have been validated to predict heart disease in numerous studies.

The ultimate goal of this article is to analyze how different ML method performs relatively with LDA and QDA in terms both on how they work all together alongside a dataset containing heart disease relational data. This aims to find out which one of those ML methods can predict if someone has heart disease and is the better diagnostic method. This study aims to provide insights in improving predictive models of the medical diagnostics by tracking results performance, accuracy and also investigating ML computational efficiency that could lead to better treatment outcomes for patients with heart disease.

The primary contributions of this study lie in the comprehensive comparative analysis of advanced machine learning models namely XGBoost, Random Forest, Gradient Boosting, and Extra Trees alongside classical discriminant analysis techniques, including LDA and QDA, for the task of heart disease prediction using two publicly available benchmark datasets. The evaluation encompasses a wide range of performance metrics such as accuracy, precision, recall, F1-score, error rate, and training time to ensure a holistic assessment of each model's diagnostic capability. Among the machine learning methods, XGBoost emerges as the most effective, achieving an outstanding accuracy of 98.54% and significantly outperforming traditional statistical approaches. While both discriminant analysis models are outperformed by ensemble learning techniques, LDA consistently demonstrates superior results compared to QDA, indicating its potential as a computationally efficient alternative in resource-constrained

clinical settings. Overall, the study offers actionable insights and practical guidance for implementing machine learning-based diagnostic systems aimed at enhancing the accuracy and reliability of early heart disease detection in healthcare environments.

Section 2 reviews recent related work on heart disease prediction using ML and DA techniques. Section 3 details the proposed methodology, including datasets, preprocessing, and the machine learning and discriminant models applied. Section 4 presents and analyzes the experimental results. Section 5 discusses the implications of the findings and compares our results to existing studies. Finally, Section 6 concludes the study and outlines directions for future research, including potential extensions using deep learning and explainable AI (XAI) approaches.

2. Related works

Heart Failure (HF) risk variables and survival were analyzed using ML in [13]. The study included five supervised ML methods: Decision Tree (DT), DT Regressor, Random Forest (RF), XGBoost, and Gradient Boosting (GB). RF had the best accuracy (97.78%). Prognostic variables include serum creatinine, age, ejection fraction, platelets, and phosphokinase. Serum creatinine, age, and ejection fraction clustered to predict HF prognosis. The results indicate that these ML models may reliably predict survival and can be used to screen HF patients.

Authors in [14] suggested that given the growing global prevalence and mortality rate of heart disease, ML can be used to improve prediction. Recognizing the importance of effective diagnostics, this study investigates several ML models are Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Naive Bayes (NB), DT, RF, and Logistic Regression (LR) that categorize risk levels of patients and predict whether they are at risk for heart disease. According to the paper, DT gives slightly greater accuracy approximately 98% this very huge data and complexity that is a really good news for ML methods in healthcare.

Researchers in [15] proposed advancing the application of ML to improve early detection of heart disease, which is responsible for nearly one-third of all deaths globally. They used the Classification and Regression Tree (CART) algorithm, a supervised ML method, to predict heart illness and create decision rules that explain input-output correlations. The study also classified heart disease risk factors by relevance, supporting the model's 87% accuracy. Additionally, the study's decision principles can be applied clinically to simplify diagnosis without additional knowledge.

Researchers in [16] investigated cardiovascular diseases using a dataset with 11 heart disease-relevant features. They evaluated three ML classifiers: LR, KNNs, and RF across five datasets including Cleveland and Hungarian. Both LR and RF demonstrated a notable 88% accuracy, underscoring their effectiveness in detecting heart disease.

Researchers in [17] proposed using data analytics and ML to improve heart disease detection and diagnosis, noting that 90% of heart diseases are preventable. The study applied algorithms are DT, NB, RF, SVM, KNN, and LR.

RF emerged as the most effective model, achieving 83.52% accuracy, an F1-score of 84.21%, AUC of 88.24%, and precision of 88.89%, demonstrating its superior predictive performance.

Researchers in [18] proposed using LR to classify cardiovascular disease, a leading cause of 17 million deaths globally. The study applied LR to the UCI dataset, with pre- processing steps including cleaning, handling missing values, and selecting highly correlated features. Among various train-test splits, the 90:10 split achieved the highest accuracy of 87.10%.

In [19], researchers suggested utilizing DL algorithms to predict coronary artery disease, a common cardiac issue characterized by plaque formation. Preventing disease development requires early diagnosis. DTs, NB, and ANN are used to identify heart abnormalities in the study. The ANN model was most accurate at 84.4%, followed by SVM (83.33%), RF (81.67%), and DT (73.33%).

Table 1 illustrate the summarization of this literature review.

Table 1 - Applications in heart disease

References	Methods	Dataset	Results
[13]	Decision Tree, DT Regress or, Random Forest, XGBoost, Gradient Boosting	Heart Failure Patients	RF achieved highest accuracy of 97.78%
[14]	SVM, KNN, Naive Bayes, DT, RF, Logistic Regression	–	DT shows accuracy of approx. 98%
[15]	Classification and Regression Tree (CART)	–	Accuracy of 87%
[16]	Logistic Regression, KNNs, RF	Cleveland and Hungarian datasets	LR and RF both showed 88% accuracy
[17]	DT, NB, RF, SVM, KNN, LR	–	RF had 83.52% accuracy, 84.21% F1-score, 88.24% AUC, 88.89% precision
[18]	Logistic Regression	UCI dataset	Highest accuracy of 87.10% in a 90:10 split
[19]	DTs, NB, ANN, SVM, RF	–	ANN highest accuracy at 84.4%, followed by SVM, RF, and DT

3. Proposed heart disease approach

Figure 1 depicts the proposed ML-based heart disease prediction model. Clean and partition the heart disease dataset into training and testing sets. ML methods like XGBoost, Extra Trees, RF, and GB construct prediction models from training data. Comparative analysis also uses LDA and QDA. Testing the trained models yields the final heart disease prediction.

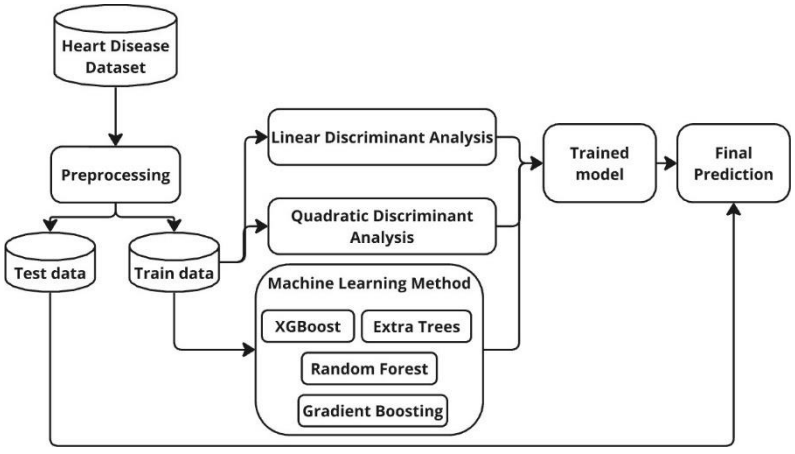

Fig. 1- Proposed Heart Disease Approach

3.1 Heart Disease Dataset Description

Kaggle 1's 1988 "Heart Disease Dataset" was updated five years ago. Cleveland, Hungary, Switzerland, and Long Beach V data take up 1025 rows and 14 columns. Most research focused on 14 of the 76 properties in the dataset, Table 2 shows the details description of this dataset.

Table 2- Attributes and descriptions in the first heart disease dataset

Attribute	Description
Age	Age of the patient

Sex	Sex of the patient (1 = male; 0 = female)
chest pain type	Type of chest pain (4 values)
resting blood pressure	Resting blood pressure (in mm Hg)
serum cholesterol	Serum cholesterol level (in mg/dl)
fasting blood sugar 120 mg/dl	Fasting blood sugar >120 mg/dl (1 = true; 0 = false)
resting electrocardiographic results	Resting ECG results (values: 0, 1, 2)
Attribute	Description
maximum heart rate achieved	Maximum heart rate achieved
exercise induced angina	Exercise-induced angina (1 = yes; 0 = no)
Old peak	ST depression induced by exercise relative to rest
slope of peak exercise ST segment	The slope of the peak exercise ST segment
number of major vessels (0-3) colored	Number of major vessels (0-3) colored by fluoroscopy
Thal	Thalassemia (0 = normal; 1 = fixed defect; 2 = reversible defect)

Age, sex, chest pain kind, blood pressure, cholesterol, and more are used to diagnose heart disease. Dummy personal identifiers are used in anonymized data.

Histograms of major numerical features from the heart disease dataset show critical variable distribution in Figure 2. The age distribution shows that most patients are 50–60, indicating a higher heart disease prevalence in this age range. Resting blood pressure (trestbps) is mostly 120–140 mm Hg. Concentrations of cholesterol (chol) range from 200 to 300 mg/dl, with few exceeding 400 mg/dl. The maximum heart rate (thalach) is normal, reaching between 150 and 170 bpm.

Finally, the old peak feature, which assesses ST depression, is severely skewed, with most values near 0, indicating that most patients have low ST depression during exercise.

These distributions provide an important understanding of the dataset's characteristics and can guide model training and feature selection in the predictive analysis of heart disease.

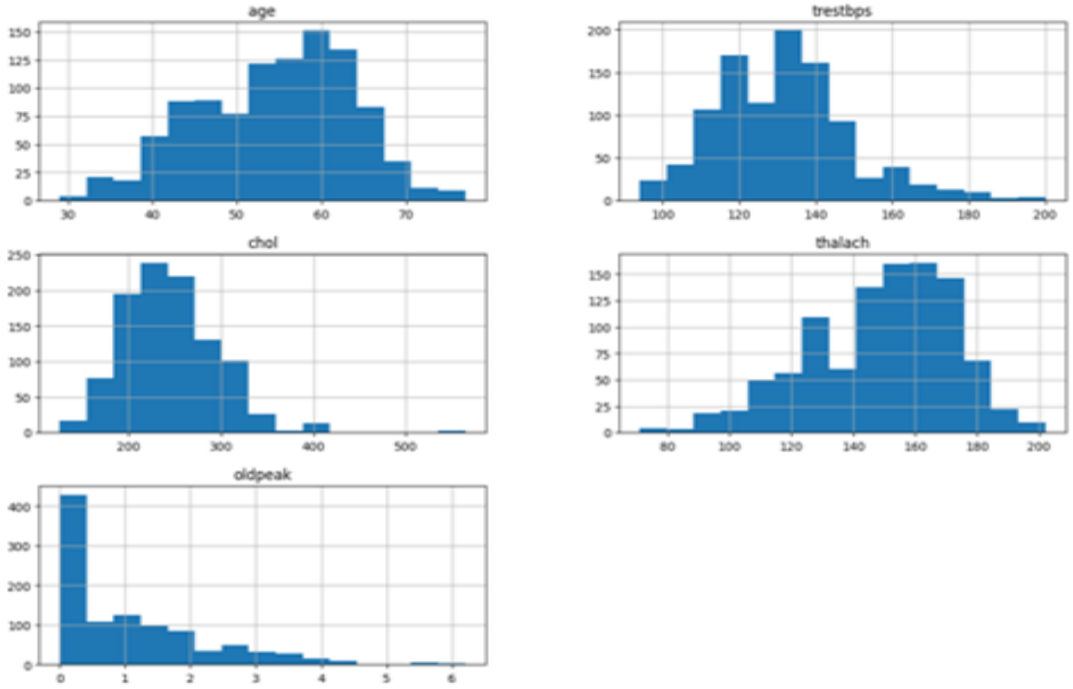

Fig. 2- Histograms of key numerical features in the First Heart Disease Dataset

Figure 3 presents a correlation heatmap illustrating the relationships between various features in the heart disease dataset. However, the categorical attribute chest pain type (cp) has a high positive correlation of 0.43 with target class making it one of most important features while predicting heart disease. Similarly, it is seen that the maximum heart rate achieved (thalach) does positively correlate with target (0.42). In contrast, exercise-induced angina (exang) and old- peak both have negative target correlations (-0.44), meaning that higher values of these attribute are indicative for a lower risk having heart disease; From these insights, it becomes possible to identify the predictive variables for training and model inference.

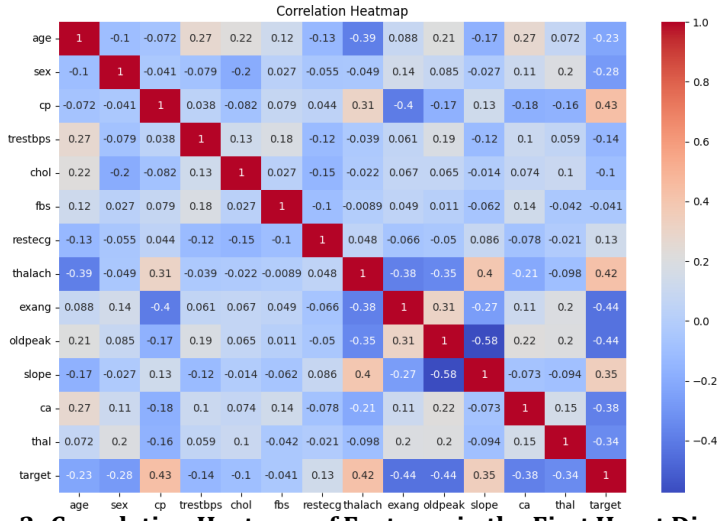

Fig. 3- Correlation Heatmap of Features in the First Heart Disease Dataset

The second heart disease dataset from Kaggle 2 is a comprehensive collection formed by combining five widely used heart disease datasets that were previously available independently. These datasets—Cleveland, Hungarian, Switzerland, Table 3 illustrate it decryption.

Table 3- Attributes and descriptions in the second heart disease dataset

Attribute	Description
age	Age in years
sex	Sex (1 = male; 0 = female)
chest pain type	Type of chest pain (1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, 4 = asymptomatic)
resting blood pressure	Resting blood pressure (in mm Hg)
serum cholesterol	Serum cholesterol level (in mg/dl)
fasting blood sugar ¿ 120 mg/dl	Fasting blood sugar >120 mg/dl (1 = true; 0 = false)
resting electrocardiogram results	Resting ECG results (0 = normal, 1 = ST-T wave abnormality, 2 = left ventricular hypertrophy)
maximum heart rate achieved	Maximum heart rate achieved (71-202 bpm)
exercise induced angina	Exercise-induced angina (1 = yes; 0 = no)
Old peak	ST depression induced by exercise relative to rest
slope of peak exercise ST segment	Slope of the peak exercise ST segment (1 = upsloping, 2 = flat, 3 = down sloping)
class	Target (1 = heart disease, 0 = normal)

Long Beach VA, and Statlog—were merged to create the largest heart disease dataset available for research, consisting of 1,190 instances and 11 common features. This combination offers a more diverse and extensive resource for studying heart disease, particularly Coronary Artery Disease (CAD). The dataset includes key features

such as age, sex, chest pain type, resting blood pressure, cholesterol levels, and more, all of which are critical for building predictive models.

Figure 4 presents the histograms of key numerical features from the second heart disease dataset. The age distribution shows that the majority of patients are between 40 and 60 years old, indicating a higher prevalence of heart disease within this age range. The resting blood pressure (resting bp s) is concentrated around 125 mm Hg, with most patients falling within a relatively normal distribution. Cholesterol levels demonstrate a peak between 200 and 300 mg/dl, reflecting common cholesterol ranges in heart disease cases, although there are a few extreme values exceeding 400 mg/dl. The maximum heart rate achieved is normally distributed, with the majority of patients reaching a heart rate between 120 and 170 bpm, indicative of typical exercise response ranges. Lastly, the old peak feature, representing ST depression, shows a highly skewed distribution, where most values are centered near zero, suggesting that many patients experience minimal ST depression during exercise. These visualizations offer valuable insights into the characteristics of the patient data, helping to inform feature selection and model development in heart disease prediction research.

Figure 5 presents the correlation heatmap for the second heart disease dataset, visualizing the relationships between various features. The heatmap reveals that chest pain type (0.46), exercise-induced angina (0.48), and ST slope (0.51) have the highest positive correlations with the target variable, making them key indicators for predicting heart disease. Conversely, maximum heart rate (-0.41) and cholesterol (- 0.20) show negative correlations, suggesting that lower heart rate and cholesterol levels are linked to a higher likelihood of heart disease. These insights are valuable for identifying the most influential features in developing predictive models for heart disease.

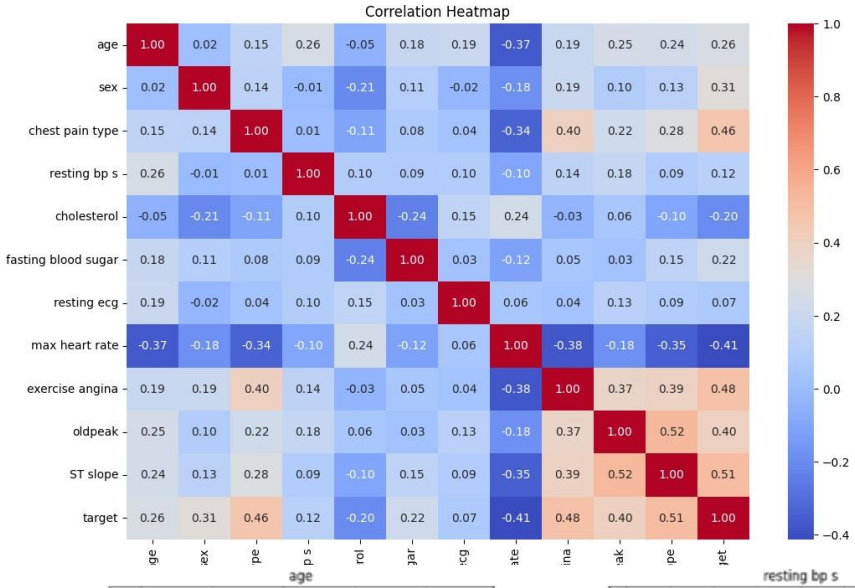

Fig. 4- Histograms of Numerical Features in the Second Heart Disease Dataset

Fig. 5- Correlation Heatmap of Features in the Second Heart Disease Dataset

3.2 Data preprocessing

Our model was requiring a lot of preprocessing work on the both data before it could be used for training to increase its quality, and homogeneity between input values. Preprocessing is the first and crucial task in ML lifecycle as it converts raw data to a clean, well-structured format required for building an optimum model.

One hot encoding applied on features which includes 'sex','cp', 'fbs', restecg,'exang', slope, ca and thal. This process translates categorical features into a numeric format by generating binary columns for each category thus permitting ML algorithms to handle such non-numeric attributes.

Only numerical features on age, trestbps, chol, thalach, and old peak should be measured normalize with standard scaler. This standardizes feature values, preventing a feature from dominating when it has a bigger scale, which would influence model performance.

An optional step was also available for Outlier detection using Z-score method. Z-score quantifies how far a value deviates from the mean. When the score is high than a threshold, this value might be an outlier and should not count in model robustness.

Final training and testing sets were 80:20. Testing on new data follows training set model creation. Figure 6 shows a well-balanced target variable (heart disease presence or absence) to avoid model bias and provide an accurate assessment.

3.3 Classification Machine Learning Techniques

Classification ML techniques, such as Extra Trees, RF, XGBoost, and GB, are widely utilized for their robust performance in predictive modeling tasks. These algorithms excel in handling complex datasets, especially when there are numerous features and potential interactions.

- The Extra Trees, or Extremely Randomized Trees [20] is another very efficient and effective algorithm commonly used on classification problems such as heart disease. Extra Trees works similarly to a normal DT, except that the algorithm builds multiple unpruned trees by randomly choosing splits for each feature rather than deterministic splits on parameter-tweaked data, so as not only diversify selection but also reduce overfitting. For heart disease, the max depth of 7 and added robustness in combination with efficiency ensures that tens of estimators are sufficient to learn on complex interactions. This is very useful for future outcome prediction of heart disease [21].

- RFs [22] is an ensemble learning method that creates many DTs during training and outputs class mode for classification issues. When used for heart disease detection, Random Forest Classifier also improves interpretability of the model by making it easier to understand how their predictions are made without sacrificing model complexity in this case using 30 estimators and a max depth of 8 so that intricate patterns can still be captured despite overfitting. RF aims to organize tree nodes by randomly selecting a subset of features from the input before training (10 in this case). It has proved to be a very successful classification technique to model the complexity of risk factors and outcomes based on medical data, especially for heart disease.

- XGBoost (extreme Gradient Boosting) classifier [23] is a state-of-art ML algorithm that belongs to the category of ensemble methods and it is known for its efficiency and accuracy in classification tasks such as predicting heart disease. XGBoost is an optimized, parallelized implementation of GB that's designed to be highly efficient, flexible and portable across platforms while also being computationally scalable using a topology-based data structure. It is well known for its regularization methods which prevent overfitting and performance in speed wise, due to parallel processing capabilities. This makes XGBoost an excellent tool for determining the existence of heart disease from different clinical features and in general XGboost is very good especially when it comes to medical data because we usually have a lot of missing values.

- Gradient Boosting Classifier (GBC) [24] is a powerful ML method and often applied to the classification problem, such as forecasting heart disease just like what we did. It creates a sequence of DTs, which correct errors made by its predecessor tree in the series, as such that prediction errors are minimized. This makes it especially well-suited to medical data, which typically has more complex phenotypic variables and non-linear predictors. The algorithm learns from its mistakes, and we improved our heart disease prediction by iteratively detecting misclassified cases.

- It is a powerful technique for overfitting reduction by adjusting rates of learning and regularization, so it can be moulded as the perfect approach to represent risk factors and outcomes related to heart disease.

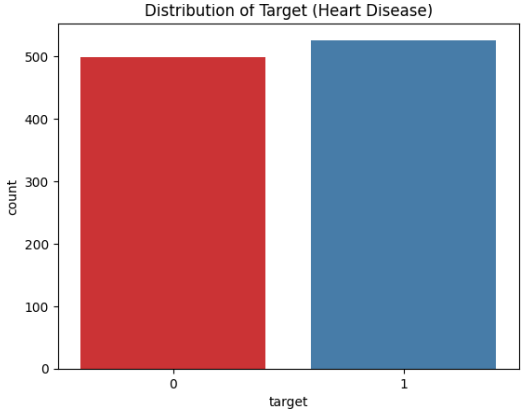

Fig. 6- Balancing data set

3.4 Discriminant analysis (DA)

Multivariate statistical method discriminant analysis (DA) classifies observations and distinguishes between two or more categories based on numerous independent variables [25] [26]. The objective is to combine a linear expression of the independent variables into a single discriminant function, by which researchers can divide new observations between predetermined categories and classify elaborated upon. Despite the advent of many discriminant functions, they all seek to serve a singular goal. The most common parametric forms of DA include LDA, and QDA.

1) Linear Discriminant Analysis: LDA [27] utilizes a linear combination of independent variables to maximize the ratio of intergroup to intragroup variation in discriminant scores. One of the most commonly applied functions in discriminant analysis is Fisher's discriminant function, while another approach involves deriving a ranking

rule by minimizing the expected cost function (ECM) [28] [29]. In LDA, numerous independent variables are compared to the class response variable. Two-group response variables are the simplest. Here, the linear discriminant function, which passes through the means (centroids) of the two groups, is used to differentiate between them. When there are multiple groups, k1 discriminant functions are required, where k represents the number of classes.

For example, if $\bar{x}_1$ and $\bar{x}_2$ are the means of the first and second groups, respectively, and S is the pooled variance-covariance matrix, Fisher's discriminant function separates the groups based on the following criteria:

- If xi belongs to the first group:

$$y = (\bar{x}_1 - \bar{x}_2)' S^{-1} X_i \geq (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 + \bar{x}_2) \quad (1)$$

- If xi is the member of the second group:

$$y = (\bar{x}_1 - \bar{x}_2)' S^{-1} X_i \geq (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 + \bar{x}_2) \quad (2)$$

LDA assumes multivariate normality, variance-covariance matrix homogeneity, linearity, and no independent variable multi collinearity [30]. Tabachnick and Fidell found that the linear discriminant function remains robust even when outliers or mismatched variance-covariance matrices are present.

2) Quadratic Discriminant Analysis: QDA [31] [32] is a discriminant analysis technique similar to LDA in that it creates classification functions based on independent variables. However, unlike LDA, the classification functions in QDA are non-linear. In certain situations where linear functions fail to adequately separate groups, Quadratic Discriminant Functions (QDFs) may offer a more effective solution. The choice between LDA and QDA depends on specific assumptions: QDA, unlike LDA, does not require the assumption of homogeneous variance-covariance matrices between groups. This makes QDA more suitable when there is heterogeneity in the variance-covariance structure across groups.

If the following equation applies, QDA groups observation i into group one and group two otherwise. The first and second groups' means are X1 and x2, and their variance-covariance matrices are S1 and S2:

$$\frac{1}{2} x' S_1^{-1} - S_2^{-1} x + \bar{x}_1' S_1^{-1} - \bar{x}_2' S_2^{-1} x - k \geq \ln \frac{c_{12}\, p_2}{c_{21}\, p_1} \quad (3)$$

Where:

$$k = \frac{1}{2} \ln \frac{|S_1|}{|S_2|} + \frac{1}{2} \bar{x}_1' S_1^{-1} \bar{x}_1 - \bar{x}_2' S_2^{-1} \bar{x}_2 \quad (4)$$

The Mahalanobis Distance (MD) is a key metric in QDA.

This function is given by:

$$f_1'(x) = D_1^2(x) - D_2^2(x) + \ln \frac{|S_1|}{|S_2|} - 2 \ln \frac{p_2}{p_1} \quad (5)$$

The MD value, represented as,

$$D_1^2(x) - D_2^2(x)$$

reduces to a linear function when S1 = S2.In this study, due to the lack of multivariate normality and homogeneous variance-covariance matrices, various types of discriminant analysis were explored to identify the most suitable model.

4. RESULTS AND DISCUSSION

4.1. Evaluation Metrics

In the evaluation of predictive models, various metrics are employed by researchers to assess and demonstrate the efficacy of their methods. Each statistic shows model performance differently. We briefly outline each metric and provide formulae for their computation below:

1) Accuracy: This indicator measures the percentage of cases accurately predicted. Formula for calculating it:

$$\text{Accuracy} = \frac{\text{True Positives + True Negatives}}{\text{Total Instances}}$$

In imbalanced datasets, accuracy may not be enough to determine correctness.

2) Precision: evaluates the relevance of the positive pre- dictions made by the model. It is calculated as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}} \quad (7)$$

This measure shows the percentage of correct positive identifications.

3) Recall: evaluates the model's ability to find all relevant occurrences. As expressed,

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \quad (8)$$

This metric measures the proportion of actual positives that were correctly identified.

4) F1 Score: Balances precision and recall with a harmonic mean. As given by:

$$\text{F- Measure} = \frac{2\text{xPrecision} \times \text{Recall}}{\text{Precision + Recall}} \quad (9)$$

This metric is particularly useful when dealing with imbalanced datasets, as it combines both precision and recall into a single value.

Among these metrics, accuracy is the most commonly re- ported in the literature. Our review article focuses on categorizing, comparing, and analyzing previous research based on accuracy, given its widespread application and fundamental role in performance evaluation.

4.2. Evaluation of the Machine Learning models

This study used Tensor Flow and Keras to build and train machine learning architectures. These frameworks were chosen for their extensive libraries, community support, and ability to scale with dataset size and complexity. Accuracy, sensitivity (recall), precision, and F1-score were used to evaluate the networks, which are important in medical diagnostics where erroneous positives and negatives have serious effects. The evaluation results of the ML models with the first dataset demonstrate varying levels of performance in terms of accuracy, recall, precision, F1 score, error rate, and training time as presents in Table 4.

Table 4- Comparison of machine learning models on the first dataset

Model	Accuracy	Recall	Precision	F1 Score	Error Rate	Training Time (s)
XGBoost	0.985366	0.970874	1.000000	0.985222	0.014634	0.739639
GB	0.980488	0.980583	0.980583	0.980583	0.019512	0.313116
RF	0.960976	0.980583	0.943925	0.961905	0.039024	0.118720
Extra Trees	0.907317	0.961165	0.868421	0.912442	0.092683	0.033382

XGBoost is the top-performing model with an accuracy of 98.54%, perfect precision (1.000), and an F1 score of 0.985, indicating exceptional predictive capability for heart disease. Its low error rate (0.0146) highlights strong performance, though it has the longest training time of 0.7396 seconds. GB follows closely with 98.05% accuracy and balanced metrics (recall, precision, F1 score of 0.9806) while offering faster training of 0.3131 seconds, making it ideal for high accuracy with reduced training time. RF achieves 96.10% accuracy, with slightly lower precision of 0.9439 but faster training of 0.1187 seconds. Extra Trees, with the lowest accuracy of 90.73% and highest error rate of 0.0927, compensates with the shortest training time of 0.0334 seconds. XGBoost and GB excel overall, while RF is a solid choice for quicker training, and Extra Trees is best suited for speed-focused scenarios.

The evaluation results of the machine learning models with the second dataset demonstrate varying performance levels as presents in Table 5.

Table 5- Comparison of machine learning models on the second dataset

Model	Accuracy	Recall	Precision	F1 Score	Error Rate	Training Time (s)
RF	0.949580	0.969466	0.940741	0.954887	0.050420	0.418750
XGBoost	0.928571	0.954198	0.919118	0.936330	0.071429	0.083599
Extra Trees	0.924370	0.923664	0.937984	0.930769	0.075630	0.337383
GB	0.915966	0.931298	0.917293	0.924242	0.084034	0.232399

RF is the best-performing model with an accuracy of 94.96%, high recall of 0.9695, precision of 0.9407, and an

F1 score of 0.9549. Despite a moderate training time of 0.4188 seconds, its strong predictive power makes it highly effective. XGBoost, with slightly lower accuracy at 92.86%, performs well with a recall of 0.9542, precision of 0.9191, and an F1 score of 0.9363. It compensates for its slightly higher error rate with a significantly faster training time of 0.0836 seconds, making it efficient. Extra Trees, with 92.44% accuracy and a recall of 0.9237, is reliable but less efficient than XGBoost, with a training time of 0.3374 seconds. GB, the lowest in accuracy at 91.60%, provides balanced performance but lags behind in efficiency and accuracy compared to Random Forest and XGBoost. Random Forest

provides the highest accuracy and predictive performance, while XGBoost offers a strong trade-off between accuracy and training time, making both models the most suitable for heart disease prediction with this dataset.

In the evaluation of models across both datasets, XGBoost in the first dataset significantly outperforms Random Forest in the second dataset. XGBoost achieved an accuracy of 98.54%, along with perfect precision (1.000) and a high F1 score of 0.985, indicating superior performance in heart disease prediction. In the second dataset, Random Forest was still effective but had a lower accuracy of 94.96%, precision of 0.9407, and F1 score of 0.9549. These results demonstrate XGBoost's superior predictive power and efficiency over Random Forest, particularly in accuracy and precision.

4.3. Evaluation of the Discriminant analysis models

Discriminant analysis, including both Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA), serves as a statistical approach for classifying a set of observations into predefined classes. The fundamental principle behind LDA involves projecting the data onto a line where classes are best separated by the means of their projections, assuming equal covariance matrices among classes. QDA uses quadratic surfaces instead of equal covariance for a more flexible decision boundary. These methods were chosen for their simplicity and efficacy with

well-defined, regularly distributed data. Simple interpretability and processing economy make them valuable in comparative investigations.

a line where classes are best separated by the means of their projections, assuming equal covariance matrices among classes. QDA uses quadratic surfaces instead of equal covariance for a more flexible decision boundary. These methods were chosen for their simplicity and efficacy with well-defined, regularly distributed data. Simple interpretability and processing economy make them valuable in comparative investigations.

Python's scikit-learn library provides robust implementations of both LDA and QDA, which were utilized to perform the discriminant analysis in this research, enabling rapid iterations and evaluations.

The evaluation results of the Discriminant Analysis models in the first dataset reveal significant differences in performance between LDA and QDA as presents in Table 6.

Table 6 - Performance comparison of discriminant analysis models on the first dataset

Model	Accuracy	Recall	Precision	F1 Score	Error Rate	Training Time (s)
LDA	0.824390	0.902913	0.781513	0.837838	0.175610	0.007396
QDA	0.541463	0.961165	0.523810	0.678082	0.458537	0.008487

In general, LDA classifies heart disease cases well (82.44%). Its recall score of 0.9029 shows most positives are accurately identified, while its precision of 0.7815 implies modest false positives. The F1 score of 0.8378 balances precision and recall, making LDA reliable for heart disease categorization. LDA is accurate and computationally efficient with a 0.1756 error rate and 0.0074-second training time.

In comparison, QDA has a significantly lower accuracy of 54.15%, meaning it struggles to classify cases correctly. However, with a recall of 0.9612, QDA excels at identifying true positive cases, even more so than LDA. At 0.5238, precision is much poorer, indicating a large false positive rate. QDA is less trustworthy than LDA, as seen by its F1 score of 0.6781. The increased error rate of 0.4585 emphasizes its lesser accuracy. Although slower at 0.0085 seconds, QDA's training is efficient.

LDA outperforms QDA across most metrics, offering a better balance between precision, recall, and overall classification accuracy. While QDA has a higher recall, its lower precision and accuracy make it a less reliable model for predicting heart disease in this context.

The evaluation results of the Discriminant Analysis models with the second dataset demonstrate notable differences in performance between LDA and QDA, as presents in Table 7.

Table 7- Performance comparison of discriminant analysis models on the second dataset

Model	Accuracy	Recall	Precision	F1 Score	Error Rate	Training Time (s)
LDA	0.873950	0.908397	0.868613	0.888060	0.126050	0.156104
QDA	0.449580	0.000000	0.000000	0.000000	0.550420	0.006885

LDA accurately identifies 87.40% of positive cardiac disease cases with a recall of 0.9084. LDA's precision of 0.8686 and F1 score of 0.8881 make it a reliable cardiac disease classification model. Though its training time of 0.1561 seconds is longer than other models, its error rate of 0.1260 is minimal.

In contrast, QDA exhibits very poor performance, with an accuracy of only 44.96%. The model records zero values for recall, precision, and F1 score, highlighting its complete failure to correctly classify any positive cases. The high

error rate of 0.5504 further emphasizes its lack of reliability for this dataset, though it does have a faster training time of 0.0069 seconds. These results suggest that QDA is unsuitable for this particular dataset, while LDA proves to be a much more effective option for heart disease prediction.

The second dataset's Linear Discriminant Analysis (LDA) outperforms the first. In the second dataset, the LDA model had 87.40% accuracy, 0.9084 recall, and 0.8881 F1 score, suggesting greater prediction skills. In contrast, LDA in the first dataset had 82.44% accuracy, 0.9029 recall, and 0.8378 F1 score. These results demonstrate that LDA in the second dataset predicts more accurately and balances precision and recall better.

5. DISCUSSION

In the comparative evaluation of machine learning models and discriminant functions using the data from the first and second datasets, illustrate in Table 8, the analysis reveals significant variations in performance across different methodologies, as outlined in the comparative. Starting with the discriminant functions, LDA and QDA, we see that LDA generally performs better than QDA. LDA shows a more stable perfor- mance with an accuracy of 82.44%, a high recall of 90.29%, and an F1 score of 83.78%, indicating its reliability under the assumption of Gaussian distributions among the classes. This makes LDA particularly useful in medical diagnostics where the underlying data distributions often approximate normality.

Table 8- Comparative analysis of machine learning models including discriminant functions

Model Type	Model	Accuracy	Recall	Precision	F1 Score	Notes
Discriminant Function	LDA	82.44%	90.29%	78.15%	83.78%	Performs well under Gaussian assumption
Discriminant Function	QDA	54.15%	96.12%	52.38%	67.81%	Flexible, but sensitive to class covariance
Machine Learning	XGBoost	98.54%	97.09%	100%	98.52%	High performance, computationally intensive
Machine Learning	Random Forest	94.96%	96.95%	94.07%	95.49%	Robust to overfitting, good for large datasets

On the other hand, QDA, although highly sensitive in detecting true positives (as evidenced by its recall of 96.12%), suffers significantly in terms of precision and over- all accuracy, which are crucial for avoiding false positives in medical diagnostics. Comparatively, machine-learning models like XGBoost and Random Forest demonstrate superior performance across both datasets. XGBoost, in particular, show cases exceptional results with an accuracy of 98.54%, perfect precision, and an F1 score nearing perfection. This model's ability to handle complex, multi-dimensional data with high efficiency and accuracy renders it highly suitable for the nuanced requirements of heart disease prediction, despite its computational intensity. Random Forest also exhibits strong performance with an accuracy of 94.96%, robust recall, and high precision. It provides a good balance between performance and overfitting resistance, making it ideal for large datasets typical in medical research. The superiority of machine learning models over discriminant functions in this analysis is evident from their higher precision and accuracy, which are critical in clinical applications where the cost of misdiagnosis can be high. Machine learning models, particularly XGBoost, not only excel in handling the intrinsic complexities of medical data but also in adapting to varied data structures without the stringent assumptions required by traditional statistical methods like LDA and QDA. This adaptability, coupled with their ability to integrate and learn from large-scale data dynamically, underscores the shift towards more advanced analytical techniques in medical diagnostics. The evaluation and comparison of these models underscore the advancing role of machine learning in enhancing diagnostic accuracies beyond the capabilities of traditional discriminant functions. This evolution is pivotal as the healthcare industry continues to embrace data-driven methodologies to improve patient outcomes and treatment efficacies. The presented table not only highlights the quantitative superiority of machine learning models but also aligns with the qualitative shift towards more sophisticated, reliable, and scalable solutions in the realm of medical diagnostics.

In comparison to existing studies as presents in Table 9, our proposed model surpasses previous works in terms of accuracy. For instance, researchers in achieved the highest accuracy of 97.78% using RF for HF prediction, while reported an accuracy of 87% using the CART algorithm for heart disease detection. Furthermore, employed DL methods, achieving an accuracy of 84.4% with ANN, while other models, such as SVM and DTs, recorded even lower accuracies.

Table 9- Comparison of methods, datasets, and accuracies from related works

Reference	Method	Dataset	Accuracy
[13]	RF	Heart Failure Dataset	97.78%
[15]	CART	Heart Disease Dataset	87.00%
[19]	ANN	Coronary Artery Disease Dataset	84.40%
Proposed Model	XGBoost	Heart Disease Dataset	98.54%

Our model predicts heart disease better than existing studies, with XGBoost being the most successful method. Our model is more trustworthy and robust than existing approaches for early cardiac disease diagnosis and risk prediction in clinical settings due to its high accuracy, precision, and F1 score.

6. CONCLUSION

Heart disease is still one of the global causes of mortality, and there is need to have timely and accurate diagnostic tools that can help in the lives of patients. This article shows the overall comparison between machine learning models and classical discriminant-based techniques, which include the LDA and the QDA in predicting heart disease through two standard data sets. The paper proves that ensemble machine learning algorithms, especially XGBoost, exceed the classical discriminant ones by far and reach the highest accuracy of 98.54 per cent and perfect precision on the first data. The results both theoretically and empirically prove the excellence of databased model over statistical classifiers in medical classification. The applied machine learning tools, as suggested, provide interpretable, clinically implementable, and high-performing and scalable solutions. The high accuracy of XGBoost and Random Forest and the scalability of LDA to computation needs makes it clear that the models will be flexible in real-world care settings. These models will decrease the time of the diagnostic process, increase the treatment planning, and eventually the prognosis of patients. Nevertheless, the research has limitations in spite of its strengths. The result could only be tested on two publicly available datasets, which might lack diversity and heterogeneity of the real clinical populations. In addition, the models showed high performance on the structured data, but this work fails to encode unstructured clinical data like medical imaging and EHR that is essential in cardiology.

To overcome these terminology in future researches, various dimensions need to be addressed. One point of improvement will be the involvement of more and wider data using various clinical sources to enhance the generalizability of the models. Second, unstructured data may offer a better overview of patient health. Third, one could go further with this model and include XAI methods to achieve better transparency of the model, which could increase the model adoption among healthcare professionals and lead to a positive user reaction. Additionally, the combination of machine learning with the structure of deep learning has a potential to further enhance the accuracy of prediction and capture more complicated data dynamics. Finally, the study makes an important contribution to emerging research on AI-aided cardiology that could be implemented in practice because of the diagnostic capacity of ensemble machine takes. As long as they are developed further and validated, such models may have a future application in clinical decision-making and the transformation of preventive cardiovascular effort.

References

[1] M. M. Hussain, U. Rafi, A. Imran, M. U. Rehman, and S. K. Abbas, “Risk factors associated with cardiovascular disorders: Risk factors associated with cardiovascular disorders,” Pakistan BioMedical Journal, vol. 6, no. 07, (2024), pp. 03–10.

[2] C. Wang, Y. Sun, D. Jiang, C. Wang, and S. Liu, "Risk-attributable burden of ischemic heart disease in 137 low-and middle-income countries from 2000 to 2019," *Journal of the American Heart Association*, vol. 10, no. 19, (2021), p. e021024.
[3] J. E. Brush, J. Sherbino, and G. R. Norman, "Diagnostic reasoning in cardiovascular medicine," *BMJ*, vol. 376, (2022).
[4] F. Yasmin, S. M. I. Shah, A. Naeem, S. M. Shujauddin, A. Jabeen, S. Kazmi, and H. M. Lak, "Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future," *Reviews in Cardiovascular Medicine*, vol. 22, no. 4, (2021), pp. 1095–1113.
[5] D. Mohajan and H. K. Mohajan, "Obesity and its related diseases: a new escalating alarming in global health," *Journal of Innovations in Medical Research*, vol. 2, no. 3, (2023), pp. 12–23.
[6] B. Bhusal, "Poor diet leading to the increasing risk of atherosclerosis in the world," *Journal of Cardiology and Cardiovascular Medicine*, vol. 9, no. 3, (2024), pp. 142–147.
[7] L. Laranjo, F. Lanas, M. C. Sun, D. A. Chen, L. Hynes, T. F. Imran, and C. K. Chow, "World Heart Federation roadmap for secondary prevention of cardiovascular disease: 2023 update," *Global Heart*, vol. 19, no. 1, (2024), p. 8.
[8] M. Allahbakhshi, A. Sadri, and S. O. Shahdi, "Diagnosis of parkin- son's disease using eeg signals and machine learning techniques: A comprehensive study," arXiv preprint arXiv:2405.00741, (2024).
[9] F. Yasmin, S. M. I. Shah, A. Naeem, S. M. Shujauddin, A. Jabeen, S. Kazmi, S. A. Siddiqui, P. Kumar, S. Salman, S. A. Hassan et al., "Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future," *Reviews in Cardiovascular Medicine*, vol. 22, no. 4, (2021), pp. 1095–1113.
[10] H. Li, M. Jia, and Z. Mao, "Dynamic feature extraction-based quadratic discriminant analysis for industrial process fault classification and diagnosis," *Entropy*, vol. 25, no. 12, (2023), p. 1664.
[11] M. Abdelsayed, E. J. Kort, S. Jovinge, and M. Mercola, "Repurposing drugs to treat cardiovascular disease in the era of precision medicine," *Nature Reviews Cardiology*, vol. 19, no. 11, (2022), pp. 751–764.
[12] Y. Zhang, J. Liu, and W. Shen, "A review of ensemble learning algorithms used in remote sensing applications," *Applied Sciences*, vol. 12, no. 17, (2022), p. 8654.
[13] M. M. Ali, V. S. Al-Doori, N. Mirzah, A. A. Hemu, I. Mahmud, S. Azam, K. F. Al-Tabatabaie, K. Ahmed, F. M. Bui, and M. A. Moni, "A machine learning approach for risk factors analysis and survival prediction of heart failure patients," *Healthcare Analytics*, vol. 3, (2023), p. 100182.
[14] A. M. Tushar, A. Wazed, E. Shawon, M. Rahman, M. I. Hossen, and M. Z. H. Jesmeen, "A review of commonly used machine learning classifiers in heart disease prediction," in *Proceedings of the 2022 IEEE 10th Conference on Systems, Process & Control (ICSPC)*, (2022), pp. 319–323.
[15] M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," *Healthcare Analytics*, vol. 3, (2023), p. 100130.
[16] D. Shrestha, "Advanced machine learning techniques for predicting heart disease: A comparative analysis using the Cleveland Heart Disease Dataset," *Applied Medical Informatics*, vol. 46, no. 3, (2024).
[17] P. Sujatha and K. Mahalakshmi, "Performance evaluation of super- vised machine learning algorithms in prediction of heart disease," in 2020 IEEE international conference for innovation in technology INOCON. (2020), pp. 1–7.
[18] G. Ambrish, B. Ganesh, A. Ganesh, C. Srinivas, K. Mensinkal, and K. Shashikala, "Logistic regression technique for prediction of cardiovascular disease," *Global Transitions Proceedings*, vol. 3, no. 1, (2022), pp. 127–130.
[19] M. S. Gangadhar, K. V. S. Sai, S. H. S. Kumar, K. A. Kumar, M. Kavitha, and S. Aravinth, "Machine learning and deep learning techniques on accurate risk prediction of coronary heart disease," in *Proceedings of the 2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, (2023), pp. 227–232.
[20] T. E. Mathew, "An optimized extremely randomized tree model for breast cancer classification," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 16, (2022), pp. 5234–5246.
[21] P. Rahi and S. S. Kang, "A novel paradigm in cardiovascular disease risk prediction through hybrid machine learning," *Системная инженерия и информационные технологии*, (2025), pp. 49–66.
[22] A. Almotairi, S. Atawneh, O. A. Khashan, and N. M. Khafajah, "Enhancing intrusion detection in IoT networks using machine learning-based feature selection and ensemble models," *Systems Science & Control Engineering*, vol. 12, no. 1, (2024), p. 2321381.
[23] H. A. Al-Jamimi, "Early prediction of heart disease risk using extreme gradient boosting: a data-driven analysis," *International Journal of Biomedical Engineering and Technology*, vol. 45, no. 4, (2024), pp. 296–313.
[24] M. Feng, X. Wang, Z. Zhao, C. Jiang, J. Xiong, and N. Zhang, "Enhanced heart attack prediction using extreme gradient boosting," *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 04, (2024), pp. 9–16.
[25] L. Qu and Y. Pei, "A comprehensive review on discriminant analysis for addressing challenges of class-level limitations, small sample size, and robustness," *Processes*, vol. 12, no. 7, (2024), p. 1382.
[26] R. Johnston, "Discriminant analysis," in *Encyclopedia of Quality of Life and Well-Being Research*, (2024), pp. 1845–1847.
[27] S. Zhao, B. Zhang, J. Yang, J. Zhou, and Y. Xu, "Linear discriminant analysis," *Nature Reviews Methods Primers*, vol. 4, no. 1, (2024), p. 70.
[28] R. R. Isnanto, I. Rashad, and C. E. Widodo, "Classification of heart disease using linear discriminant analysis algorithm," in *E3S Web of Conferences*, vol. 448, (2023), p. 02053.
[29] E. Sivari and S. Sürücü, "Prediction of heart attack risk using linear discriminant analysis methods," *Journal of Computer, Electrical and Electronics Engineering Sciences*, vol. 1, no. 1, (2023), pp. 5–9.
[30] A. Rajarathinam, "Discriminant analysis for heart disease attributes," *International Journal of Statistics and Applied Mathematics*, vol. 9, no. 2, (2024), pp. 180–188.
[31] M. U. Aslam, S. Xu, S. Hussain, M. Waqas, and N. L. Abiodun, "Machine learning-based classification of valvular heart disease using cardiovascular risk factors," *Scientific Reports*, vol. 14, no. 1, (2024), p. 24396.
[32] H. P. Wibowo, M. Anshori, and M. S. Haris, "The discriminant analysis function was implemented to predict the presence of diabetes," *Journal of Enhanced Studies in Informatics and Computer Applications*, vol. 1, no. 2, (2024), pp. 47–55.