ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Artificial intelligence model for prediction of cardiovascular disease: An empirical study

Buhari Ugbede Umar[1], Lukman Adewale Ajao[1]*, Eustace Mananyi Dogo[1], Falilat Jumoke Ajao[2], and Micheal Atama[1]

[1]Department of Computer Engineering, Federal University Technology, Minna, Niger State, Nigeria

[2]Department of Computer Science, Kwara State University, Malete-Ilorin, Kwara State, Nigeria

Abstract

Cardiovascular disease (CVD) is a disease related to the heart and blood vessels. Prediction of CVD is essential for early detection and diagnosis, which is however compounded by the complex interplay between medical history, physical examination outcomes, and imaging results. While the existing automated systems are fraught with the usage of irrelevant and redundant attributes, artificial intelligence (AI) helps in the identification of potential CVD populations by prediction models. This work aims at developing an AI model for predicting CVD using different classifications of machine learning techniques. The CVD dataset was obtained from the UCI repository containing about 76 cardiac attributes for training in various machine learning models, which include a hybrid of artificial neural network-genetic algorithm (ANN-GA), artificial neural network, support vector machine (SVM), K-means, K-nearest neighbor (KNN), and decision tree (DT). The performance of the models was measured in terms of accuracy, means square error, sensitivity, specificity, and precision. The results showed that the hybrid model of ANN-GA performs better with an accuracy of 86.4%, compared to the SVM, K-means, KNN, and DT measured at 84.0%, 59.6%, 79.0%, and 77.8%, respectively. It was observed that the system performs better as the number of datasets increases in the database, with a fewer selection of attributes using genetic algorithm for selection. Thus, the ANN-GA model is recommended for CVD prediction and diagnosis.

Keywords: Artificial neural network; Cardiovascular disease; Genetic algorithm; Machine learning; Support vector machine

***Corresponding author:**
Lukman Adewale Ajao
(ajao.wale@futminna.edu.ng)

Citation: Umar BU, Ajao LA, Dogo EM, *et al.*, 2024, Artificial intelligence model for prediction of cardiovascular disease: An empirical study. *Artif Intell Health*, 1(1): 1746.
https://doi.org/10.36922/aih.1746

Received: September 1, 2023

Accepted: November 14, 2023

Published Online: December 1, 2023

1. Introduction

Cardiovascular disease (CVD) is a general term encompassing a broad category of diseases affecting the components of the circulatory system, such as heart, blood vessels, and coronary artery[1]. In coronary artery disease, a blockage caused by the buildup of fatty material in the coronary artery limits the pumping capacity and the blood flow of the heart, thereby result in heart or coronary failure. The CVDs present in several forms, including coronary heart conditions, cerebrovascular disease (stroke), high blood pressure, peripheral artery disease, rheumatic heart disease, inborn cardiomyopathy disease, inflammatory heart condition, and coronary failure[2,3]. Globally, the CVDs stand as the leading cause of morbidity and mortality and its prevalence is increasing, posing a growing challenge to the public healthcare system[4]. According to the World Health Organization, an estimated 17 million people die from CVDs each year, particularly heart attacks and strokes[5]. The statistics are projected to reach 22.2 million people by 2030, if CVDs remain the leading cause of death and disability worldwide[6]. CVDs are no longer considered an affluent nation's disease because the statistics of more than 80% CVD-related deaths were contributed by low-and-middle-income countries[7-10]. The current clinical guideline for primary prevention of this vascular disease is to identify the asymptomatic patient by prognostication and early diagnosis, which help with the reduction of high-risk CVD patients[11-13]. The heart diseases are rapidly increasing in terms of number of cases among the young generation but are disproportionately affecting men with a mortality rate twice as high than in women[14]. Approximately 88,000 deaths can be traced to coronary heart disease, while stroke caused about 50,000 deaths and blood circulatory disease 49,000 deaths[15-17].

The major risk factors associated with these vascular diseases are age, sex, smoking, family history, cholesterol, poor diet, high blood pressure, diabetes, obesity, physical inactivity, and alcohol intake[18]. Some symptoms such as chest pain and tiredness are symptomatic indications aiding in the detection of CVD. However, diagnosing heart diseases is a difficult task in the field of medicine because a thorough analysis of patient clinical data and health history is required. Thus, an intelligent automated system for medical diagnosis can help enhance health-care observation and analysis, thereby facilitating the administration of suitable treatment for lifesaving purposes[19,20]. For the prediction of heart disease, several algorithms and diagnostic models have been created to predict heart disease by considering its risk factors, such as arterial high blood pressure, high cholesterol, diabetes mellitus, smoking, obesity, unhealthy diet, and a family history of heart disease[21-23].

The cardiac dataset obtained from the clinic is voluminous and contains many irrelevant and redundant attributes. These attributes of CVD risk factors may include inactivity, unhealthy eating, smoking, diabetes, age, and family history. Therefore, there is a need for attribute selection to reduce data redundancy and accuracy improvement using machine learning techniques to predict the output from existing data[24]. This emergence of machine learning algorithms or intelligent automated systems is widely applicable to several fields of studies for the simplification and counteraction of their natural and artificial challenges as it usually yields optimum performance. Such emergent technology systems in health care could be an expert system for different purposes, such as diagnosis and drug prediction, wearable devices or implanted sensors for the patient condition monitoring, or an artificial intelligence (AI) system using neural network (NN), genetic algorithm (GA), fuzzy logic, and many other techniques for the prediction and medical diagnosis[25].

The Tabu search algorithm is an optimization technique that uses adaptive memory to improve the local search for global optimum performance[26]. This search algorithm works using a Tabu list to prevent cycling and aspiration criteria for a globally optimal solution and to prevent the repetition of solutions[27]. The present study aims to analyze the performance of various machine learning techniques, such as artificial NNs (ANNs), artificial NN-genetic algorithms (ANN-GA), support vector machines (SVMs), K-means, K-nearest neighbor (KNN), and decision trees (DT), in yielding the best model for CVD prediction. The motivation of this research is to discover the most effective machine learning method for detecting CVDs with greater precision, sensitivity, and accuracy. The main contributions of this work to the literature are:

(i) Development of a hybrid model of ANN-GA for the early prediction and diagnosis of CVD.
(ii) Comparative analysis of CVD prediction and diagnosis model using different types of supervised machine learning techniques, such as ANN, K-means, KNN, and DT.
(iii) Development of an expert system-based application program interface (API) for CVD prediction and diagnosis.
(iv) Evaluation of the performance of different machine learning algorithms used for the prediction and diagnosis of CVDs.

The remainder of the paper is structured as follows: the review of related work is presented in Section 2; the research methodology is described in Section 3, and the result and discussion are presented in Section 4. Finally, the conclusion of the research work is presented in Section 5.

2. Related works

This section presents the strengths and weaknesses of various works conducted by other researchers, through which we can identify the research gaps. There are different types of vascular heart diseases, including coronary heart disease, congestive heart failure, cardiomyopathy, congenital heart disease, arrhythmias, deep vein thrombosis and pulmonary embolism, and many others. Heart disease has been identified as one of the leading causes of death globally. One of the reasons for heart disease-related deaths is that the risk is not recognized at an earlier stage. Manually calculating the chances of developing heart disease based on risk factors is difficult, but machine learning methods can be efficiently used to predict the outcome from the existing dataset. Predicting a dependent variable from the values of independent variables is one of the applications of these machine learning techniques due to large data resources that are difficult to manage manually as in the health-care sector. Some of the techniques used for these prediction problems are SVM, NN, DT, regression, and Naive Bayes classifiers. An ensemble classification method for improving the accuracy of a weak classifier of heart disease was developed by combining multiple classifiers. The results of the ensemble techniques in the bugging and boosting are effective in improving the prediction accuracy of weak classifiers[28].

Chowdhury *et al.* used the multilayer NN and the backpropagation learning algorithm with the heart disease dataset[29]. The initialization of NN weights was optimized using a GA. An accuracy of about 98% was obtained. Due to the limited dynamism in patterns and associations among the data mining techniques used[30], a feature subset selection method was applied to medical data. This method used a Naive Bayes classifier to select 5 attributes from 15 attributes. It was able to find a critical nugget, which reduced the irrelevant attribute, and found the top critical nuggets. The principal limitation of this method is that it can use only one data mining technique.

Srivenkatesh proposed CVD prediction using machine learning algorithms, such as SVM, random forest (RF), Naive Bayes classifier, and logistic regression, for vascular presumption, with the logistic regression showing a better accuracy of 77.06% when compared with another machine learning algorithm[31]. Sharma and Parmar proposed a heart disease prediction method using a deep learning NN model[32]. The dataset used in their works was obtained from the UCI repository for deep training, and the results obtained are promising for CVD prediction[32]. Separately, Mohan *et al.* predicted heart disease using machine learning models, such as logistic regression, KNN, SVM, DT, and RF, during the dataset training[33]. The accuracy values obtained for the K-neighbor classifier are 0.95619%, SVM 0.9561945%, DT 0.91050%, RF classifier 0.95404%, and LR 0.95592%[33]. Chowdary used RF, LR, and ANN, with activation by KNN, Gaussian Naive Bayes (GNB), and rectified linear unit (ReLu) for prediction of the heart disease infection[34]. The average performance accuracy and precision obtained were 89% and 91.6%, respectively[34,35].

Rabbi *et al.* evaluated the performance of data mining classification techniques for heart disease prediction using three popular classification techniques, such as KNN, SVM, and ANN, achieving 82.963%, 85.1852%, and 73.3333% in accuracy, respectively[36]. Empirical performance analysis of various machine learning techniques has been carried out to predict CVD using SVM, RF, DT, and KNN. The results are explicitly discussed with the DT classification accuracy of 73%[37]. Khourdifi and Bahaj proposed a machine learning algorithm for heart disease prediction and classification using particle swarm optimization (PSO) and ant colony optimization (ACO). The classification average accuracy for PSO and ACO is 99.65%[38].

Shah *et al.* proposed a methodology for the prediction of heart disease infection using Naive Bayes, DT, KNN, and RF algorithms on a dataset with 303 instances and 76 attributes[39]. The evaluation performance results showed that the KNN algorithm had the highest accuracy score of 90.789%[39]. An SVM classifier and GA have been combined to improve the performance of the SVM classifier in predicting heart disease based on risk factors. A system with an accuracy of 95% was obtained[40].

Haq *et al.* developed a machine learning model for CVD risk prediction in accordance with a dataset that contains 11 features used to forecast CVD[41]. The dataset was collected from Kaggle on CVD with approximately 70,000 patient records used for CVD prediction. This Kaggle dataset has plenty of training and validation records. The machine learning models used are NNs, RF, Bayesian networks, C5.0, and QUEST. The results acquired have a high prediction accuracy of 99.1%, which is significantly superior to previous methods[41].

Taylan *et al.* utilize machine learning to predict, classify, and improve the diagnostic accuracy of CVDs using support vector regression (SVR), multivariate adaptive regression splines, M5Tree model, and NNs for the training process[42]. While the KNN, Naive Bayes, and adaptive neuro-fuzzy inference system (ANFIS) were used to predict 17 CVD risk factors, the mixed-data transformation and classification methods were employed for categorical and continuous variables which predict CVD risk. However, the result obtained outperformed the well-known statistical and machine learning approaches, a clear indication of their versatility and utility in CVD classification. The investigation indicates that the prediction accuracy of ANFIS for the training process is 96.56%, and SVR is 91.95%[42].

Tran *et al.* developed a prediction mortality model for patients with CVDs to support health-care services[43]. The dataset used was obtained from the Medicare Benefits Scheme and Pharmaceutical Department, Australia, between 2004 and 2014. The dataset contains about 346,201 patient records. Some of the AI algorithms used in prediction include LR, RF, extra trees (ET), gradient boosting trees (GBT), and deep learning algorithms. However, some of the minority deceased patient records contained in the dataset were experimented separately using the synthetic minority oversampling technique (SMOTE) to enrich the data. Regarding model performance in terms of discrimination, GBT and RF were the models with the highest receiver operating characteristic curve of 97.8% and 97.7%, respectively. The discriminative powers of ET, LR, and DNN were 96.8%, 96.4%, and 95.3%, respectively, with the latter exhibiting the least discriminative power. In terms of reliability, LR predictions were the least calibrated compared with the other four algorithms. Thus, despite increasing the training time, SMOTE was proven to further improve the model performance of LR, while algorithms like GBT and DNN performed well with class-imbalanced data[43].

Khan *et al.* utilized a machine learning algorithm for the accurate prediction and decision-making for CVD patient[44]. Simple random sampling was used to select heart disease patients from the Khyber Teaching Hospital and Lady Reading Hospital, Pakistan. Some of the machine learning methods involved are DT, RF, and LR. The performance of the proposed machine learning algorithm was estimated using numerous conditions to recognize the most suitable machine learning algorithm in the class of models. The RF algorithm has the highest accuracy of prediction, sensitivity, and recursive operative characteristic curve of 85.01%, 92.11%, and 87.73%, respectively, for CVD prediction. It also has the least specificity and misclassification errors of 43.48% and 8.70%, respectively. The overall performance evaluation result showed that the RF algorithm is the most appropriate algorithm for CVD classification and prediction[44]. A summary of the authors' focus, methods, and strengths is presented in Table 1.

3. Materials and methods

3.1. Implementation of ANN-GA for CVDs model

In this research, a CVD prediction model was developed using a combination of techniques based on the evolution theory – GA and ANN. This ANN-GA model was used for the correlated attribute selection, training, and optimization of the CVD selected features to achieve better prediction accuracy. The ANN utilized a multilayer perceptron to receive the transmitted signal, process the signal (neurons), and compute the output of each neuron for weighted using back-propagation to adjust the NN model parameters for a reduced mean squared error (MSE). While the GA is an optimization searching agent in the evolution theory, that is used to optimize the output performance of a classifier model using a supervised machine learning algorithm (SMLA). The function of GA evolution initiates the process from a randomly generated population for an individual which changes at every iteration in the loops to formulate an offspring called generation[45] The fitness (best chromosome) of every individual generated within the population was appraised, and the fittest was chosen (selection of chromosome) by computing the probability for being selected from among the best chromosome in the population in proportion to the sum of the fitness among individuals as given in Equation I[46]. Nevertheless, the generation of the individual population can be modified by crossover and mutation to form a newly evolved generation candidate as presented in Figure 1. The overview of the ANN-GA flowchart is presented in Figure 2.

Table 1. Summary of the related works in terms of the authors' focus, methods, and strength

Authors' focus	Methods	Performance accuracy
Proposed CVD prediction model[35]	Using machine learning algorithms such as SVM, RF, Naive Bayes classifier, and LR	K-neighbor 95.62%, SVM 95.62%, DT 91.05%, RF 95.40%, and LR 95.59%
Performance of data mining classification techniques for heart disease prediction[36]	Using three popular classification techniques, such as KNN, SVM, and ANN	KNN 82.96%, SVM 85.18%, and ANN 73.33%
Proposed machine learning algorithm for heart disease prediction and classification[38]	Using PSO and ACO	The average accuracy for PSO and ACO is 99.65%
Machine learning approaches for predicting, classifying, and improving the diagnosis accuracy of CVDs[42]	SVR, multivariate adaptive regression splines, M5Tree model neural networks, adaptive neuro-fuzzy, and KNN, and Naive Bayes classifiers	ANFIS 96.56%, and SVR 91.95%
Nurturing clinicians for accurate prediction of mortality among CVD patients[43]	LR, RF, ET, and GBT	GBT 97.8%, RF 97.7%, ET 96.8%, LR 96.4%,
Accurate prediction and decision-making for CVD classification[44]	DT, RF, LR	DT 85.01%, RF 92.11%, and LR 87.73%

Abbreviations: ACO: Ant colony optimization; ANFIS: Adaptive neuro-fuzzy inference system; ANN: Artificial neural network; CVD: Cardiovascular disease; DT: Decision tree; ET: Extra tree; GBT: Gradient boosting tree; KNN: K-nearest neighbor; LR: Logistic regression; PSO: Particle swarm optimization; RF: Random forest; SVR: Support vector regression.

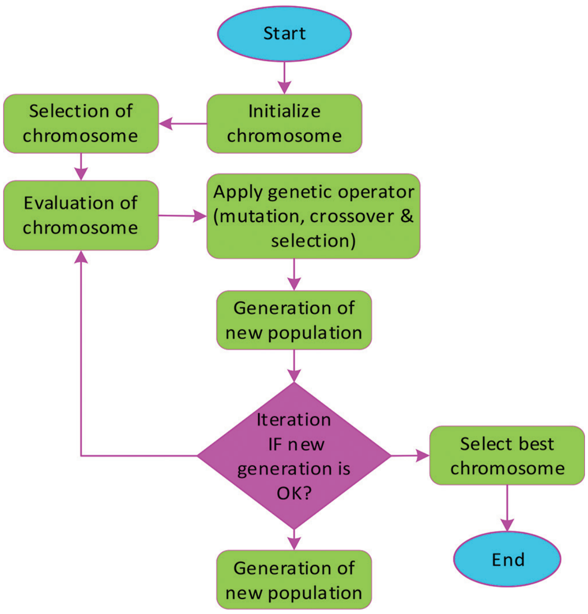

Figure 1. Overview of genetic algorithm for feature selection.

$$\rho_k = \frac{f_k}{\sum_{k=1}^{\delta s} f_k} \quad \text{(I)}$$

Where ρ_K is the probability of an individual selection, f_k is an individual fitness, and δs is the population size.

3.2. Machine learning algorithm for CVD prediction model

Similarly, some of the popular SMLAs, such as K-means, KNN, SVM, and DT, were further utilized for the training and prediction of CVDs to carry out a comparative analysis of the prediction model. The CVD dataset was obtained from the UCI repository, which contains about 76 cardiac attributes for the training in various machine learning models mentioned. This CVD dataset consists of 14 attributes, including age, sex, chest pain type, resting blood pressure, serum cholesterol (mg/dl), fasting blood sugar, resting electrocardiographic result, maximum heart rate, exercise-induced angina, oldpeak, slope of the peak,

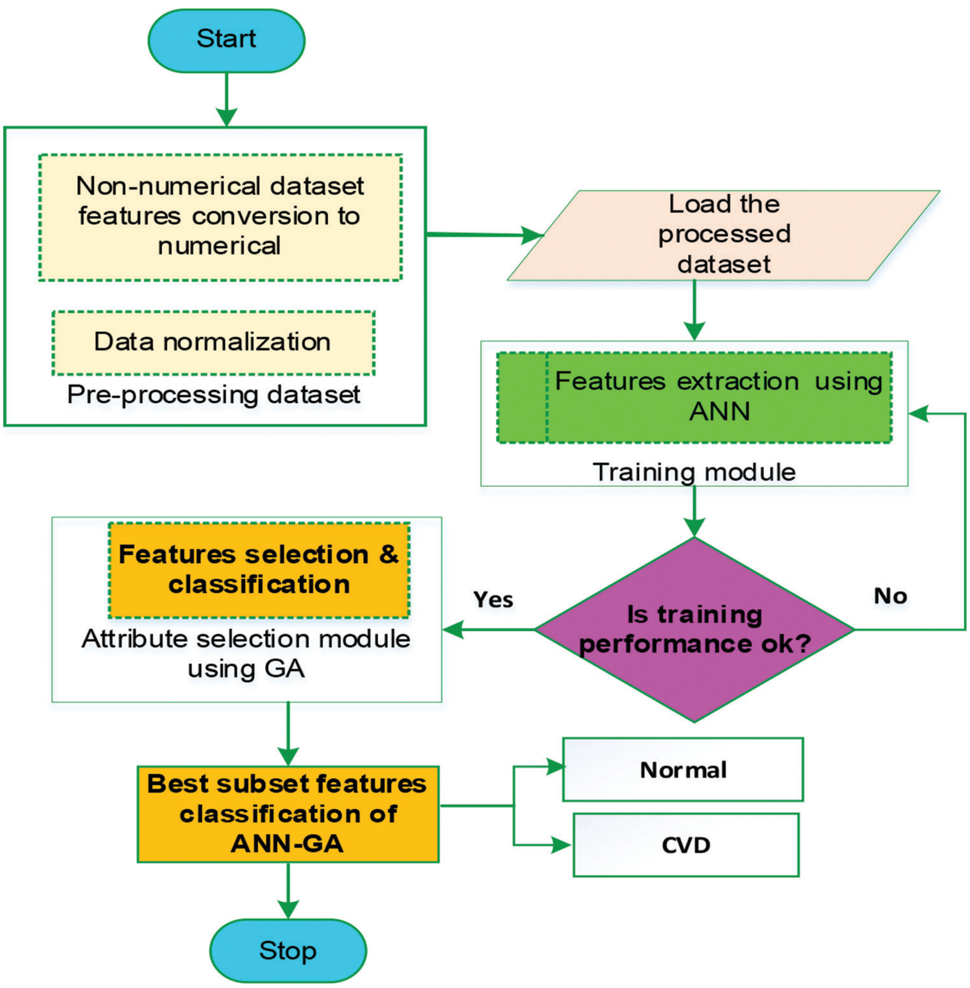

Figure 2. The flowchart for artificial neural network-genetic algorithm training model.

number of major vessels colored by fluoroscopy, and defect type (Thal) (Table 2). The workflow of the machine learning model for vascular disease prediction and classification is given in Figure 3.

However, machine learning is a method of analyzing data samples and drawing main conclusions using mathematical and statistical approaches, which allows machines to learn without having to be programmed. The collected data were pre-processed using the data imputation of the KNN method to fill in the missing values and data normalization before the classification process. The attribute average (mean) for a column with a missing value was calculated and this was used to fill in the missing values. This technique is direct and simple to implement using Equation II.

$$X^{-1} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad \text{(II)}$$

where x[1] is the normalized size, x is the original value, and min(x) and max(x) are the minimum and maximum values of the original image, respectively. The normalization considered the input values for each attribute obtained using the min-max normalization technique.

The normalization ensures that the values stay in the range of 0 – 1 to prevent a surge to higher values. Min-max normalization technique was used because of the NN activation function with a recommended range of 0.1 – 0.9 to avoid saturation. The data pre-processed was then divided into two: a training set of 70% and a testing set of 30% ratio. The training allows the NN to develop a relationship between the input data and the target. The testing set was used to test how efficiently and accurately the NN can predict the target. The number of attributes was reduced from 14, 10, and 8, and also, GA was used for data selection. Furthermore, the number of datasets varied from 2000, 1500, 1000, and 500, and their performance was evaluated.

A multilayer feed-forward back propagation NN was used with 14, 10, and 8 input layers, various hidden layers,

Table 2. CVD dataset attributes

Key attribute	Patient attribute ID
Age	Year
Sex	Value 1: Male; value 0: Female
Chest pain type	Value 1: Typical type 1 angina; value 2: Typical type angina; value 3: Non-angina pain; value 4: Asymptomatic
Fasting blood sugar	Value 1: >120 mg/dl; value 0: <120 mg/dl
Resting electrographic results (Restecg)	Value 0: Normal; value 1: Having ST-T wave abnormality; value 2: Showing probable or definite left ventricular hypertrophy
Exercise-induced angina (Exang)	Value 1: Yes; value 0: No
Slope of the peak exercise ST segment (Slope)	Value 1: Unsloping; value 2: Flat; value 3: Down-sloping
Number of major vessels colored by fluoroscopy (CA)	Value 0–3
Defect type (Thal)	Value 3: Normal; value 6: Fixed defect; value 7: Reversible defect
Resting blood pressure	The blood pressure of patients admitted to the hospital was measured in mmHg
Serum cholesterol	Measured in mg/dl
Thalach	Maximum heart rate achieved
Oldpeak – ST	ST depression induced by exercise

Abbreviation: CVD: Cardiovascular disease.

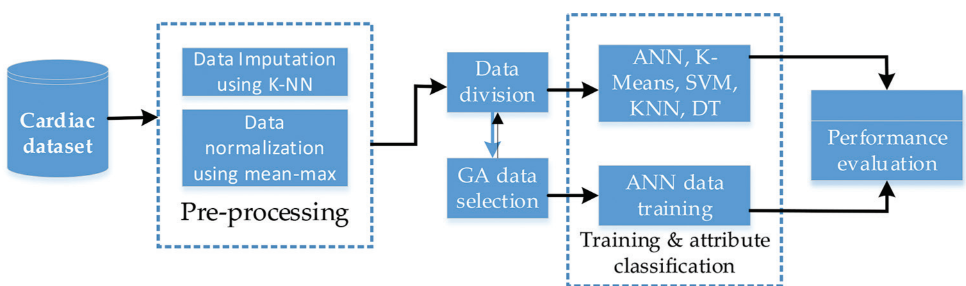

Figure 3. The overview of the machine learning model for vascular disease prediction.

and two output layers. The effect of the number of hidden neurons on the network performance was carried out, and five different numbers of hidden neurons (10, 20, 25, and 30) were investigated. In each case, the network was trained several times and the best performance was recorded using the Levenberg algorithm. The number of training cycles required for proper generalization of outputs was determined through the trial-and-error method. Furthermore, the threshold function was determined through trial and error (0.5). The NN was trained and tested using 76 samples, of which 70% samples for training and 30% samples were for testing 14, 10, and 8 attributes, and the number of datasets varied between 2000, 1500, 1000, and 500, respectively. The data collected for training was inputted into the training network designed. The training performance was evaluated, and the MSE was observed. The ANN and GA were combined to evaluate the performance of the training and testing samples. The GA was used for attribute selection, and the obtained attributes were then trained and tested with the number of datasets between 2000, 1500, 1000, and 500.

3.3. API for the CVD prediction

An expert CVD prediction model was developed in MATLAB for the experimentation and prediction observation of the CDV parameters. The CVD model interface involved the load data tab, the pre-processing data tab, and the diagnoses patient tab. This model allows a patient to enter the medical information required in the input data directory. The "Load data" button is ticked to load data for processing in the CVD system model. The "Processing data" button is ticked for pre-processing (data normalization) and displaying the data on the interface. The "Diagnoses patient" button is ticked to initiate information processing by the system and generate results on the message box as either "Normal" or "CVD". The graphical user interface developed in MATLAB is illustrated in Figure 4.

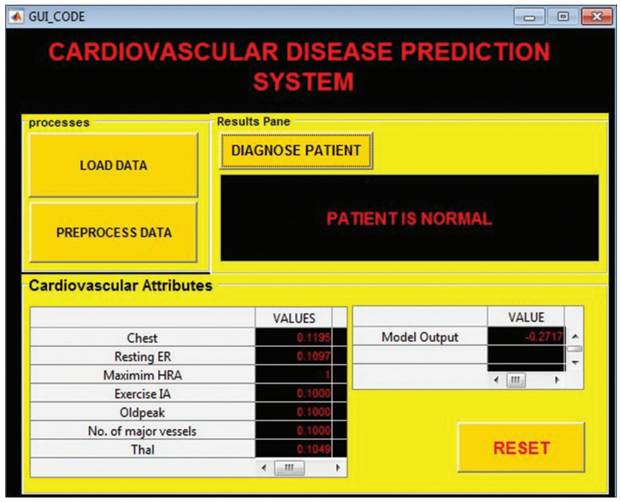

Figure 4. Graphical user interface for cardiovascular disease prediction system.

4. API for CVD prediction

The CVD dataset was pre-processed, trained, and predicted using ANN, hybridized ANN-GA, K-means, KNN, SVM, and DT classification techniques. The results obtained from the various AI techniques for CVD with their performance were evaluated using the cardiac disease attributes and varying the number of datasets: 2000, 1500, 1000, and 500. This number of selected attributes of the dataset varied from 14, 10, and 8, and their performance on the different classifiers of the machine learning model was observed and presented in Section 4.1 and Section 4.2.

4.1. Performance evaluation results of the selected number of attributes using the ANN model

The performance of the ANN classification technique using different numbers of attributes of 14, 10, and 8 was experimented. The result of ANN performance was evaluated in terms of accuracy, mean square error (MSE), sensitivity, specificity, and precision, measuring 82.70%, 0.1728, 87.50%, 75.80%, and 84.00%, respectively. It is observed that when the cardiac disease attributes were reduced to 10, the results obtained from the training were 80.20%, 0.1975, 91.70%, 63.60%, and 78.60%, for the same set of performance parameters (as in the above order). For the 8 attributes, the results of the same set of performance parameters were 79.0%, 0.2099, 79.20%, 78.80%, and 84.40%, respectively. Therefore, it is observed that as the number of attributes was reduced from 14 to 10 and 8, the accuracy obtained decreased, the MSE value increased, and the sensitivity, specificity, and precision decreased (Figure 5).

The performance evaluation of combined ANN-GA for the vascular disease classification was analyzed. GA was used for the subset of selected attribute. The seven best attributes correlated with cardiac disease classification are presented in Table 3. ANN was used to train the selected attributes, and GA was used for the attribute selection and classification. The results obtained showed 86.4%, 0.1358, 91.70%, 78.80%, and 86.3% for accuracy, MSE, sensitivity, specificity, and precision, respectively, as illustrated in Figure 6.

The K-means algorithm was used in the training of different selected attributes (14, 10, and 5). The results obtained showed 59.60%, 0.4030, 64.50%, 51.90%, and

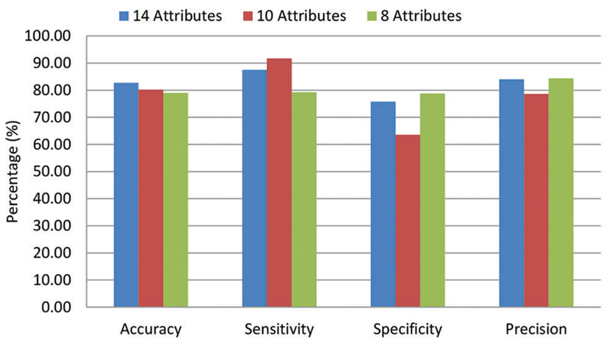

Figure 5. Performance analysis of artificial neural network algorithm on the selected attributes.

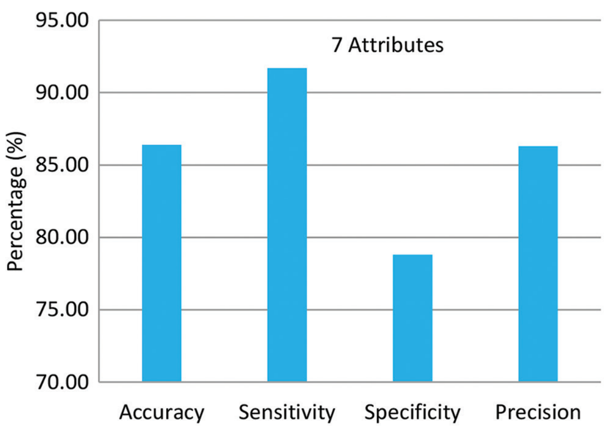

Figure 6. Performance analysis of artificial neural network-genetic algorithms on the selected seven attributes.

68.20% for accuracy, MSE, sensitivity, specificity, and precision, respectively. From the analysis of the results, the system performed better on 14 attributes. When the cardiac disease attributes were reduced to 10, the classification accuracy, MSE, sensitivity, specificity, and precision obtained were 58.90%, 0.4111, 63.90%, 51.00%, and 67.50% respectively. For 8 attributes, the same set of performance parameters was measured at 58.90%, 0.4111, 63.90%, 51.00%, and 67.50%. The performance of K-means classification, as the number of attributes was reduced from 14, 10, and 8, is illustrated in Figure 7.

The performance evaluation of the system using the KNN classification technique with 14, 10, and 8 attributes showed that the accuracy, MSE, sensitivity, specificity, and precision are 79.0%, 0.21, 95.80%, 54.60%, and 75.40%, respectively. Based on the result, the system performed better when 14 attributes were selected. It was also observed that when the selected attributes were reduced to 10, the classification accuracy, MSE, sensitivity, specificity, and precision were 75.30%, 0.247, 93.80%, 48.50%, and

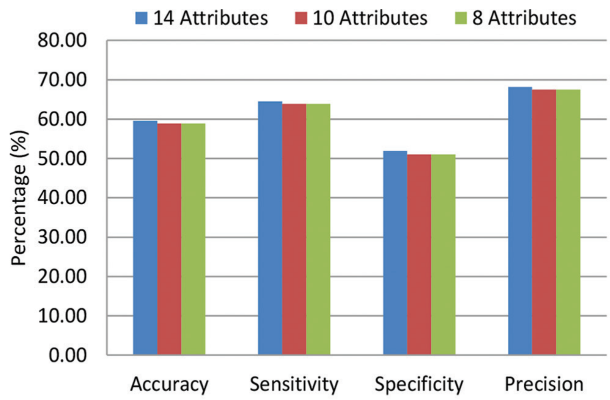

Figure 7. Performance analysis of the K-means algorithm on the selected attributes.

Table 3. GA selected attributes

S. No.	Seven attributes selected with GA
1	Chest pain type
2	Resting electrocardiographic result
3	Maximum heart rate
4	Exercise-induced angina
5	Oldpeak
6	Number of major vessels colored
7	Thal

Abbreviation: GA: Genetic algorithm.

72.6%, respectively, which were obtained. For 8 attributes, the results for accuracy, MSE, sensitivity, specificity, and precision were 71.60%, 0.2840, 97.90%, 35.32%, and 67.70%, respectively. However, when the number of attributes was reduced from 14 to 10, the sensitivity decreased, and as the number of attributes reduced to 8, the sensitivity slightly increased (Figure 8). This shows that 8 selected attributes contributed to better classification performance in this scenario. The MSE also increased as the number of attributes was reduced from 14, 10, and 8 attributes.

The performance analysis of the SVM classification technique with 14, 10, and 8 selected attributes was performed. The results obtained for the SVM classification based on the accuracy, MSE, sensitivity, specificity, and precision were 84.10%, 0.1600, 93.80%, 67.70%, and 81.80%, respectively. It was observed that when the attributes were reduced to 10, the performance accuracy, MSE, sensitivity, specificity, and precision obtained were 77.80%, 0.222, 77.10%, 78.80%, and 84.10%, respectively. For 8 attributes, the results obtained for the same set of performance parameters were 79.0%, 0.210, 93.6%, 58.82%, and 75.86%, respectively. The MSE, specificity, and precision were observed to increase as the number of attributes reduced from 14 to 10 and 8. However, the accuracy and sensitivity decreased as the number of attributes reduced from 14 to 10 and then increased as the number of attributes was further reduced to 8. The results demonstrated that the system performed better with 14 attributes, compared to other selected attributes such as 10 and 8, as illustrated in Figure 9.

The performance analysis of using DT algorithm on the selected attributes of 14, 10, and 8 was conducted. The results obtained for the accuracy, MSE, sensitivity, specificity, and precision were 77.80%, 0.222, 91.67%, 57.6%, and 75.9%, respectively. The results presented in Figure 10 showed that a better result was achieved when 14 attributes were selected. But when the selected attributes were reduced to 10, the classification accuracy, MSE, sensitivity, specificity, and precision obtained were 70.4%, 0.296, 77.10%, 60.60%, and 74.0%, respectively. For the selected 8 attributes, the obtained accuracy, MSE, sensitivity, specificity, and precision were 77.80%, 0.222, 97.9%, 50.00%, and 73.0%, respectively. However, when the number of attributes was reduced from 14 to 10, the accuracy and sensitivity decreased but increased when 8 was used, whereas the MSE and specificity increased and slightly decreased when the number of attributes was reduced to 8. The details of the selected attributes for 14, 10, and 8 are presented in Tables 4-6.

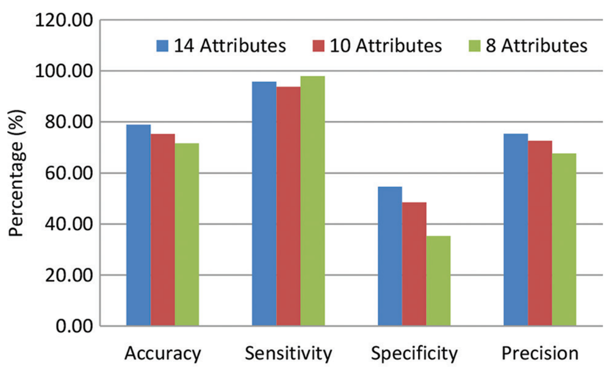

Figure 8. Performance analysis of the K-nearest neighbor algorithm on the selected attributes.

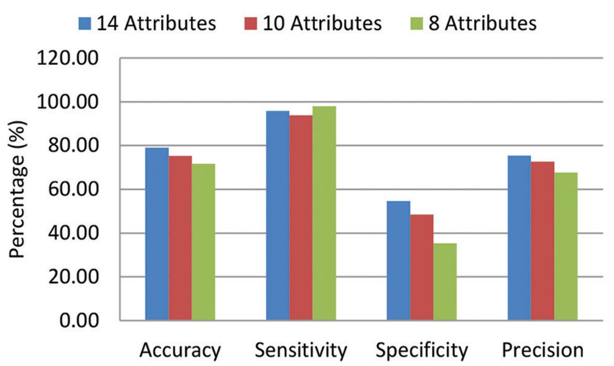

Figure 9. Performance analysis of support vector machine algorithm on the selected attributes.

4.2. Performance evaluation analysis of classification algorithm on the varied number of datasets

The number of the dataset was varied from 2000, 1500, 1000, and 500. The corresponding performance on different classifiers was observed. The performance analysis result of various models using 2000 sets of data with ANN is presented in Figure 11. The results obtained for the accuracy, MSE, sensitivity, specificity, and precision were 73.30%, 0.2667, 77.80%, 66.70%, and 77.80%, respectively. The ANN classification showed better performance with a high range of datasets; hence, 2000 datasets were adopted. When the cardiac disease dataset was reduced to 1500, the accuracy, MSE, sensitivity, specificity, and precision obtained were 71.10%, 0.2889, 82.60%, 59.10%, and 67.90%, respectively. Considering 1000 CVD datasets, the results obtained for accuracy, MSE, sensitivity, specificity, and precision were 83.30%, 0.1667, 90.00%, 70.00%, and 85.70%, respectively. Finally, for the 500 CVD datasets, the accuracy, MSE, sensitivity, specificity, and precision results obtained were 66.70%, 0.3330, 83.30%, 55.60%, and 66.70%, respectively. However, when the selected CVD dataset number was reduced from 2000 to 1500, the accuracy, specificity, and precision decreased and increased sharply at 1000 datasets before decreasing again when the number of cardiac dataset was 500. The sensitivity increased when the dataset was reduced from 2000 to 1000 and decreased only slightly when the dataset number was 500. Meanwhile, the MSE increased at 1500 CVD datasets using ANN, decreased when the dataset used was 1000, and increased slightly when the number of cardiac dataset dropped to 500.

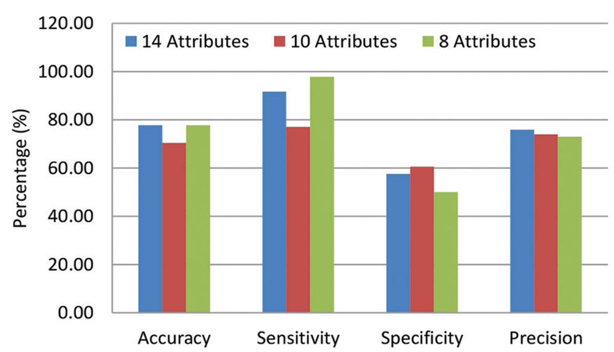

Figure 10. Performance analysis of using decision tree algorithm on the selected different attributes.

 https://doi.org/10.36922/aih.1746

Table 4. A decision tree with 14 attributes selection

S. No.	14 Attributes
1	Age
2	Sex
3	Chest pain type
4	Resting blood pressure
5	Serum cholesterol
6	Fasting blood sugar
7	Resting electrocardiographic result
8	Maximum heart rate
9	Exercise-induced angina
10	Oldpeak
11	Slope
12	Number of major vessels colored
13	Thal
14	Class

Table 5. A decision tree with 10 selected attributes

S. No.	Ten attributes
1	Resting blood pressure
2	Serum cholesterol
3	Fasting blood sugar
4	Resting electrocardiographic result
5	Maximum heart rate
6	Exercise-induced angina
7	Oldpeak
8	Slope
9	Number of major vessels colored
10	Thal

Table 6. A decision tree with 8 selected attributes

S. No.	Eight attributes
1	Chest pain type
2	Resting blood pressure
3	Serum cholesterol
4	Fasting blood sugar
5	Max. heart rate
6	Exercise-induced angina
7	Oldpeak
8	Number of major vessels colored

After classification using the K-means technique, the performance of the selected datasets was evaluated in terms of accuracy, MSE, sensitivity, specificity, and precision, measuring 59.50%, 0.4050, 64.10%, 51.40%,

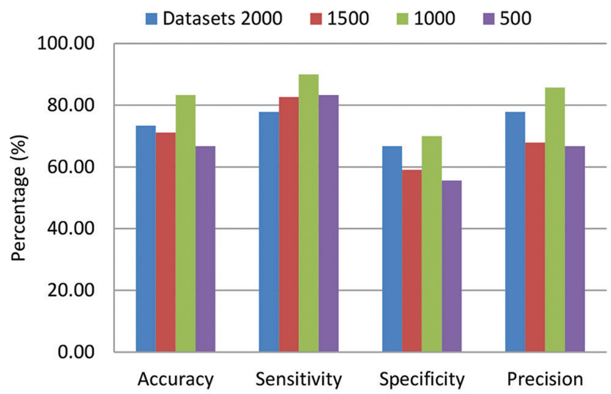

Figure 11. Performance analysis of artificial neural network algorithms on datasets of different sizes.

and 70.10%, respectively, for 2000 CVD datasets. When the dataset number was reduced to 1500, the accuracy, MSE, sensitivity, specificity, and precision obtained were 42.00%, 0.5800, 48.20%, 38.30%, and 61.80%, respectively (Figure 12). For the 1000 CVD datasets, the accuracy, MSE, sensitivity, specificity, and precision obtained were 58.00%, 0.4200, 62.30%, 48.40%, and 72.90%, respectively. The accuracy, MSE, sensitivity, specificity, and precision of 500 CVD datasets were 54.00%, 0.4600, 55.00%, 50.00%, and 81.50%, respectively. When the CVD dataset was reduced from 2000 to 1500, the accuracy and sensitivity decreased and increased at 1000 CVD datasets before decreasing when the number of dataset equals to 500. The MSE increased when the dataset was reduced from 2000 to 1500, then decreased at 1000 before increasing when the dataset was further reduced to 500, whereas the specificity and precision decreased as the dataset was reduced from 2000 to 1500, then increased as the dataset was further reduced.

The performance of the ANN-GA was analyzed on the selected CVD datasets of 2000, 1500, 1000, and 500 (Figure 13). GA was used for the selection of a subset of attributes. The seven best attributes that are more correlated are responsible for cardiac disease. After training and testing the ANN with the obtained attributes, the performance of 2000 CVD datasets in accuracy, MSE, sensitivity, specificity, and precision were 80.00%, 0.2000, 97.20%, 54.20%, and 76.10%, respectively. For 1500 CVD datasets, the accuracy, MSE, sensitivity, specificity, and precision were 77.80%, 0.2220, 95.70%, 59.10%, and 71.00%, respectively. For 1000 CVD datasets, accuracy, MES, sensitivity, and precision of 73.30%, 0.2667, 90.00%, 40.00%, and 75.00% were obtained respectively, and for 500 CVD datasets, the same set of performance parameters was 73.30%, 0.2667, 83.00%, 66.70%, and 62.500%, respectively. The accuracy and sensitivity decreased steadily as the number of datasets

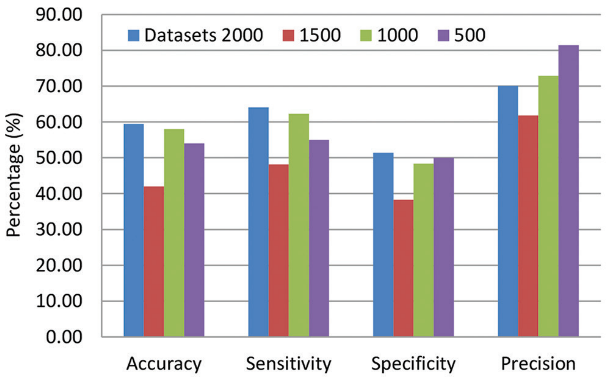

Figure 12. Performance analysis of K-means algorithms on datasets of different sizes.

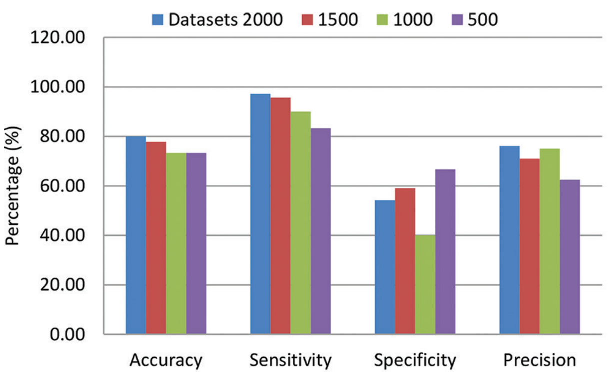

Figure 13. Performance analysis of artificial neural network-genetic algorithms on datasets of different sizes

reduced from 2000 to 500. MSE increased with a reduced number of datasets. Specificity increased as the number of datasets was reduced from 2000 to 1500, decreased at 1000 CVD datasets, but increased sharply when the number of datasets was reduced to 500. Meanwhile, the precision decreased as the number of datasets was reduced from 2000 to 1500, increased at 1000 datasets, and decreased as the number of datasets was reduced to 500.

The performance of the KNN classification technique using 2000, 1500, 1000, and 500 datasets is presented in Figure 14. The classification accuracy, MSE, sensitivity, specificity, and precision were 71.70%, 0.2838, 100%, 70.00%, and 67.90%, respectively, for 2000 datasets. Based on the results, the system performed better at 2000 datasets when compared with 1500, 1000, and 500 datasets, and hence, 2000 datasets were adopted for the system. When the dataset number was reduced to 1500, the accuracy, MSE, sensitivity, specificity, and precision were 55.60%, 0.4440, 100%, 100%, and 55.60%, respectively. For 1000 datasets, the accuracy, MSE, sensitivity, specificity, and precision were 56.70%, 0.4330, 100%, 100%, and 56.70%, respectively. For 500 datasets, the accuracy, MSE, sensitivity, specificity,

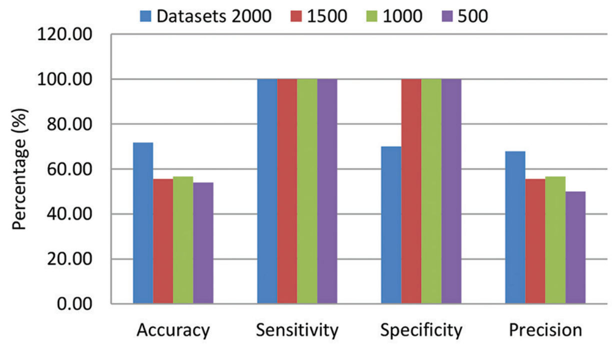

Figure 14. Performance analysis of K-nearest neighbor algorithms on datasets of different sizes.

and precision were 54.00%, 0.4600, 100%, 100%, and 50%, respectively. The results showed that KNN has excellent sensitivity and specificity of 1000% even at a reduced number of datasets. Accuracy and precision decreased gradually as the number of datasets reduced from 2000 to 500. On the other hand, MSE increased steadily with a reduced number of datasets.

The performance results of the model using the SVM classification technique on 2000, 1500, 1000, and 500 datasets are presented in Figure 15. The classification accuracy, MSE, sensitivity, specificity, and precision were 78.30%, 0.2170, 91.70%, 58.30%, and 96.10%, respectively, for 2000 datasets. According to the results, the system performed better at 2000 datasets, hence adopted for the system. When the dataset number was reduced to 1500, the accuracy, MSE, sensitivity, specificity, and precision were 71.10%, 0.2890, 84.00%, 55.00%, and 70.00%, respectively. For 1000 datasets, the accuracy, MSE, sensitivity, specificity, and precision were 73.30%, 0.2670, 100%, 75.00%, and 85.70%, respectively, and for the 500 datasets, the same set of performance parameters was measured at 58.30%, 0.4170, 66.70%, 50.00%, and 57.10%, respectively. The accuracy, sensitivity, specificity, and precision decreased as the number of datasets reduced from 2000 to 1500, then increased as the number of datasets was 1000 before decreasing at 500 datasets. While the MSE increased when datasets reduced to 1500, it decreased at 1000 datasets before increasing to its peak at 500 datasets.

The performance results of the system using the DT algorithm for 2000, 1500, 1000, and 500 datasets are presented in Figure 16. For 2000 datasets, the classification accuracy, MSE, sensitivity, specificity, and precision were 75.00%, 0.2500, 88.90%, 54.20%, and 74.40%, respectively. It was observed that the system performed better at 2000 datasets; therefore, it was adopted for the system. Based on the results, when the datasets were reduced to 1500, the accuracy, MSE, sensitivity, specificity, and precision were

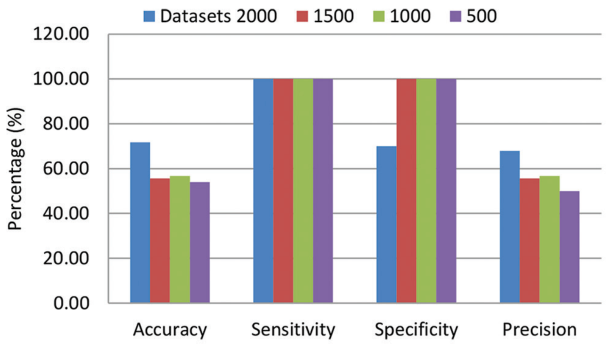

Figure 15. Performance analysis of support vector machine algorithms on datasets of different sizes.

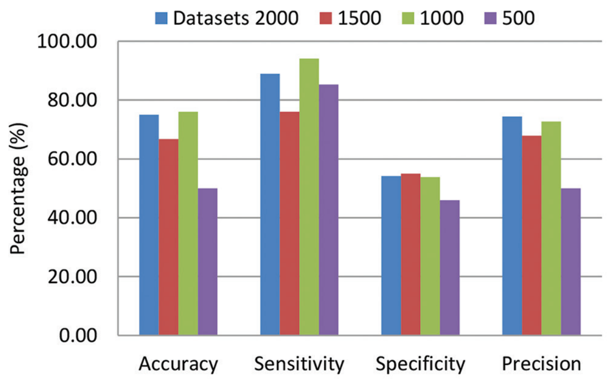

Figure 16. Performance analysis of decision tree algorithms on datasets of different sizes.

66.70%, 0.3300, 76.00%, 55.00%, and 67.90%, respectively. For 1000 datasets, the accuracy, MSE, sensitivity, specificity, and precision were 76.70%, 0.3400, 94.10%, 53.80%, and 72.70%, respectively. For 50 datasets, the accuracy, MSE, sensitivity, specificity, and precision were 50.00%, 0.5000, 85.30%, 46.00%, and 50.00%, respectively. Using the DT algorithm, the classification accuracy, sensitivity, and precision decreased as the number of datasets reduced to 1500 and then increased at 1000 datasets before decreasing again at 500 datasets. MSE was increased as the number of datasets was reduced from 2000 to 500. On the other hand, the specificity increased at first as the number of dataset reduced from 2000 to 1500 before a steady decrease with the reduction in the number of datasets.

The formulas for determining sensitivity, specificity, accuracy, precision, and MSE are given in the Equations III to VII:

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

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

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

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

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}\left(S_t - S'_t\right)^2 \quad \text{(VII)}$$

where TP is true positive that correctly classified positive cases; TN is true negative that correctly classified negative cases; FP is false positive that incorrectly classified positive cases; FN is false negative that incorrectly classified negative cases; MSE is mean squared error; n is classifications; $S_{(t)}$ is actual classification; and $S'_{(t)}$ is predicted classification.

5. Conclusion

Various classification techniques and a comprehensive understanding of the hidden correlations between attributes that play pivotal role in CVD are instrumental for cost-effective, automatic, and early prediction of the disease to reduce the mortality rate. This study worked on different CVD attributes from patients using ANN, ANN-GA, K-means, KNN, SVM, and DT in a MATLAB environment. Given the diversity in the attributes and dataset number, GA was employed for the selection of correlated attributes that contribute to CVD. The purpose of this work is to shed light on different classifiers with a better predictive ability (precision) since wrong and late diagnosis may lead to death. The performance of the classifiers was evaluated in terms of accuracy, MSE, sensitivity, specificity, and precision. Based on the results, the ANN model combined with GA performs better with an accuracy of 86.4% as compared to SVM at 84.0%, K-means at 59.6%, KNN at 79.0%, and DT at 77.8%. Thus, the ANN-GA model is therefore recommended for CVD diagnosis and prediction. This research shows that better accuracy is obtained when a larger number of datasets are used. Future research work should focus on expert system development for the CVD prediction, diagnosis, and prescription of drugs. Furthermore, more robust AI and optimization algorithms should be developed for the optimal performance of CVD prediction and diagnosis.

Acknowledgments

The authors appreciate the management of Federal University of Technology, Minna, Nigeria, for the given access to the AI and Embedded System Laboratory in Computer Engineering Department, where thorough research investigation and coding were carried out. The authors also extend their sincere appreciation to the

Academic Chief Technologist of the Laboratory for his effort and validation of the results.

Funding

None.

Conflict of interest

The authors declare that they have no competing interest.

Author contributions

Conceptualization: Lukman Adewale Ajao, Buhari Ugbede Umar
Formal analysis: Eustace Mananyi Dogo, Falilat Jumoke Ajao
Investigation: Micheal Atama, Lukman Adewale Ajao
Methodology: Micheal Atama, Lukman Adewale Ajao, Buhari Ugbede Umar
Writing – original draft: Micheal Atama, Lukman Adewale Ajao, Buhari Ugbede Umar
Writing – review & editing: Eustace Mananyi Dogo, Falilat Jumoke Ajao, Lukman Adewale Ajao
All authors equally contributed to the work.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The dataset used was collected from Heart Disease UCI Machine Learning Repository (http://archive.ics.uci.edu).

References

1. Dinesh KG, Arumugaraj K, Santhosh KD, *et al.*, 2018, Prediction of Cardiovascular Disease Using Machine Learning Algorithms. In: International Conference on Current Trends towards Converging Technologies (ICCTCT). Coimbatore, India, p1–7.

   https://doi.org/10.1109/ICCTCT.2018.8550857

2. Wong KK, Fortino G, Abbott D, 2020, Deep learning-based cardiovascular image diagnosis: A promising challenge. *Future Gener Comput Syst*, 110: 802–811.

   https://doi.org/10.1016/j.future.2019.09.047

3. Masethe HD, Masethe MA, 2014, Prediction of heart disease using classification algorithms. *Proc World Congr Eng Comput Sci*, 2: 25–29.

4. Crosier R, Austin PC, Ko DT, *et al.*, 2021, Intensity of guideline-directed medical therapy for coronary heart disease and ischemic heart failure outcomes. *Am J Med*, 134: 672–681.

   https://doi.org/10.1016/j.amjmed.2020.10.017

5. Jibril IZ, Agajo J, Ajao LA, *et al.*, 2018, Development of a medical expert system for hypertensive patients diagnosis: A knowledge-based rules. *Adv Electron Eng J*, 1: 23–29.

6. Wang S, Shen B, Martin C, 2021, Signal Analysis of Heart Rate Variability and Applications on the Diagnosis of Cardiovascular Diseases. In: MOL2NET, International Conference Series on Multidisciplinary Sciences, p1–3.

7. Martin-Isla C, Campello VM, Izquierdo C, *et al.*, 2020, Image-based cardiac diagnosis with machine learning: A review. *Front Cardiovasc Med*, 7: 1–3.

8. Tarawneh M, Embarak O, 2019, Hybrid Approach for Heart Disease Prediction Using Data Mining Techniques. In: International Conference on Emerging Internetworking, Data and Web Technologies, p447–454.

9. Khourdifi Y, Bahaj M, 2018, Applying Best Machine Learning Algorithms for Breast Cancer Prediction and Classification. In: International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), p1–5.

10. Boateng D, Agyemang C, Beune E, *et al.*, 2018, Cardiovascular disease risk prediction in Sub-Saharan African populations comparative analysis of risk algorithms in the RODAM study. *Int J Cardiol*, 254: 310–315.

11. Alaa AM, Bolton T, Di Angelantonio E, *et al.*, 2019, Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PloS One*, 14: e0213653.

    https://doi.org/10.1371/journal.pone.0213653

12. Mohan S, Thirumalai C, Srivastava G, 2019, Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 7: 81542–81554.

    https://doi.org/10.1109/ACCESS.2019.2923707

13. Rubini PE, Subasini CA, Katharine AV, *et al.*, 2021, A cardiovascular disease prediction using machine learning algorithms. *Ann Rom Soc Cell Biol*, 25: 904–912.

14. Sanchis-Gomar F, Perez-Quilis C, Leischik R, *et al.*, 2016, Epidemiology of coronary heart disease and acute coronary syndrome. *Ann Transl Med*, 4: 256.

    https://doi.org/10.21037/atm.2016.06.33

15. Yang L, Wu H, Jin X, *et al.*, 2020, Study of a cardiovascular disease prediction model based on random forest in Eastern China. *Sci Rep*, 10: 5245.

    https://doi.org/10.1038/s41598-020-62133-5

16. Patel J, TejalUpadhyay D, Patel S, 2015, Techniques. In: International Conference on Smart Homes and Health Telematics. Cham: Springer, p299–306.

17. Abdeldjouad FZ, Brahami M, Matta N, 2020, A Hybrid

Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques. In: International Conference on Smart Homes and Health Telematics, p299–306.

18. Latha CB, Jeeva SC, 2019, Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Inf Med Unlocked*, 16: 100203.

    https://doi.org/10.1016/j.imu.2019.100203

19. Subhadra K, Vikas B, 2019, Neural network-based intelligent system for predicting heart disease. *Int J Innov Technol Exploring Eng*, 8: 484–487.

20. Aliyu A, Ajao LA, Agajo J, *et al.*, 2018, Human vital physiological parameters monitoring: A wireless body area technology based internet of things. *J Teknol Sistem Komput (JTSiskom)*, 6: 115–121.

    https://doi.org/10.14710/jtsiskom.6.3.2018.115-121

21. Waghulde NP, Patil NP, 2014, Genetic neural approach for heart disease prediction. *Int J Adv Comput Res*, 4: 778.

22. Ngabo D, Wang D, Iwendi C, *et al.*, 2021, Blockchain-based security mechanism for the medical data at fog computing architecture of internet of things. *Electronics*, 10: 1–17.

    https://doi.org/10.3390/electronics10172110

23. Salihu BA, Ajao LA, Audu SA, *et al.*, 2020, Design of Robotic Wearable Shoes for Locomotion Assistance System. In: Information and Communication Technology Application (ICTA 2020), Germany: Springer, p689–702.

    https://doi.org/10.1007/978-3-030-69143-1_52

24. Kavitha R, Kannan E, 2014, A methodology for heart disease diagnosis using data mining technique. *Res J App Sci Eng Technol*, 8: 1350–1354.

    https://doi.org/10.19026/RJASET.8.1106

25. Jindal H, Agrawal S, Khera R, *et al.*, 2021, Heart disease prediction using machine learning algorithms. *IOP Conf Ser Mater Sci Eng*, 1022: 012072.

    https://doi.org/10.1088/1757-899X/1022/1/012072

26. Çavuşlar G, Çatay B, Apaydin MS, 2012, A tabu search approach for the NMR protein structure-based assignment problem. *IEEE/ACM Trans on Comput Biol Bioinform*, 9: 1621–1628.

    https://doi.org/10.1109/TCBB.2012.122

27. Ateş A, Yeroglu C, 2016, Optimal fractional order PID design via tabu search based algorithm. *ISA Trans*, 60: 109–118.

    https://doi.org/10.1016/j.isatra.2015.11.015

28. Diwakar M, Tripathi A, Joshi K, *et al.*, 2021, Latest trends on heart disease prediction using machine learning and image fusion. *Mater Today Proc*, 37: 3213–3218.

    https://doi.org/10.1016/j.matpr.2020.09.078

29. Chowdhury MN, Ahmed E, Siddik MA, 2021, Heart disease prognosis using machine learning classification techniques. In: 2021 6th International Conference for Convergence in Technology (I2CT), p1–6.

    https://doi.org/10.1109/I2CT51068.2021.9418181.

30. Peter TJ, Somasundaram K, 2012, Study and development of novel feature selection framework for heart disease prediction. *Int J Sci Res*, 2: 1–7.

31. Srivenkatesh M, 2020, Prediction of cardiovascular disease using machine learning algorithms. *Int J Eng Adv Technol*, 9: 1–7.

    https://doi.org/10.35940/ijeat.B3986.029320

32. Sharma S, Parmar M, 2020, Heart diseases prediction using deep learning neural network model. *Int J Innov Technol Exploring Eng*, 9: 2244–2248.

    https://doi.org/10.35940/ijitee.C9009.019320

33. Mohan S, Thirumalai C, Srivastava G, 2019, Effective heart disease prediction using hybrid machine learning techniques, *IEEE Access*, 7: 81542–81554.

    https://doi.org/10.1109/ACCESS.2019.2923707

34. Chowdary GJ, 2020, Effective prediction of cardiovascular disease using cluster of machine learning algorithms. *J Crit Rev*, 7: 1865–1875.

35. Krittanawong C, Virk HU, Bangalore S, *et al.*, 2020, Machine learning prediction in cardiovascular diseases: A meta-analysis. *Sci Rep*, 10: 16057.

    https://doi.org/10.1038/s41598-020-72685-1

36. Rabbi MF, Uddin MP, Ali MA, *et al.*, 2018, Performance evaluation of data mining classification techniques for heart disease prediction. *Am J Eng Res*, 7: 278–283.

37. Chandralekha M, Shenbagavadivu N, 2018, Performance analysis of various machine learning techniques to predict cardiovascular disease: An empirical study. *Appl Math Inf Sci*, 12: 217–226.

38. Khourdifi Y, Bahaj M, 2019, Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. *Int J Intell Eng Syst*, 12: 242–252.

    https://doi.org/10.22266/ijies2019.0228.24

39. Shah D, Patel S, Bharti SK, 2020, Heart disease prediction using machine learning techniques. *SN Comput Sci*, 1: 1–6.

    https://doi.org/10.1007/s42979-020-00365-y

40. Kaur R, Kaur S, 2015, Prediction of heart disease based on risk factors using genetic SVM classifier. *Int J Adv Res Comput Sci Softw Eng*, 5: 205–208.

41. Haq IU, Chhatwal K, Sanaka K, *et al.*, 2022, Artificial intelligence in cardiovascular medicine: current insights and future prospects. *Vasc Health Risk Manag*, 18: 517–528.

    https://doi.org/10.2147/VHRM.S279337

42. Taylan O, Alkabaa AS, Alqabbaa HS, *et al.*, 2023, Early prediction in classification of cardiovascular diseases with machine learning, neuro-fuzzy and statistical methods. *Biology* (*Basel*), 12: 117.

https://doi.org/10.3390/biology12010117

43. Tran L, Chi L, Bonti A, *et al.*, 2021, Mortality prediction of patients with cardiovascular disease using medical claims data under artificial intelligence architectures: Validation study. *JMIR Med Inform*, 9: e25000.

https://doi.org10.2196/25000

44. Khan A, Qureshi M, Daniyal M, *et al.*, 2023, Novel study on machine learning algorithm-based cardiovascular disease prediction. *Health Soc Care Community*, 2023: 1406060.

https://doi.org/10.1155/2023/1406060

45. Bi H, Lu F, Duan S, *et al.*, 2020, Two-level principal-agent model for schedule risk control of IT outsourcing project based on genetic algorithm. *Eng Appl Artif Intell*, 91: 103584.

https://doi.org/10.1016/j.engappai.2020.103584

46. Ajao LA, Apeh ST, 2023, Secure edge computing vulnerabilities in smart cities sustainability using petri net and genetic algorithm-based reinforcement learning. *Intell Syst Appl*, 18: 200216.

https://doi.org/10.1016/j.iswa.2023.200216