Machine Learning Based Cardiovascular Disease Detection Using Least Absolute Shrinkage and Selection Operator Method

Waqas Ahmad*, Narub Iqbal†, Shairose‡, Muhammad Nadeem§

Abstract

Early detection of cardiovascular disease is paramount as it stands among the most fatal and devastating illnesses globally. Despite extensive research efforts, crucial evaluation parameters like the area under the receiver operating characteristic curve (AUC-ROC), pivotal for diagnostic model assessment, have often been overlooked. AUC-ROC, accuracy, sensitivity, specificity, precision, and other assessment metrics have also been considered in a few research, however their conclusions have not been successful. This paper introduces a model utilizing the Least Absolute Shrinkage and Selection Operator (LASSO) feature selection technique within artificial neural networks (ANN). Alongside ANN, several machine learning classifiers including Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and K Nearest Neighbors (KNN) are employed to compare their efficacy in identifying cardiovascular illness, using a dataset of 70,000 medical records from Kaggle. According to the experimental findings, the suggested approach using ANN obtained accuracy, sensitivity, specificity, precision, and ROC score of 90.39%, 90.92%, 89.97%, 87.79%, and 0.95, respectively. In contrast, KNN, SVM, RF, and DT achieved cardiovascular diagnosis accuracies of 76.47%, 83.40%, 84.98%, and 79.58%, respectively. The findings of the classifiers utilized in this study clearly show that ANN performed better at identifying cardiovascular illness.

Keywords: Machine Learning; Artificial Neural Network; Cardiovascular Disease; Least Absolute Shrinkage and Selection Operator.

Introduction

Cardiovascular disease, commonly known as heart disease, stands as one of the most lethal and impactful health conditions worldwide. (Update, 2017) revealed that heart disease is responsible for the loss of 17.9 million lives globally, representing approximately 31% of all annual deaths. In 2017, the WHO released research stating that 85% of deaths are linked to cardiac disease, encompassing heart attacks and strokes. Traditionally regarded as a predominantly male issue, heart disease has historically received less attention in women's healthcare. However, in

---

*Corresponding Author: Department of Computer Science and Technology, University of Science and Technology Beijing, Beijing 10083, China, waqasahmad9107@gmail.com

†Department of Computer Science, Alhamd Islamic University, Islamabad 44000, Pakistan, narubiqbal.36982@gmail.com

‡Department of Computer Science, Alhamd Islamic University, Islamabad 44000, Pakistan, shairoseshamoon55@gmail.com

§ Department of Computer Science and Technology, University of Science and Technology Beijing, Beijing 10083, China, nadeem72g@gmail.com

1999, the American Heart Association released the first treatment recommendations tailored specifically for women for using cardiovascular disease prevention strategies. It boosted women's awareness of heart disease from 30% to 54% between 1997 and 2009 (Ng et al., 2014).

The landscape surrounding cardiovascular disease (CVD) is intricate and influenced by several interconnected factors closely tied to patients' lives (Kasbe & Pippal, 2018). Consequently, all these factors demand comprehensive consideration for accurate heart disease diagnosis. In its recently released 2020 strategic plan, the American Health Association (AHA) concluded that merely 5% of Americans had "excellent cardiovascular fitness." Unfortunately, an upsurge in death rates due to cardiac illnesses in recent years has been attributed significantly to misdiagnosis. Misdiagnosis, a critical issue, stems from two pivotal factors: the patient's lack of awareness regarding the severity of the illness and the healthcare provider's insufficient understanding of the circumstances. Identifying and addressing these factors become crucial in tackling misdiagnoses within the realm of cardiovascular disease (Jiang et al., 2021).

The development of an artificial neural network with a LASSO optimization-based system is anticipated to provide a more effective solution to the current difficulties that people and medical professionals face when identifying heart sickness (Rani et al., 2018). LASSO is likely chosen for feature selection due to its ability to perform both regularization and feature elimination, making it ideal for reducing overfitting while maintaining interpretability by retaining only the most relevant original features. Unlike PCA, which transforms features into new components that may lose domain relevance, or Recursive Feature Elimination (RFE), which is computationally intensive, LASSO provides a more efficient and interpretable solution, especially for large datasets with potential feature redundancy. This makes it particularly suitable for medical applications like cardiovascular disease detection, where understanding the contribution of individual features is crucial (Ghosh et al., 2021).

This system is poised to significantly elevate the accuracy of diagnoses while concurrently reducing the number of deaths attributed to heart disease. The ultimate goal is to prevent medical professionals from misdiagnosing patients and potentially eradicate the illness through early preventive measures. This study relies on a specific dataset comprising diagnostic information concerning individuals previously diagnosed with potential heart conditions to construct this model. Understanding the various diagnostic stages – training, testing and validation – enables this study to ascertain the prevalence of heart disease among those diagnosed.

Related Work

Numerous research endeavors have delved into the early detection of heart diseases. For instance, a study conducted by (Rippe & Angelopoulos et al., 2019) delved into modeling heart disease using data mining classification methodologies. To refine and assess the proposed model, data sourced from the Cleveland heart disease dataset is utilized. Subsequently, classification and clustering analyses are performed on the dataset under investigation. Interestingly, data mining methods have been found to be less effective and slower than logistic regression. By assessing specificity, sensitivity, and accuracy, the authors assessed under discussion study.

A deep neural network (DNN) model is presented by (Mienye, Sun et al., 2020) to assist patients and healthcare providers in diagnosing cardiovascular disease. DNN training classification and deep neural network prediction are the two subsystems that comprise the DNN model. At the end of the deep neural network training process, deep neural network diagnosis is performed using the final weights obtained after training. The suggested model is evaluated using the AUC-ROC, accuracy, sensitivity, specificity, and precision metrics. For a diagnostic model that can save lives, the published figures of 83.67% accuracy and 72.86% specificity must be raised.

Research is done by (Maini et al., 2021) to create an automated classification model for the diagnosis of heart disease using KNN and RF classifiers. When assessing the suggested model, ROC score, sensitivity, accuracy, and specificity are not considered. While RF achieved 81.967% accuracy, KNN generated 86.88%. For crucial medical diagnostic systems, the accuracy levels produced by RF and KNN need to be increased and strengthened.

Persons with cardiac disorders have been discovered to have a number of dependent and independent variables. Age, sex, blood sugar level, and other independent variables can be useful in the diagnosis of heart disease (Kumar et al. '2021). The experiment's 80% accuracy is below that of a typical diagnostic system, and it required improvement. Additionally, factors like sensitivity, specificity, and AUC-ROC that are essential for assessing the effectiveness of any medical diagnostic model are not considered.

A thorough investigation on the use of contemporary technology to identify cardiovascular illnesses is carried out by (Gharehchopogh & Abdollahzadeh, 2022). They suggested utilizing the trained recurrent fuzzy neural network (RFNN) in conjunction with a genetic algorithm (GA) to identify heart disorders. The suggested model underwent an accurate, sensitive, particular, and precise evaluation. The suggested model's performance when exposed to large-scale datasets may be impacted because it is developed and tested on a small-scale dataset.

Yang et al. (2020) utilize the random forests (RF) and classification and regression tree (CART) models for automatic detection of different cardiovascular diseases. The review of existing work shows the significant advancement in the application of machine learning techniques for cardiovascular disease detection. In assessing the suggested model, accuracy, sensitivity, and specificity are not considered. The ROC score for CART is 0.7025, while the ROC score for RF is 0.7872.

Numerous methods, including neural networks, decision trees, and support vector machines, have demonstrated encouraging outcomes in improving prediction efficiency and diagnostic accuracy. The investigation does, however, highlight shortcomings in addressing model interpretability, including reliable datasets, and creating generalized strategies appropriate for a range of demographics. In order to close these gaps, this study intends to provide a revolutionary machine learning framework that makes use of different machine learning approaches and least absolute shrinkage and selection operator method and extensive dataset, resulting in more precise and scalable cardiovascular disease detection solution.

Proposed Methodology

Firstly, dataset is retrieved from the source. Once the dataset is retrieved then it is explored for data preparation. Following this, feature selection is performed using the LASSO approach. Subsequently the model is trained by using characteristics of the dataset. Finally, the performance of the dataset is assessed.

Dataset Acquisition and Pre-processing

The Kaggle Repository's dataset is obtained for this purpose. The dataset's characteristics are divided into groups for objective, subjective, and test purposes. Objectives are information that is true, subjective aspects are information that the patient provides, and examination characteristics are derived from the results of medical tests. Dataset used in this study is a publicly available dataset. Using this dataset ensures transparency and allows other researchers to validate the results or build upon the work. Figure 1 illustrates the research flowchart.

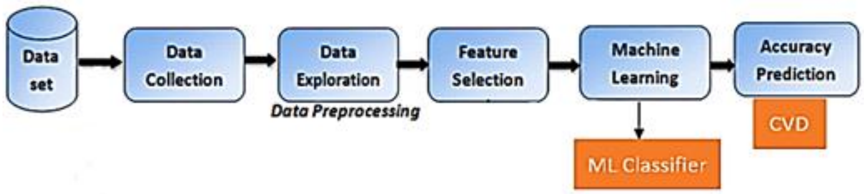


Figure 1: Research Flowchart

The dataset is verified, and it is found that there are no entries with null, missing and duplicate values. The patients' ages are stated as the number of days, which are subsequently changed to years to aid in analysis. The gender column is further divided into "male" and "female" columns. For a female patient, 0 is entered in the male field, and vice versa. The patient's height and weight are used to construct a new column in the dataset called BMI (body mass index). The Average age is 53, the percentage of Male is 35%, Female 65%, Smokers 8%, Alcohol consumers 5%, and Physical activity is 80%. Five BMI outliers are recognized, and individuals with a BMI of 60 or below are removed from the sample. Based on the systolic and diastolic blood pressure readings, the dataset is examined and split into five groups. BP categorical groups are shown in Figure 2.

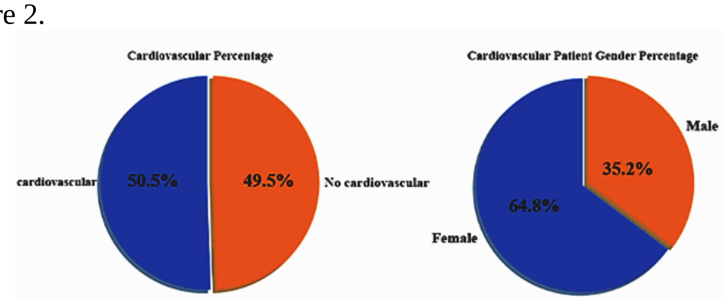


Figure 2: BP Categorical groups

The dataset set's gender-based (males and females) and patient-based (patients with cardiovascular ailment or not) breakdowns are presented below. According to the report, cardiovascular disease affects 65% of women and 35% of men, respectively.

Elder people are more likely to experience cardiovascular disease because of the correlation between age and the illness. Furthermore, there is a link between cardiovascular diseases and BMI, meaning that those with higher BMIs are more susceptible to these ailments.

Additionally, since there is a connection between glucose levels and cardiovascular disease, 60% of persons who have glucose levels that are substantially higher than normal have an increased risk of developing the disease. Because cardiovascular disease and high cholesterol are closely related, having high cholesterol is linked to an increased risk of developing cardiovascular diseases (Baghdadi et al., 2023). It has been discovered that there is no significant increase in the risk of cardiovascular disease from drinking or smoking. Sports and physical activity participation may have a tenuous link to cardiovascular disease, as the disease can also strike inactive individuals. Cardiovascular disease and high blood pressure are strongly correlated, with the development of heart disease being most likely in those with high blood pressure. Table 1 illustrates the dataset composition. Correlation between Age/MBI and cardiovascular disease, and dataset attributes composition with respect to presence of cardiovascular disease are shown in Figure 3 and Figure 4 respectively.

Table 1: Dataset composition w.r.t presence of cardiovascular disease

Blood Pressure Category	SYSTOLIC mm Hg (upper number)	AND/OR	DIASTOLIC mm Hg (Lower Number)
Normal	LESS THAN 120	AND	LESS THAN 70
Elevated	120-129	AND	LESS THAN 80
High Blood Pressure (HYPERTENSION) STAGE 1	130-139	OR	80-89
High Blood Pressure (HYPERTENSION) STAGE 2	140 OR HIGHER	OR	90 or HIGHER
HYPERTENSIVE CRISES (Consult your doctor immediately)	HIGHER THAN 180	AND/OR	HIGHER THAN 120

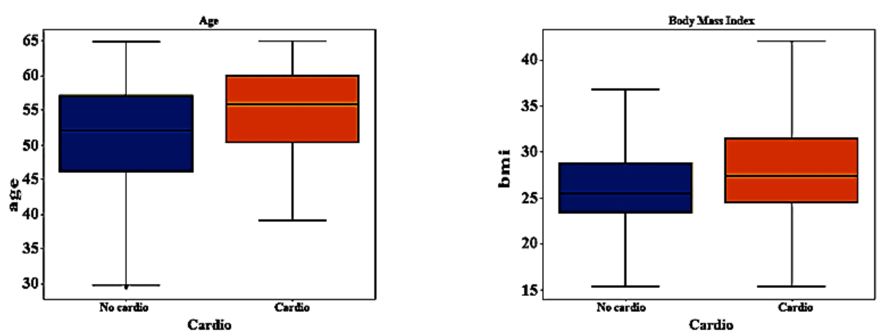


Figure 3: Correlation between Age/MBI and cardiovascular disease

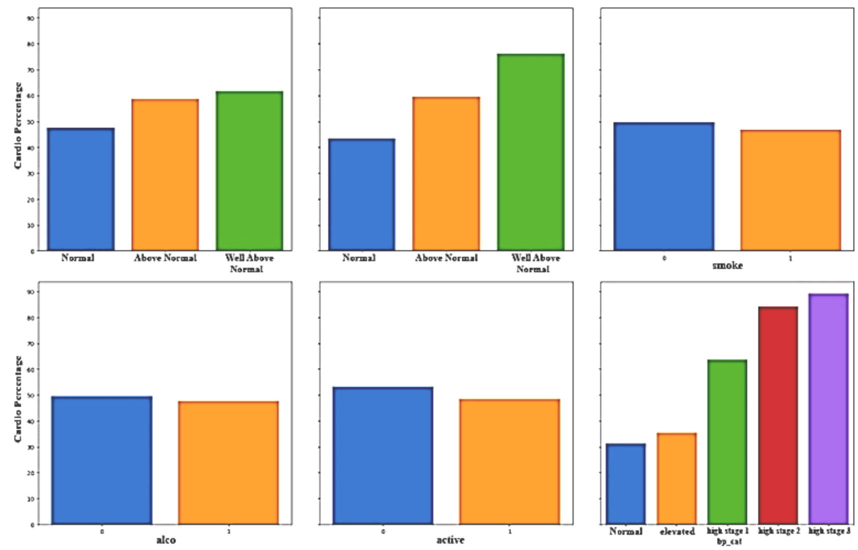


Figure 4: Dataset attributes composition with respect to presence of cardiovascular disease

Machine Learning-Based Cardiovascular Disease Detection

An imitation of a decision-making computer system that can categories data to detect cardiovascular illness is part of the detection model architecture. Such as NN-based expert system uses human indications as inputs and tries to identify the disease using statistical reasoning based on the information given. A traditional expert system functions by employing if-then-else logic rather than procedural code. In the workplace, expert systems are developed using machine learning algorithms rather than if-then-else rules (Walek & Fajmon, 2023). The classification methods used in this study help to identify cardiovascular disease. The cardiovascular disease detection model is based on the prediction model, which is composed of numerous machine learning approaches. These techniques employ a range of classifiers, such as NN, SVM, and KNN.

The dataset consists of 70000 numerical records, each of which has 12 characteristics, including the patient's gender, height, weight, age, systolic blood pressure, glucose levels, alcohol intake, smoking, diastolic blood pressure, physical activity, cholesterol levels, and diagnostic outcome. Data are preprocessed to reduce noise and missing values in order to improve the proposed research's accuracy and dependability. Given that feature selection is essential for the detection of cardiovascular diseases, the LASSO is employed to address this issue. In order to prevent the suggested model from having to work too hard, LASSO chooses the characteristics. Additionally, it aids in preventing the diagnostic model from being over fit. It is crucial to exclude any traits or factors that are unhelpful for detecting cardiovascular illness since here is where LASSO comes into play. The two kinds of the category property are healthy (0) and sick (1). The resulting dataset is subsequently put to use in the creation of the smart cardiovascular diagnostic model's training and testing datasets. The next step in creating a diagnosis model for cardiovascular disease is applying a machine learning algorithm to a cleaned-up dataset. In the suggested study, the ANNBP and LASSO techniques are used to increase the validity and dependability of the research. The suggested model's research approach diagram is shown in the Figure 5.

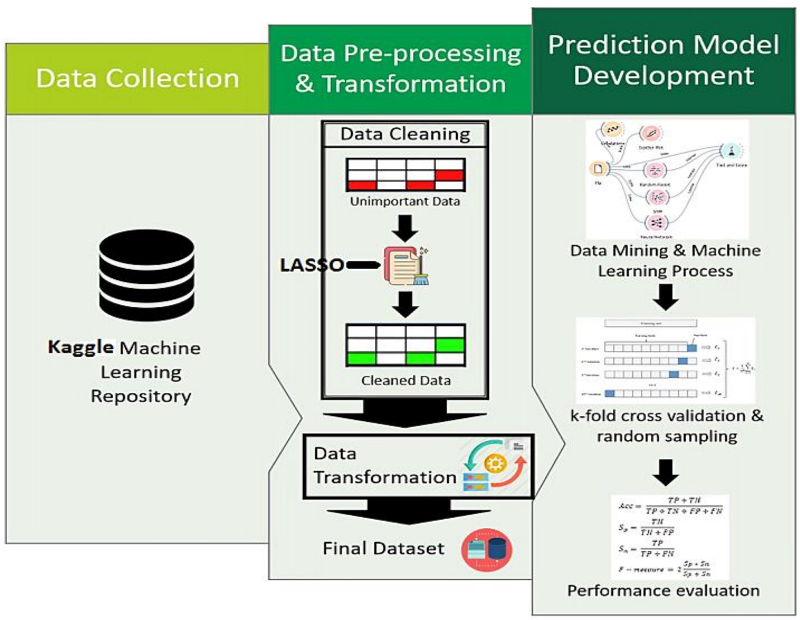


Figure 5: Proposed Methodology for Expert System

Artificial Neural Networks

A mathematical description is given of an expert system that uses neural networks to forecast cardiovascular disease. The computation of bias, feed forwarding of the accumulations, update of weights in back propagation, and initial weight selection for inputs are all covered in the NN-based solution. The activation function of neurons in the buried layer is f(x) = sigmoid (x). Equation 1 presents this function for the input layer, while Equation 2 describes the sigmoid function of the proposed model. The algorithm's learning component is represented by λ_F . Figure 6 shows structure of artificial neural network model.

$$x_{\mathrm{m}} = \mathrm{B}_1 + \sum_{i=l}^{a} (w_{l,m} \times r_l) \quad (1)$$

$$y_m = \frac{1}{1 + e^{-x_m}} \; ; m = 1,2,3, \dots, n \quad (2)$$

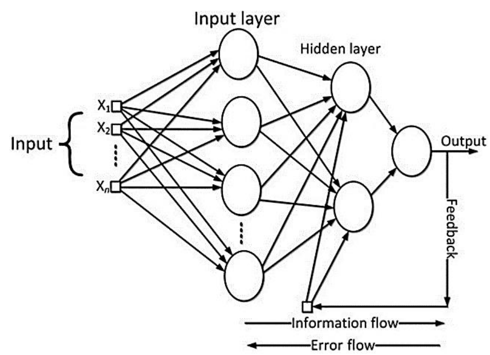


Figure 6: Artificial Neural Network Model Structure

K-Nearest Neighbors (KNN) Classifier

k nearest neighbor's classifier locates the k extra data points in the vector space that are closest to a given data point (Polat, Şahan et al. 2007). Data samples with the same objective value are combined from multiple sources using KNN supervised learning. When one of the classes has to be categorized, it uses a distance metric to give it a new value (Tu, Shin et al. 2009). Euclidean distance is one of the various distance metrics that may be used, the steps involved in KNN are:

1. Both training and testing datasets have been loaded.
2. Choosing the nearest neighboring data points, indicated by the number k.
3. The following is a distance calculation using the Euclidean formula.

$$\sqrt{\langle l_1 - m_1\rangle^{2} + \langle l_2 - m_2\rangle^{2} + \langle l_3 - m_3\rangle^{2} + \cdots + \langle l_z - m_z\rangle^2}$$

$$\sum_{i=1}^{z} \langle l_i - m_i\rangle^2$$

Here, “l” and “m” are instances.

Support Vector Machine (SVM)

The support vector machine modeling technique is used in the classification and regression analyses to interpret the data and pinpoint relationships (Latif et al., 2022). SVM classifies data by choosing the appropriate hyper-plane, which categorizes all data points into two groups. Given the larger difference between the two classes, the model performs better. The performance of the support vector machine is improved by large data sets. Figure 7 shows structure of SVM model.

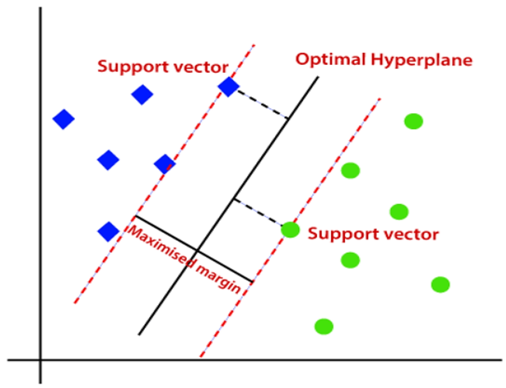


Figure 7: Support Vector Machine Model Structure

Decision Tree (DT)

A type of supervised learning approach that is most frequently employed for resolving classification-related challenges. The population is divided into two or more sets of the same sort using the most significant features as the foundation for the split (Shokrzade et al., 2021). Decision trees are predictive models that use basic binary concepts to predict a target variable's value. They separated a dataset into smaller, more manageable pieces, with similar elements in each. By fitting a sine curve to the data, decision trees are used to determine the rules for regression or classification. The structure of DT model is shown in Figure 8.

Random Forest (RF)

The random forest categorization technique evaluates the average forecast of several separate decision trees. It contains ensemble methods as well as supervised learning (Shah et al., 2020). It increases the model's unpredictability as the trees get larger. Rather of looking for the most important property, when dividing a node, it seeks for the best feature out of a group of randomly picked features. Figure 9 shows structure of RF model.

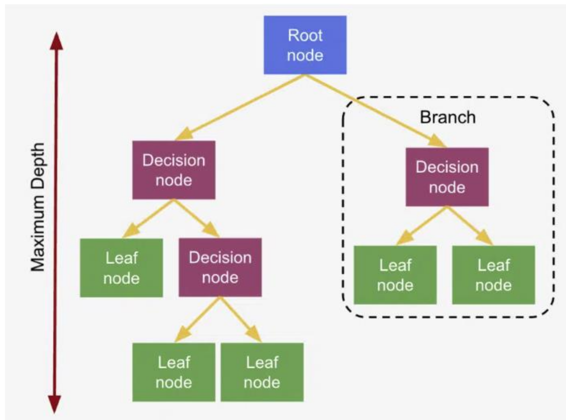


Figure 8. Decision Tree Model Structure

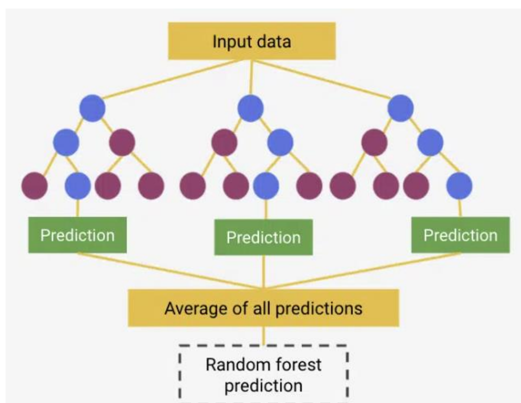


Figure 9. Random Forest **Model Structure**

Results

Confusion Matrix

The confusion matrix, which considers both the training and validation data, can be used to characterize the classifier's performance (Arias-Duart et al., 2023). It also provides details about the test's accuracy, sensitivity, specificity, and precision. It indicates the proportion of instances where the system diagnoses a condition correctly or incorrectly. The confusion matrix for the ANN, DT, RF, SVM, and KNN classifiers are shown in Tables 2 to 6.

Table 2: Confusion matrix of ANN

N=69832		Predicted	
		True $_{Positive}$=**30895**	True $_{Negative}$=**38937**
Actual	Predicted *Positive*	True $_{Positive}$ 25018	False $_{Positive}$ 8413
	Predicted Negative	False $_{Negative}$ 5837	True $_{Negative}$ 30524

Table 3: Confusion matrix of DT

N=69832		Predicted	
		True $_{Positive}$=**30895**	True $_{Negative}$=**38937**
Actual	Predicted *Positive*	True $_{Positive}$ 28089	False $_{Positive}$ 3905
	Predicted Negative	False $_{Negative}$ 2806	True $_{Negative}$ 35032

Table 4: Confusion matrix of RF

N=69832		Predicted	
		True $_{Positive}$=**30895**	True $_{Negative}$=**38937**
Actual	Predicted *Positive*	True $_{Positive}$ 27016	False $_{Positive}$ 6613
	Predicted Negative	False $_{Negative}$ 3879	True $_{Negative}$ 32324

Table 5: Confusion matrix of SVM

N=69832		Predicted	
		True $_{Positive}$=**30895**	True $_{Negative}$=**38937**
Actual	Predicted *Positive*	True $_{Positive}$ 26016	False $_{Positive}$ 6883
	Predicted Negative	False $_{Negative}$ 4679	True $_{Negative}$ 32054

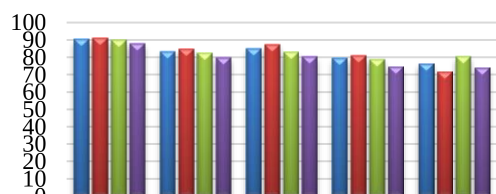
Table 6: Confusion matrix of KNN

N=69832		Predicted	
		True $_{Positive}$=**30895**	True $_{Negative}$=**38937**
Actual	Predicted *Positive*	True $_{Positive}$ 22159	False $_{Positive}$ 7692
	Predicted Negative	False $_{Negative}$ 8736	True $_{Negative}$ 31245

Table 7: Tabular comparison of ANN, SVM, RF, DT, and KNN classifiers

ML Classifier	Accuracy	Sensitivity	Specificity	Precision
ANN	90.39	90.92	89.97	87.79
AVM	83.4	84.76	82.32	79.8
RF	84.98	87.44	83.02	80.34
DT	79.58	81.08	78.79	74.83
KNN	76.47	71.72	80.25	74.23

Results include an analysis of the suggested design's performance using artificial neural networks. The classification accuracy, sensitivity, precision, and specificity measures are also used to assess performance. In addition to these indicators, ROC analysis is assessed to improve the accuracy of performance analysis. The suggested implementation of the neural network algorithm consisted of 13 neurons, with 13 neurons generated from the dataset file for the input layer, 13 neurons using back propagation for the hidden layer, and one neuron for the output layer. Table 7 provides a tabular comparison of the output from the ANN, SVM, RF, DT, and KNN classifiers. Figure 10 provides a graphical comparison of the outcomes provided by the ANN, SVM, RF, DT, and KNN classifiers. With a neural network-based model, the classification module is found to operate with the highest degree of accuracy and precision. ANN's highest accuracy is 90.39% when an automated feature is used for classification. RF achieved 84.98%, SVM 83.40%, DT 79.58% and KNN achieved 76.47% accuracy.



	ANN	SVM	RF	DT	KNN
Accuracy	90.39	83.4	84.98	79.58	76.47
Sensitivity	90.92	84.76	87.44	81.08	71.72
Specificity	89.97	82.32	83.02	78.79	80.25
Precision	87.79	79.8	80.34	74.83	74.23

Accuracy Sensitivity Specificity Precision

Figure 10: Results produced by ANN, SVM, RF

Examining the Algorithm's Sensitivity and Specificity

System's specificity and sensitivity are assessed when employed for disease diagnosis. A much greater specificity of 89.97 % and sensitivity of 90.92 % are achieved using the recommended strategy using ANN. The ROC points show how sensitive the algorithm is to a certain decision criterion. The graph's upper left corner displays the curve that an algorithm with 100% sensitivity and specificity would produce. (Hoo et al., 2017) found that when the ROC curve approaches the left upper corner, the algorithm's accuracy rises as shown in Figure 11.

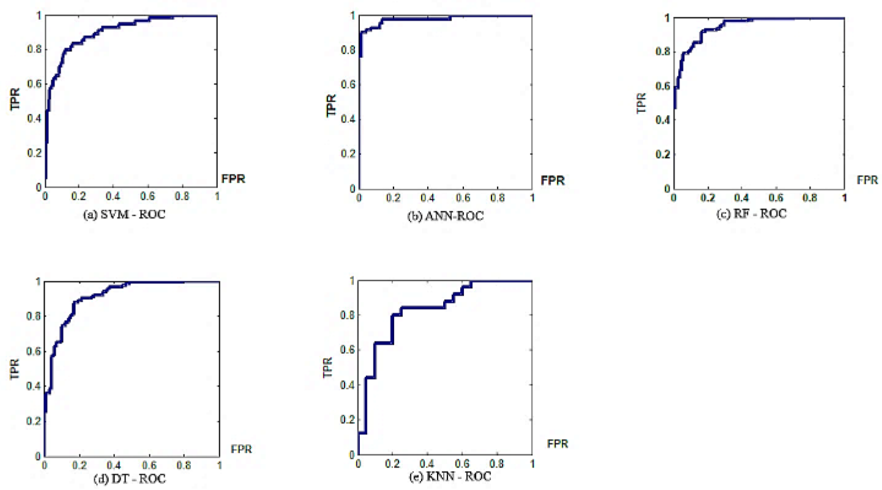


Fig 11. Receiver Operating Characteristic

Conclusion and Future Work

The proposed method facilitates precise and accurate diagnosis of cardiovascular disease and produced better outcomes in terms of sensitivity, specificity, and AUC-ROC score. The study begins with an in-depth introduction to artificial neural networks and an overview of the practical applications of machine learning techniques. It proceeds to a comprehensive review of relevant studies within the field. The third section details the methodologies and datasets employed in this research, while the fourth section comprehensively presents the experimental outcomes and system hardware specifications utilized. The fifth section highlights the importance of further research. Extensive analysis and comparisons are made regarding precision, recall/sensitivity, specificity, and accuracy against the most recent findings in the field. Notably, the proposed model outperformed the rival model in terms of accuracy, sensitivity, as evidenced by the comparative analysis. However, future research could benefit from testing the model on additional datasets, such as those from UCI repositories or real-world hospital data, to further enhance the model's reliability and generalizability across diverse populations and clinical settings. A detailed explanation of the hyperparameter tuning process for the ANN and other models, such as SVM and KNN, would be helpful to enhance the replicability and performance optimization of the study.

References

Arias-Duart, A., Mariotti, E., Garcia-Gasulla, D., & Alonso-Moral, J. M. (2023). A confusion matrix for evaluating feature attribution methods. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3709-3714).

Baghdadi, N. A., Farghaly Abdelaliem, S. M., Malki, A., Gad, I., Ewis, A., & Atlam, E. (2023). Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. *Journal of Big Data*, *10*(1), 144.

Gharehchopogh, F. S., & Abdollahzadeh, B. (2022). An efficient harris hawk optimization algorithm for solving the travelling salesman problem. *Cluster Computing*, *25*(3), 1981-2005.

Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F. J. M., Ignatious, E., ... & De Boer, F. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access*, *9*, 19304-19326.

Hoo, Z. H., Candlish, J., & Teare, D. (2017). What is an ROC curve?. *Emergency Medicine Journal*, *34*(6), 357-359.

Jiang, Y., Zhang, X., Ma, R., Wang, X., Liu, J., Keerman, M., ... & Guo, H. (2021). Cardiovascular disease prediction by machine learning algorithms based on cytokines in Kazakhs of China. *Clinical epidemiology*, 417-428.

Kasbe, T., & Pippal, R. S. (2018). Enhancement in diagnosis of coronary artery disease using fuzzy expert system. *Int. J. Sci. Res. Comput. Sci. Eng. Informat. Technol*, *3*(3), 1324-1331.

Kumar, A., Das, S., Tyagi, V., Shaw, R. N., & Ghosh, A. (2021). Analysis of classifier algorithms to detect anti-money laundering. Computationally intelligent systems and their applications, 143-152.

Latif, G., Ben Brahim, G., Iskandar, D. A., Bashar, A., & Alghazo, J. (2022). Glioma Tumors' classification using deep-neural-network-based features with SVM classifier. Diagnostics, 12(4), 1018.

Maini, E., Venkateswarlu, B., Maini, B., & Marwaha, D. (2021). Machine learning–based heart disease prediction system for Indian population: An exploratory study done in South India. medical journal armed forces india, 77(3), 302-311.

Mienye, I. D., Sun, Y., & Wang, Z. (2020). Improved sparse autoencoder based artificial neural network approach for prediction of heart disease. Informatics in Medicine Unlocked, 18, 100307.

Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., Margono, C., ... & Gakidou, E. (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. The lancet, 384(9945), 766-781.

Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. Diagnostics, 14(2), 144.

Polat, K., Şahan, S., & Güneş, S. (2007). Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour) based weighting preprocessing. Expert Systems with Applications, 32(2), 625-631.

Rani, K. S., Chaitanya, M. S., & Kiran, G. S. (2018). A Heart Disease Prediction Model Using Logistic Regression By Cleveland DataBase. International Journal of Trend in Scientific Research and Development, 2(3), 1467-1470.

Rippe, J. M., & Angelopoulos, T. J. (2019). Lifestyle strategies for risk factor reduction, prevention and treatment of cardiovascular disease. Lifestyle Medicine, Third Edition, 19-36.

Shah, D., et al. (2020). "Heart disease prediction using machine learning techniques." SN Computer Science 1(6): 345.

Shokrzade, A., Ramezani, M., Tab, F. A., & Mohammad, M. A. (2021). A novel extreme learning machine based kNN classification method for dealing with big data. Expert Systems with Applications, 183, 115293.

Tu, M. C., Shin, D., & Shin, D. (2009, October). Effective diagnosis of heart disease through bagging approach. In *2009 2nd international conference on biomedical engineering and informatics* (pp. 1-4). IEEE.

Update, A. S. (2017). "Heart disease and stroke statistics–2017 update." Circulation 135: e146-e603.

Walek, B. and P. Fajmon (2023). "A hybrid recommender system for an online store using a fuzzy expert system." Expert Systems with Applications 212: 118565.

Yang, L., Wu, H., Jin, X., Zheng, P., Hu, S., Xu, X., ... & Yan, J. (2020). Study of cardiovascular disease prediction model based on random forest in eastern China. *Scientific reports*, *10*(1), 5245.