Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Heart Disease Prediction of Cleveland Clinic Patients using Advanced Machine Learning Algorithms

Lubna A. Alharbi[1,*]

[1] Department of Computer Science, University of Tabuk, Tabuk 47512, Saudi Arabia

ARTICLE INFO

Article history:
Received 22 March 2024
Received in revised form 17 December 2024
Accepted 14 March 2025
Available online 28 March 2025

Keywords:
Heart disease patient; prediction; machine learning algorithms; XGBoost

ABSTRACT

Globally, cardiovascular diseases (CVDs) constitute the primary cause of morbidity and mortality worldwide. Early diagnosis of those at risk of CVDs may lower the number of avoidable fatalities. It has been shown that machine learning (ML) is helpful in anticipating cardiac issues. Adoption of a prediction system that can detect cardiac diseases before they deteriorate would offer people worldwide enormous hope and help in decision-making. ML has become a popular technique for generating predictions from enormous real-world datasets. It has also been discovered that many ML classifiers contain issues and flaws. However, the latest ML algorithm from the boosting family, i.e., XGBoost, may enhance performance and assist in exact prediction. As a result, this study will compare XGBoost to other prominent classifiers in terms of their capacity to anticipate and improve performance. ML classifiers, such as Multilayer Perception, K-nearest neighbours (K-NN), Support Vector Classifier, CART and XGBoost algorithms, are used to differentiate between healthy and CVD patients. When compared to competing classifiers, the XGB classifier achieves 89% accuracy, 87% precision, 94% sensitivity, 94% specificity, 90.2% F1 score, 81.2 % ROC, 78.90% Mathew coefficient and 3.87% log loss. In the future other ML classifiers such as Random Forest, Multilayer Perceptron, K-nearest Neighbor, Extra Tree Classifier, Extreme Gradient Boosting, Support Vector classifier, Stochastic Gradient Descent, AdaBoost, Classification and Regression Tree and Gradient Boosting along with these algorithms could be applied to compare the MLA efficiency.

1. Introduction

Over one-third of all yearly fatalities worldwide are caused by cardiovascular disease, which is mostly caused by heart disease. According to a World Health Organization (WHO) estimate, 17.9 million deaths worldwide in 2019 were attributable to heart disease (CVDs) [1]. This accounts for 32% of all fatalities worldwide and having a death rate that exceeds 17.7 million per year [2]. It is projected that by 2030, there would be 22 million deaths worldwide if nothing is done.

A heart attack or stroke may be caused by plaques on the artery walls that impede blood flow. Numerous risk factors, including a poor diet, inactivity and heavy alcohol and tobacco use, contribute

* *Corresponding author*
E-mail address: lualharbi@ut.edu.sa

https://doi.org/10.37934/ard.126.1.114

to heart disease [3,4]. A lifestyle choice that includes eating less salt, eating more fruits and vegetables, exercising regularly, giving up alcohol and tobacco and cutting down on smoking and other unhealthy habits all help to lower the risk of heart disease [5].

Predicting the incidence of several contributing risk factors, such as diabetes, high blood pressure, excessive cholesterol, irregular pulse rate and other variables, has become quite difficult in recent times. Many machine learning (ML) techniques have been used to forecast the severity of cardiovascular disease in the general population [12].

Many methods, including Multilayer Perceptron, k-Nearest Neighbor (k-NN), Support Vector Machines (SVM) and Decision Tree (DT), are used to classify the severity [22]. Because the severity of the disease is so complex, it requires more careful treatment. Although the condition has no symptoms, it may sometimes cause unexpected death. For the purpose of forecasting various metabolic disorders, the broader research views of ML algorithms and medical science are used [21].

It is very difficult to diagnose and treat heart disease when cutting edge technology and medical professionals are not accessible. A sound diagnosis and course of therapy may save the lives of a great number of individuals [13]. Heart disorders are diagnosed by a doctor based on an assessment of the patient's medical history, the results of the physical examination and an analysis of any worrisome symptoms. Nevertheless, the results of this diagnostic approach are not enough to identify people with cardiac disease [14].

Moreover, its analysis is computationally demanding and expensive. An expert judgement system based on artificial fuzzy logic and ML classifiers effectively diagnoses heart disorders. Consequently, the death ratio experiences a decrease [6,7]. The Cleveland heart disease dataset has been utilised by many studies. The ML prediction models need the right data for testing and training. ML classifier accuracy may be increased using a revised dataset for testing and training.

Many researchers have used a combination of ML approaches such as KNN, MLP, XGBoost, SVC, CART separately as well as combined as shown in Table 1. However, using all the resources none of them have adopted all ML approaches on single dataset. In this work, different ML approaches have been implemented on Cleveland clinic dataset. The ultimate research objective is to check the diversity of ML approaches as well as prediction accuracy.

An extensive series of trials were conducted in this research endeavour to establish a novel method for predicting heart illness that is exceptionally capable of distinguishing it from other ailments. The subsequent sections are structured as follows: Section 2 provides a comprehensive examination of the existing literature about the prediction of heart disease. Section 3 delineates the study approach. Discussions and outcomes comprise Section 4. The conclusion, which includes future research directions, is Section 5.

2. Related Work

According to Srivastava *et al.,* [10], ML described can be described as "Computer virus that learn from previous experience and from a few other functions, as measured, improves knowledge". Once ML algorithms have established a link, the model might utilise that relationship to forecast future events or to generate intriguing patterns.

Arthur Samuel created ML in 1959 and it changed our thoughts. ML is the study and creation of algorithms that allow us to make predictions based on data and learn from user input. Process statistics, which also focuses on the generation of computerised estimates, is most closely associated with machine learning [8]. Mathematical efficiency and cubic centimetre have a close link that highlights the approaches, theories and backgrounds used in the area. When training and testing of data is a major emphasis of the subterranean storage area, ML is often integrated with data processing.

Within the domain of structured data, ML is a sub-method that is used for intricate models and algorithms aimed at anticipating statistics. Through the use of historical relationships and data analysis, this analytical methodology has helped researchers, data engineers, data scientists and data analysts generate high-quality repeatable alternatives and outcomes as well as uncover hidden patterns [9].

Heart disease prediction has been one of the main uses of ML in recent years and it has shown some effectiveness with several methods. Table 1 describes the many ML techniques that researchers have used for the prediction of heart disease.

Table 1
Existing studies

Study	Dataset	Method	Accuracy	Limitations
Srivastava *et al.,* [10]	UCI	KNN	85	Website was not fully functional
Yang *et al.,* [11]	Heart Disease Dataset	XGBoost KNN	93.44 91.77	Sensitivity, Mathew and loss was not evaluated
Doki *et al.,* [12]	Cleveland	XGBoost	85.96	AUC and Accuracy was calculated only
Saboor *et al.,* [13]	Cleveland	CART XGB	83.66 91.80	Sensitivity, Mathew and loss was not evaluated
Nagavelli *et al.,* [14]	Heart Disease Dataset	XGBoost	95.90	Sensitivity, Mathew and loss was not evaluated
Garg *et al.,* [15]	Heart Disease Dataset	KNN Random Forest	86.88 81.96	Other algorithms are not tested
Kriplani *et al.,* [16]	Heart Disease Dataset	KNN XGB	80.85 91.90	Sensitivity, Mathew and loss was not evaluated
Rindhe *et al.,* [17]	UCI	SVC	84	Sensitivity, Mathew and loss was not evaluated
Kaushik *et al.,* [18]	Institute of Cardiology	XGBoost	88	Other algorithms are not tested
Jani *et al.,* [19]	Cleveland	KNN	75	Sensitivity, Mathew and loss was not evaluated
Rahman [20]	Heart Disease Dataset	KNN XGBoost	89 95	Mathew and log loss was not used to check the loss
Boukhatem *et al.,* [21]	Heart Disease Dataset	MLP	81.67	Mathew and log loss was not used to check the loss

Srivastava *et al.,* [10] used KNN approach on UCI dataset, their approach was 85% accurate as well as the website developed by them was not functional for all the stakeholders. Whereas Yang *et al.,* [11] adopted feature extraction method for prediction of heart diseases using XGBoost and KNN approaches. The accuracy of the heart disease dataset was 93.44% and 91.77%. Whereas Doki *et al.,* [12] implemented XGBoost on Cleveland dataset and it was found that the accuracy of this ML approach was able to predict 86%. Saboor *et al.,* [13] used CART and XGB models on Cleveland datasets and their models performed up to 92%.

Also, Jani *et al.,* [19] used Cleveland dataset and adopted KNN which produced 75% accuracy, whereas Nagavelli *et al.,* [14] adopted XGB only on HDD and found 96% accurate as well as Garg *et al.,* [15] and Kriplani *et al.,* [16] adopted KNN-RF and KNN-XGB respectively on HDD and found similarity in KNN accuracy, but the XGB found more accurate than RF. In addition to performance of ML models, a study was conducted by Rindhe *et al.,* [17] and in that he used SVC on UCI dataset which produced 84% accurate in performance. Another study was conducted by Kaushik *et al.,* [18] on dataset developed by institute of cardiology using XGB algorithm. The performance on XGB algorithms turns into 88% accurate. Rahman [20] used HDD and adopted KNN and XGB dataset and it produced more accurate than Kaushik *et al.,* [18] output. Lastly, Boukhatem *et al.,* [21] used MLP on HDD and found 82% accurate. Also, few studies elaborate on practices for safety [23-26].

Based on the existing techniques used in Table 1 for classification of heart disease, some of them are limitations such as use of one or more algorithms, while others have training and testing ratio as well as few studies have limitations in terms of dataset (no EDA). To overcome the barriers or limitation faced by the previous studies, this research has proposed a study that can classify heart disease using five different ML algorithms using EDA on Cleveland Clinic Patients dataset.

3. Methodology

Methodology comprises of seven stages for prediction of heart disease from the selected data using ML algorithms. The steps are self-explanatory on its own as they are described through pictures and equations. The seven stages of the methodology are shown in Figure 1.

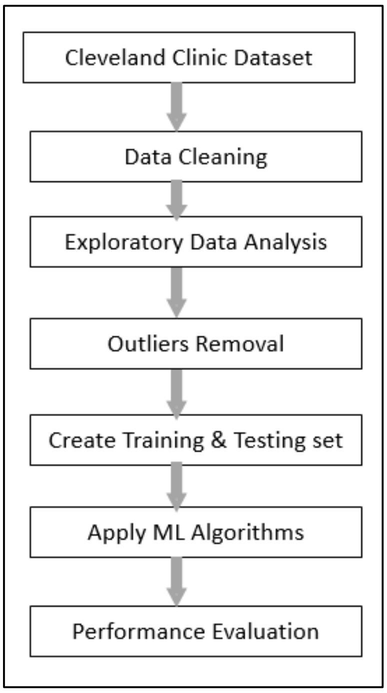

Fig. 1. Methodology adopted

3.1 Dataset Description

Cleveland Clinic dataset is selected for experiments as it was also used by few studies in past but using different ML algorithms [12,18]. 1189 values totalling data on both healthy and cardiac disease patients make up the dataset. Eleven characteristics and a target variable make up the chosen dataset. Table 2 lists six nominal variables and five numerical variables from it.

3.2 Data Cleaning and Pre-processing

After selecting the dataset, the second stage of methodology is to clean and balance the dataset. The initial step in this process is to rename the columns in accordance with accepted naming practices. A few of the columns have unique naming conventions. Furthermore, the characteristics are classified into categorical factors like normal or cardiac disease.

Table 2
Dataset description

Sex	Gender (Male - 1, Female - 0)
Age	Age (Numbers)
Resting BP	BP level measured in mm/HG during rest mode
Cholesterol	Serum cholesterol in mg/dl
Fasting blood sugar	Fasting Blood sugar are measured with values more than > 120 mg/dl and if it is selected than it means True and if not then False
Resting ECG	Values for electrocardiogram during rest time are 0 : Normal, 1: Abnormality in ST-T
Wave 2	Left ventricular hypertrophy
Max heart rate	Heart rate at high level
Exercise angina	Angina brought by exercise if it is 0 which shows NO and 1 shows Yes
Old peak	Exercise brought by ST-depression compared to the resting condition
ST slope	ST segment measured in 4 categories, if it is slope during peak exercise 0: Normal if not than 1: Upsloping and 2 means Flat and lastly 3: Downsloping
Target	The target variable is one that we must predict: 1 indicates that the patient is at risk for cardiac problems and 0 indicates that the patient is normal.
Chest Pain Type	Type of chest pain experienced by patient categorized into 1 typical, 2 typical angina, 3 non- anginal pain, 4 asymptomatic (Nominal)

3.3 Exploratory Data Analysis

Once the data in dataset is normalized in normal heart disease patient, the third stage of methodology comes into play. This step seeks to distribute the attributes of the chosen dataset in several ways, including the distribution of heart disease, the distribution of gender and age, the distribution of the types of chest pain, the distribution of resting electrocardiograms and the distribution of numerical variables:

i. Distribution of Heart disease (target variable): Figure 2 illustrates how the dataset is balanced in terms of heart disease distribution, with 628 heart disease patients and 561 normal people.

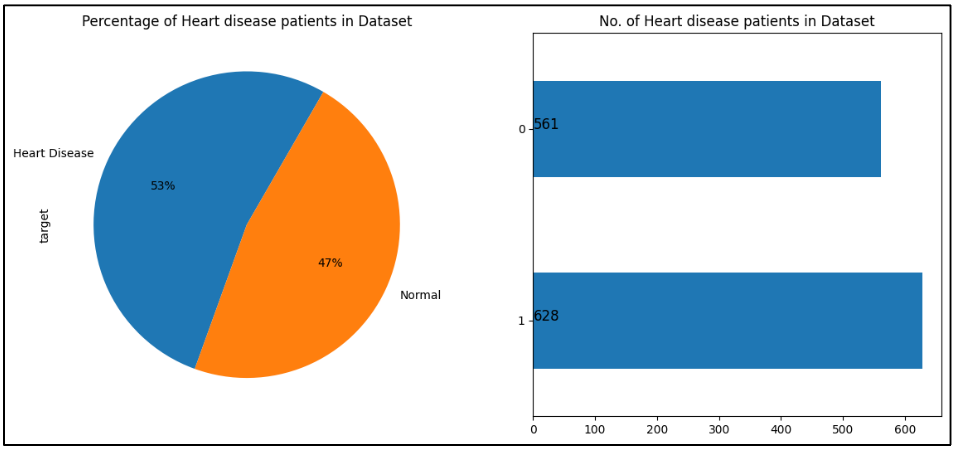

Fig. 2. Dataset distributions in terms of normal and patient

ii. Checking Gender and Age-wise Distribution: In gender-wise distribution, the dataset is having 76% male and 24% female ratio as shown in Figure 3.

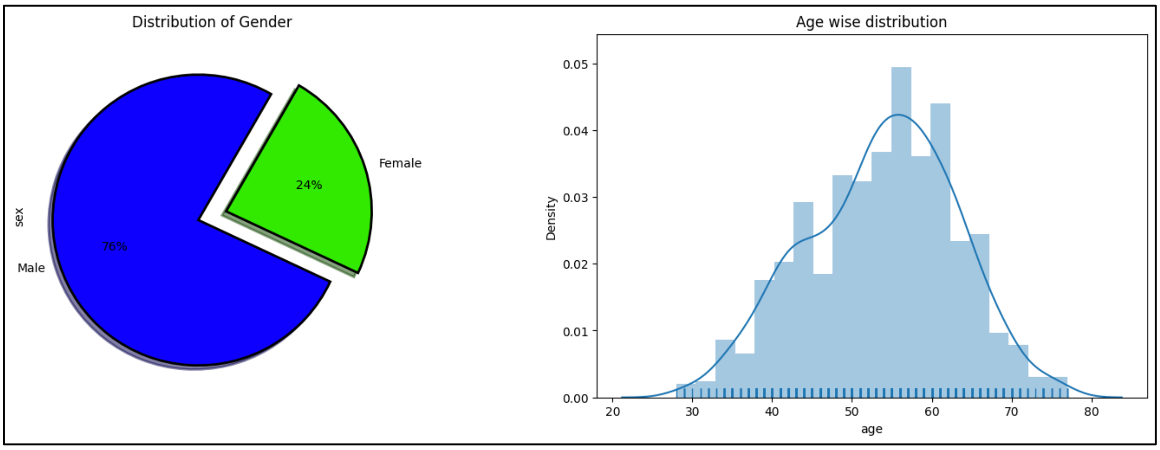

Fig. 3. Dataset distributions in terms of patient's gender

Whereas the age-wise distribution of 350 male and 200 female are part as a normal patient as shown in Figure 4.

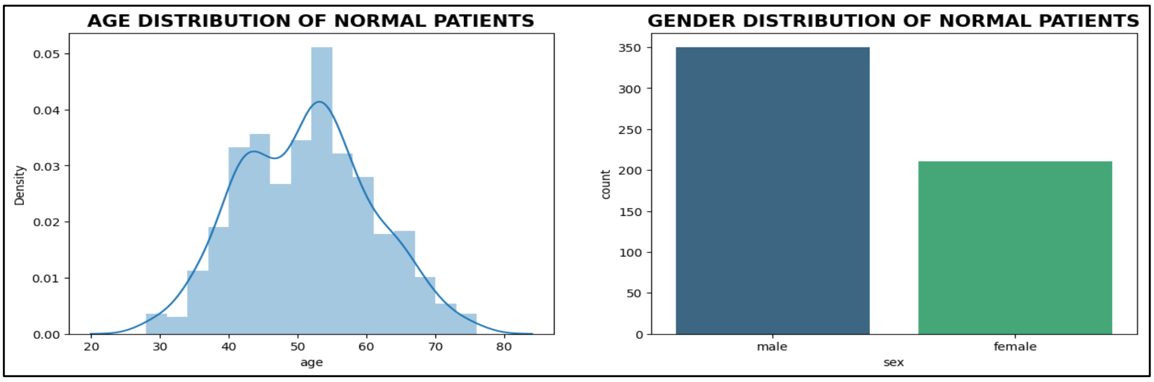

Fig. 4. Age distributions of normal people

As well as heart disease patient are 550 male patient and 50 female patients as shown in Figure 5. The graphic above illustrates how, despite the average patient age of 55, the proportion of men in this collection is much greater than that of women.

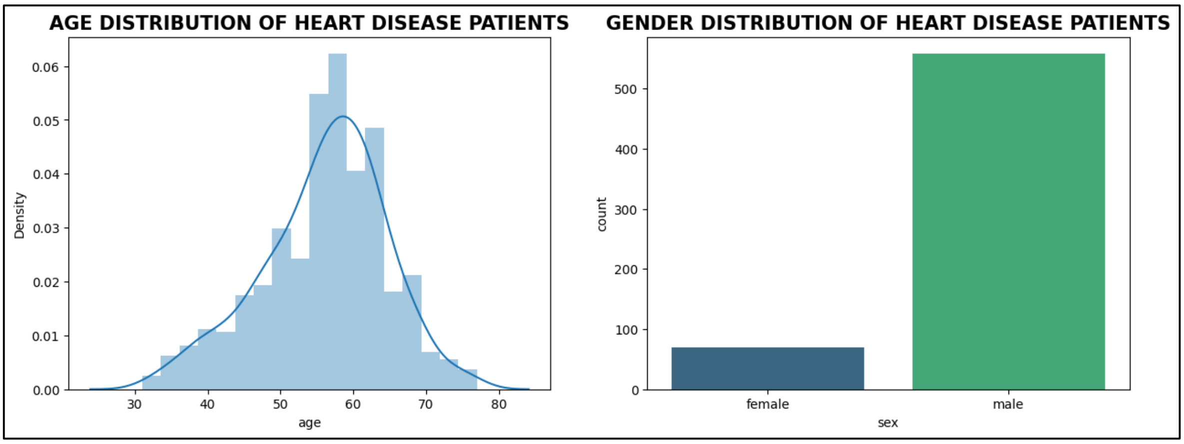

Fig. 5. Age distribution of heart patients

The above figure shows that there are more male patients with heart disease than female patients and that the average age of heart disease patients is between 58 and 60 years old.

iii. Distribution of Chest Pain Type: Typical angina, atypical angina, non-anginal pain and asymptomatic type are all part of the chest pain distribution in the dataset, which includes both normal and heart disease individuals. Figure 6 shows that 76% of people with chest discomfort as a sign of heart disease really do not experience any symptoms at all.

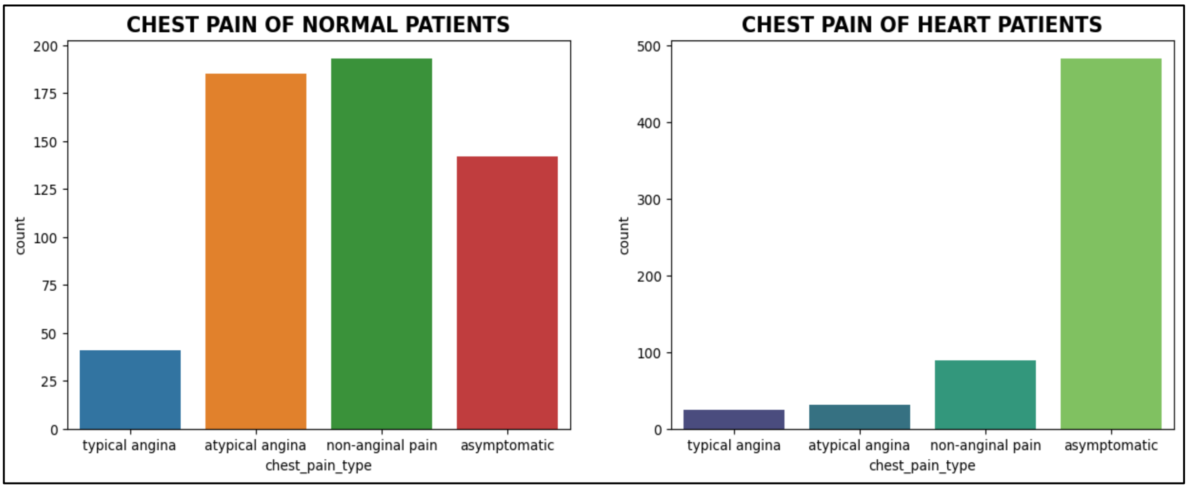

Fig. 6. Distribution of chest pain type in normal people

iv. Distribution of Rest ECG: The heart's electrical impulses may be captured by an electrocardiogram. In many cases, it's the first line of defence against heart issues and a standard tool for monitoring cardiac health. Electrocardiograms – also termed ECGs. It detects the heart's rhythm and pace but cannot diagnose artery blockages. For that reason, as seen in Figure 7, around 52% of patients with cardiac disease in our sample had normal electrocardiograms.

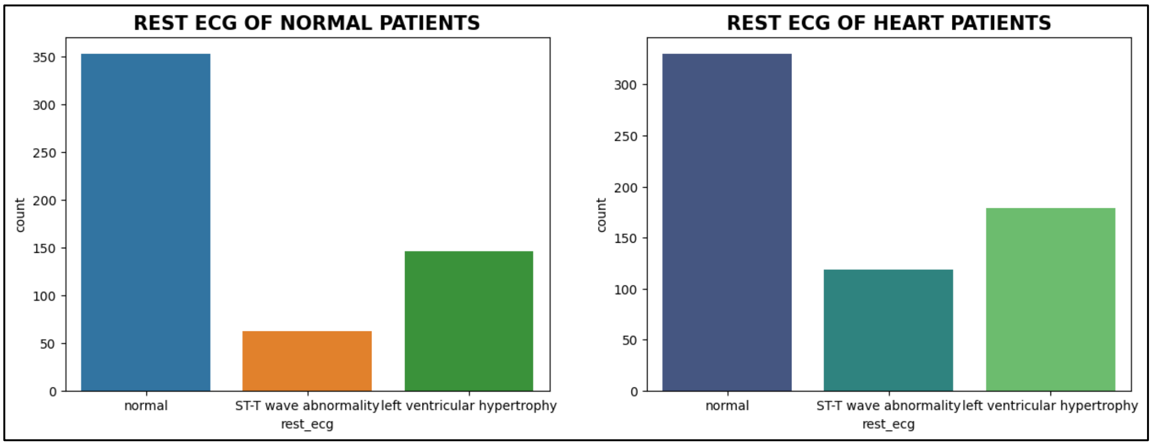

Fig. 7. Distribution of ECG in normal people

v. Distribution of Numerical features: From the below plot of numerical feature of heart disease patients, As can be shown in Figure 8, the likelihood of heart disease rises in direct correlation with age. It also illustrated the distribution of cholesterol, resting blood pressure and age. The distribution shows that the yellow colour means more critical stage of CVD, cholesterol which is totally dependent on age.

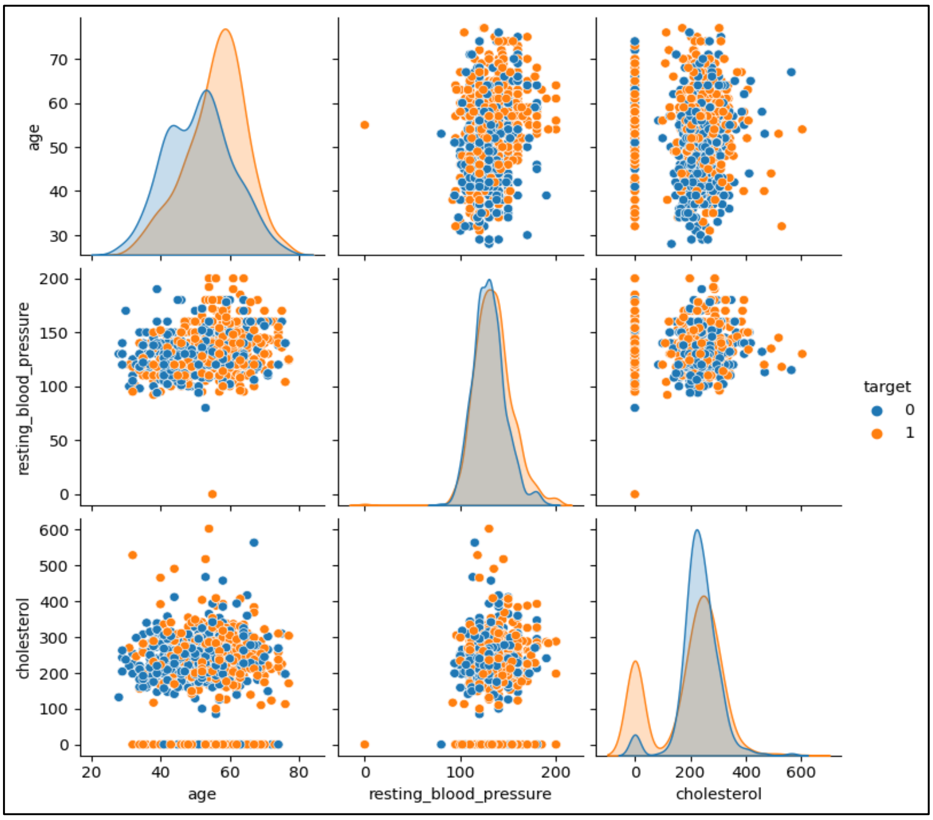

Fig. 8. Distribution of numerical features in normal and patients

3.4 Outliers

It is evident from the above EDA map that there are several outliers; for example, some patients have zero cholesterol and one patient has zero resting blood pressure; this might be because some records in the dataset are missing. In this stage of the methodology, the selected dataset is processed for removing anomalies using z-score shown in Figure 9.

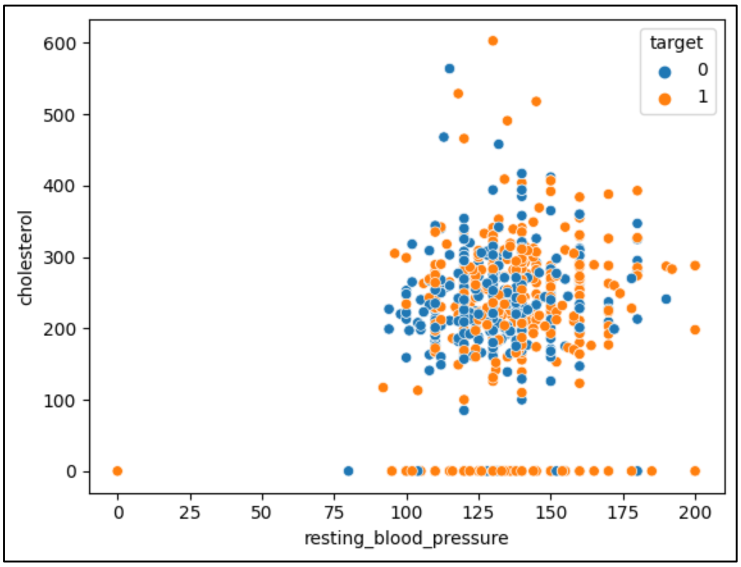

Fig. 9. Visualization of outliers in dataset

3.5 Training and Testing

After anomalies removal, the dataset is now in balanced format which can be used for training and testing of ML algorithms. The chosen dataset is partitioned into two distinct phases, namely, training and testing. Twenty percent of the data is assigned for testing purposes and eighty percent is used for training. ML algorithms selected for heart disease prediction on selected dataset are MLP, KNN, SVC, CART and XGBoost.

3.6 Model Evaluation

During this final phase of the methodology, the performance of the ML algorithm is assessed through the utilization of various performance measurement techniques, including but not limited to Accuracy, Precision, Sensitivity, Specificity, F1-measure, ROC AUC curve, Log loss and Matthew correlation coefficient as shown in Eqs. (1) – (6):

$$\text{Accuaracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \tag{1}$$

$$\text{Precision } \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

$$\text{F1} = \frac{\text{2PR}}{\text{P+R}} \tag{3}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{4}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{5}$$

$$\text{MCC} = \frac{\text{TP.TN}-\text{FP.FN}}{\sqrt{\text{(TP+FP)(TP+FN).(TN+FP).(TN+FN)}}} \tag{6}$$

4. Results and Discussion

As mentioned before, the purpose of this research was to forecast cardiac disease by applying several ML algorithms to a specific dataset. By carrying out each of the procedures outlined in Figure 1. This section aims to prove that the addition of XGB algorithm along with other algorithms will help in better prediction rate. To begin with this section, it illustrates the results of different ML algorithms and comparison with existing studies. In addition to the result, this study aims to highlight on performance metrics such as log loss and Mathew correlation co-efficient as many studies are not calculating the loss as well as correlation factor which is important for healthcare application and prediction models.

4.1 Machine Learning Algorithms

ML algorithms adopted in this study are: Multilayer Perceptron, K-nearest Neighbor, Support Vector classifier, Decision Tree Classifier and Extreme Gradient Boosting. The performance of each algorithm is described below.

4.1.1 Multilayer perception

Figure 10 illustrates the performance of Multilayer perceptron on Cleveland clinic dataset is promising. MLP scored 82% Accuracy, 79% Precision, 89% Sensitivity, 74% Specificity, 84% F1 score, 82% ROC, 6.901 Log loss and 0.618 Mathew Co-relation Coefficient.

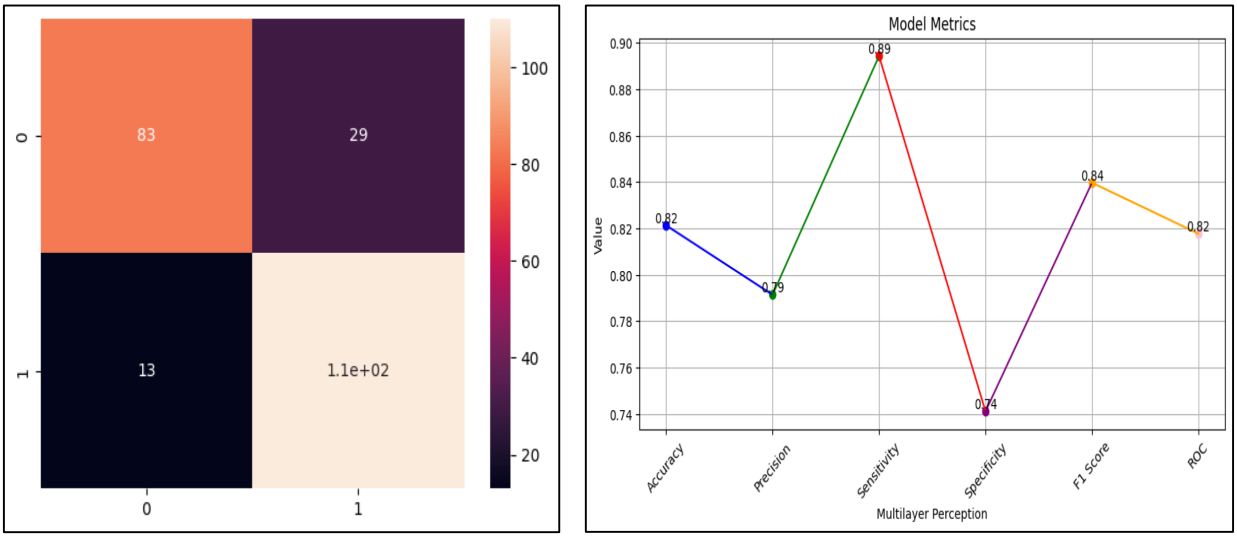

Fig. 10. Confusion matrix and model metrices plot

4.1.2 K-Nearest neighbor

Figure 11 illustrates the performance of K- Nearest Neighbor on Cleveland clinic dataset is encouraging. KNN scored 81% Accuracy, 79% Precision, 87% Sensitivity, 74% Specificity, 83% F1 score, 81% ROC, 6.441 Log loss and 0.645 Mathew Co-relation Coefficient. There are slightly changes in performance as compared to MLP. Accuracy, sensitivity, F1 and ROC score reduced but the log loss and Mathew correlation performance increased.

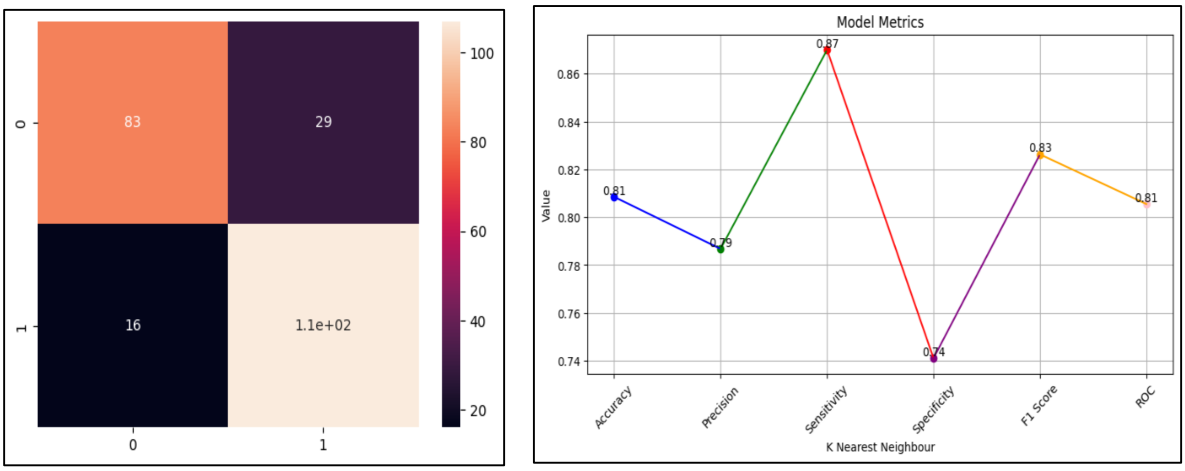

Fig. 11. Confusion matrix and model metrices plot

4.1.3 Support vector classifier

Figure 12 illustrates the performance of support- vector classifier on Cleveland clinic dataset is promoting. KNN scored 83% Accuracy, 80% Precision, 89% Sensitivity, 76% Specificity, 84% F1 score, 82% ROC, 6.288 Log loss and 0.652 Mathew Co-relation Coefficient. There are slightly changes in performance as compared to KNN. Accuracy, sensitivity, F1 and ROC, log loss and Mathew correlation performance increased.

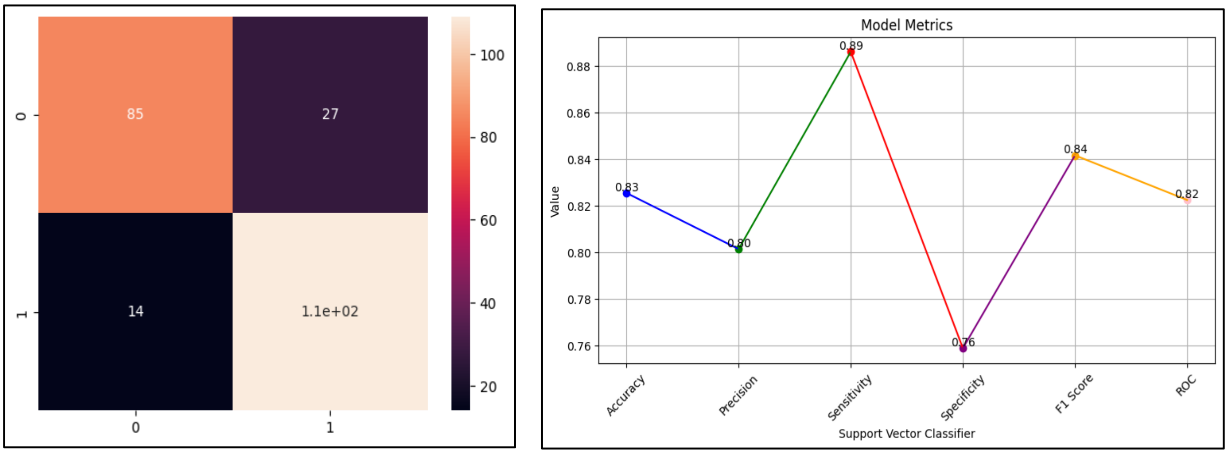

Fig. 12. Confusion matrix and model metrices plot

4.1.4 Decision tree classifier (CART)

Figure 13 illustrates the performance of Decision Tree Classifier (CART) on Cleveland clinic dataset is encouraging. CART scored 84% Accuracy, 83% Precision, 87% Sensitivity, 80% Specificity, 85% F1 score, 82% ROC, 5.828 Log loss and 0.675 Mathew Co-relation Coefficient. There are slightly changes in performance as compared to SVC. Accuracy, F1 and ROC, log loss and Mathew correlation performance increased whereas sensitivity decreased by 2%. Also, it can be noted that the log loss score jumps to best score till now.

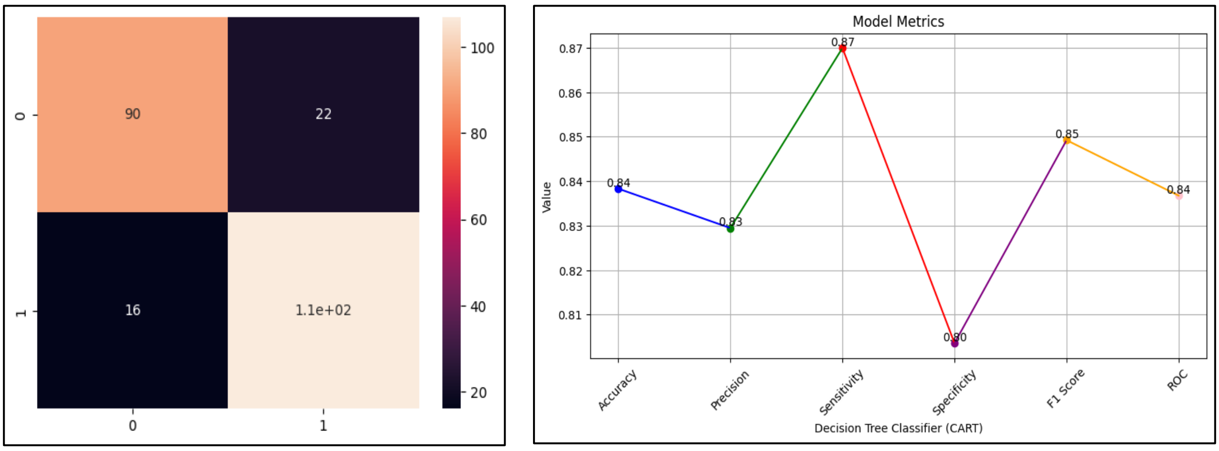

Fig. 13. Confusion matrix and model metrices plot

4.1.5 Extreme gradient boosting

Figure 14 illustrates the performance of Extreme Gradient Boosting (XGB) on Cleveland clinic dataset is encouraging. XGB scored 89% Accuracy, 87% Precision, 94% Sensitivity, 84% Specificity, 90% F1 score, 89% ROC, 3.834 Log loss and 0.789 Mathew Co-relation Coefficient. There are marginal changes in performance as compared to CART. Accuracy, precision, Sensitivity, F1 and ROC, log loss and Mathew correlation performance increased. Also, it can be noted that the log loss score jumps to best score now as compared to CART.

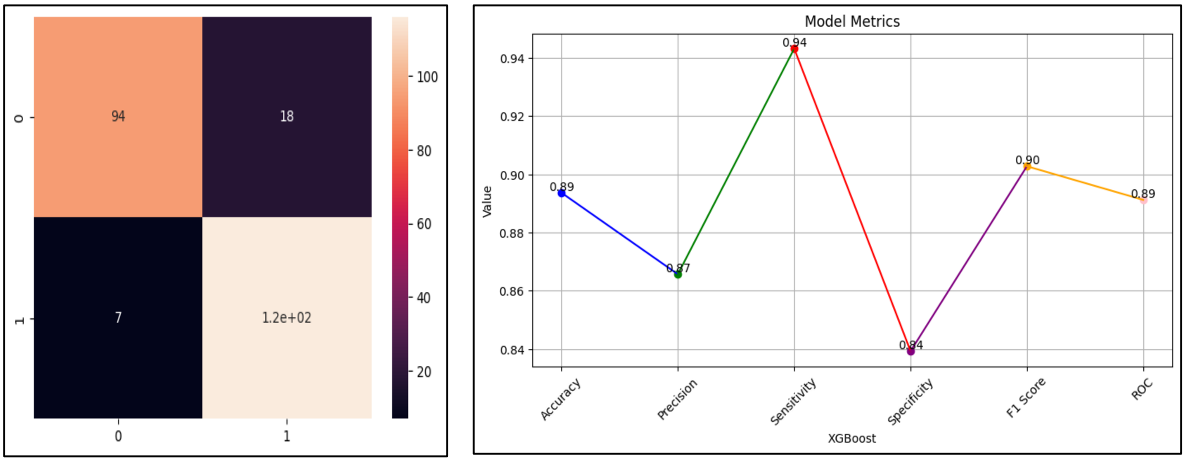

Fig. 14. Confusion matrix and model metrices plot

The overall performance of different ML algorithms on Cleveland datasets is shown in Table 3. It can be observed that the performance of XGB is better than among all participating algorithms. The main factors which need to be more emphasized are *Log_Loss* and *Mathew correlation co-efficient* values.

Table 3
Performance measurement of ML algorithms for CVD disease detection

Sr.	Model	Accuracy	Precision	Sensitivity	Specificity	F1 Score	ROC	Log_Loss	mathew_corrcoef
1	KNN	0.808	0.786	0.869	0.741	0.826	0.805	6.901	0.618
2	MLP	0.821	0.791	0.894	0.741	0.839	0.817	6.441	0.645
3	SVC	0.825	0.801	0.886	0.758	0.841	0.822	6.288	0.652
4	CART	0.838	0.829	0.869	0.803	0.849	0.836	5.828	0.675
5	XGB	0.893	0.865	0.943	0.839	0.902	0.891	3.834	0.789

4.2 Comparison of XGB Performance on Cleveland Dataset

Following an evaluation of the performance of certain ML models, this section examines the performance of the amount of prior research on the Cleveland dataset is limited in comparison to the present study and as indicated in Table 4, alternative performance measuring techniques have not been used in previous investigations.

Table 4
Analysis of XGB performance on Cleveland dataset

Performance	Srichand *et al.*, [12]	Shubham and Birok, [18]	This Study
Accuracy	0.859	0.88	0.893
Precision	-	-	0.865
Sensitivity	-	-	0.943
Specificity	-	-	0.839
F1 Score	-	-	0.902
ROC	-	-	0.891
Log-Loss	-	-	3.834
Mathew-corr-coeff.	-	-	0.789

From the analysis it can be seen that this study not only improved the accuracy of XGB model but as well as it shows other good performance using measurements techniques such as Precision, Sensitivity, Specificity, F1Score, ROC, Log-loss and MCC which makes this study more novel and unique from existing ones.

5. Conclusion and Future Recommendations

In terms of both mortality and morbidity, cardiovascular illnesses rank first worldwide. Predicting who is at risk of cardiovascular diseases early on may help cut down on needless fatalities. Predicting cardiac issues using ML has shown to be effective. Results showed that the Multilayer Perception, K-nearest neighbours (K-NN), Support Vector Classifier, CART and XGBoost algorithms outperformed the others when trained and evaluated on the Cleveland dataset. The following performance metrics were achieved by the XGB classifier in comparison to competing classifiers: 89% accuracy, 87% precision, 94% sensitivity, 94% specificity, 90.2% F1 score, 81.2 % ROC, 78.90% Mathew coefficient and 3.87% log loss. In the future other ML classifiers such as Random Forest, Multilayer Perceptron, K-nearest Neighbor, Extra Tree Classifier, Extreme Gradient Boosting, Support Vector classifier, Stochastic Gradient Descent, AdaBoost, Classification and Regression Tree and Gradient Boosting along with these algorithms could be applied to compare the MLA efficiency.

References

[1] World Health Organization. “Cardiovascular Diseases.” https://www.who.int/health-topics/cardiovascular-Diseases
[2] Dwivedi, Amit Krishna, Syed Anas Imtiaz and Esther Rodriguez-Villegas. "Algorithms for automatic analysis and classification of heart sounds–a systematic review." *IEEE Access* 7 (2018): 8316-8345. https://doi.org/10.1109/ACCESS.2018.2889437
[3] Tao, Rong, Shulin Zhang, Xiao Huang, Minfang Tao, Jian Ma, Shixin Ma, Chaoxiang Zhang *et al.,* "Magnetocardiography-based ischemic heart disease detection and localization using machine learning methods." *IEEE Transactions on Biomedical Engineering* 66, no. 6 (2018): 1658-1667. https://doi.org/10.1109/TBME.2018.2877649
[4] Mohan, Senthilkumar, Chandrasegar Thirumalai and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." *IEEE access* 7 (2019): 81542-81554. https://doi.org/10.1109/ACCESS.2019.2923707
[5] Spencer, Robinson, Fadi Thabtah, Neda Abdelhamid and Michael Thompson. "Exploring feature selection and classification methods for predicting heart disease." *Digital health* 6 (2020): 2055207620914777. https://doi.org/10.1177/2055207620914777
[6] Ansarullah, S. Immamul and Pradeep Kumar. "A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method." *Int. J. Recent Technol. Eng* 7, no. 6S (2019): 1009-1015.
[7] Nazir, Shah, Sara Shahzad, Saeed Mahfooz and Muhammad Nazir. "Fuzzy logic based decision support system for component security evaluation." *Int. Arab J. Inf. Technol.* 15, no. 2 (2018): 224-231.
[8] Krishnan, Santhana and S. Geetha. "Prediction of heart disease using machine learning algorithms." In *2019 1st international conference on innovations in information and communication technology (ICIICT)*, pp. 1-5. IEEE, 2019.
[9] Mohan, Senthilkumar, Chandrasegar Thirumalai and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." *IEEE access* 7 (2019): 81542-81554. https://doi.org/10.1109/ACCESS.2019.2923707
[10] Srivastava, Asmit and Ashish kumar Singh. "Heart disease prediction using machine learning." In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 2633-2635. IEEE, 2022. https://doi.org/10.1109/ICACITE53722.2022.9823584
[11] Yang, Jian and Jinhan Guan. "A heart disease prediction model based on feature optimization and smote-Xgboost algorithm." *Information* 13, no. 10 (2022): 475. https://doi.org/10.3390/info13100475

[12] Doki, Srichand, Siddhartha Devella, Sumanth Tallam, Sai Sujeeth Reddy Gangannagari, P. Sampathkrishna Reddy and G. Pradeep Reddy. "Heart disease prediction using xgboost." In *2022 third international conference on intelligent computing instrumentation and control technologies (ICICICT)*, pp. 1317-1320. IEEE, 2022. https://doi.org/10.1109/ICICICT54557.2022.9917678

[13] Saboor, Abdul, Muhammad Usman, Sikandar Ali, Ali Samad, Muhmmad Faisal Abrar and Najeeb Ullah. "A method for improving prediction of human heart disease using machine learning algorithms." *Mobile Information Systems* 2022, no. 1 (2022): 1410169. https://doi.org/10.1155/2022/1410169

[14] Nagavelli, Umarani, Debabrata Samanta and Partha Chakraborty. "Machine Learning Technology-Based Heart Disease Detection Models." *Journal of Healthcare Engineering* 2022, no. 1 (2022): 7351061. https://doi.org/10.1155/2022/7351061

[15] Garg, Apurv, Bhartendu Sharma and Rijwan Khan. "Heart disease prediction using machine learning techniques." In *IOP Conference series: materials science and engineering*, vol. 1022, no. 1, p. 012046. IOP Publishing, 2021. https://doi.org/10.1088/1757-899X/1022/1/012046

[16] PriyankaKriplani, Prof, Prateek Gupta and Nagendra Singh. "Heart Disease Prediction Using Feature Selection and XGBoost."

[17] Rindhe, Baban U., Nikita Ahire, Rupali Patil, Shweta Gagare and Manisha Darade. "Heart disease prediction using machine learning." *Heart Disease* 5, no. 1 (2021). https://doi.org/10.48175/IJARSCT-1131

[18] Kaushik, Shubham and Rajesh Birok. "Heart Failure prediction using Xgboost algorithm and feature selection using feature permutation." In *2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pp. 1-6. IEEE, 2021. https://doi.org/10.1109/ICECCT52121.2021.9616626

[19] Jani, Rafsun, Md Shariful Islam Shanto, Md Mohsin Kabir, Md Saifur Rahman and M. F. Mridha. "Heart disease prediction and analysis using ensemble architecture." In *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, pp. 1386-1390. IEEE, 2022. https://doi.org/10.1109/DASA54658.2022.9765237

[20] Rahman, Md Mahbubur. "A web-based heart disease prediction system using machine learning algorithms." *Network Biology* 12, no. 2 (2022): 64.

[21] Boukhatem, Chaimaa, Heba Yahia Youssef and Ali Bou Nassif. "Heart disease prediction using machine learning." In *2022 Advances in Science and Engineering Technology International Conferences (ASET)*, pp. 1-6. IEEE, 2022. https://doi.org/10.1109/ASET53988.2022.9734880

[22] Karthikeyan, G., G. Komarasamy and S. Daniel Madan Raja. "An Efficient Method for Heart Disease Prediction Using Hybrid Classifier Model in Machine Learning." *Annals of the Romanian Society for Cell Biology* (2021): 5708-5717.

[23] Mohamad, Muhammad Arif and Muhammad Aliif Ahmad. "Handwritten Character Recognition using Enhanced Artificial Neural Network." *Journal of Advanced Research in Computing and Applications* 36, no. 1 (2024): 1-9. https://doi.org/10.37934/arca.36.1.19

[24] Aburashed, Laila, Marah AL Amoush and Wardeh Alrefai. "SQL Injection Attack Detection using Machine Learning Algorithms." *Semarak International Journal of Machine Learning* 2, no. 1 (2024): 1-12. https://doi.org/10.37934/sijml.2.1.112

[25] Tan, Catherine Hwee Fung, Nurul Aini Jaafar, Hui Qi Goo, Hui Ni Hew, Qian Yee Law, Joe Yinn Wong and Mallinath Dhange Dhange. "StepSafe Cane: Empowering Confident Steps for a Healthier Future." *Semarak International Journal of Electronic System Engineering* 3, no. 1 (2024): 56-64. https://doi.org/10.37934/sijese.3.1.5664

[26] Kemal, Edwar and Suhaida Omar. "Exploring the Linguistics Features of Code-Switching Advertisement as Public Information: A Textual Analysis of Critical Discourse Analysis (CDA)." *Journal of Advanced Research in Technology and Innovation Management* 12, no. 1 (2024): 27-38. https://doi.org/10.37934/jartim.12.1.2738