scientific reports

Check for updates

OPEN

An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database

Saravanan Srinivasan[1], Subathra Gunasekaran[2], Sandeep Kumar Mathivanan[3], Benjula Anbu Malar M. B[4], Prabhu Jayagopal[4] & Gemmachis Teshite Dalu[5]✉

Heart disease is a significant global cause of mortality, and predicting it through clinical data analysis poses challenges. Machine learning (ML) has emerged as a valuable tool for diagnosing and predicting heart disease by analyzing healthcare data. Previous studies have extensively employed ML techniques in medical research for heart disease prediction. In this study, eight ML classifiers were utilized to identify crucial features that enhance the accuracy of heart disease prediction. Various combinations of features and well-known classification algorithms were employed to develop the prediction model. Neural network models, such as Naïve Bayes and Radial Basis Functions, were implemented, achieving accuracies of 94.78% and 90.78% respectively in heart disease prediction. Among the state-of-the-art methods for cardiovascular problem prediction, Learning Vector Quantization exhibited the highest accuracy rate of 98.7%. The motivation behind predicting Cardiovascular Heart Disease lies in its potential to save lives, improves health outcomes, and allocates healthcare resources efficiently. The key contributions encompass early intervention, personalized medicine, technological advancements, the impact on public health, and ongoing research, all of which collectively work toward reducing the burden of CHD on both individual patients and society as a whole.

The healthcare industry generates a lot of data about patients, illnesses, and diagnoses, but it isn't being used correctly to produce the desired results. Heart disease and stroke are two of the main causes of death. According to a WHO report, cardiovascular diseases directly kill more than 17.8 million people every year. Because there isn't enough analysis, the healthcare industry's huge amounts of patient, illness, and diagnosis data don't have the effect on patient health that was hoped for[1]. Heart and blood vessel diseases, or CVDs, include coronary artery disease, myocarditis, vascular disease, and other conditions. Stroke and heart disease kill 80% of all people who die from CVD. Three-quarters of all people who die are under the age of 70. The main things that put you at risk for cardiovascular disease are your gender, smoking, age, family history, poor diet, lipids, lack of physical activity, high blood pressure, weight gain, and drinking alcohol[2]. High blood pressure and diabetes are two examples of things that can be passed down and make you more likely to get cardiovascular disease. Some of the other things that raise the risk are being inactive, being overweight, not eating well, having back, neck, and shoulder pain, being very tired, and having a fast heartbeat. Most people have chest pain, shoulder pain, arm pain, shortness of breath, and a general sense of weakness. As it has been for a long time, chest pain is the most common sign that the heart isn't getting enough blood[3]. This kind of chest pain is called angina in medicine. Some tests, like X-rays, Magnetic Resonance Imaging (MRI), and angiography, may help figure out what is wrong. On the other hand, sometimes important medical equipment is not easily accessible, which limits what can be done in an emergency. When it comes to figuring out what's wrong with your heart and treating it, every second counts[4]. Heart disease diagnostics aren't as good as they could be, and there is a huge need for better big-data analysis in cardiovascular system redesign and patient outcomes. But noise, incompleteness, and irregularities

[1]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, India. [2]Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India. [3]School of Computing Science and Engineering, Galgotias University, Greater Noida 203201, India. [4]School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India. [5]Department of Software Engineering, College of Computing and Informatics, Haramaya University, POB 138, Dire Dawa, Ethiopia. ✉email: gemmachis.teshite@haramaya.edu.et

in the data make it hard to draw clear, accurate, and well-grounded conclusions from them. Because of recent improvements in technologies like big data, information storage, and retrieval, computerised intelligence plays an important role in cardiology. In order to draw conclusions from the data mined with different ML models, researchers used pre-processing techniques[5]. Using a common set of algorithms and their variations, which are used to keep track of hereditary cardiac disorders and healthy controls, it is possible to predict when the first stage of heart failure will start. Classification technique, DT, SVC, LR, and RF machines are all types of algorithms that can be used to predict cardiac arrest. When it comes to machine learning, there are three main ways to think: The three main types of machine learning are task-driven supervised ML (classification/regression), data-driven unsupervised ML (clustering), and error-driven reinforcement learning (RL). Coronary artery disease is a very common disease of the main blood vessels that bring blood to the heart muscle. Plaques, which are made up of lipoproteins, can build up in the arteries of the heart, which can lead to coronary artery disease. Atherosclerosis is the name for the buildup of these plaques[6]. Atherosclerosis slows the flow of blood through the veins to the chest and other organs. It goes up if you have heart disease, angina, or a stroke. Men and women may have different warning signs and symptoms of coronary artery disease. For example, men are more likely than women to have chest pain. In addition to chest pain, women are more likely to experience shortness of breath, nausea, and sudden exhaustion. Heart failure, chest tightness, chest pressure, and chest pain can all be signs of coronary artery disease[7]. The Heart Disease Prediction System incorporates the Naive Bayesian Classification technique to assist in making decisions. By analyzing a vast database of past heart disease cases, the system uncovers valuable insights. This model is highly efficient in identifying patients at risk of heart disease. It possesses the ability to respond to intricate queries, showcasing its strengths in terms of interpretability, access to comprehensive information, and accuracy[8]. Making accurate and timely decisions is crucial in the medical field, especially when treating patients. Machine learning (ML) techniques play a significant role in predicting diseases by leveraging the extensive data generated by the healthcare industry. In India, heart disease is a leading cause of mortality, and the World Health Organization (WHO) emphasizes the importance of timely intervention to predict and prevent strokes. This paper focuses on predicting cardiovascular disease with enhanced accuracy by employing ML techniques such as Decision Tree and Naïve Bayes, in conjunction with risk factors. The dataset utilized in this study is the Heart Failure Dataset, which comprises 13 attributes[9]. The author investigated how well two algorithms, Support Vector Machine (SVM) and Naive Bayes, performed in predicting the occurrence of heart disease and the survival status of patients. The algorithms were applied to a dataset that included sixteen attributes from the University of California, Irvine's Centre for Machine Learning and Intelligent Systems. To assess the models' performance, a confusion matrix was used to visualize metrics like accuracy, recall, precision, and error. Additionally, statistical analysis was carried out by utilizing the receiver operating characteristic (ROC) curve and calculating the area under the curve to demonstrate the accuracy of the models[10]. In this research paper, a system is introduced that employs a radial basis function neural network to accurately predict eight different types of cardiac arrhythmias. The primary focus of the study is the analysis of heart rate time series data, and the proposed algorithm is specifically designed to predict specific arrhythmias, namely Left bundle branch block, Atrial fibrillation, Normal Sinus Rhythm, Right bundle branch block, Sinus bradycardia, Atrial flutter, Premature Ventricular Contraction, and Second-degree block. The heart rate time series data utilized in the study is sourced from the MIT-BIH arrhythmia database. Both linear and nonlinear features are extracted from the heart rate time series of each individual arrhythmia. Training of the radial basis function neural network (RBFN) is conducted using 70% of the feature datasets, while the remaining 30% is dedicated to predicting the occurrence of the eight cardiac diseases. The proposed approach demonstrates an impressive overall prediction accuracy of 96.33%, surpassing the performance of existing methods documented in the literature[11]. A novel method known as Radial Basis Classification is introduced for the classification of heart disease using clinical databases. Conventional classifiers that involve multiple attributes tend to have a large number of parameters, making it difficult to determine the ideal attributes. To address this, the concept of Multivariate Function Classifier Ideas is proposed, aiming to encourage a more cohesive stochastic trend and minimize the likelihood of errors or unforeseen results. This formula proves beneficial for arranging multidimensional data and enhancing the accuracy of grouping in the analysis phase. The results of the study indicate that the suggested calculation method offers higher precision compared to previous approaches[12]. The backpropagation neural network has demonstrated satisfactory performance in predicting accuracy. However, to further enhance accuracy and determine the specific type of heart disease, the paper integrates the CBR technique with the ANN. By leveraging historical patient records, a level of accuracy reaching 97% is attained. This research not only utilizes CBR to enhance accuracy but also to predict the type of heart disease. The CBR output encompasses both the identified type of heart disease and the recommended medication. This enables a comparison between the original medication and the medication suggested by the RBF (Radial Basis Function). The medication prescribed using this approach exhibits a comparative accuracy of 98%[13]. Symptoms include trouble breathing, pain in the upper back, neck, jaw, or throat, and pain, numbness, weakness, or a chill in the limbs. Due to the narrowing of blood vessels in certain parts of the body, it is possible to have coronary artery disease and not know it until you have a heart attack, angina, stroke, or heart failure. Keep an eye out for signs of heart problems, and if you're worried, talk to your doctor. If you get checked out often, heart (cardiovascular) disease may be found earlier[14]. This proposed method uses supervised ML classifiers to show how different models can predict the presence of cardiovascular disease and evaluate the performance of these classifiers, such as the random forest, decision tree, support vector machine, XGBoost, radial basis function, k-nearest neighbour, naïve bayes andlearning vector quantization.

The goal of predicting Cardiovascular Heart Disease is to develop accurate and reliable models that can assess an individual's risk of developing various cardiovascular conditions, enabling early intervention, personalized treatment, and ultimately reducing the burden of heart disease on public health.

The remaining sections of the paper are structured as follows: Section "Literature overview" provides a comprehensive review of the relevant literature. Section "Proposed methodology" presents the proposed methodology

in detail. The experimental results are analyzed and discussed in Section "Experimental results and discussion". Section "Conclusion" presents the conclusion of the study, while Section "Future work" outlines future work and potential research directions.

Literature overview

Heart rate variability (HRV) has emerged as a reliable predictor for congestive heart failure (CHF). However, challenges remain in effectively extracting temporal features and efficiently classifying high-dimensional HRV representations. To address these challenges, this study proposes an ensemble method that utilizes short-term HRV data and deep neural networks for CHF detection. The research incorporates five publicly available databases: BIDMC CHF database (BIDMC-CHF), CHF RR interval database (CHF-RR), MIT-BIH normal sinus rhythm (NSR) database, fantasia database (FD), and NSR RR interval database (NSR-RR). Three different lengths of RR segments (N = 500, 1000, and 2000) are employed to evaluate the proposed method. Initially, expert features are extracted from the RR intervals (RRIs). Subsequently, a network based on long short-term memory-convolutional neural networks is constructed to automatically extract deep-learning (DL) features. Finally, an ensemble classifier is used to detect CHF using the aforementioned features. Blindfold validation is conducted on three CHF subjects and three normal subjects, resulting in accuracies of 99.85%, 99.41%, and 99.17% for N = 500, 1000, and 2000 length RRIs, respectively, utilizing the BIDMC-CHF, NSR, and FD databases[15]. In this publication, there is a summary of past studies and an analysis of how well the algorithm works. Before training and testing different algorithms, the suggested architecture processes the data that comes in first. The author suggests using Adaboost because it makes every ML method look better. Also, the author agreed that settings could be fine-tuned to improve accuracy. Researchers came up with a deep learning strategy for analysing and spotting cardiac conditions by using the UCI dataset. They went on to say that deep neural networks could help improve the analysis and diagnosis of cardiovascular disease as a whole. Compared to other ways to improve model performance, they found that the Talos Hyper process worked the best[16]. The KNN, RF, SVM, and DT algorithms were studied as ML models for predicting heart disease with high accuracy, high recall, and high precision. As shown in their estimation method for cardiac disorders, which is hosted on the UCI ML library, SVM-based categorization was the most accurate. We looked at the results of four machine learning techniques and one neural network (NN) for spotting heart disease. This study compared algorithms for predicting cardiac dose based on things like reliability, recall, accuracy, and F1. The Deep NN algorithm was able to spot heart problems 98% of the time. In order to show that the algorithm is useful for predicting illness, they focused on how it could be used with a medical dataset. The researchers came to the conclusion that boosting and bagging are good ways to improve the performance of classifiers that aren't very good at predicting the risk of heart disease. The results showed that the accuracy of predictions went up a lot after feature selection was used, which improved the procedure[17]. Ensemble approaches were used to improve the accuracy of bad classifiers by no more than 7%. In recent years, ML algorithms have gotten a lot of praise for how accurate and useful they have become at making predictions. It is critical to be able to create and recommend models with the greatest accuracy and efficiency possible[18]. Since hybrid models use many ML techniques and data systems, they may be able to accurately predict health problems. Weedy classifiers worked better when they used bagging and boosting, and their ability to predict cardiovascular disease risk was rated well when they worked together. They made the hybrid model by using majority voting with the Bayes Net, NB, C4.5, MLP, and RF classifiers[19]. With 85.48 percent of the time, the model that was made is right. In addition to learning models, the UCI cardiovascular disease dataset has recently been used with ML methods like RF and SVM. Accuracy went up when a lot of classifiers were added to the voting-based model[20]. Based on the data, using the weak classifiers led to an increase of 2.1% in accuracy. We used ML classification methods to figure out how people with long-term conditions would do. They found that the Hoeffding classifier can predict coronary disease with an accuracy of 88.56 percent. Overall, they found that when the hybrid model was used with the desired features, it was 87.41% accurate. We used an SVM model and the Fisher score method to choose features based on the mean[21].

We used a lot of different classification methods and feature sets to make this one-of-a-kind prediction model. The proposed HRFLM used an ANN with a deep network and 13 clinical features as inputs. Data mining techniques like DT, SVM, NN, and KNN were also looked into. Researchers have found that it's helpful to use SVM to predict who will get sick. There was a new method called "vote," and a hybrid method that combines LR and NB was talked about. The HRFLM strategy worked out to be 88.7% effective[22]. We were able to make a model to predict death from cardiac failure that takes into account a wider range of risk factors by improving the random survival forest[23]. The IRSF used a split criterion and a stop criterion that were new to the field to tell the difference between survivors and people who didn't make it. Data mining has also been used to find out if someone has a cardiovascular disease[24]. Heart diseases are still diagnosed using Bayesian, DT classifiers, NN, association law, KNN, SVM, and ML algorithms. SVM was right 99.3% of the time. Several classifiers based on machine learning have been made to predict how long a patient will live[25]. Characteristics that were linked to the most important risk factors were rated, and the results were compared to traditional bio statistical testing. Researchers came to the conclusion that serum creatinine levels and ejection fraction are the two most important things to look at when trying to make accurate predictions[26]. The ML algorithm was used to make a model for finding CVD. In this study, we cleaned and looked at the data in four different ways. The DT and RF methods got an accuracy rate of 99.83%, while the SVM and KNN methods only got accuracy rates of 85.32% and 84.49%, respectively. Using the ensemble method, another study predicted CHF by looking at HRV and using deep neural networks to fill in knowledge gaps in unrelated areas. Overall, the method suggested was 99.85% right. In a recent publication[27], different types of data were used to make an intelligence framework. These were principal component analyses and RF-based MLA. The FAMD was applied to RF in order to value the relevant properties and predict illness. The suggested method is correct 93.44% of the time, sensitive 89.28% of the time, and specific 96.74% of the

time. In order to test their theory, the authors used a set of 303 cases that were made by adding to the Cleveland dataset. In tests, the suggested DT algorithm did 75.5% better than the baseline algorithm. Heart disease is often referred to as "cardiovascular disease"[28]. Several researchers are trying to make it easier to tell if someone has heart disease. Their research on heart disease covers a lot of ground. The author used data from the Hungarian and Statlog sets to classify CVD using the reduced error pruning tree (REP tree), R tree, M5P tree, logistic regression (LR), J48, naive bayes (NB), and JRIP. People use random forest (RF), decision tree (DT), and linear regression (LR). Support vector machine (SVM), CART, linear discriminant analysis (LDA), gradient boosting (XGB), and random forest (RF) are all used[29]. The goal of this study is to find a way to figure out how likely someone is to get heart disease. The results show that SVM does better than LR because it gets 96% accuracy while LR only gets 92% accuracy. The author says that the DT model always does better than the NB model and the SVM model. SVM has been shown to be 87% accurate, DT to be 90% accurate, and LR to be the most accurate at predicting when heart disease will happen, compared to DT, SVM, NB, and k-nearest neighbour (KNN). Table 1, represents the overall performance metric comparison of state-of-the-art methods.

The RF-based method is 97% accurate at predicting congenital heart disease, with a specificity of 88% and a sensitivity of 85%. They were able to find CVD with 94% accuracy, 95% specificity, and 93% sensitivity by using LR, MARS, EVF, and CART-ML. RF was used to predict drug targets in host-host and host–pathogen interactions related to CVD caused by microorganisms. Several ensembles and hybrid representations have been put forward to solve the problem of predicting heart disease. Based on the suggested method[30], CVD from the Mendeley Institute, the Cleveland datasets, and the IEEE Port are all processed with a high level of accuracy (96%, 88.24%, and 93%, respectively). The author put together the LR and RF algorithms to predict heart disease and got an accuracy of 88.7%. In this study, researchers want to find out more about how calcium in the coronary arteries and plaque in the carotid arteries are related. Both are linked to a higher risk of heart disease, but they may not be causing any symptoms yet. Machine learning and the internet of things are often used to predict and diagnose illnesses right now. The author was able to predict heart problems 94% of the time with the help of mobile devices and the deep learning method. The author employs machine learning classifiers and the Internet of Things to predict heart infections before they occur[31]. At the end of the day, we want to show that ML could be a good way to solve the problem at hand. We can use ML to look at cases related to illnesses and health problems by looking at hundreds of healthcare datasets. Researchers have worked on sophisticated computer perception for reliable healthcare to find out how machine vision practises help human needs, such as psychosocial health, specific movement, exposure-induced fatigue, frequently having to watch live actions, image analysis, deep learning, pattern classification, and how language understanding and computer animation work with robotics[32]. The authors noticed and wrote about how users learn about sharp interfaces and virtual reality tools, which leads to the development of complex restorative systems that can do human activities and recognise them. The work backs up the direct method of machine vision in the healthcare sector. This includes the technology behind intelligent wheelchairs, possible help for the visually impaired, and other object tracking solutions that have recently been used to monitor health and safety[33]. Scientists used support vector machines, generalised boosting machines, logistic regression, light boosting machines, and random forests to see how likely someone was to get cardiovascular disease. RF was the best way to predict who would get heart disease. It was right 88% of the time. Our method is put up against the current study. This is the first and only study to compare the accuracy of seven different ML classifiers for predicting cardiovascular illness. These methods include the most cutting-edge ones like learning vector quantization, RBF neural networks, and logistic regression. So, it is now possible to use a system that is both accurate and useful for predicting heart problems. Also, we suggest using the best machine learning classifier when making smart systems for predicting CHD[34]. The key features of cardiovascular illnesses include high morbidity, disability, and death, and the etiology of heart disease remains an unresolved worldwide issue. Therefore, accurate early prediction of anticipated outcomes in individuals affected by cardiac illness is necessary. In this work, we employed ML modelling to predict cardiac disease. This study focuses on predicting heart disease using ML classifiers. The authors first address the dataset problem, and subsequently enhance and standardize it for tokenization and lowercase conversion. The datasets were then utilized to train and test the classifiers, assessing their performance to achieve the highest level of accuracy. These algorithms must meet strict

Year	Author Name	Online Database	Classification Type	Performance Metric	Accuracy
2022	[20]	IoT based data	K-NN, DT, RF, MLP, NB, L-SVM	Accuracy, sensitivity, F1 score	96.12
2022	[21]	Di-ScRi database	Evimp functions, Multivariate adaptive regression	Accuracy, Specificity, Sensitivity, F1 score	91.2
2022	[22]	Hungarian-Statlog database	LR, NB, RF REP, M5P Tree, J48, JRIP	RMSE, MAE	89.7
2022	[23]	UCI repository	KNN, DT, LR, NB, SVM	Accuracy, Sensitivity, F1-Score, Specificity	93.23
2022	[24]	Congenital heart disease database of 3910 Singleton	RF-fetal echocardiography	RMSE, MAE	95.02
2022	[25]	Pathogen, Host feature	LR, KNN, SVM, RF	Accuracy, sensitivity, F1 score	94.08
2022	[26]	Heart Disease (Kaggle Repository)	KNN, RF, ANN, Ada, GBA	RMSE, MAE	90.91
2021	[27]	Heart Cleveland (UCI repository)	LR, DT, RF, SVM, HRFLM	Accuracy, Sensitivity, F1-Score, Specificity	96.22
2021	[28]	UCI Cleveland database	RF, DT, LR	Accuracy, sensitivity, F1 score	94.21
2021	[29]	UCI repository	SVM, NB, DT	Sensitivity, accuracy	94.11

Table 1. Literature review state-of-art method (metric comparison).

admission criteria, including modernity, representativeness, and high maturity. Previously, we employed Naive Bayes and Radial Basis Functions by examining the works of prior researchers. We investigated whether these approaches had been utilized on the UCI heart dataset by earlier researchers.

The proposed work contributions:

i. The authors commence by discussing datasets, which are subsequently standardized and enhanced. These datasets are then employed to train and test several classifiers to determine the one with the highest accuracy.
ii. Subsequently, the authors utilize the correlation matrix to classify the optimal values or features.
iii. The third step involves applying the ML classifiers to the pre-processed dataset, aiming for the highest achievable accuracy through parameter modifications.
iv. In the fourth and final step, the suggested classifiers are assessed for accuracy, precision (specificity), recall (sensitivity), and F-Measure.

Ultimately, the suggested classifiers outperform the state-of-the-art classifiers presented in Table 1 in terms of accuracy.

Proposed methodology

With the utilization of the heart dataset, we employed ML classifiers to predict the presence of coronary heart disease. The dataset was obtained from the UCI repository, and feature engineering was applied for data preprocessing before selecting the features. Subsequently, we divided it into training and test datasets, using around 70% of the total data for training and the remaining portion for testing. The training dataset is used to create a model that predicts heart disease, while the test dataset is utilized to evaluate the classifiers. Prior to transforming categorical variables into numerical values for classification, a thorough dataset analysis was conducted. The dataset was labelled as "normal" and "diseased" in Step 1. The "diseased" label indicates the presence of heart disease, while the "normal" label indicates the absence of heart disease. In Step 2, data cleaning was performed during the training phase. Data pre-processing involved handling missing values by calculating the mean due to the presence of partial and missing values. Step 3 involved data visualization using Exploratory Data Analysis (EDA) to examine relationships between various attributes. Notably, we identified that the correlation for FBS is relatively low. Moving to Step 4, ML classifiers were applied to the pre-processed dataset, and the classifiers' performance was evaluated using a variety of parameters. As previously mentioned, the dataset was split into test and training sets to respectively assess the classifiers and develop the model. The employed classifiers demonstrated varying levels of accuracy in detecting the presence of heart disease. Figure 1 illustrates the stages of our proposed working approach.

Dataset availability. We used the publicly available cardiovascular disease data sets from the UCI database. There are 503 cases in all, with multivariate features represented by 10 attributes, and a range of integer, category, and real values. The data set is described in Table 2. Database: https://archive.ics.uci.edu/ml/datasets/heart+disease.

Proposed model overview. Using the heart dataset and ML classifiers, we were able to make accurate predictions on the presence of coronary heart disease. The dataset was obtained from the UCI-repository, and material that was previously carried out was carried out prior to feature engineering being used to pick the features. We then split it up into two portions, one for training and one for testing, with the former containing typically 75% of the total data and the latter the remainder. The training dataset is used to make predictions about cardiovascular illness, while the test information is used to evaluate classifiers. Before transforming categorical variables to quantitative data for classification, we first analyse the dataset.

Phase 1: The dataset was annotated with "normal" and "abnormal" labels. Both the "healthy" and "sick" labels indicate that the respective individuals are free of any heart-related issues. Phase 2: There was some tidying up of the data that we did. Due to the partial and missing data in the dataset, we averaged the remaining values to complete the phase. Phase 3: We used exploratory data analysis to visualise the data and look for patterns in the relationships between variables. Our research showed that the association between FBS and anything else was quite modest. Phase 4: We next examined the performance of the ML classifiers on the pre-processed dataset using a variety of metrics. As was previously said, the dataset is often divided into testing and training sets, the former of which is used to assess the efficacy of the classifiers and the latter to educate the model. Classifiers used to make predictions about cardiac health have varying degrees of success. Figure 1 depicts the stages of our suggested working method.

Learning vector quantization: cardiovascular classification. Learning vector quantization is a network that is based on competition and uses supervised learning. We could say that it is a method of organizing patterns into groups, in which each transfer function is a group. Since it uses a learning algorithm, the system will be given a collection of learning patterns with recognised classifications and a preliminary allocation of the output variable. After the training is done, LVQ would then categorise an input vector by placing it in the same class as the output channel. The architecture of LVQ is shown in the following Fig. 2. As we can see, there are "n" units serving as input, and "m" units serving as output. The layers are completely attached to one another and have weights placed on them. The following respective parameters have been used for LVQ training operations for cardiovascular classification:

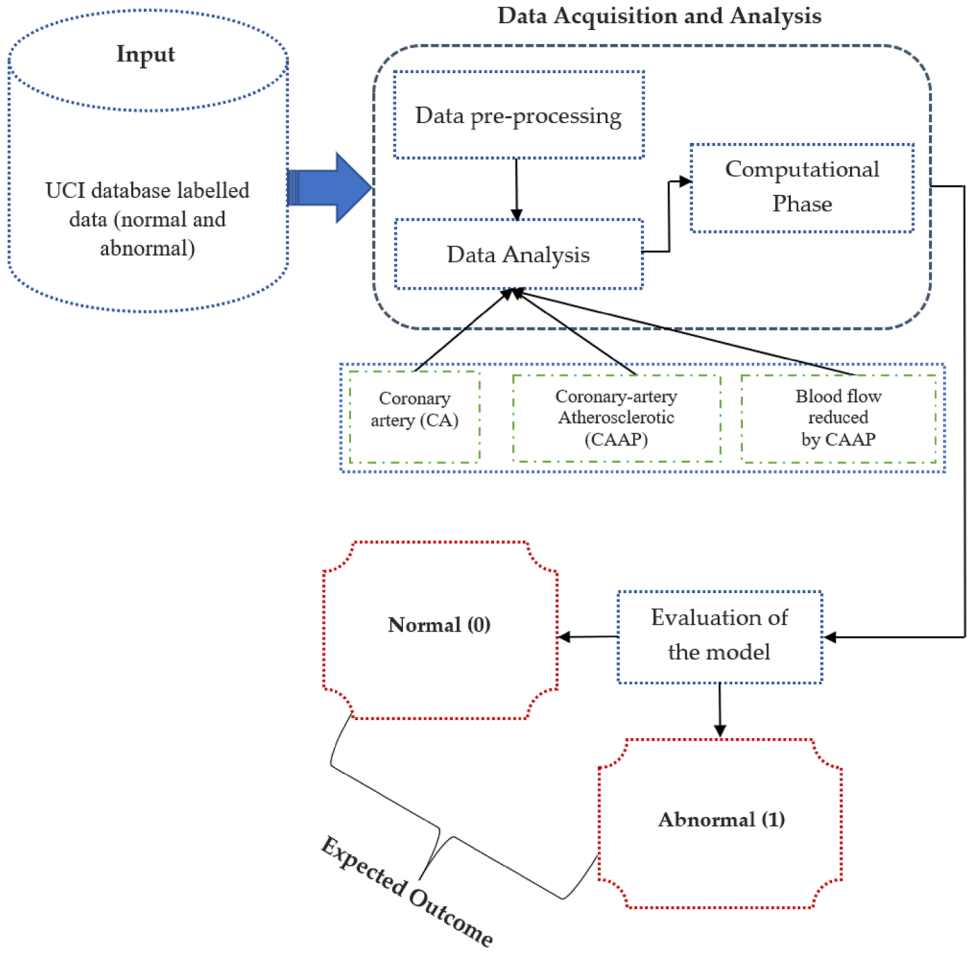

Figure 1. Proposed system operation overview.

S. no	Features	Attributes	Character
1	Gender	Sex	Category (male or female)
2	Age-group	Uses the integer value	Fundamental aspect to divide the patient history
3	Cardiac status	Non-anginal, typical angina and asymptomatic	Creates a severe chest pain
4	Blood-pressure monitoring	Floating and integer range values	Multiple organ failure may induce
5	Fat	Floating and integer range values	High-density, low-density lipoproteins, and triglycerides
6	Diabetic-average	Binary value to represent the true or false	Exceeded the recommended value blood-sugar
7	Electrocardiogram	Representation of different normal and abnormal hypertro-phy	Electrocardiogram evaluation
8	Obtained pulse rate peak value	Floating and integer range values	Exceeded the recommended heart rate value
9	Angina pectoris	Binary value to represent the true or false	It can be reduced by doing exercise
10	Obsolete peak	Floating and integer range values	Relaxation is required to compete the stress

Table 2. Dataset attributes and characters.

$$a \rightarrow \textit{is a suggested trainig vector } (a_1 \ldots a_i \ldots a_n)$$

$$T_v \rightarrow \textit{training vectro class for } 'x'$$

$$W_j \rightarrow \textit{vector weight for outcome uinit of } j^{th}$$

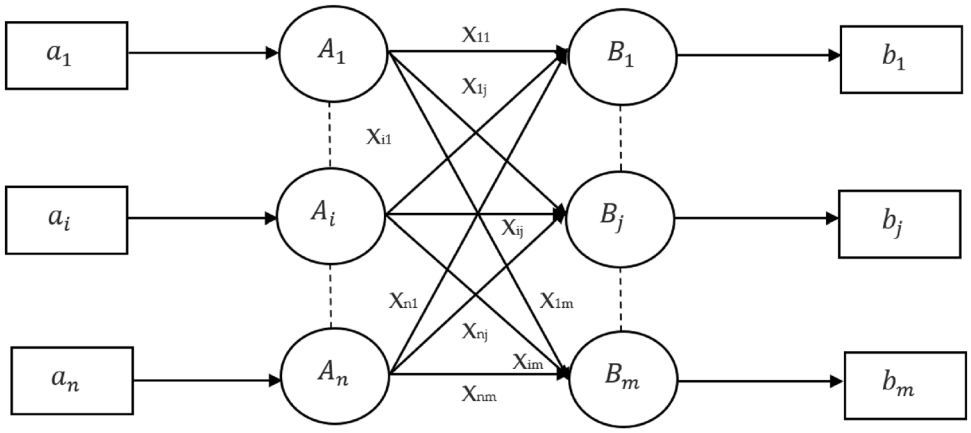

Figure 2. Learning Vector Quantization architecture for Cardiovascular Classification.

$$D_j \rightarrow j^{th} \textit{class associated outcome unit}$$

Learning vector quantization—cardiovascular classification

Step 1: Start.

Step 2: Reference vector initialization based on the training vectors and denotes 'm' is the cluster numbers and it can be used as a weight vector. Rest of the vectors will be assigned for training mode.

Step 3: Randomly assigning the initial classification and its corresponding weights.

Step 4: Initializing K-means clustering technique.

Step 5: The reference vector β is assigned.

Step 6: Computing the square of Euclidean distance for, i and j (1-to-m and 1-to-n) respectively.

$$ED(j) = \sum_{i=1}^{n}\sum_{j=1}^{m}(a_i - X_{i,j})^2 \tag{1}$$

Step 7: To compute and achieve the raising unit J where ED is locally minimum.

Step 8: Compute the initial weight of the raising unit using the relative conditions,

$$If\, S = R_j\, then\, X_j(new) = X_j(old) + \beta[a - X_j(old)] \tag{2}$$

$$If\, S \neq R_j\, then\, X_j(new) = X_j(old) - \beta[a - X_j(old)] \tag{3}$$

Step 9: Lessen the β learning rate.

Step 10: Initiate the stopping condition of testing.

Step 11: Stop.

Clear coronary arteries, coronary arteries with atheromatous lesions, and coronary arteries with reduced blood flow due to blockage are all shown by the coronary artery contour. The degree and direction of a linear connection between two quantitative variables may be described by examining their correlation. Table 3 shows

S. no	Features	Range
1	Gender	0.31
2	Age-group	0.26
3	Cardiac status	0.41
4	Blood-pressure monitoring	0.038
5	Fat1	0.082
6	Diabetic-average	0.39
7	Electrocardiogram	0.16
8	Obtained pulse rate peak value	0.41
9	Angina pectoris	0.38
10	Obsolete peak	0.31

Table 3. Correlation-matrix value.

the relationships between the various columns. Most columns have some correlation with the "number" variable, but "BS-F" have very little.

$$\widehat{x}_{a_i} = \alpha_{a_0} + \alpha_{a_1 y_1} + \alpha_{a_2 y_2} + \alpha_{a_3 y_3} + \ldots \alpha_{a_n y_n} \tag{4}$$

From the Eq. (1) Correlation coefficients α between one explanatory variable (y) and another (x) are represented by a string of characters in this formula (x). The value of α1 indicates the strength of the relationship between variable (y) and independent (x) variables, and so on. Figure 3 depicts a heat map embedded inside a correlation matrix. A heatmap is a visual representation of the relationship between independent characteristics and dependent values. In addition, it is clear which characteristics have the strongest link to the supplementary characteristic's variable. The end product is shown in Fig. 2. To better understand the data, we will now plot the characteristic of the cardiovascular disease dataset against the number. Statistic graphics and other forms of data visualisation are common tools in exploratory data analysis, which is used to examine datasets in order to identify and describe their most salient features.

$$s = \frac{\sum \left(ay_i - \overline{ay}\right)\left(ax_i - \overline{ax}\right)}{\sqrt{\sum \left(ay_i - \overline{ay}\right)^2 \left(ax_i - \overline{ax}\right)^2}} \tag{5}$$

Equation (5), the overall correlation between two variables in a sample population is given by this equation. This would be the connection between the independent variable and the dependent variable in basic linear regression. Table 4, illustrates the various age-group cardiovascular analysis.

Out of a maximum of 503 cases of illness, we determined that 305 individuals had some kind of heart disease issue. Malignant is represented by 1, benign by 0, and 198 of the total patients are considered healthy. Given these

Gender	1	0.054	0.08	0.018	0.024	0.0071	-0.28	0.029	0.089	0.092	0.081
Age-group	0.054	1	0.15	0.031	0.0065	0.066	-0.1	0.12	0.22	0.31	0.078
Cardiac status	0.08	0.15	1	0.081	0.0565	0.116	-0.05	0.17	0.27	0.36	0.128
BPM	0.018	0.031	0.081	1	0.0765	0.136	-0.03	0.19	0.29	0.38	0.148
Fat1	0.024	0.0065	0.0565	0.0765	1	0.206	-0.04	0.26	0.36	0.45	0.218
Diabetic average	0.0071	0.066	0.116	0.136	0.206	1	-0.05	0.0061	0.0056	0.0047	0.081
Electrocardiogram	-0.28	-0.1	-0.05	-0.03	-0.04	-0.05	1	-0.04	-0.13	-0.21	-0.33
Obtained pulse rate peak value	0.029	0.12	0.17	0.19	0.26	0.0061	0.0078	1	0.12	0.14	0.21
Angina pectoris	0.089	0.22	0.27	0.29	0.36	0.0056	0.0089	0.12	1	0.11	0.17
Obsolete peak	0.092	0.31	0.36	0.38	0.45	0.0047	0.012	0.14	0.11	1	0.31
Fat2	0.081	0.078	0.128	0.148	0.218	0.081	0.15	0.21	0.17	0.31	1

Figure 3. Heat map-correlation matrix.

Age-group	Exemplify	Age-group	Exemplify
30	97	48	55
31	96	49	58
32	52	50	54
33	96	51	78
34	78	52	71
35	57	53	78
36	88	54	61
37	90	55	37
38	69	56	50
39	60	57	47
40	78	58	43
41	37	59	36
42	56	60	33
43	66	61	16
44	60	62	38
45	45	63	36
46	56	64	61
47	58	65	48

Table 4. Different age-group cardiovascular analysis.

results, we may infer that 53.36 percent of patients have cardiac issues and that 46.64% do not. We also looked at other characteristics in the dataset, including gender, age group, cardiac status, blood pressure monitoring, fat, and smoking. Diabetic-average, the electrocardiogram obtained the pulse rate peak value for angina pectoris and the obsolete peak. As can be seen in Fig. 4, the sex property accepts two values: 0 for women and 1 for men. According to the results, women have a higher risk of developing cardiovascular disease than men do. Figure 5 shows the age distribution of the dataset, demonstrating that the risk of heart disease is independent of age group. Both age and the desired percentage are shown by the x- and y-axes, respectively. Chest discomfort is common among those who suffer from heart disease. Chest pain can be experienced by cardiovascular patients. However, the chest pain can be divided into different categories, such as non-anginal, asymptomatic, non-typical angina, and typical angina. Figure 6 depicts the different categories of chest pain that may occur. According to Fig. 6, patients with non-typical angina may have the highest risk of cardiac arrest. Blood sugar during fasting (BS-F) cannot have a significant impact on the development of heart disease.

We performed an analysis on the information in which the value 1 (true) is assigned to the case in which the patient's fasting blood sugar level is more than 120 mg/dL, indicating that they are at risk for the condition; otherwise, the number 0 (false) is assigned to the case, as depicted in Fig. 7. According to the findings, there is nothing particularly remarkable about this method for predicting the existence of heart disease. Electrocardiogram readings are 0, 1, and 2. The results demonstrate that those whose ECG values are "1" or "0" have an increased risk of developing heart disease in comparison to people whose ECG values are "20," as seen in Fig. 8. Figure 9 depicts the ECG analysis of cardiovascular possibility. Table 5, represents the various categories of cardiovascular occurrence. Table 6, illustrates the chances of cardiovascular found from ECG scrutiny.

As shown in Fig. 10, those who suffer from angina have a much lower risk of developing cardiac issues. If the score of workout angina is 1, it indicates that the patient does in fact have a heart issue; on the other hand, if it is 0, it indicates that the patient does not have a heart problem and is thus less likely to develop heart problems.

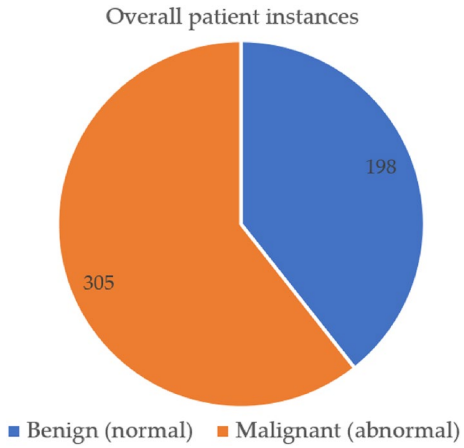

Figure 4. Overall heart-disease patient instances.

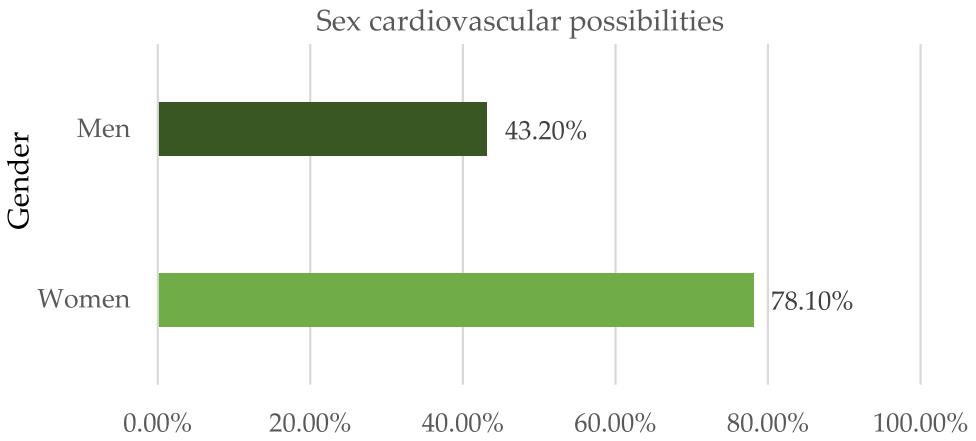

Figure 5. Sex categorization based cardiovascular possibilities.

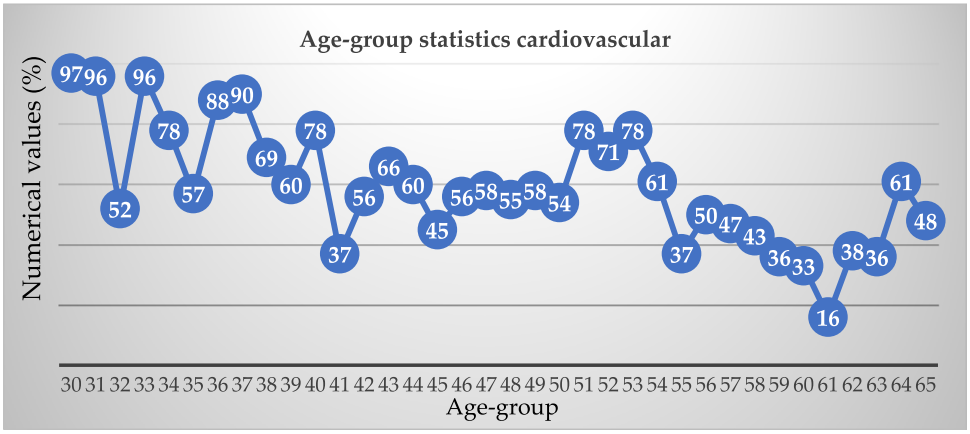

Figure 6. Age-group statistics cardiovascular.

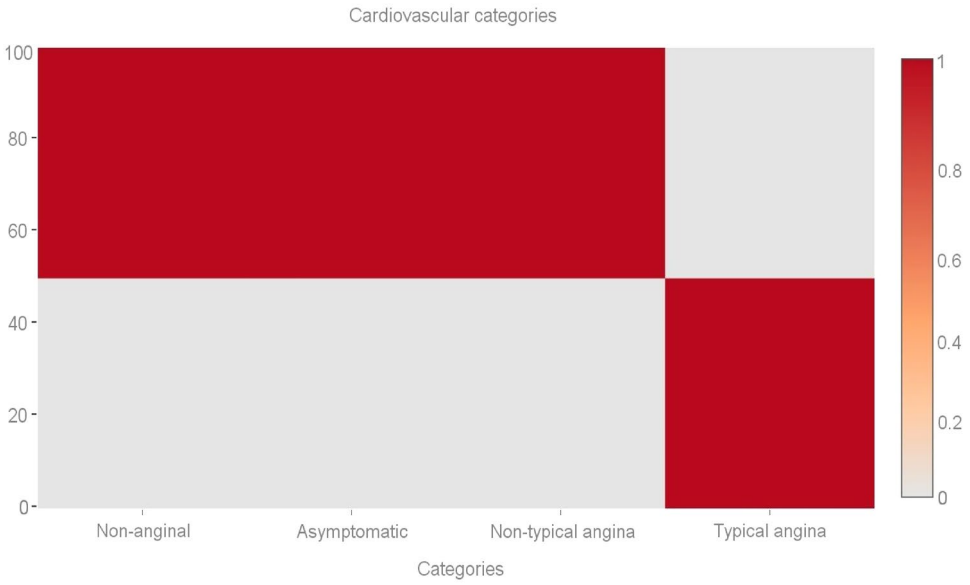

Figure 7. Different cardiovascular types.

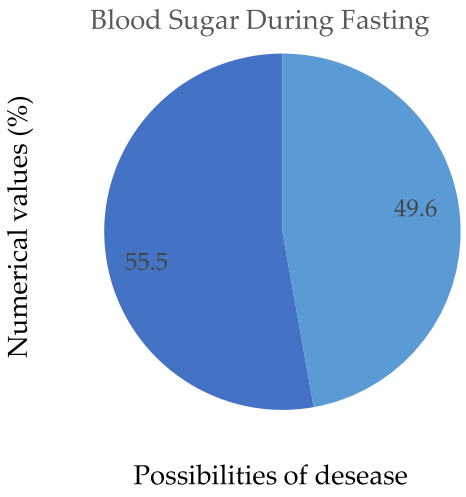

Figure 8. Possibility of disease during fasting.

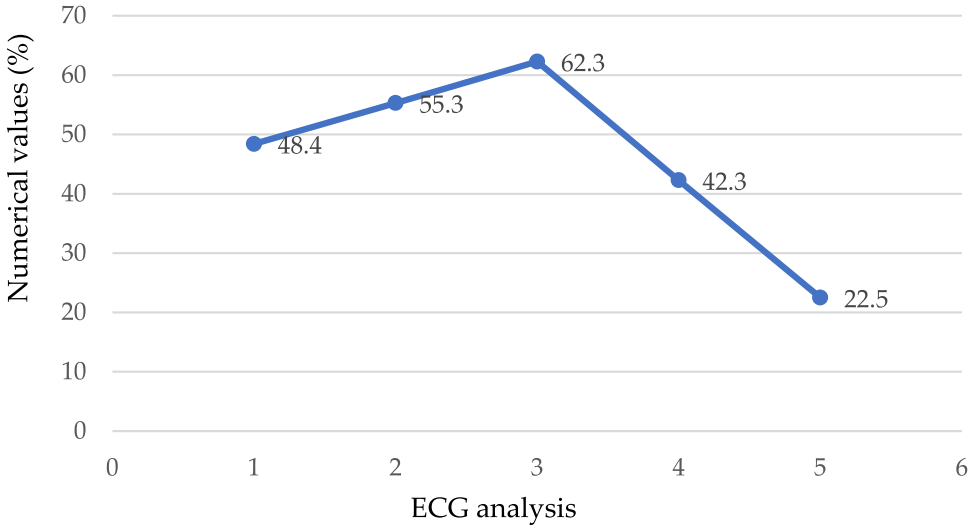

Figure 9. Analysis of ECG of cardiovascular possibility.

Category	Occurrence
Non-anginal	79
Asymptomatic	75
Non-typical angina	83
Typical-angina	32

Table 5. Various categories of cardiovascular.

Type	ECG analysis
A0	48.4
A(0,1)	55.3
A1	62.3
A(1,2)	42.3
A2	22.5

Table 6. Chances of cardiovascular from ECG scrutiny.

Experimental results and discussion

In this chapter, the results of ML classifiers on various evaluation requirements, such as accuracy, recall, and F-measure, are addressed. Examples of these evaluation constraints include: In addition to this, the performance of machine learning classification models is assessed using the dataset, which includes information on heart disease. k-NN did not do very well, although RBF, NB, and LVQ fared better than the other classifiers when compared to their overall performance. As can be seen in Table 7, the most important assessment criteria that were taken into consideration in this study to evaluate the performance of the ML classifier are the sensitivity, accuracy, specificity, recall, precision, and F-measure ratings. As a consequence of this, the specificity and sensitivity of the targeted class are calculated in order to evaluate the accuracy with which the given method is projected to perform. The "TP" (true positive), "TN" (true negative), "FN" (false negative), and "FP" (false positive) rates are used to compute the accuracy, precision, recall, and F measure in ML. These measures are determined by the quality of the data. Each correct positive and correct negative prediction is further subdivided into correct positive and correct negative forecasts. Every model correctly predicted the TP, TN, FP, and FN outcomes. The letters TP stand for diseased, which means infected. FN is an illness that is not believed to be related to cardiovascular disease. The FP illness is one that has been predicted but has never been seen in humans. In the actual world, TN does not exist as a disease, and this is not anticipated to change in the foreseeable future. The performance of ML approaches in terms of accuracy is listed in Table 7. By associating the performances of these classifiers, we observed that radial basis functions, naive bayes, and learning vector quantization, as well as their relatedness to other ML classifiers, led these models to achieve almost 90.06%, 94.16%, and 98.07% accuracy, respectively, as shown in Fig. 11.

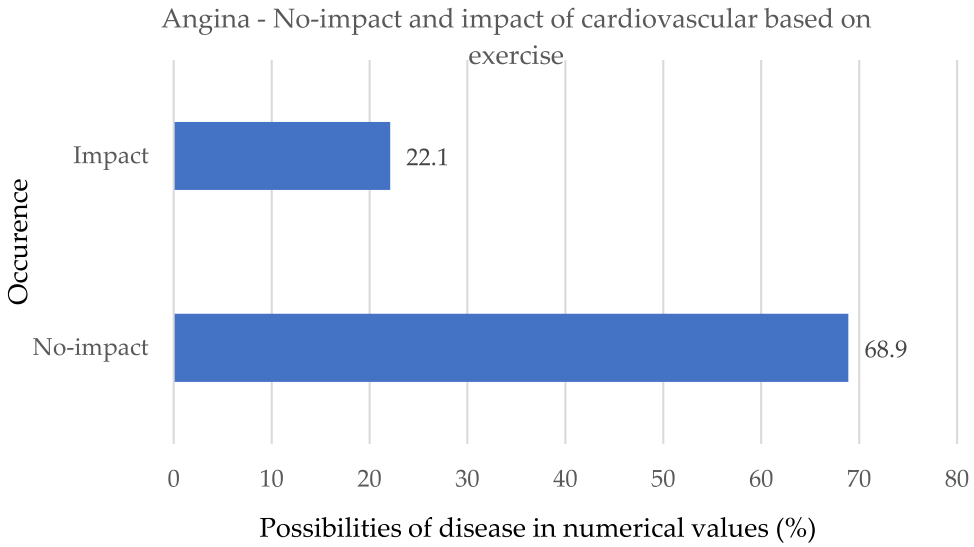

Figure 10. No-impact and impact of the occurrence based on exercise.

Classification techniques	Performance metric parameters					
	Precision	Accuracy	Sensitivity	Specificity	Recall	F-measure
Random forest	88.07	88.78	87.91	87.1	85.31	87.89
Proposed learning vector quantization	**98.07**	**98.78**	**97.91**	**97.1**	**95.31**	**97.89**

Table 7. Parameter metric comparison of RF and LVQ. Significant values are in bold.

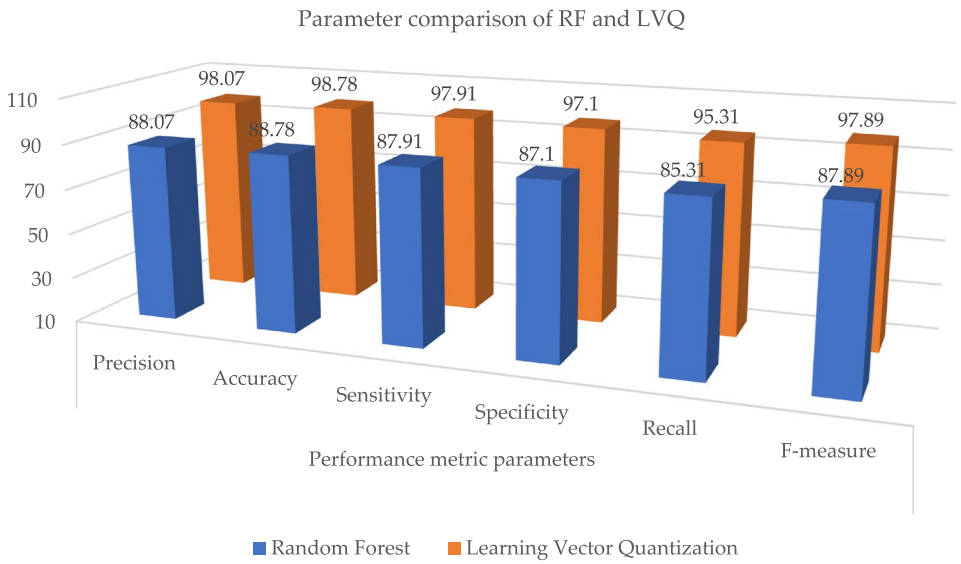

Figure 11. Graphical illustration of parameter metric comparison of RF and LVQ.

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \tag{6}$$

$$Sensitivity = \frac{True\ positive}{True\ positive + False\ negative} \tag{7}$$

$$Accuracy = \frac{True\ positive + True\ negative}{True\ positive + False\ positive + True\ negative + False\ negative} \quad (8)$$

$$Specificity = \frac{True\ negative}{True\ negative + False\ positive} \quad (9)$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \quad (10)$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (11)$$

Table 7 and Fig. 11, illustrates the paramenter metric (precision, accuracy, sensitivity, specificity, recall and f-measure) comparison outcome of random forest (RF) and learning vector quantization (LVQ). The outcome shows that the LVQ obtained better outcome accuracy of 98.78%. Table 8, the paramenter metric (precision, accuracy, sensitivity, specificity, recall and f-measure) comparison outcome of decision tree (DT) and learning vector quantization (LVQ).

The outcome shows that the LVQ obtained better outcome accuracy of 98.78%, and the graphical illustration is shown in Fig. 10. Tables 9 and 10, illustrates that the proposed system outcome is better than the XGBoost and KNN methods, and graphical view representation shown in Figs. 11 and 12 respectively. Table 11, depicts the paramenter metric (precision, accuracy, sensitivity, specificity, recall and f-measure) comparison outcome of support vector machine (DT) and learning vector quantization (LVQ). Then, Table 12, shows the performance metric parameter comparison of various classifiers such as, DT, KNN, RF, SVM and XGBoost. From the Table 12, the proposed system achieved (Pre 98.07%, Acc 98.78%, Se 97.91%, Sp 97.1%, Recall 95.31% and Fm 97.89%) better outcome in all parameters than the other conventional techniques. Figure 13, depicts the graphical illustration of parameter metric comparison of XGBoost and LVQ. Figure 14, represents the graphical illustration of parameter metric comparison of KNN and LVQ.

When compared to the other classification techniques, the two corresponding techniques, radial basis function and nave Bayes, produced the best results. So, its respective parameters are taken, and it is compared with the proposed system shown in Table 13; the resultant shows that the proposed system parameter outcomes are better than those two outcomes, as illustrated in Fig. 15. Figure 16, depicts the proposed method performance metric parameter comparison of classification accuracy. Figure 17, illustrates the performance metric comparison of RBF, NB and LVQ classifiers. The receiver operating characteristics of the learning vector quantization are illustrated in Fig. 18.

Classification techniques	Performance metric parameters					
	Precision	Accuracy	Sensitivity	Specificity	Recall	F-measure
Decision tree	89.07	89.78	88.91	88.1	86.31	88.89
Proposed learning vector quantization	**98.07**	**98.78**	**97.91**	**97.1**	**95.31**	**97.89**

Table 8. Parameter metric comparison of DT and LVQ. Significant values are in bold.

Classification techniques	Performance metric parameters					
	Precision	Accuracy	Sensitivity	Specificity	Recall	F-measure
XGBoost	87.07	87.78	86.91	86.1	84.31	86.89
Proposed Learning vector quantization	**98.07**	**98.78**	**97.91**	**97.1**	**95.31**	**97.89**

Table 9. Parameter metric comparison of XGBoost and LVQ. Significant values are in bold.

Classification techniques	Performance metric parameters					
	Precision	Accuracy	Sensitivity	Specificity	Recall	F-measure
K-Nearest Neighbour	79.07	79.78	78.91	78.1	76.31	78.89
Proposed learning vector quantization	**98.07**	**98.78**	**97.91**	**97.1**	**95.31**	**97.89**

Table 10. Parameter metric comparison of KNN and LVQ. Significant values are in bold.

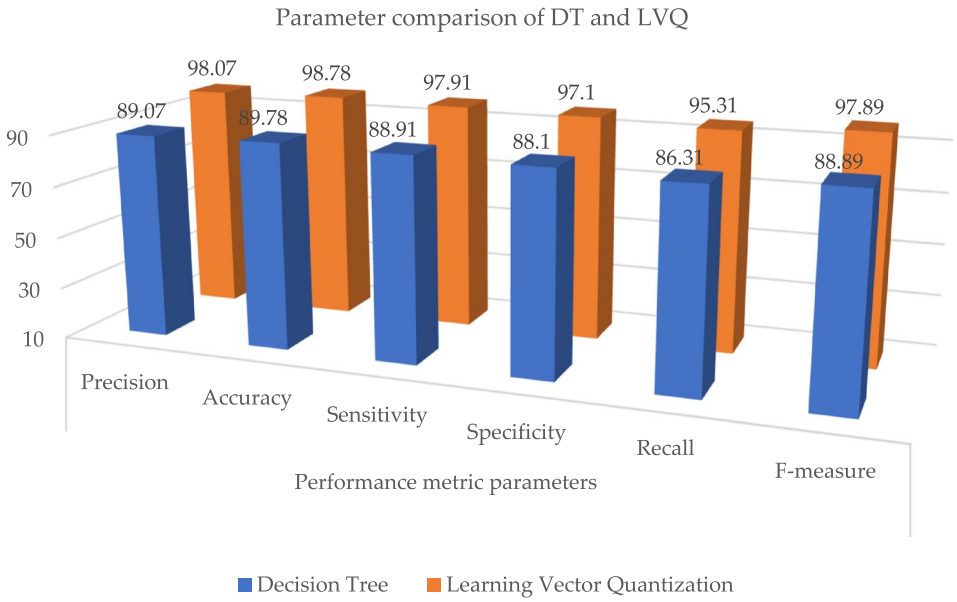

Figure 12. Graphical illustration of parameter metric comparison of DT and LVQ.

Classification techniques	Performance metric parameters					
	Precision	Accuracy	Sensitivity	Specificity	Recall	F-measure
Support vector machine	86.07	86.78	85.91	85.1	83.31	85.89
Proposed learning vector quantization	**98.07**	**98.78**	**97.91**	**97.1**	**95.31**	**97.89**

Table 11. Parameter metric comparison of SVM and LVQ. Significant values are in bold.

Classification techniques	Performance metric parameters					
	Precision	Accuracy	Sensitivity	Specificity	Recall	F-measure
Random forest	88.07	88.78	87.91	87.1	85.31	87.89
Decision tree	89.07	89.78	88.91	88.1	86.31	88.89
Support vector machine	86.07	86.78	85.91	85.1	83.31	85.89
XGBoost	87.07	87.78	86.91	86.1	84.31	86.89
Radial basis functions	90.07	90.78	89.91	89.1	87.31	89.89
K-nearest neighbour	79.07	79.78	78.91	78.1	76.31	78.89
Proposed learning vector quantization	**98.07**	**98.78**	**97.91**	**97.1**	**95.31**	**97.89**
Naive Bayes	94.07	94.78	93.91	93.1	91.31	93.89

Table 12. Performance metric comparison of various classifiers. Significant values are in bold.

Conclusion

In this study, machine learning classifiers are utilised to determine whether or not a patient has heart problems. The dataset was taken from the repository at UCI. Following data collection, they will go through cleaning and pre-processing steps. Following this step, machine learning models are used for predictive analysis. We investigated the potential of these eight applied machine learning methods for making accurate predictions about cardiac disease. The inclusion criteria for these algorithms are that they be mature, representative, and at the state of the art in their respective fields. We have previously used the Naive Bayes and RBF neural networks, but other scholars have not used them on the UCI cardiovascular disease dataset. As a result, we have achieved a higher level of accuracy than they have, as shown in the table titled "state of the art," which compares our results to those of other researchers. The final findings demonstrate that when the learning machine classifiers were put to use, the Naive Bayes and RBF neural networks achieved an accuracy of 94.78% when attempting to forecast the presence of coronary cardiovascular disease. However, the Learning Vector Quantization method achieved

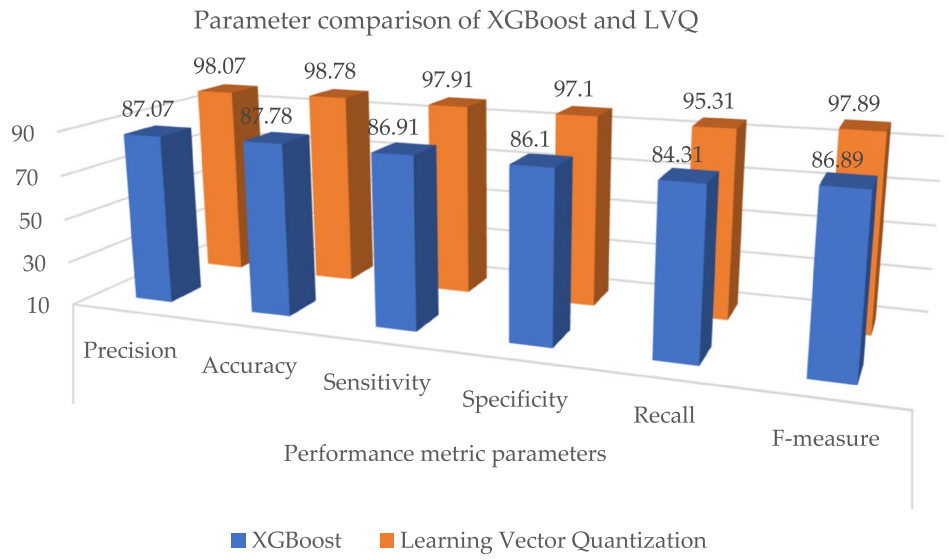

Figure 13. Graphical illustration of parameter metric comparison of XGBoost and LVQ.

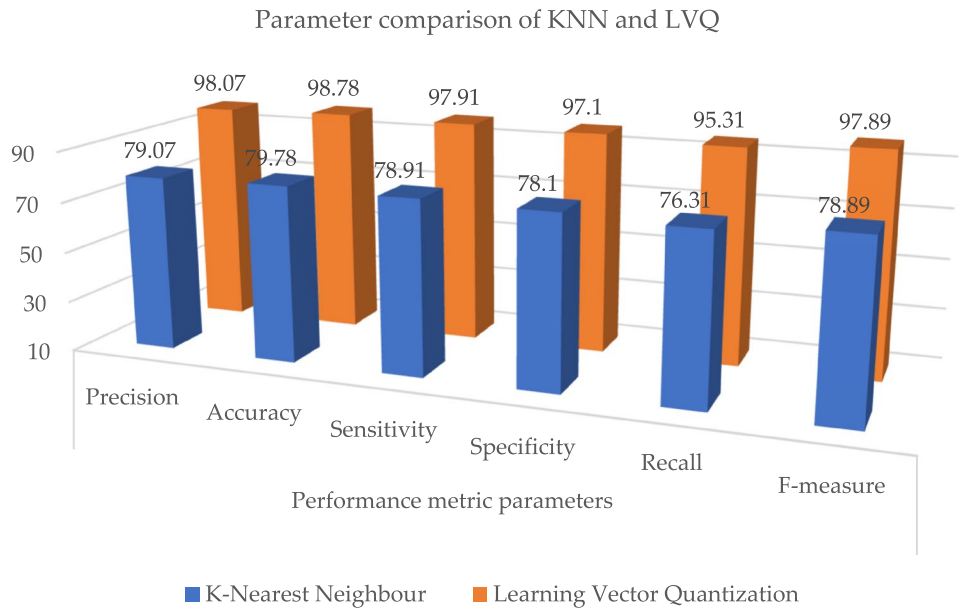

Figure 14. Graphical illustration of parameter metric comparison of KNN and LVQ.

Classification techniques	Performance metric parameters					
	Precision	Accuracy	Sensitivity	Specificity	Recall	F-measure
Radial basis functions	90.07	90.78	89.91	89.1	87.31	89.89
Naive Bayes	94.07	94.78	93.91	93.1	91.31	93.89
Proposed learning vector quantization	**98.07**	**98.78**	**97.91**	**97.1**	**95.31**	**97.89**

Table 13. Parameters comparison for three respective classifiers. Significant values are in bold.

the highest categorization accuracy of 98.78%, with a specificity of 97.1% and sensitivity of 97.91%, a precision of 98.07% and 95.31%, and 97.89% F1score and F-measure values, respectively.

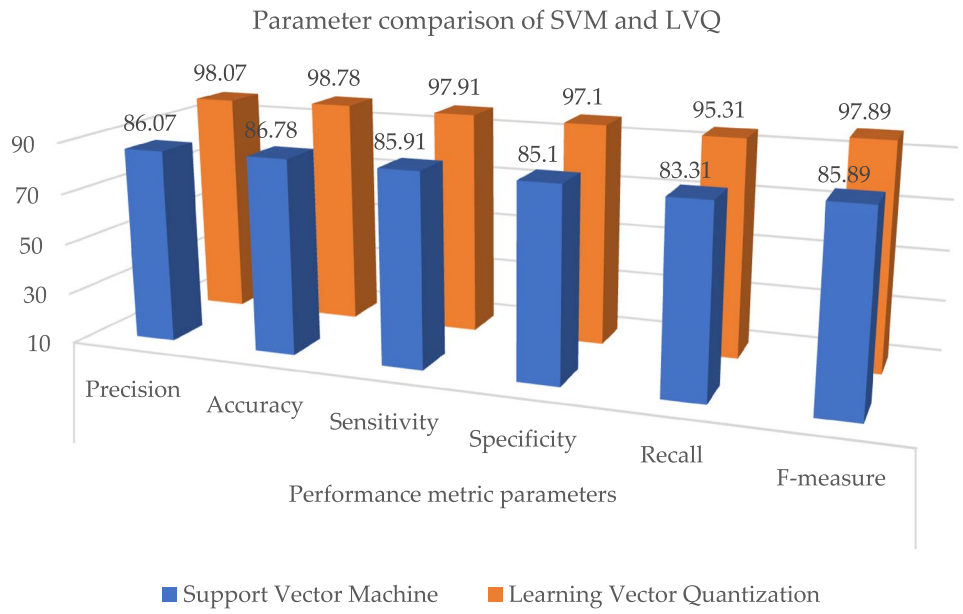

Figure 15. Graphical illustration of parameter metric comparison of SVM and LVQ.

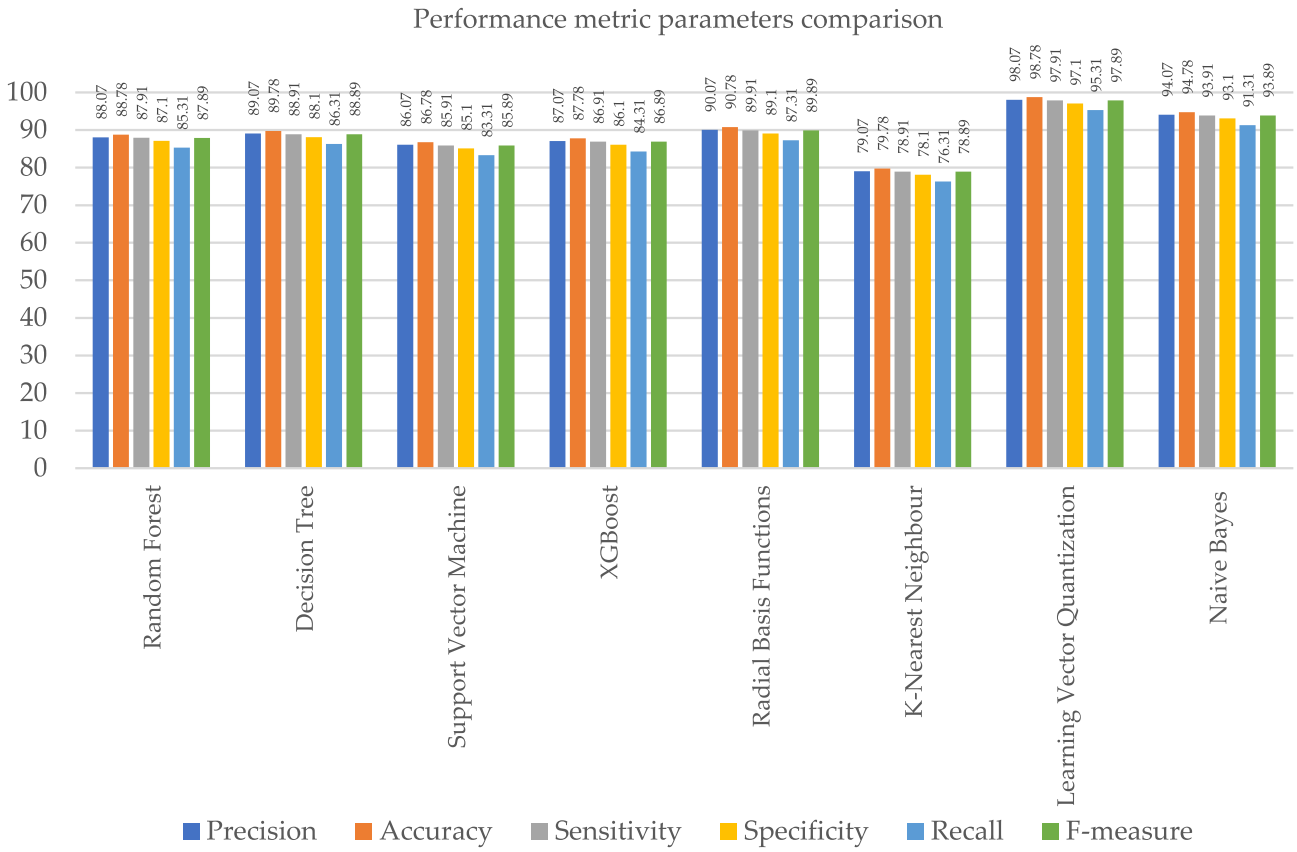

Figure 16. Classification accuracy—performance metric parameter comparison.

Future work

In the future, our research aims to further enhance the reliability of our conclusions by incorporating additional datasets. We will explore the use of metaheuristic techniques and nature-inspired algorithms to optimize the parameters of machine learning classifiers and deep learning methods. This optimization process will enable us to more effectively evaluate the presence of heart disease across various heart disease-related datasets. Additionally, we will focus on improving the accuracy of existing algorithms to enhance their performance in detecting

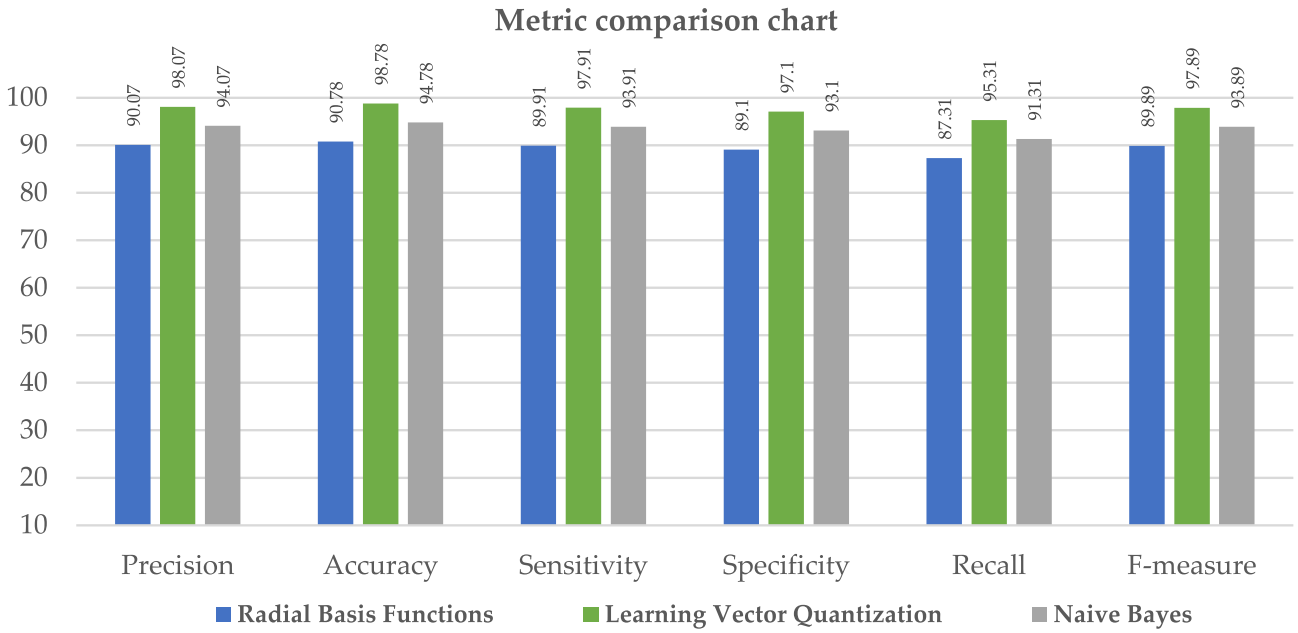

Figure 17. Performance metric comparison of RBF, NB and LVQ classifiers.

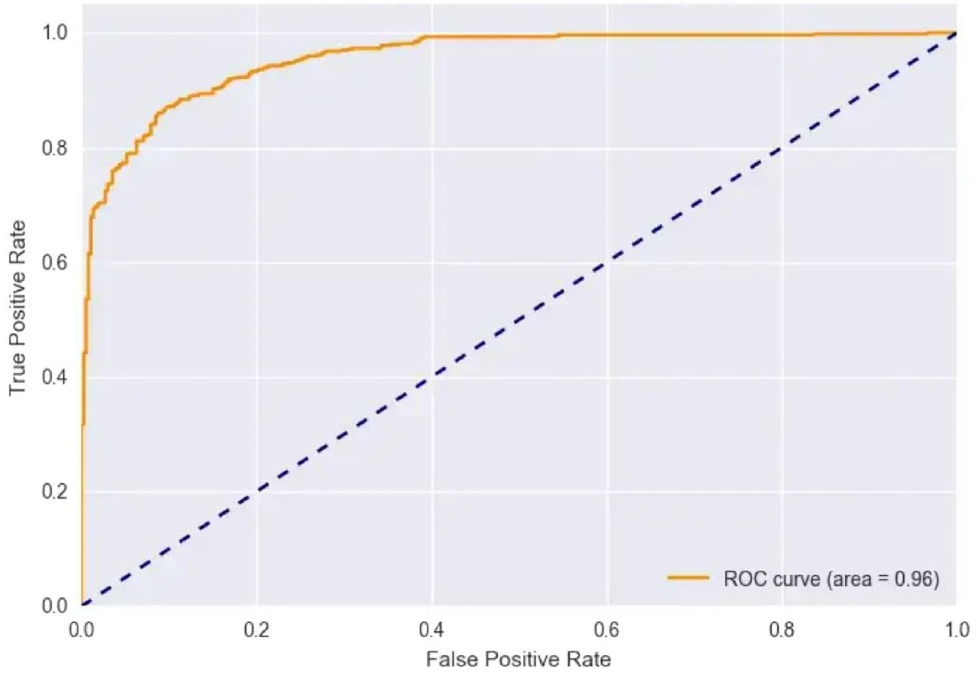

Figure 18. Learning vector quantization ROC.

heart disease. By leveraging these advancements, we aim to provide more robust and accurate methods for the diagnosis and evaluation of heart disease.

Data availability

Used publicly available database, and no human data/sample used in the study" https://archive.ics.uci.edu/ml/datasets/heart+disease.

Received: 25 December 2022; Accepted: 16 August 2023
Published online: 21 August 2023

References

1. Gour, S., Panwar, P., Dwivedi, D. & Mali, C. A machine learning approach for heart attack prediction. *Intell. Sustain. Syst.* **2555**(1), 741–747 (2022).
2. Juhola, M. *et al.* Data analytics for cardiac diseases. *Comput. Biol. Med.* **142**(1), 1–9 (2022).
3. Alom, Z. *et al.* Early-stage detection of heart failure using machine learning techniques. *Proc. Int. Conf. Big Data IoT Mach. Learn.* **95**, 75–88 (2021).
4. Sharma, S. & Parmar, M. Heart-diseases prediction using deep learning neural network model. *Int. J. Innov. Technol. Explor. Eng.* **9**(3), 2244–2248 (2020).
5. Ravindhar, N. & Hariharan Ragavendran, S. Intelligent diagnosis of cardiac disease prediction using machine learning. *Int. J. Innov. Technol. Explor. Eng.* **9**(11), 1417–1421 (2019).
6. Arunpradeep, N. & Niranjana, G. Different machine learning models based heart disease prediction. *Int. J. Recent Technol. Eng.* **8**(6), 544–548 (2020).
7. Ravindhar, N., Anand, H. & Ragavendran, G. Intelligent diagnosis of cardiac disease prediction using machine learning. *Int. J. Innov. Technol. Explor. Eng.* **8**(11), 1417–1421 (2019).
8. Subulakshmi, G. Decision support in heart disease prediction system using Naive Bayes. *Indian J. Comput. Sci. Eng.* **2**(2), 170–176 (2011).
9. Sai Krishna Reddy, V., Meghana, P., Subba Reddy, N. V. & Ashwath Rao, B. Prediction on cardiovascular disease using decision tree and naïve bayes classifiers. *J. Phys.* **2161**, 1–8 (2022).
10. Patel, T. S., Patel, D. P., Sanyal, M. & Shrivastav, P. S. Prediction of heart disease and survivability using support vector machine and Naive Bayes algorithm. *bioRxiv* https://doi.org/10.1101/2023.06.09.543776 (2023).
11. Kelwade, P. & Salankar. 2016. Radial basis function neural network for prediction of cardiac arrhythmias based on heart rate time series. *IEEE First International Conference on Control, Measurement and Instrumentation* 454–458. https://doi.org/10.1109/CMI.2016.7413789 (2016).
12. Kumar, S. Heart disease detection using radial basis function classifier. *ICTACT J. Data Sci. Mach. Learn.* **1**(4), 105–108 (2020).
13. Jothikumar, R., Sivakumar, N. & Ramesh, P. S. Heart disease prediction system using ANN, RBF and CBR. *Int. J. Pure Appl. Math.* **117**(21), 199–217 (2017).
14. Saravanan, S. & Thirumurugan, P. Performance analysis of glioma brain tumor segmentation using ridgelet transform and CANFES methodology. *J. Med. Imaging Health Inform.* **10**(11), 2642–2648 (2020).
15. Latha, C. & Jeeva, S. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Inform. Med. Unlock* **16**(1), 1–9 (2019).
16. Abdeldjouad, F., Brahami, M. & Matta, N. A hybrid approach for heart disease diagnosis and prediction using machine learning techniques. In *International Conference on Smart Homes and Health Telematics* 299–306 (Springer, 2020).
17. Tarawneh, M. & Embarak, O. Hybrid approach for heart disease prediction using data mining techniques. *Acta Sci. Nutr. Health* **3**(7), 147–151 (2019).
18. Javid, I., Alsaedi, A. & Ghazali, R. Enhanced accuracy of heart disease prediction using machine learning and recurrent neural networks ensemble majority voting method. *Int. J. Adv. Comput. Sci. Appl.* **11**(3), 540–551 (2020).
19. Kumar, N. & Sikamani, K. Prediction of chronic and infectious diseases using machine learning classifiers: A systematic approach. *Int. J. Intell. Eng. Syst.* **13**(4), 11–20 (2020).
20. Saqlain, S. *et al.* Fisher score and Matthew's correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines. *Knowl. Inf. Syst.* **58**, 139–167 (2019).
21. Mohan, S., Thirumalai, C. & Srivastava, G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access* **7**, 81542–81554 (2019).
22. Miao, F., Cai, Y., Zhang, Y. & Li, Y. Predictive modeling of hospital mortality for patients with heart failure by using an improved random survival forest. *IEEE Access* **6**, 7244–7253 (2018).
23. Chicco, D. & Jurman, G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Med. Inform. Decis. Mak.* **20**(16), 1–16 (2020).
24. Ahmad, E., Tiwari, A. & Kumar, A. Cardiovascular diseases (CVDs) detection using machine learning algorithms. *Int. J. Res. Appl. Sci. Eng. Technol.* **8**(6), 2341–2346 (2020).
25. Wang, L., Zhou, W., Chang, Q., Chen, J. & Zhou, X. Deep ensemble detection of congestive heart failure using short-term RR intervals. *IEEE Access.* **7**, 69559–69574 (2019).
26. Gupta, A., Kumar, R., Arora, H. & Raman, B. MIFH: A machine intelligence framework for heart disease diagnosis. *IEEE Access* **8**, 14659–14674 (2019).
27. Rashmi, G. & Kumar, U. Machine learning methods for heart disease prediction. *SN Comput. Sci.* **8**, 220–223 (2019).
28. Nadakinamani, R. *et al.* Clinical data analysis for prediction of cardiovascular disease using machine learning techniques. *Comput. Intell. Neurosci.* **2022**, 1–13 (2022).
29. Hossen, M. *et al.* Supervised machine learning-based cardiovascular disease analysis and prediction. *Math. Probl. Eng.* **2021**, 1–10 (2021).
30. Saboor, A. *et al.* A Method for improving prediction of human heart disease using machine learning algorithms. *Mobile Inf. Syst.* **2022**, 1–11 (2022).
31. Arumugam, K. *et al.* Multiple disease prediction using Machine learning algorithms. *Mater. Today* **2021**, 1–10 (2021).
32. Gupta, C., Saha, A., Reddy, N. & Acharya, U. Cardiac Disease Prediction using Supervised Machine Learning Techniques. *J. Phys: Conf. Ser.* **2161**, 1–12 (2022).
33. Truong, V. *et al.* Application of machine learning in screening for congenital heart diseases using fetal echocardiography. *Int. J. Cardiovasc. Imaging* **38**, 1007–1015 (2022).
34. Abdalrada, A., Abawajy, J., Al-Quraishi, T. & Islam, S. Machine learning models for prediction of co-occurrence of diabetes and cardiovascular diseases: A retrospective cohort study. *J. Diabetes Metab. Disord.* **21**, 251–261 (2022).

Author contributions

Conceptualization, S.S. and S.G.; methodology, S.K.M.; validation, B.A.M.M.B.; resources, P.J.; data curation, G.T.D.; writing—original draft preparation, S.S. and S.G.; writing—review and editing, S.K.M., B.A.M.M.B., P.J. and G.T.D.; visualization, S.K.M., B.A.M.M.B., P..J and G.T.D.; supervision S.K.M., B.A.M.M.B., P.J. and G.T.D.; project administration.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to G.T.D.