Advances in Artificial Intelligence and Machine Learning; Research 5 (4) 4594-4617 Received 23-08-2025; Accepted 11-11-2025; Published online 18-11-2025

Comparative Evaluation of Hybrid Multi-Classification Algorithms for Diagnosing Diabetes and Cardiovascular Diseases

Noura M. Alotaibi Nmalotaibi@uj.edu.sa
Computer Science and Artificial Intelligence Department,
University of Jeddah,
Jeddah, 21959 Saudi Arabia

Manal M. Khayyat mmkhayat@uqu.edu.sa
Department of Computer Science and Artificial Intelligence, College of Computing,
Umm Al-Qura University,
Makkah, 24382, Saudi Arabia

Araek Tashkandi astashkandi@uj.edu.sa
Department of Information Systems and Technology, College of Computer Science and Engineering,
University of Jeddah,
Jeddah, 21959, Saudi Arabia

Enas Jambi eajambi@uj.edu.sa
Computer Science and Artificial Intelligence Department,
University of Jeddah,
Jeddah, 21959, Saudi Arabia

Safa Habibullah shhabiballah@uj.edu.sa
Department of Information Systems and Technology, College of Computer Science and Engineering,
University of Jeddah,
Jeddah, 21959, Saudi Arabia

Mashael M. Khayyat Mkhayyat@uj.edu.sa
Department of Information Systems and Technology, College of Computer Science and Engineering,
University of Jeddah,
Jeddah, 21959, Saudi Arabia

Kaouther Omri KOmri@uj.edu.sa
Department of Computer and Network Engineering, College of Computer Science and Engineering,
University of Jeddah,
Jeddah, 21959, Saudi Arabia

Corresponding Author: Noura M. Alotaibi

Abstract

The integration of artificial intelligence (AI) into medical decision support systems (MDSS) offers transformative potential for enhancing diagnostic accuracy, particularly in resource-constrained healthcare environments. The aim of this research is to provide a comparative

Citation: Noura M. Alotaibi, et al. Comparative Evaluation of Hybrid Multi-Classification Algorithms for Diagnosing Diabetes and Cardiovascular Diseases. **Advances in Artificial Intelligence and Machine Learning**. 2025;5(4):255. https://dx.doi.org/10.54364/AAIML.2025.54255

study of features selection approaches in combination with multi-classification algorithms in aiding the diagnosis of diabetes and cardiac diseases- two of the most prevalent and life-threatening chronic conditions worldwide. The performance of four features selection approaches and six classification algorithms are evaluated. The proposed features selection strategies are full feature set, Minimum Redundancy Maximum Relevance (MRMR), KNN-based ranking, and univariate ranking using the Chi-square test. These approaches are evaluated under six classifiers which are support vector machine (SVM), k-nearest neighbors (KNN), linear discriminant analysis (LDA), decision tree, diagonal linear discriminant analysis (DLDA), and Naive Bayes (NB). The results demonstrated that KNN based features selection and univariate ranking with decision tree classifier were provided best performance in cardiac prediction reaching up 85.6%. While the best performance achieved in diabetes prediction reached to 78.7% using NB and DLDA models and KNN-based feature selection approach. These findings indicate that the hybrid models approach are effective in cardiac and diabetes prediction and shed light in the importance feature selection.

Keywords: Cardiovascular diseases, Diabetes, KNN, Medical decision support systems, MRMR, Multi-classification, Univariate features ranking.

1. INTRODUCTION

Recent advancements in artificial intelligence (AI) and its techniques have led to the rapid growth of AI-driven applications that support human experts, especially in tasks involving complex decision-making. Medical field is one of the most advanced area in AI adoption, with applications ranging from disease prediction and diagnosis to clinical decision support. Medical Decision Support Systems (MDSS) play an important role in clinical decision-making by reducing human error, improving outcomes, and enabling long-term data tracking [1].

Currently, medical decisions such as disease diagnosis and assessment of health status of the patient are made based on reasoning skills and experiential knowledge of medical professionals. In poor countries, providing accurate diagnosis to the boarder population remains a major challenge due to the limited numbers of healthcare providers. Furthermore, many diagnostic procedures require costly medical tests, posing a financial burden on low-income individuals. Human decision-making is also inherently error-prone, potentially leading to problems such as over diagnosis which consequently providing unnecessary treatments that negatively affecting both patient health outcomes and economic stability [2].

Therefore, this study aims to investigate the efficiency of different features selection approaches in combination with several classifiers for the diagnosis of diabetes and cardiovascular diseases, using these conditions as case studies to explore the application of AI in medical diagnostics. The selection of diabetes and cardiovascular disease, in particular, was guided by the rising global incidence of diabetes, which has become one of the most prevalent and costly chronic conditions. Among its many complications, cardiovascular disease is the leading cause of death in individuals with diabetes.

Currently, diabetes and cardiovascular diseases are diagnosed based on clinical assessments. However, in recent years, there has been growing interest in employing AI technologies in the medical field. Viloria et al. (2020) [3], proposed a decision support system for predicting diabetes using a support vector machine (SVM) with a radial basis function (RBF) kernel on dataset containing 500 subjects. The model classified the subjects into one of three classes: diabetic, prediabetic or non-diabetic. The decision was made on the basis of three features: age, body mass index (BMI), and blood glucose concentration and achieved accuracy reached to 65.5% on the test dataset. Another study by Maniruzzaman et al. (2020) [4], used a logistic regression alongside four classifiers–naive Bayes (NB), decision tree, AdaBoost, and random forest to predict the diabetes from a dataset of 6,561 subjects (657 diabetic and 5,904 non-diabetic). The classification was based on seven features selected using logistic regression: age, education level, BMI, systolic blood pressure, diastolic blood pressure, direct cholesterol, and total cholesterol. Feeding these features into random forest classifier yielded an accuracy of 94.25%. Islam et al. (2025) [5] compared the performance of various machine learning models–K-Nearest Neighbours (KNN), random forest, decision tree, and ExtraTrees classifier–incorporating random oversampling and hyperparameter tuning techniques. These models were trained to predict diabetes on two different datasets: the BRFSS dataset (which includes multiple classes: no diabetes, prediabetes, and diabetes) and the Diabetes 2019 dataset (binary classification: diabetic or non-diabetic). The study also proposed several explainability methods to enhance model interpretability, including SHapley Additive Explanations (SHAP), Partial Dependence Plots (PDP), and Local Interpretable Model-Agnostic Explanations (LIME). The ExtraTrees classifier performed exceptionally well, achieving 97.23% accuracy on the multi-class dataset and 97.45% on the binary dataset. Ensemble classifiers such as bagging, boosting, and voting have also been employed in several studies to predict diabetes [6–9]. In addition, deep learning models have been investigated for diabetes prediction with promising results [10–13].

On the other hand, several studies have employed AI for the prediction of cardiovascular diseases. In particular, machine learning techniques have been utilized to predict patients at risk of heart failure [14–18]. Ensemble methods have also been explored to enhance the prediction of heart disease risk, as demonstrated in recent works [19–22]. Beyond diagnosis, machine learning has been applied to identify patients at greater risk of adverse outcomes such as mortality and hospital readmission. A recent review by Hajishah et al. (2025) [23], assessed the effectiveness of machine learning models in predicting heart failure-related outcomes.

Alongside machine learning, deep learning models have also been employed for heart disease prediction. Zhou et al. (2025) [24], conducted a comprehensive systematic review evaluating the contributions and limitations of deep learning algorithms, extended deep learning approaches, and hybrid models that integrate deep learning with other technologies. Sivasubramaniam and Balamurugan (2025) [25], proposed a multimodal deep learning algorithm combining convolutional neural network (CNNs) and long short-term memory (LSTM) networks for the early detection and prediction of heart disease.

The literature on AI-based prediction of diabetes and cardiovascular diseases highlights the significant potential of AI to improve diagnostic accuracy and clinical outcomes. However, despite these advancements, the real-world adoption of AI tools in clinical practice remains limited. For instance, no FDA-approved AI-based medical device has yet been introduced in diagnosing practice, largely due to issues related to model accuracy, generalizability, and lack of robust validation.

In this study, we conducted a comparative evaluation of a muli-classification approaches: SVM, KNN, linear discriminant analysis (LDA), decision tree, NB, and diagonal linear discriminant analysis (DLDA). Additionally, we conducted an experiment with five different coding schemes: ternary complete, binary complete, ordinal coding, one-vs-one, and one-vs-all. Furthermore, we explored four feature selection strategies: full feature set, Minimum Redundancy Maximum Relevance (MRMR), KNN-based ranking, and univariate ranking using the Chi-square test. The aim is to assess the performance of different classifiers, coding schemes and features selection combinations in order to identify the most effective hybrid algorithm for predicting diabetes and cardiovascular diseases.

This paper is organized into four sections, Section II gives an overview of the methodology adopted in this research. Section III provides the results of the study. Section 4 offers an in-depth discussion of the findings, while Section 5 concludes the study.

2. METHODOLOGY

2.1 Dataset

In this study, two datasets were used: one for diabetes and another for cardiac disease. The UCI Machine Learning Repository [26], a publicly accessible online resource intended for machine learning and data mining research, was the source of both datasets.

The diabetes dataset used is the Pima Indians diabetes dataset, which has 768 samples and eight attributes. The dataset, which includes a binary response variable with ”1” denoting a positive diagnosis of diabetes and ”0” denoting a negative instance, was gathered from women of Pima Indian descent living in Arizona, USA, who were 21 years of age or older. The features of the dataset are described in depth in TABLE 1.

Table 1: Description of Diabetes Dataset Features

Feature Name	Description
Number of pregnancies	Described in integer numbers range between 0 and 17
Plasma glucose concentration	Integer number measured 2-hours after an oral glucose tolerance test described in milligrams per deciliter (mg/dL)
Diastolic blood pressure	The lower integer number in a BP reading described in millimeters of mercury (mmHg)
Triceps skinfold thickness	Integer numbers described in millimeters (mm), that is indicator of subcutaneous fat
2-Hour serum insulin	Integer number in conventional insulin concentration units (muU/ml) that measures insulin secretion or deficiency
BMI (Body Mass Index)	Fractional numbers calculated by dividing the body weight over its height ($\mathrm{kg/m^2}$)
Diabetes pedigree function	Verification function of diabetes presence described in fractional numbers range between 0.078 and 2.42
Patient’s age	Described in integer numbers range between 21 and 81 years
Class	Binary number (0 = No diabetes, 1 = Diabetes)

The cardiac disease dataset used in this study consists of 304 patient records and 13 clinical features. These features include demographic variables such as age and sex, as well as several physiological and diagnostic indicators. The chest pain type (CP) attribute is categorized into four classes: 0 (asymptomatic), 1 (atypical angina), 2 (non-anginal pain), and 3 (typical angina). Resting blood pressure (Trestbps) represents the patient's baseline arterial pressure, measured in mm Hg at the time of hospital admission. Although serum cholesterol (Chol) corresponds to total cholesterol measured in mg/dL. Fasting blood sugar (Fbs) and resting electrocardiographic attributes (Restecg) are also included, with Restecg classified as 0 (probable/definitive left ventricular hypertrophy according to Estes' criteria), 1 (normal) or 2 (ST-T-wave abnormalities involving T-wave inversion and/or significant elevation or depression of ST > 0.05 mV). Additional clinical variables comprise the maximum heart rate achieved (Thalach); presence of exercise-induced angina (Exang); depression in the ST segment induced by exercise relative to rest (Oldpeak); the slope of the ST segment with maximum exercise (Slope); the number of major vessels visualized by fluoroscopy (Ca); and the status of thalassemia (Thal).

The original dataset includes class labels ranging from 0 to 4, representing varying degrees of heart disease. For the purpose of this study, the dataset was reformulated into a binary classification task: presence of heart disease (originally labeled as 1, 2, 3, or 4, and recoded as 1) versus absence of heart disease (originally labeled as 0). The TABLE 2, provides a detailed description of the dataset features.

Table 2: Description of Cardiac Dataset Features

Feature Name	Description
Age	Described in years
Sex	1=Male, o=Female
CP	Chest pain (1=Typical angina, 2=Atypical angina, 3=Non-anginal pain, 4=Asymptomatic)
Trestbp	Resting blood pressure (mm Hg)
Chol	Serum cholesterol (mg/d)
Fbs	Fasting Blood Sugar > 120mg/dl (1 = True, o = False)
Restecg	Resting electrocardiographic (0=Normal,1=ST-T wave abnormality, 2=Left ventricular hypertrophy)
thalach	Maximum heart rate
exang	Exercise induced angina (1 = Yes, O = No)
oldpeak	ST-Segment Depression
Slope	The slope of exercise ST segment peak (1=Upsloping, 2=Flat, 3=Downsloping)
Ca	Count of major vessels (0-3)
Thal	Value 0: NULL, Value 1: fixed defect (no blood flow in some part of the heart), Value 2: normal blood flow Value 3: reversible defect (a blood flow is observed but it is not normal)
Target	Heart disease (1 = no, 0 = yes)

Two file formats were used in this study: Comma-Separated Values (CSV) for the diabetes dataset and a generic data file format (DAT) for the cardiac dataset. Each dataset was split into two subsets, with 70% allocated for training and 30% for testing. The diabetes dataset consisted of five files: two feature files containing eight predictor columns with varying numbers of samples, and three label files, each containing a single column of response data. Similarly, the cardiac dataset included two feature files with 13 predictor columns and three label files, each containing a single column corresponding to the class labels.

2.2 Preprocessing

The importance of the preprocessing stage is to address incomplete or missing values in a data set as well as feature selection. In this research, the cardiac data set has six instances of missing values. Thus, understanding of missing data was very important for dealing with as they will significantly impact the validity of the analysis. Moreover, examining the nature of the data and the distribution of missing values is important to enhance the robustness and reliability of the data processing pipeline. In this study, the following strategies were explored to handle incomplete or missing values:

- Ignorance procedure: proceeding with the algorithm while allowing missing values to remain in the dataset.
- List-wise deletion (LWD): reducing the dimensionality of the problem by removing any rows that contains one or more missing values, thereby focusing only on complete cases.

2.3 Feature Selection

In this study, two feature selection approaches were employed:

- Utilization of the full set of predictors - All available features were used without any filtering or ranking.
- Utilization of top-ranked predictors - A subset of features was selected based on their importance as determined by feature ranking algorithms. Three features ranking methods were investigated in this study namely MRMR, KNN ranking method and univariate feature ranking.
  - MRMR method ranks the importance of dataset predictors based on their correlation with the output class, rather than the correlation among the predictors themselves [27, 28]. MRMR relies on the computation of Mutual Information (MI) between two variables, as shown in Equation 1:

    $$I(X:Z) = \sum P(x,z) \log \frac{P(x,z)}{P(x)P(z)} \tag{1}$$

    where $X = (x_1, x_2, .., x_i)$, $Z = (z_1, z_2, .., z_i)$, and P is the probability of a given variable. The goal of MRMR algorithm is find the optimal subset of features that maximize V_S, representing the relevance of the feature subset S with respect to the response variable y, while minimizing W_S, which measures the redundancy among the features in S. The relevance and redundancy are defined with mutual information I:

$$V_S = \frac{1}{|S|} \sum_{x \in S} I(x, y), \tag{2}$$

$$W_S = \frac{1}{|S|^2} \sum_{x,z \in S} I(x, z), \tag{3}$$

where $|s|$ denotes the number of features. The MRMR ranks the features using mutual information quotient (MIQ) value as following:

$$MIQ_x = \frac{V_x}{W_x} \tag{4}$$

– KNN ranking method ranks predictors based on the proximity of representative observations for each class [29]. Considering a random observation x_r, x_q a query observation, and a categorical variable y, the weight W_j is defined according to Equation 5 for observations belonging to the same class, and Equation 6 for observations from different classes: where i represents the update step in the learning process (i is the current step and $i-1$ is the previous step). Note that, each step updates the weight W_j based on the difference in observations. Thus, W_j^i is the value of weight W_j after the i^{th} update step.

$$W_j^i = W_j^{i-1} - \frac{\Delta j\ (x_r, x_a)}{m} \cdot d_{ra} \tag{5}$$

$$W_j^i = W_j^{i-1} + \frac{\left(P_{x_q}\right) \Delta j\ \left(x_r x_g\right)}{1 - P_{y_r} \quad m} \cdot d_{rq} \tag{6}$$

Let P_{y_r} and P_{y_q} denote the class membership probabilities of observations x_r and x_q, respectively. Let m be the number of iterations, and let $\Delta j(x_r, x_q)$ represent the difference in the value of the j-th predictor between the two observations. The distance function d_{r_q} is calculated as shown in Equation 7, where k_a represents the number of nearest neighbors:

$$d_{rq} = \frac{sd_{rq}}{\sum_{l=1}^{k} sd_{rl}} \tag{7}$$

Here, sd_{rq} refers to the scaled distance between observations r and q, and sd_{rl} denotes the scaled distance between observation r and its l-th nearest neighbour. The scaled distance sdr_q is calculated as described in Equation 8, based on the position of the q-th observation among the closest neighbors of the r-th observation:

$$sd_{rq} = e^{-(rank(r,q)/\mathrm{sigma})^2} \tag{8}$$

– Univariate feature ranking ranks the importance of predictive features based on a statistical approach that relies on the Chi-square test, which assesses the independence of each feature [30]. The score is calculated as shown in Equation 9:

$$x^2 = \frac{\sum (f_o - f_e)^2}{f_e} \tag{9}$$

where f_o represents the observed frequency and f_e denotes the expected frequency.

2.4 Classification

This study employed multi-classification approaches to predict the health status of individuals with either diabetes or heart disease. This approach was utilized to compare of multi classifiers combined with different features selection approaches. The implementation of parameters of this approach were configured as follows:

- The classification learner variable was iteratively modified using one of the following options: SVM, KNN,decision tree, LDA, DLDA, or NB.
- The coding design variable was systematically modified using one of the following configuration options, based on the number of classes K:
  - Ternary complete: This design partitions the classes into three groups, as described in [31, 32], with binary learner calculated as:

$$\left(3^K - 2^{K+1} + 1\right) \div 2 \tag{10}$$

  - Binary complete: This design partitions the classes into two groups [32], with the binary learners calculated as:

$$2^{K-1} - 1 \tag{11}$$

  - All pairs or one vs. one: Considering samples by case against case [33], with a number of binary learners calculated as:

$$K * (K - 1) \div 2 \tag{12}$$

  - One vs. All: This design compares each class against all remaining classes, with the number of binary learners equal to the number of classes.
  - Ordinal: This design incorporates class order into the classification process [34], with the number of binary learners equal to:

$$K - 1 \tag{13}$$

2.5 Evaluation

The classifiers' performance was measured using seven metrics: accuracy (ACC), macro average precision (MAP), weighted average precision (WAP), macro average recall (MAR), weighted average recall (WAR), macro average F1-score (MAS-F1), and weighted average F1-score (WAS-F1). These weighted metrics were employed to guarantee fair evaluation among different classes.

ACC is defined as the ratio of true predicted classes (TPC) to the total number of actual class instances (ACI), and is calculated as:

$$\text{Accuracy} = \frac{\left(\sum_{i=1}^{n} TPC_i\right)}{\text{ACI}} \tag{14}$$

MAP is a measure used to evaluate a model's performance for each label individually and is calculated as follows:

$$MAP = \frac{\left(\sum_{i=1}^{n} P_i\right)}{n} \tag{15}$$

where p denotes precision, which is calculated as the ratio of TPC to the total number of predicted class instances (PCI) as shown in the following equation:

$$P = \frac{TPC}{PCI} \tag{16}$$

WAP is calculated using the equation shown below, where the precision of each class is multiplied by its corresponding weight:

$$WAP = \sum_{i=1}^{n} P_i w_i \tag{17}$$

where w_i represents the weight of the i-th class.

MAR is computed as follows:

$$MAR = \frac{\left(\sum_{i=1}^{n} R_i\right)}{n} \tag{18}$$

where R_i represents recall which is defined as the ratio of TPC to the total number of ACI, also referred to as the sensitivity, and is computed as follows:

$$Recall = \frac{TPC}{ACI} \tag{19}$$

WAR is calculated using the following equation:

$$WAR = \sum_{i=1}^{n} R_i w_i \tag{20}$$

MAS-F1 is another metric used to assess the performance of a classifier and is defined as:

$$MAS - F1 = \frac{\left(\sum_{i=1}^{n} F_1 - score_i\right)}{n} \tag{21}$$

where the F1-score measures the accuracy of a classifier by combining precision and recall through their harmonic mean. It is calculated as shown in the equation below:

$$F_1 - score = \frac{P.R}{P + R} \tag{22}$$

The seventh evaluation metric, known as the $WAS - F1$, is defined in the equation below:

$$WAS - F_1 = \sum_{i=1}^{n} F_1 - score_i w_i \tag{23}$$

The weight (w_i) used in the weighted metrics was assigned to each class according to its relative frequency in the dataset. The weight for class i is calculated as shown in the equation below:

$$w_i = \frac{n_i}{N} \tag{24}$$

In the above equation, n_i denotes to the number of samples belonging to class i, while N represents the total number of samples across all classes. This weighting method ensures that classes with a larger presence in the dataset have a proportionally greater influence on the overall metric, while maintaining balanced evaluation between majority and minority classes.

3. RESULTS

In the results, we present the proposed the approaches for predicting individuals' health status. As we mentioned that we have used two distinct datasets: one for diabetes and one for heart disease. The study evaluated various versions of the proposed hybrid classification model using:

- Six classifiers (SVM, KNN, decision tree, LDA, DLDA, NB),
- Five coding designs (One-vs-All, One-vs-One, Ternary Complete, Binary Complete, Ordinal).
- Four feature ranking techniques (full features set, MRMR, KNN-based and uniivariate features ranking using Chi-square test).

Performance was primarily assessed using ACC, MAP, WAP, MAR, WAR, MAS-F1, and WAS-F1. Weighted metrics (WAP, WAR, and WAS-F1) were computed using class-proportional weights as described in Section 2.5 to ensure balanced evaluation across classes. In the following subsection, the results of the various experiments applied on diabetes dataset and cardiac dataset were illustrated.

3.1 Diabetes Dataset

The several experiments employed three features selection algorithms with six classification algorithms. TABLE 3, shows the performance of the six classifiers using the full predictors. These set of predictors namely plasma glucose concentration, BMI, number of pregnancy, 2-Hour serum insulin, triceps skinfold thickness, age, diastolic blood pressure, and diabetes pedigree function.

Table 3: Performance of multi-classification using the full feature set for diabetes disease prediction

Classification Learner	Coding Design	ACC	MAP	WAP	MAR	WAR	MAS-F1	WAS-F1
SVM	All	78.3	0.78	0.782	0.73	0.783	0.743	0.77
KNN	All	70.4	0.675	0.697	0.662	0.704	0.666	0.698
Decision tree	All	66.5	0.639	0.672	0.645	0.665	0.642	0.668
LDA	All	78.7	0.788	0.787	0.734	0.787	0.747	0.776
DLDA	All	77.8	0.766	0.774	0.735	0.778	0.745	0.771
NB	All	77.0	0.754	0.765	0.728	0.769	0.736	0.763

As shown in TABLE 3, the full set of diabetes predictors was used to evaluate the baseline performance of six classification algorithms. The results show that the LDA and SVM classifiers achieved the highest accuracies of 78.7% and 78.3%, respectively, suggesting that the dataset exhibits a largely linear relationship between the predictors and diagnostic outcomes. Furthermore, the DLDA and NB models also performed competitively, with accuracies above 77%, indicating their robustness and suitability for datasets with limited dimensionality. In contrast, the KNN and decision tree classifiers showed lower performance, likely due to their sensitivity to data distribution and potential overfitting. We think that these findings help to establish a performance benchmark which reflects that linear and probabilistic models are effective for the prediction of diabetes when utilizing all features.

In the second experiment, the MRMR feature selection approach was employed to evaluate the relative importance of each predictor in the diabetes dataset. The features were ranked in the following order: plasma glucose concentration, BMI, number of pregnancies, 2-Hour serum insulin, triceps skinfold thickness, age, diastolic blood pressure, and diabetes pedigree function (ranked as the least important predictor). FIGURE 1 presents the feature ranking scores, revealing a substantial gap between the top two predictors and the remaining features. These two features exhibited markedly higher importance scores compared to all others. Consequently, only the top two features were selected as inputs for multiple classifiers to predict diabetes. Moreover, empirical evaluations using additional features (beyond the top two or three) did not result in noticeable improvements in predictive performance but increased model complexity. Therefore, the chosen cutoffs were determined to achieve an optimal balance between model accuracy and simplicity. TABLE 4, shows the performance of the six classifiers using the top two features ranked according to MRMR.

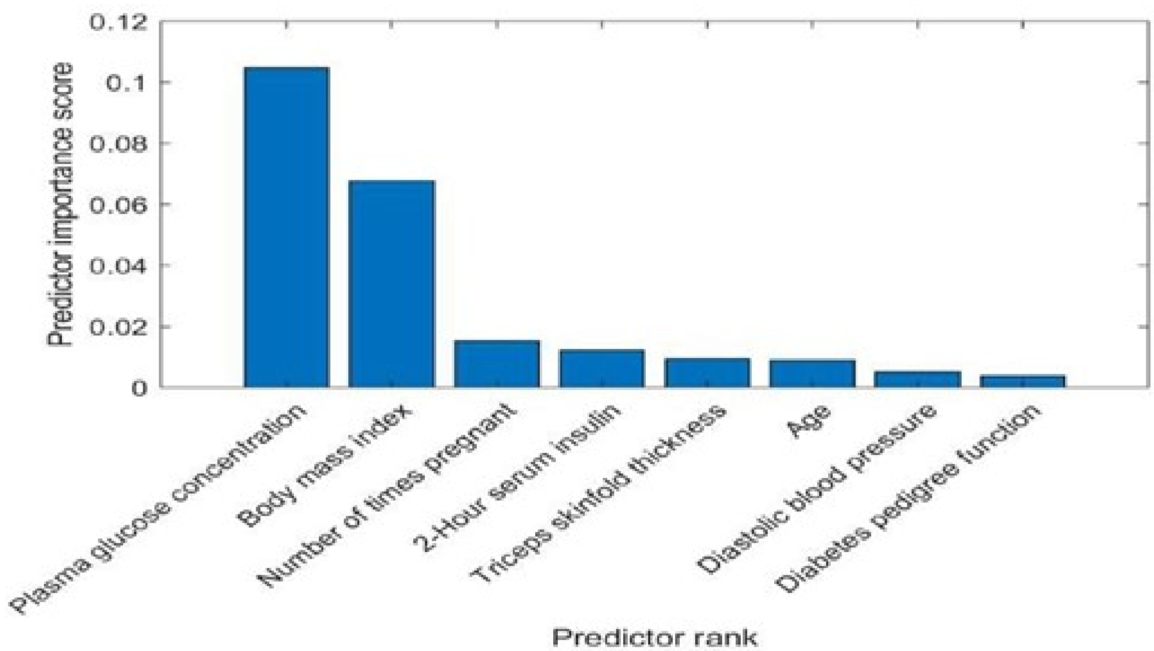

Figure 1: Rank of diabetes features according to MRMR ranking approach

We also explored other features selection approach which was KNN-based features ranking. The ranking function nominated the three features plasma glucose concentration, age, and BMI as the top three features. The selection of these features was guided by empirical evaluation, as including additional lower-ranked features did not yield statistically significant improvements in performance. Followed by the following features number of pregnancies, diastolic blood pressure, diabetes pedigree function, 2-Hour serum insulin, and triceps skinfold thickness. FIGURE 2, shows the ranking of the features according to the KNN algorithm. Table 5, shows the performance of the six classifiers using the top three features ranked according to KNN.

Table 4: Performance of multi-classification using the top two features according to MRMR for diabetes disease prediction

Classification Learner	Coding Design	ACC	MAP	WAP	MAR	WAR	MAS-F1	WAS-F1
SVM	All	76.5	0.763	0.764	0.706	0.765	0.718	0.751
KNN	All	71.3	0.685	0.707	0.674	0.713	0.678	0.708
Decision tree	All	68.7	0.653	0.674	0.632	0.687	0.636	0.675
LDA	All	76.1	0.756	0.758	0.703	0.761	0.714	0.747
DLDA	All	76.5	0.76	0.763	0.709	0.765	0.72	0.753
NB	All	77.0	0.771	0.77	0.709	0.769	0.722	0.755

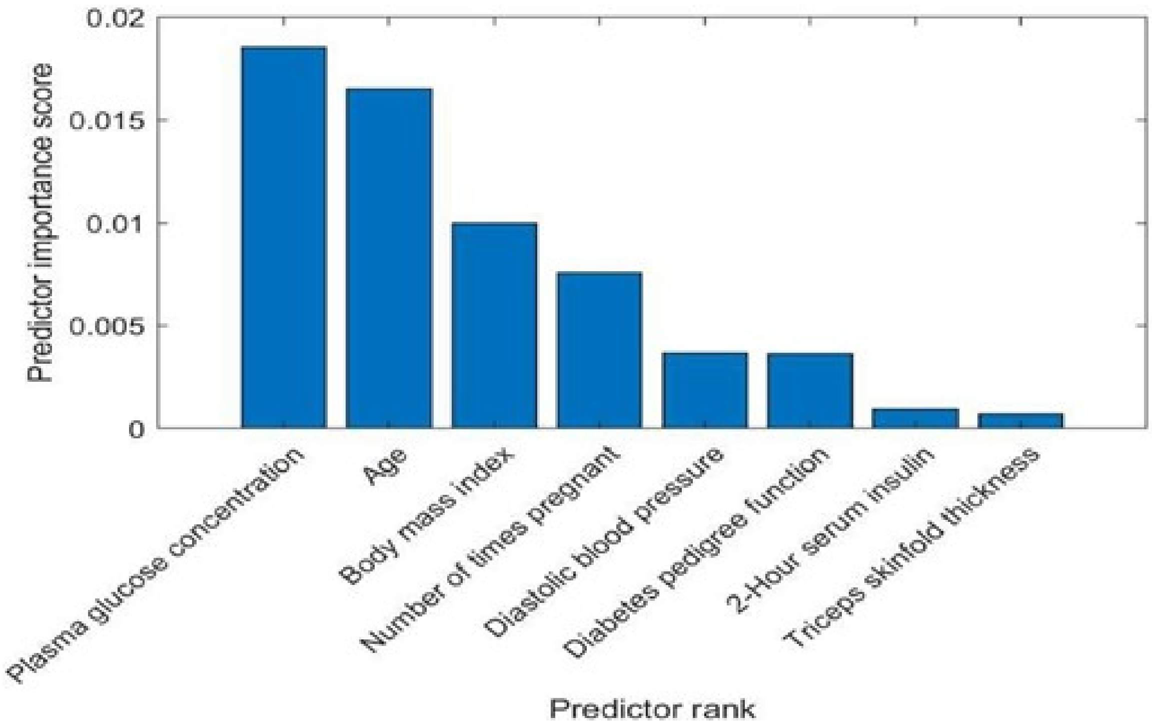

Figure 2: Rank of diabetes features according to KNN ranking approach

In the fourth experiment, a univariate feature ranking approach based on the Chi-square test was employed. This method evaluates the strength of association between each independent variable and the target variable. The resulting feature ranking obtained using this method was as follows: plasma glucose concentration, BMI, age, 2-hour serum insulin, number of pregnancies, triceps skinfold thickness, diastolic blood pressure, and diabetes pedigree function. FIGURE 3, shows the ranking score of each feature. The performance of the six classifiers using top three features ranked according to chi-square test is shown in TABLE 6. The best classification performance was reached up 76.5% using SVM classifier.

Table 5: Performance of multi-classification using the top three features according to KNN approach for diabetes disease prediction

Classification Learner	Coding Design	ACC	MAP	WAP	MAR	WAR	MAS-F1	WAS-F1
SVM	All	78.3	0.787	0.785	0.725	0.783	0.739	0.769
KNN	All	68.7	0.658	0.685	0.656	0.686	0.656	0.686
Decision tree	All	75.2	0.73	0.748	0.72	0.752	0.724	0.749
LDA	All	78.3	0.787	0.785	0.725	0.783	0.739	0.769
DLDA	All	78.7	0.785	0.785	0.737	0.787	0.749	0.777
NB	All	78.7	0.791	0.789	0.731	0.787	0.745	0.775

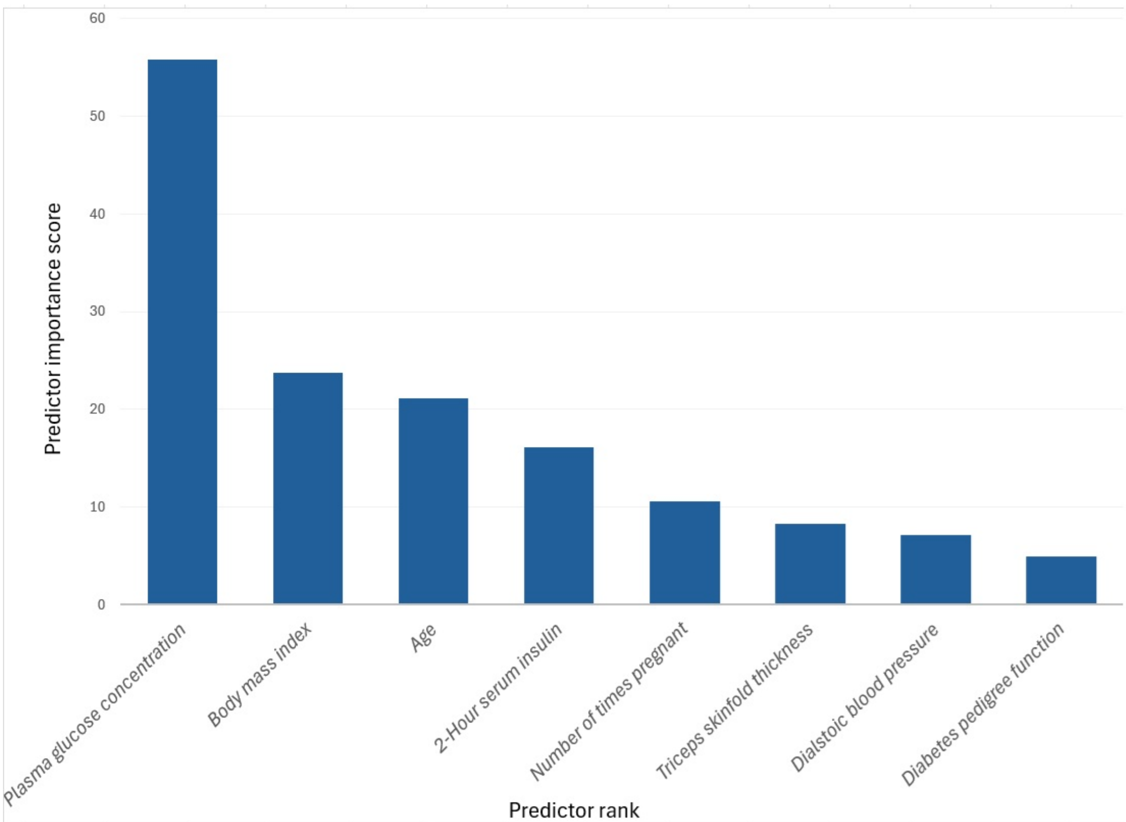

Figure 3: Rank of diabetes features according to univariate feature ranking approach based on the Chi-square test

To sum up, the best performance with full predictors was achieved using LDA and SVM, attaining an accuracy of 78.7%. When applying the MRMR feature selection method (top two features), the highest accuracy of 77% was obtained using the NB classifier. In the experiment employing

Table 6: Performance of multi-classification using the top three features according to univariate feature ranking approach based on the Chi-square test for diabetes disease prediction

Classification Learner	Coding Design	ACC	MAP	WAP	MAR	WAR	MAS-F1	WAS-F1
SVM	All	76.5	0.744	0.76	0.724	0.765	0.731	0.76
KNN	All	66.1	0.631	0.667	0.635	0.661	0.632	0.661
Decision tree	All	66.2	0.662	0.697	0.67	0.687	0.664	0.691
LDA	All	76.1	0.737	0.757	0.726	0.761	0.731	0.758
DLDA	All	74.8	0.722	0.745	0.716	0.748	0.719	0.746
NB	All	76.1	0.737	0.757	0.726	0.761	0.731	0.758

the KNN-based feature ranking (top three features), the best performance was achieved with both DLDA and NB, each reaching 78.7% accuracy. Conversely, adopting the Chi-square test for feature ranking yielded the lowest performance, with an accuracy of 76.1%.

3.2 Cardiac Disease Dataset

The cardiac dataset contained several missing values. To address this issue, two approaches were implemented: the ignorance procedure, where missing values were replaced with the keyword ”NAN,” and the LWD approach, where entries with missing values were removed entirely. The results of applying the ignorance procedure using the full set of 13 features are presented in TABLE 7.

Table 7: Performance of multi-classification using ignorance preprocessing technique and full feature set for cardiac disease prediction

Classification Learner	Coding Design	ACC	MAP	WAP	MAR	WAR	MAS-F1	WAS-F1
SVM	All	85.6	0.855	0.856	0.854	0.856	0.855	0.855
KNN	All	56.7	0.563	0.564	0.561	0.567	0.56	0.564
Decision tree	All	72.2	0.727	0.731	0.726	0.722	0.722	0.723
LDA	All	84.4	0.845	0.845	0.842	0.844	0.843	0.844
DLDA	All	82.2	0.823	0.822	0.819	0.822	0.821	0.822
NB	All	82.2	0.823	0.823	0.823	0.822	0.822	0.822

Using the second feature selection approach, the MRMR ranked cp, sex, and ca as the top three features. The ranking of features from most to least important, was as follows: cp, sex, Ca, thalach, oldpeak, thal, exang, restecg, age, slope, chol, fbs, and trestbps (ranked as the least important feature). FIGURE 4 shows the score of features according MRMR. Consequently, only the two best features were fed into multiple classifiers to predict cardiac disease. Table 8 shows the performance of the six classifiers using the top two features ranked based on MRMR.

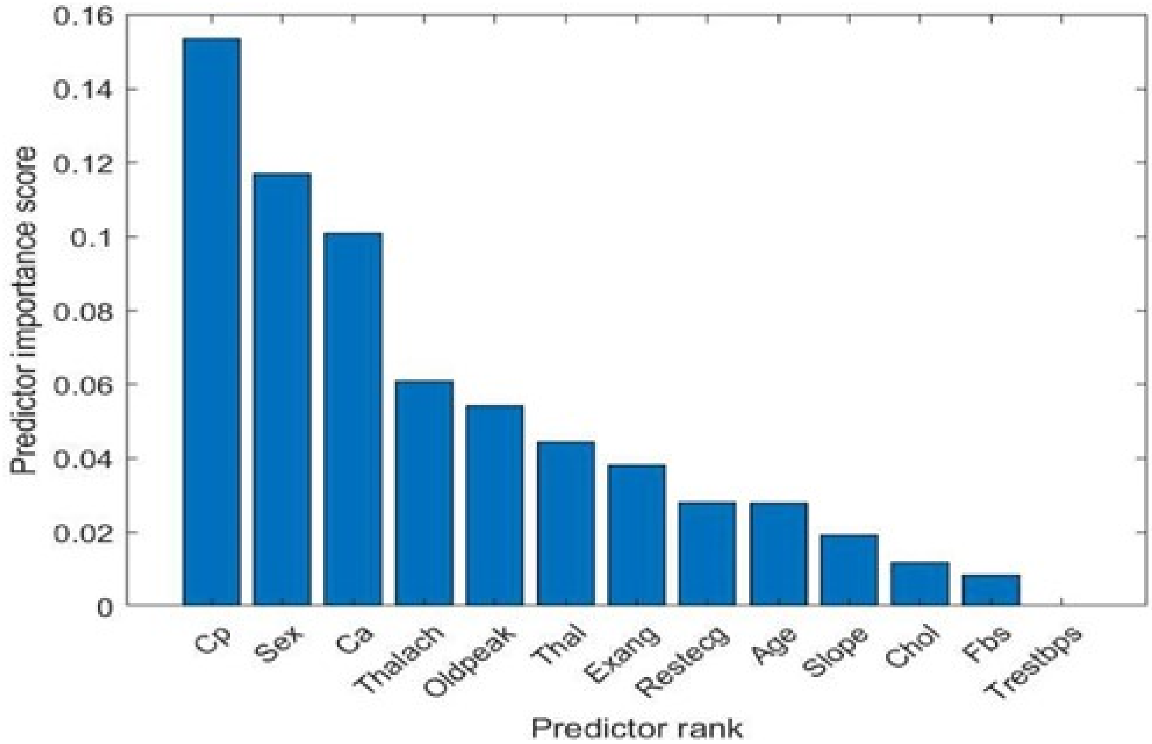

Figure 4: Rank of cardiac features according to MRMR ranking approach

Table 8: Performance of multi-classification using ignorance preprocessing technique and MRMR features selection for cardiac disease prediction

Classification Learner	Coding Design	A	MAP	WAP	MAR	WAR	MAS-F1	WAS-F1
SVM	All	77.8	0.778	0.778	0.775	0.778	0.776	0.777
KNN	All	74.4	0.799	0.792	0.729	0.744	0.723	0.728
Decision tree	All	77.8	0.778	0.778	0.775	0.778	0.776	0.777
LDA	All	77.8	0.778	0.778	0.775	0.778	0.776	0.777
DLDA	All	78.9	0.788	0.789	0.787	0.789	0.788	0.789
NB	All	78.9	0.788	0.789	0.787	0.789	0.788	0.789

The third features selection approach was univariate feature ranking using Chi-square tests. It ranked the features cp, thal, and ca as the three significant features. The ranking of features was as follows: exang, oldpeak, thalch, slope, sex, age, restecg, chol, trestbps, and fbs. FIGURE 5, presents the feature ranking according to univariate feature ranking. The top three highest-ranking features were selected and used as inputs for the proposed classifiers. TABLE 9, summarizes the performance of the proposed classifiers in predicting cardiac disease.

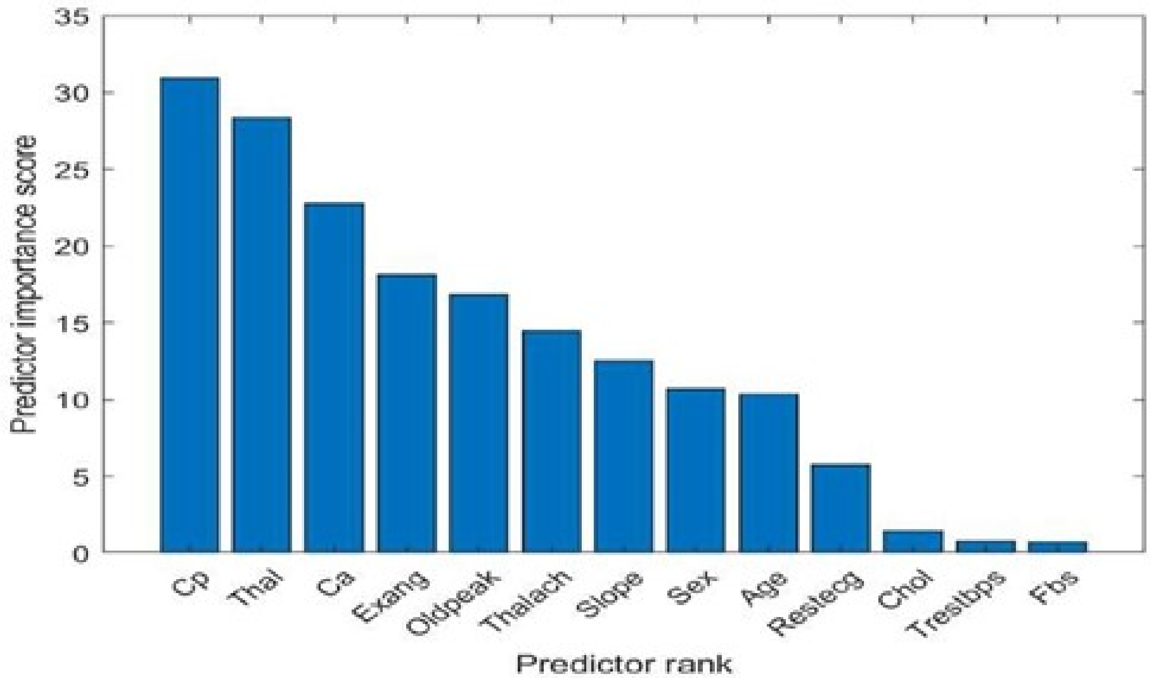

Figure 5: Rank of cardiac features according to univariate feature ranking approach

Table 9: Performance of multi-classification using ignorance preprocessing technique and univariate feature ranking approach for cardiac disease prediction

Classification Learner	Coding Design	ACC	MAP	WAP	MAR	WAR	MAS-F1	WAS-F1
SVM	All	78.9	0.788	0.789	0.787	0.789	0.788	0.789
KNN	All	77.8	0.787	0.784	0.771	0.778	0.773	0.775
Decision tree	All	85.6	0.857	0.857	0.853	0.856	0.854	0.855
LDA	All	78.9	0.788	0.789	0.787	0.789	0.788	0.789
DLDA	All	78.9	0.788	0.789	0.787	0.789	0.788	0.789
NB	All	80.0	0.799	0.8	0.799	0.8	0.799	0.8

The KNN-based feature selection method could not be applied using this preprocessing approach because the dataset contained missing values. In the KNN algorithm, the computation of the mean for each feature is required; however, the presence of missing values represented by the keyword NaN prevents the calculation of valid mean values, thereby hindering the application of this method.

The second preprocessing technique adopted to handle the problem of missing values was LWD. By using LWD, six rows from the predictor columns Ca and thal were dropped. The results of using the second preprocessing technique- LWD- with first features selection approach-full features- is presented in TABLE 10.

Table 10: Performance of multi classification using LWD preprocessing and the full feature set for cardiac disease prediction

Classification Learner	Coding Design	A	MAP	WAP	MAR	WAR	MAS-F1	WAS-F1
SVM	All	84.3	0.856	0.852	0.8346	0.843	0.838	0.84
KNN	All	55.1	0.545	0.5477	0.544	0.551	0.544	0.548
Decision tree	All	74.2	0.746	0.745	0.734	0.742	0.735	0.738
LDA	All	84.3	0.864	0.859	0.833	0.843	0.837	0.839
DLDA	All	82.0	0.839	0.835	0.81	0.82	0.813	0.816
NB	All	83.1	0.841	0.838	0.824	0.831	0.827	0.829

Using the feature ranking selection approach, the MRMR function identified cp, ca, thalach, and sex as the top four predictors. The complete features ranking, from most to least important, was as follows: cp, ca, thalach, sex, thal, oldpeak, fbs, exang, restecg, age, slope, trestbps, and chol. FIGURE 6 shows the ranks of the features. Four top-ranked featurees were fed into proposed classifiers and the results illustrated in TABLE 11.

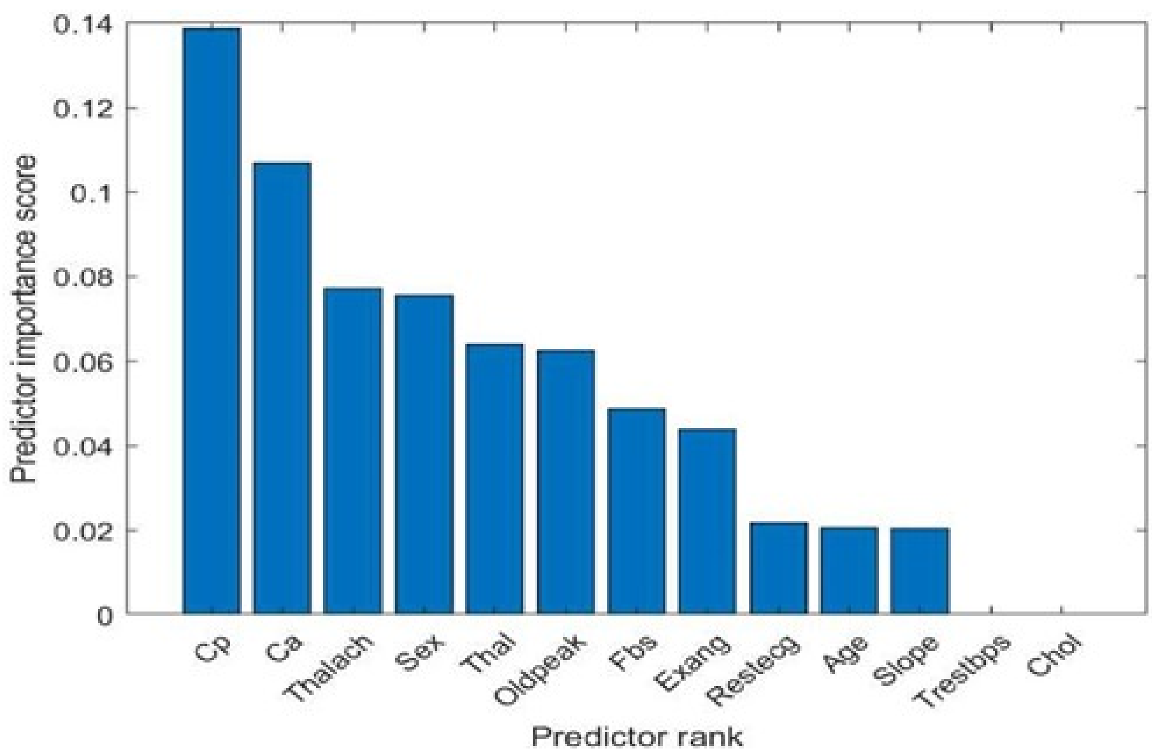

Figure 6: Rank of cardiac features according MRMR feature ranking approach and LWD preprocessing technique

Table 11: Performance of multi-classification using LWD preprocessing and the MRMR features selection approach for cardiac disease prediction

Classification Learner	Coding Design	A	MAP	WAP	MAR	WAR	MAS-F1	WAS-F1
SVM	All	79.8	0.815	0.81	0.788	0.798	0.79	0.793
KNN	All	75.3	0.76	0.758	0.744	0.753	0.746	0.749
Decision tree	All	78.7	0.789	0.788	0.781	0.787	0.783	0.785
LDA	All	79.8	0.815	0.81	0.788	0.798	0.79	0.793
DLDA	All	78.7	0.799	0.796	0.777	0.787	0.779	0.783
NB	All	80.9	0.823	0.819	0.799	0.809	0.803	0.805

In the third experiment, the KNN-based feature selection approach was employed. The features were ranked according to their importance as follows: presence of thalassemia, chest pain type, number of major vessels, exercise-induced angina, slope, maximum heart rate, cholesterol level, ST depression induced by exercise, resting electrocardiographic results, sex, age, blood pressure, and fasting blood sugar. Since the ranking of features showed a convergent pattern, and to maintain consistency with the diabetes datasets, the top three features were selected for model evaluation. FIGURE 7 illustrates the ranking of the 13 cardiac features, and TABLE 12 presents the performance of multiple classifiers using these top-ranked features.

Table 12: Performance of multi-classification using LWD preprocessing and the KNN features selection approach for cardiac disease prediction

Classification Learner	Coding Design	A	MAP	WAP	MAR	WAR	MAS-F1	WAS-F1
SVM	All	81.1	0.81	0.811	0.81	0.811	0.81	0.811
KNN	All	81.8	0.819	0.818	0.815	0.818	0.816	0.818
Decision tree	All	83.2	0.833	0.832	0.828	0.832	0.83	0.831
LDA	All	80.8	0.809	0.808	0.805	0.808	0.806	0.807
DLDA	All	81.1	0.812	0.0.812	0.808	0.811	0.809	0.811
NB	All	80.1	0.802	0.802	0.797	0.801	0.799	0.80

On the other hand, the univariate feature ranking approach identified chest pain type, number of major vessels, and presence of thalassemia as the most significant features. The remaining features were ranked in the following order: exercise-induced angina, ST depression induced by exercise, maximum heart rate, slope, sex, age, resting electrocardiographic results, fasting blood sugar, cholesterol level, and resting blood pressure. FIGURE 8 shows the ranking of the predictors. TABLE 13 shows the performance of feeding the top three features into proposed classifiers.

To summarize, the best performance using the full feature set was achieved with the SVM classifier and the ignorance preprocessing technique, yielding an accuracy of 85.6%. Applying the MRMR feature selection approach produced the highest performance of 80.9% using the NB model with the LWD preprocessing technique. In the KNN-based feature selection approach, the best result reached 83.2% using the decision tree model with the LWD preprocessing method. On another

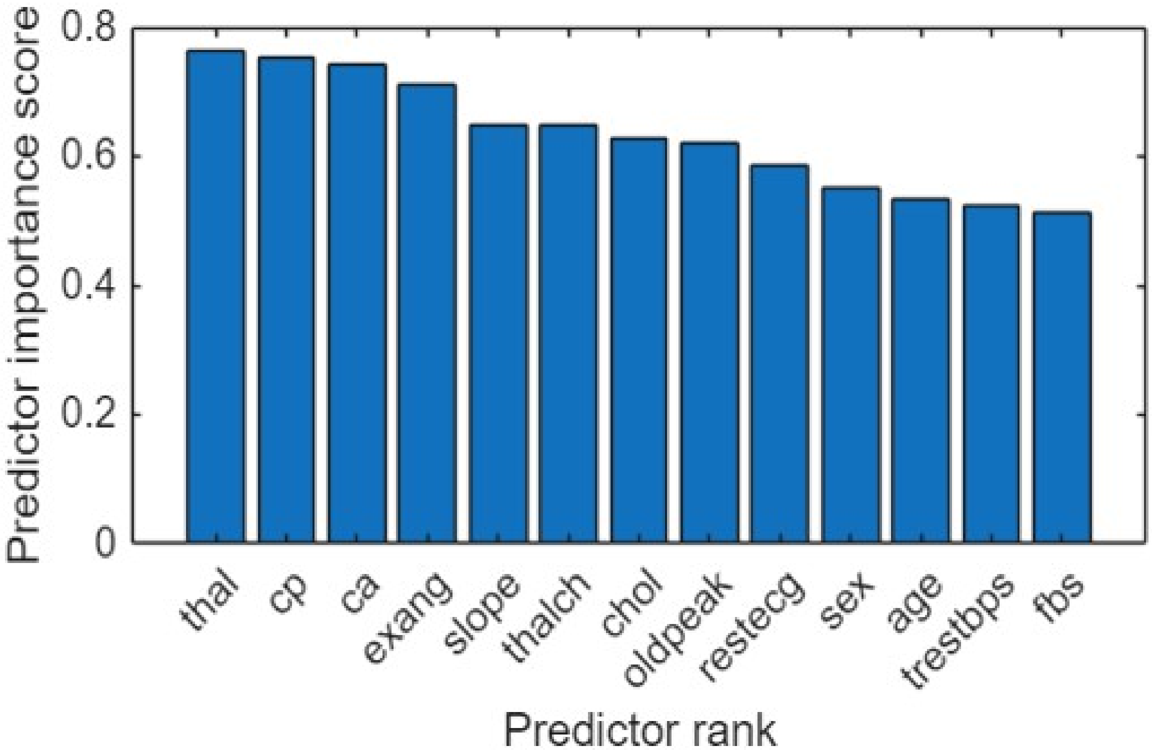

Figure 7: Rank of cardiac features according KNN feature ranking approach and LWD preprocessing techinque

Table 13: Performance of multi-classification using LWD preprocessing and the univariate features ranking approach for cardiac disease prediction

Classification Learner	Coding Design	A	MAP	WAP	MAR	WAR	MAS-F1	WAS-F1
SVM	All	82.0	0.823	0.822	0.816	0.82	0.817	0.819
KNN	All	82.0	0.823	0.822	0.816	0.82	0.817	0.819
decision tree	All	85.4	0.859	0.857	0.849	0.854	0.851	0.853
LDA	All	82.0	0.823	0.822	0.816	0.82	0.817	0.819
DLDA	All	82.0	0.823	0.822	0.816	0.82	0.817	0.819
NB	All	83.1	0.833	0.832	0.827	0.831	0.829	0.831

hand, adopting the Chi-square test for feature ranking achieved the best performance with both preprocessing techniques and the decision tree classifier, reaching an accuracy of 85.6%.

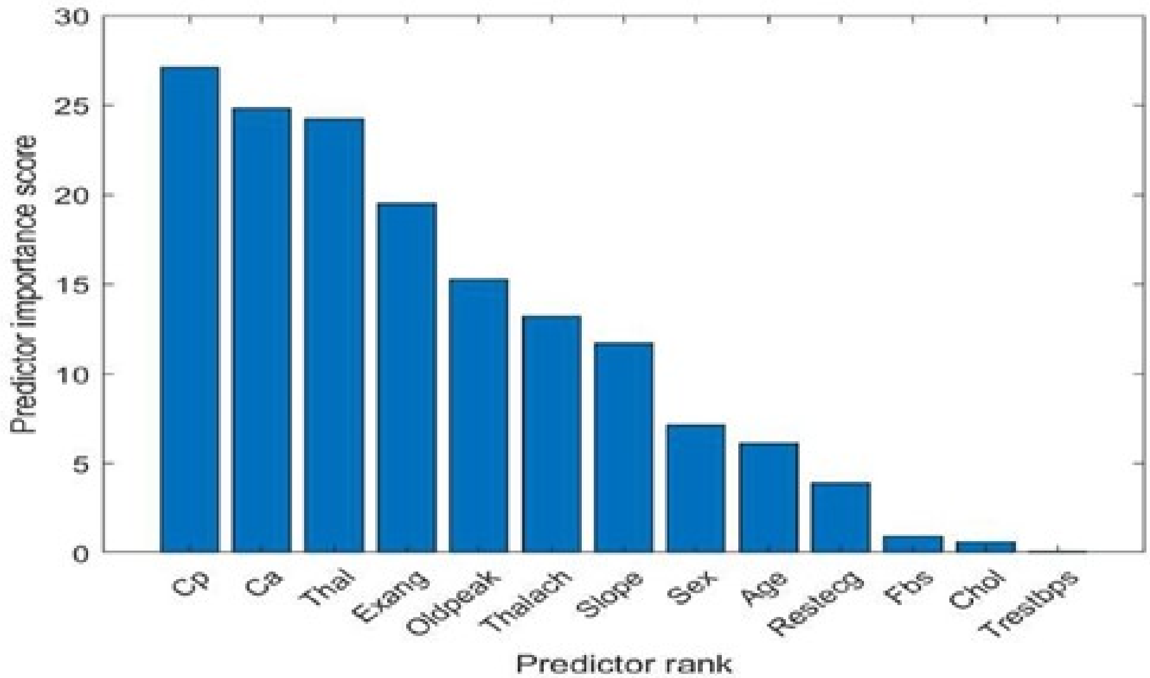

Figure 8: Rank of features according univariate feature ranking approach based on the Chi square test and LWD preprocessing techinque

4. DISCUSSION

This study explored the effectiveness of multi-classification algorithms in predicting diabetes and cardiac disease. Additionally, five different encoding schemes were examined.

The results of this study show that using the full set of predictors yielded the highest accuracy for diabetes prediction, reaching approximately 78.7% with both LDA and SVM while adopted the MRMR method with the top 2 features achieved the highest accuracy at 77% using NB model, demonstrating its robustness with a minimal feature set. In the KNN-based feature ranking (top 3 features), both DLDA and NB achieved the highest accuracy of 78.7%. According to the performance analysis, DLDA exhibited balanced precision and recall, achieving the highest F1-scores (MAS-F1: 74.9%, WAS-F1: 77.7%). NB also showed outstanding average precision across all classes, with MAP and WAP scores of 79.1% and 78.9%, respectively. On the another hand, employing univariate feature ranking using the Chi-square test achieved the highest performance of 76.1%, obtained with both the LDA and NB classifiers.

These findings suggest that KNN-based feature selection consistently enhances performance by effectively identifying the most relevant features. Regarding model performance, NB with a minimal number of features selected by MRMR and Chi-square methods performed well, indicating that the algorithm is suitable for data-constrained environments. In terms of F1-scores, the DLDA classifier achieved the highest performance when combined with KNN-based feature selection.

This demonstrates that DLDA can maintain balanced classification performance when features are selected using the KNN approach.

In the cardiac disease dataset, the SVM and LDA classifiers achieved the highest accuracy (around 85.6%) when no feature selection was applied. However, when using MRMR, the accuracy of DLDA and NB slightly decreased to 78.9%, revealing that MRMR removed some useful features for this data.

Univariate Chi-square ranking produced the best results with the decision tree classifier, which reached 85.6% accuracy and strong MAS-F1 and WAS-F1 scores. The NB classifier also yielded stable performance under this method, achieving 80% accuracy. Decision trees likely performed well because they naturally handle feature interactions and non-linear patterns in medical data.

The KNN-based feature selection method also provided good performance reached to 83.2% accuracy with the decision tree. However, this method is sensitive to missing values, which prevented its use when missing data were kept in the cardiac dataset.

The novelty of the study is lied on providing the comprehensive comparison of three feature selection methods (MRMR, KNN-based selection, and Chi-square ranking) across two real medical datasets (diabetes and cardiac disease) experimental settings. Although the results were generally similar across methods, each approach exhibited distinct strengths. MRMR effectively reduces redundant and correlated features and is characterized by robustness to multicollinearity; however, it can be computationally heavy and sensitive to dataset size due to its reliance on mutual information estimation. In addition, the MRMR is limited in scalability for high-dimensional data, which may explain its slightly lower performance in the cardiac dataset compared to other approaches. In contrast, KNN-based feature selection captures nonlinear relationships within features, but it is computationally expensive and memory-intensive and affected by noise and missing data. This sensitivity explains why KNN could not be applied under the ignorance preprocessing approach, which retained missing values in the cardiac dataset represented as NaN. On the other hand, Chi-square ranking is fast and simple, but does not consider interactions between features.

In conclusion, feature ranking improved model performance and efficiency in both datasets. decision tree, NB and DLDA performed well when fewer features were used, making them suitable for models that must remain simple or work with limited data.

5. CONCLUSION

This study presented a comprehensive comparative assessment of hybrid multi-classification algorithms to enhance the accuracy of health state diagnosis, with a particular focus on diabetes and cardiovascular diseases. The results demonstrated that the application of feature ranking notably improved both computational efficiency and classification performance. Among the evaluated approaches, the KNN algorithm and univariate feature ranking methods exhibited superior effectiveness. Furthermore, the NB and DLDA classifiers achieved optimal performance in predicting diabetic conditions, attaining accuracies of up to 78.7% even with a reduced feature set. In the case of cardiovascular data, the integration of the decision tree classifier with univariate feature ranking achieved the highest accuracy of 85.6%. These findings underscore the critical role of judicious feature selection, while also revealing the inherent limitations of relying solely on conventional tabular features for medical diagnosis. Although the proposed framework demonstrates improvements over existing studies, the results also indicate the need for further refinement. Future research should focus on expanding the dataset and sample diversity to strengthen the robustness, reliability, and generalizability of the proposed diagnostic models.

6. FUNDING

The authors thank the Deanship of Scientific Research (DSR) at the University of Jeddah, Jeddah, Saudi Arabia, for funding this project under grant ID UJ-21-DR-142.

7. ACKNOWLEDGMENT

The authors would like to thank Safiah Endargiri for her valuable preliminary contributions to this work.

References

[1] Ghaffar Nia N, Kaplanoglu E, Nasab A. Evaluation of Artificial Intelligence Techniques in Disease Diagnosis and Prediction. Discov Artif Intell. 2023;3:5.

[2] Ahsan MM, Luna SA, Siddique Z.Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. Healthcare (Basel). 2022;10:541.

[3] Viloria A, Herazo-Beltran Y, Cabrera D, Pineda OB. Diabetes Diagnostic Prediction Using Vector Support Machines. Procedia Comput Sci. 2020;170:376-381.

[4] Maniruzzaman M, Rahman MJ, Ahammed B, Abedin MM. Classification and Prediction of Diabetes Disease Using Machine Learning Paradigm. Health Inf Sci Syst. 2020;8:7.

[5] M. M. Islam MM, Rifat HR, Shahid MSB, Akhter A, Uddin MA, Uddin KM. Explainable Machine Learning for Efficient Diabetes Prediction Using Hyperparameter Tuning, Shap Analysis, Partial Dependency, and Lime. Eng Rep. 2025;7.

[6] Abnoosian K, Farnoosh R, Behzadi MH. Prediction of Diabetes Disease Using an Ensemble of Machine Learning Multi-Classifier Models. BMC Bioinform. 2023;24:337.

[7] Bhattacharya M, Datta D. Development of Predictive Models of Diabetes Using Ensemble Machine Learning Classifier. Commun Comput Inf Sci Advancements in Smart Computing and Information Security. ASCIS 2022. 2022;1759.

[8] Ganie SM, Malik MB. An Ensemble Machine Learning Approach for Predicting Type-II Diabetes Mellitus Based on Lifestyle Indicators. Healthc Anal. 2022;2:100092.

[9] Kibria HB, Nahiduzzaman M, Goni MO, Ahsan M, Haider J. An Ensemble Approach for the Prediction of Diabetes Mellitus Using a Soft Voting Classifier With an Explainable AI. Sensors. 2022;22:7268.

[10] Wee BF, Sivakumar S, Lim KH, Wong WK, Juwono FH. Diabetes Detection Based on Machine Learning and Deep Learning Approaches. Multimed Tools Appl. 2024;83:24153-24185.

[11] El-Bashbishy AE, El-Bakry HM. Pediatric Diabetes Prediction Using Deep Learning. Sci Rep. 2024;14:4206.

[12] Aslan MF, Sabanci K. A Novel Proposal for Deep Learning-Based Diabetes Prediction: Converting Clinical Data to Image Data. Diagnostics (Basel). 2023;13:796.

[13] Zhu T, Li K, Herrero P, Georgiou P. Deep Learning for Diabetes: A Systematic Review. IEEE J Biomed Health Inform. 2021;25:2744-2757.

[14] Sianga BE, Mbago MC, Msengwa AS. Predicting the Prevalence of Cardiovascular Diseases Using Machine Learning Algorithms. Intell-Based Med. 2025;11:100199.

[15] El-Sofany H, Bouallegue B, El-Latif YM. A Proposed Technique for Predicting Heart Disease Using Machine Learning Algorithms and an Explainable AI Method. Sci Rep. 2024;14:23277.

[16] Bouqentar MA, Terrada O, Hamida S, Saleh S, Lamrani D, Cherradi B et al. Early Heart Disease Prediction Using Feature Engineering and Machine Learning Algorithms. Heliyon. 2024;10:e38731.

[17] Singh A, Mahapatra H, Biswal AK, Mahapatra M, Singh D, et al. Heart Disease Detection Using Machine Learning Models. Procedia Comput Sci. 2024;235:937-947.

[18] Subramani S, Varshney N, Anand MV, Soudagar ME, Al-keridis LA, et al. Cardiovascular Diseases Prediction by Machine Learning Incorporation With Deep Learning. Front. Med. 2023;10:1150933.

[19] Al-Mahdi IS, Darwish SM, Madbouly MM. Heart Disease Prediction Model Using Feature Selection and Ensemble Deep Learning With Optimized Weight. CMES Comput Model Eng Sci. 2025;143:875-909.

[20] Kumar S, Thakur B. Heart Disease Prediction Using a Stacked Ensemble Learning Approach. SN Comput Sci. 2025;6:3.

[21] Ganie SM, Pramanik PK, Zhao Z. Ensemble Learning With Explainable AI for Improved Heart Disease Prediction Based on Multiple Datasets. Sci Rep. 2025;15:13912.

[22] Shah P, Shukla M, Dholakia NH, Gupta H. Predicting Cardiovascular Risk With Hybrid Ensemble Learning and Explainable AI. Sci Rep. 2025;15:17927.

[23] Hajishah H, Kazemi D, Safaee E, Amini MJ, Peisepar M, et al. Evaluation of Machine Learning Methods for Prediction of Heart Failure Mortality and Readmission: Meta-Analysis. BMC Cardiovasc Disord. 2025;25:264.

[24] Zhou C, Dai P, Hou A, Zhang Z, Liu L, et al. A Comprehensive Review of Deep Learning-Based Models for Heart Disease Prediction. Artif Intell Rev. 2024;57:263.

[25] Sivasubramaniam S, Balamurugan SP. Early Detection and Prediction of Heart Disease Using Wearable Devices and Deep Learning Algorithms. Multimed Tools Appl. 2025;84:6187-6201.

[26] Aha D, Murphy P, Merz C, Keogh E, Blake C, et al. UC Irvine Machine Learning Repository. 2023.

[27] Zhong K, Han M, Qiu T, Han B, Chen YW. Distributed Dynamic Process Monitoring Based on Minimal Redundancy Maximal Relevance Variable Selection and Bayesian Inference. IEEE Trans Control Syst Technol. Sep 2020;28:2037-2044.

[28] Radovic M, Ghalwash M, Filipovic N, Obradovic Z. Minimum Redundancy Maximum Relevance Feature Selection Approach for Temporal Gene Expression Data. BMC Bioinform. 2017;18:9.

[29] Guo G, Neagu D, Cronin MT. Using KNN Model for Automatic Feature Selection. In: Singh S, Singh M, Apte C, Perner P, editors. Pattern recognition and data miningICAPR. Lecture Notes in Computer Science.. Berlin, Heidelberg: Springer. 2005;3686:410-419.

[30] Sumaiya Thaseen IS, Aswani Kumar C. Intrusion Detection Model Using Fusion of Chi-Square Feature Selection and Multi Class SVM. J King Saud Univ Comput Inf Sci. 2017;29:462-472.

[31] Zhang Q, Yan L, Wu Y, Ji L, Chen Y, et al. A Ternary Classification Using Machine Learning Methods of Distinct Estrogen Receptor Activities Within a Large Collection of Environmental Chemicals. Sci Total Environ. 2017;580:1268-1275.

[32] C. Zhang, Qiu S, Wang S, He H. Target Detection Using Ternary Classification During a Rapid Serial Visual Presentation Task Using Magnetoencephalography Data. Front Comput Neurosci. 2021;15:619508.

[33] Galar M, Fernández A, Barrenechea E, Bustince H, Herrera F. An Overview of Ensemble Methods for Binary Classifiers in Multi-Class Problems: Experimental Study on One-Vs-One and One-Vs-All Schemes. Pattern Recognit. 2011;44:1761-1776.

[34] Frank E, Hall M. A Simple Approach to Ordinal Classification. In: De Raedt L, Flach P, editors. Machine learning: ECML 2001ECML. Berlin, Heidelberg: Springer. 2001;2167:145-156.