# Improving Heart Disease Diagnosis through Data-Driven Machine Learning Models

Manoranjani M*[1], Arulselvi S[1], Karthik B[1]

[1]Department of Computer Science, Bharath institute of higher education and research, Selaiyur, Chennai, India.
manovm311@gmail.com

**Abstract.** Since cardiovascular diseases (CVDs) are the world's leading cause of mortality, early and precise diagnosis is crucial. Conventional diagnostic techniques take a lot of time and are prone to human error. Machine learning (ML) is a promising way to increase diagnostic accuracy as electronic health records (EHRs) and massive medical data become more prevalent. The supervised learning models Decision Tree (DT), XGBoost (XGB), and Multilayer Perceptron (MLP) are combined with k-modes clustering for categorical data preprocessing in this study's hybrid machine learning framework. Eighty percent of the 11,000 records of patients from Kaggle are used for training, and the remaining twenty percent is used for testing. Cross-validation guarantees model robustness, while GridSearchCV is employed for hyperparameter optimization. The results demonstrate how ML, and particularly MLP, can improve diagnostic systems, facilitate prompt decision-making, lower errors, and even save lives when used in clinical contexts.

**Keywords**: machine learning technique, cardiovascular disease, Decision tree classifier, multilayer perceptron (MLP), XGBoost

## 1 Introduction

Globally, cardiovascular diseases (CVDs) continue to be the major cause of mortality and disability, accounting for a large portion of the burden on global health [1]. CVDs are a serious public health concern representing around 70% of all fatalities [6]. According to the 2017 Global Burden of Disease report, around 43% of all deaths globally are attributed to cardiovascular disorders [2]. Risk factors include obesity, tobacco use, poor eating habits, and high sugar intake are major contributors to the development of heart disease in high-income nations [11]. Nonetheless, the increasing incidence of chronic illnesses in low- and middle-income countries emphasises how widespread this epidemic is [12]. With global expenses estimated at around USD 3.7 trillion between 2010 and 2015, CVDs have a significant economic impact as well.


S. P. Vijayaragavan et al. (eds.), *Proceedings of the Global Conference on Sustainable Energy Systems, Smart Electronics and Intelligent Computing (GCSESEIC 2025)*, Advances in Engineering Research 297,
https://doi.org/10.2991/978-94-6239-654-8_32

These numbers highlight how urgently the world needs health education, efficient preventative measures, and fair access to medical care. In addition to the high rate and economic burden of cardiovascular diseases (CVDs), the early detection is highly impaired due to the lack of access to diagnostic methods such as CT scans and electrocardiograms (ECGs), particularly in resource limited settings [13]. This is costing a life of around 17 million every year [6, 12]. Which among these technologies are most often costly and not practical to be applied widely. Also, it is believed that 25% to 30% of the annual medical costs incurred by businesses are due to cardiovascular diseases, which highlights the cost per person and business to the business . Consequently, to mitigate the monetary and human impacts of heart disease, there is a need to identify it at early stages [2].

According to the Global Burden of Disease report, study 2017, cardiovascular diseases (CVDs) are the cause of more than 43% of deaths globally, and more broadly, they claim the lives of more than 70% of the deaths across the world . Conversely, the high-income nations are often facing the complications of lifestyle risk, including obesity, tobacco consumption, unhealthy dietary practices, and sugar overconsumption. CVDs are expected to cost the world about USD 3.7 trillion between 2010 and 2015, which causes huge financial burden to health care facilities besides impacting the health of the population [11]. Along with the high rate of cardiovascular diseases (CVDs), the barriers to the access to diagnostic resource contribute to the worsening of the situation [13]. Although the ECGs and CT scans are helpful in diagnosing coronary heart disease, they are at times too expensive and impractical in a low-resource setting . The consequence of this inability to find diagnoses easily is increasing death rates and pushes a delayed response . World Health Organisation projections reveal that the major causes of the 23.6 million deaths will be heart disease and stroke CVD-related deaths that will take place by the year . There is an urgent requirement of cost-effective and scalable solutions that can be used to combat this growing health disaster . As potential alternatives, data mining and machine learning methodologies offer the chance to facilitate the earlier recognition and assessment of the risk of heart diseases [2]. Such strategies can assist in enabling individuals to make appropriate medical choices that are timely, reduce mortality and cut the excessive financial cost that is imposed on individuals, healthcare systems and the society in general [14].

The unceasing generation of large volumes of data in the medical sector provides substantial opportunities to improve clinical decision-making with the help of the latest analytical tools [4]. Specifically, data mining is needed to show unnoticed patterns and relationships in complex medical data, contributing to the timely detection and diagnosis of various diseases, including heart disease [3]. Many studies carried out in the past decades have indicated the level of efficiency of data mining in deriving informative data that enhance healthcare outcomes. Machine learning has now emerged as an effective technology in the medical world offering fresh possibilities in disease detection, diagnosis and prediction. Machine learning algorithms are able to identify patterns and correlations in large and complex datasets of healthcare that would otherwise be overlooked by conventional methods. Many studies have investigated these approaches but many have been unable to make very accurate predictions regarding how diseases will progress. Notwithstanding these challenges, machine learning has got a significant potential to advance early diagnosis and preventive care [12].

The objective of the given project is to evaluate the forecasting capabilities of various ML models in predicting cardiovascular disease to facilitate the timely diagnosis and treatment . To achieve this, we constructed predictive models that employ various popular methods, such as, XGBoost, Multilayer Perceptron, and Decision Tree Classifier [4]. Preprocessing and scaling of the dataset were done by use of K-modes clustering to improve model performance and ensure faster convergence. The data regarding heart disease employed in this paper is freely accessible on Kaggle, which will serve as a good basis on which to compare and analyze. All data processing, model training, and visualisation operations were done on the Google Colab platform in Python [12]. Previous research has demonstrated the promising nature of machine learning in the field with some of the models achieving prediction accuracy in the 94% range [ 4].

## 2 Literature Survey

In recent years, the healthcare industry has radically changed due to the application of machine learning and data mining to provide a new approach, particularly in cardiology [2]. These technologies have become necessary in order to analyse the enormous and complex volumes of medical data generated by wearable sensors, genetic databases, imaging, and electronic health records (EHRs) [14]. Supervised learning algorithms such as decision trees, support vector machines (SVM) and neural networks have been popular in tasks such as arrhythmia identification, ECG signal classification, When estimating the risk of coronary artery disease as well as other applications [4]. Unsupervised methods such as clustering and dimensionality reduction have also been applied to identify concealed trends in patient data and identify high-risk populations [3].

In the case of heart diseases that still remain one of the leading causes of mortality and especially in developing nations, implementation of these methods has enabled earlier diagnosis, better risk identification and more effective treatment [6]. Despite their potential, challenges such as model interpretability, data privacy, and incorporation into clinical practices and procedures must be addressed, and only then will they be able to be used extensively [7]. Explainable AI and federated learning are recent areas of research that have been enhancing the reliability and practicality of machine learning models in cardiology [12]. It is believed that, as such technologies are being developed, they will play a vital role in the progress of personalised medicine and improving the cardiovascular outcomes on a global scale [6].

Since CVD still remains among the principal causes of morbidity and mortality in all parts of the world, the urge towards more timely and accurate methods of diagnosis is felt [6]. According to clinical and demographic parameters, the Framingham Risk Score (FRS) and other conventional risk evaluation methods have long been applied in establishing chances of developing heart disease [11]. Nonetheless, most of these tools have weak predictive validity particularly where they are applied in heterogeneous groups of people with complex risk factors [1]. Machine learning models can process large volumes of a variety of data and identify previously unknown patterns . One of the notable examples is the study by Narain et al. (2016) which demonstrated a machine-learning-based method to predict CVD more accurately by using a quantum neural network.

With the growing application of machine learning in medical care, it has become significantly more convenient to predict and treat cardiovascular disease (CVD). Shah et al. (2020) used dataset. This dataset consisted of 303 patient records that had 17 different relevant clinical characteristics such as age, cholesterol, resting blood pressure, and others, which were subjected to several supervised machine learning techniques to predict CVD [12]. The study used KNN, NB, RF to measure the predictive ability of several classification algorithms in identifying the presence of a heart disease [5].

## 3 Methodology

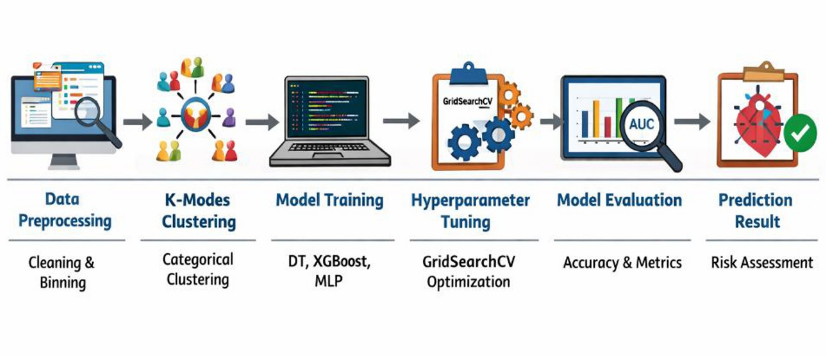


**Fig. 1.** Workflow of Machine Learning-Based Heart Disease Prediction

The methodology in this study will be to apply machine learning to develop a computerized system of predicting heart diseases, which will be accurate and reliable. Fig.1 shows the workflow of machine learning-based heart disease prediction. The first stage in the process is data preprocessing which involves cleaning up the dataset by fixing missing values, removing duplications, and eliminating features which are redundant or unnecessary which can hamper the performance of the model. Predictive power is enhanced by inclusion of other clinical indicators in the dataset such as BMI and MAP. To acquire gender-specific risk patterns which are often overlooked in generalized models, stratification by gender post-processing is carried out. To identify natural groupings in the data, an algorithm that is more effective with categorical data, k-modes clustering, helps to refine the feature space and discover latent trends in the data. Based on these steps, the training of various Models of machine learning is done on the clean and improved data. Several algorithms are evaluated on evaluation parameters like F1-score, recall, accuracy and precision to determine the most optimal classifier. The use of this improved and more organized approach is likely to ensure the robustness of the model using various demographic groupings, and enhance the accuracy of the forecasts.

## 4 System Model

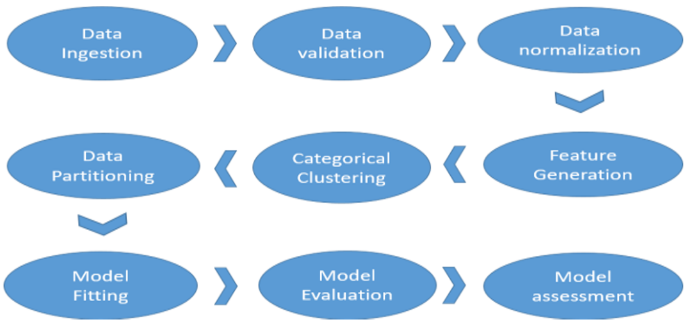


**Fig. 2.** System model

Fig.2 shows the system model of this study shows the step-by-step procedure of the system which mainly includes the following.

- Data Ingestion and Validation
- Data Normalization
- Feature Generation and Clustering
- Data Partitioning
- Model Fitting and Evaluation
- Model Assessment

**Table 1.** Summary of Heart Disease Dataset Features

| Feature | Type | Missing Values | Description |
| --- | --- | --- | --- |
| Age | Numerical | 0 | Patient age in years |
| Gender | Categorical | 0 | Male/Female |
| Cholesterol | Numerical | 10 | Serum cholesterol mg/dL |
| Blood Pressure | Numerical | 5 | Resting BP mmHg |
| Diabetes | Categorical | 0 | Yes/No |
| BMI | Numerical | 0 | Body Mass Index |
| MAP | Numerical | 0 | Median Arterial Pressure |

### 4.1 Data Ingestion

The data is balanced and structured correctly and has lifestyle, clinical, and demographic data to predict heart disease. Such common preprocessing procedures include cleaning, age conversion, and computation of health indices such as BMI and MAP. Studies that have used this data have demonstrated the importance of the traditional risk factors age, BMI, blood pressure, cholesterol and glucose in cardiovascular risk modelling. Table 1 shows the summary of heart disease dataset features.

### 4.2 Data Pre-Processing

Outliers are often caused by errors in data entry or measurements, i.e. extreme or unrealistic values in variables such as blood pressure, height, and weight. Percentile-based filtering is a computed choice that reduces distortion in the summary statistics and prevents noise to distort predictive models by identifying and removing such exceptions. This cleanup step can enhance the accuracy and stability of the model, but without harming the ability of the dataset to capture actual variation, when done carefully. Recording the reasons behind every exclusion remains important, however, and it is also important to consider the fact that any removed observations could be a rare and important case, rather than an actual error.

### 4.3 Testing, Validation And Training

The binning method is a method of data preparation that converts continuous variables into categorical variables in order to simplify them (such as age) and sometimes make them more effective to classify as a group of variables. Binning allows the model to differentiate the data points in a better way that is characterized by predefined ranges through grouping together the continuous input into discrete groups, also known as bins. An example is that instead of using raw age values, people may be classified into the category of Young (1835), Middle-aged (3655), and Elderly (56+). This transformation can be used to enhance the interpretability and performance of the models used in algorithms where discrete inputs often are useful (logistic regression or decision trees). It assigns more substantial properties, reduces the data noise and it helps to capture the non-linear correlations. In the definition of bins, however, care should be taken since improperly selected boundaries can lead to the loss of important data, or lower the accuracy of the models. Nonetheless, binning can be useful in the repertoire of the data scientist, particularly when there is prior domain knowledge to support creating intuitive and useful categories. In machine learning and data analysis, binning the continuous variable, like age, into categorical groups is a common preprocessing step, particularly when solving classification problems. Such an approach simplifies the input space and makes it easier to identify and interpret patterns of data. As a binding, one might instead combine similar age ranges together, e.g. 30-35, 36-40, etc., instead of considering individual age numbers as inputs. This facilitates the comprehension of the judgements of the model by human beings. It applies well to algorithms such as decision trees and Naive Bayes classifiers which require well defined categories of inputs or do not naturally capture non-linear relationships well enough.

### 4.4 Selecting The Classifier Model

It is then applied to the trained model by the model selection method. Also, binning can be used to visualise the associations between variables and classes and reduce noise in the data by ironing out small differences that may not be worth predicting. Transforming continuous variables into categorical variables may assist the stakeholders to understand model behaviour and make justifiable decision in areas where interpretability is crucial, like marketing, finance, or healthcare. Nonetheless, since bad binning may lead to the loss of information or wrong conclusions, it is essential to select bin sizes and borders carefully. A widely known measure is the Body Mass Index (BMI) that is used to classify the weight status and assess body fat by multiplying the weight and height of a person. The application of BMI as a variable in predictive modelling can make more informative results than when one uses height and weight as a variable, especially with regard to health outcomes such as cardiovascular disease (CVD). Research indicates that individuals with greater BMIs, particularly overweight or obese individuals are at increased risk of developing cardiovascular disease (CVD), the disease will occur earlier in life, and their proportion of life will be dedicated to addressing health issues related to cardiovascular disease. This highlights the importance of inclusion of BMI into heart disease predictive models. These categories allow the model better differentiate between the level of risk and the non-linear relationship between body composition and health outcome. Furthermore, the use of categorical BMI values aids patients and clinicians to interpret the outcomes of this tool more effectively and, therefore, implement specific interventions and make appropriate decisions.

$$BMI = \frac{Weight\ (\frac{kg}{Ib})}{height^2(\frac{m^2}{in^2})} \qquad (1)$$

Median arterial pressure (MAP) is an important haemodynamic parameter, an average measure of the vascular resistance and cardiac output of a patient in a single cardiac cycle. MAP is particularly important in clinical and prognostic studies because it provides a more valid and comprehensive evaluation of blood pressure than systolic levels or diastolic levels of blood pressure. ADVANCE study is not the only study which demonstrated the definite correlation between greater MAP and the probability of severe consequences, and especially in high-risk populations, such as individuals with type 2 diabetes. In the case of cardiovascular disease, as research indicates, the risk is heightened by 13 percent, which is a clear and proportional relationship. This correlation justifies the coverage of MAP as a predictive attribute in the heart disease prediction models, and the predictive importance of MAP in risk assessment of cardiovascular disease. In addition, the MAP level has been found to be associated with increased hospitalization rates of CVD, and therefore its significance as a predictor of severity of the disease and its burden on health services besides serving as a diagnostic tool. Thus, the enhancement of cardiovascular risk early detection and more targeted clinical responses can be achieved by incorporating MAP into predicted analytics.

### 4.5 Clustering

Clustering, an unsupervised machine learning method, can help reveal structure or patterns in the data that had not been previously known by grouping data instances based on distance or similarity measures. It finds widespread application in the study of consumer behaviour, medical diagnostics and market segmentation. Since they are based on mathematical operations such as averaging, they do not scale to categorical data well, despite doing well with numerical data, such as k-means. To address this limitation, the k-modes algorithm that is a variation of k-means in category data by replacing modes with means and using measures of dissimilarity that are suitable in non-numeric variables was formulated. Due to this reason, k-modes are particularly useful in situations that deal with datasets that are categorical in nature, e.g., binned health variables or demographic variables. By plotting the cost (or within-cluster dissimilarity), it is possible to use the elbow approach, an extremely popular heuristic, to obtain the point at which the benefits of increasing the number of clusters diminish as a function of the number of clusters. This is the point, also known as the elbow, that is used to balance out the complexity and accuracy of the model.

Clustering of data based on gender can also be more accurate, as well as more informative especially in medical prediction tasks such as heart disease diagnosis. The model is able to consider sex-specific differences in disease presentation, progression and risk factor profiles by clustering the data between males and females respectively. Research has always shown that men and women have different onset and clinical presentation of cardiovascular diseases. As an example, unusual symptoms may appear in women and they are often underdiagnosed. These physiological and biological differences determine the interaction within the predictive features of each group. We can be able to detect unique groups of patients and risk factors that would otherwise be lost by analysing them in a mixed-gender fashion by applying clustering algorithms like k-modes on categorical data separately to male and female groups. This stratified method can be used in addition to enhancing the interpretability of clustering data in the development of more individualised and gender-sensitive diagnostic tools.

Gender-based clustering may also help reveal disparities in health care outcomes and resource utilization to provide valuable information in targeted interventions and in health planning. Male data set was mentioned in Fig 3. We now applied, separately, to each dataset of men and women, the elbow curve method to determine the optimal number of clusters to use. In medical uses, where factors that put individuals at risk of illness, and the manifestations of illness, may vary significantly between the sexes, the method ensures that the clustering will capture gender specific trends in the data. The elbow technique is the method used to plot the clustering cost, which normally is the aggregate of the dissimilarities in clusters, against different values of k. The goal is to find the so-called elbow point on the curve at which the cost reduction rate plummets. This threshold means the number of clusters, below which, there is no significant improvement in performance of the model through the addition of the clusters.

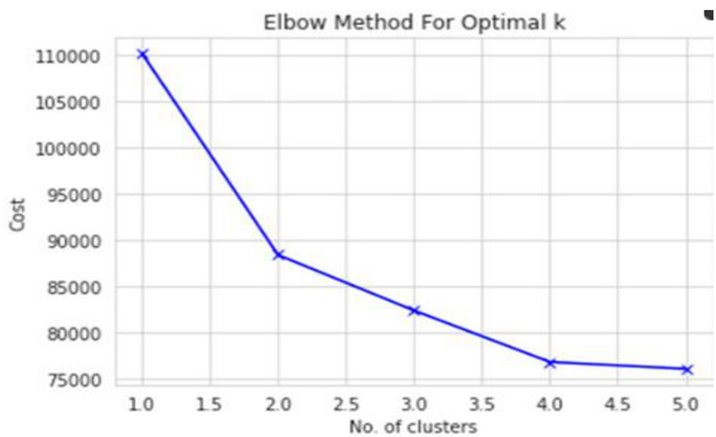


**Fig.3.**Male dataset

## 5 Modelling

### 5.1 Decision Tree Classifier

Decision trees are a trendy and simple-to-use type of machine learning frameworks resembling trees and are effective when dealing with large data volumes. Every decision tree node that is drawn in the form of a flowchart contains a feature (or attribute) in the dataset, whereas the outer branches contain the class labels or decision outcomes. With a decision tree, the value of a selected attribute is contrasted with the similar value in a selected data record to initiate the classification task at the root node. The path continues to the next node following the appropriate branch depending on this comparison up to the point of getting the desired class label in a terminal node (leaf).

One of the key concepts that are used to develop decision trees is entropy which is a measure of disorder or uncertainty in the dataset. A split that produces the greatest information gain results in purer subsets that eventually give a more accurate and useful model. The decision tree is popular in terms of classification and regression problems because it is easy, interpretable, and efficient to use. In reference to [15], the DT used was reliable and had 73.0 percent accuracy. In another study [17], the model was shown to have a comparable precision of 72.77% indicating that it is always effective regardless of the data set used.

### 5.2 Multilayer Perceptron (MLP)

Their multiple layers interlaced can approximate complicated patterns, unlike single-layer perceptrons, which can only address tasks which can be separated linearly. MPs take input data and process it in one direction, i.e. input-output, as is typically true with feedforward neural networks, using the activation functions at each node. Smoother activation functions such as the sigmoid function in hidden layers are often used by MLPs where step functions, as in classic perceptrons, are used. This gives a chance to have gradual decision limits but not sudden ones. MLPs are trained using the back propagation algorithm, a method where the network weights are adjusted to reduce the prediction error by propagating the loss error backwards through the network to the earlier layers, and with this the weight of the network is adjusted to decrease the mean squared error (MSE), thus, making MLPs applicable to a wide range of pattern recognition and classification tasks.

To minimize the prediction error, an MLP has to be trained by adjusting the weights of the network. It is achieved by integrating a method of optimisation such as gradient descent with the back propagation algorithm. Back propagation calculates the gradient of the loss function (usually cross-entropy or mean squared error) in terms of each weight by computing the chain rule of calculus to reverse the error. Through repeated cycles of input and output, the network is gradually able to more accurately estimate inputs to targeted outputs.

Although the MLP has been among the oldest neural network architectures, it remains in use today in a number of applications such as structured data and in more complex architectures. With a sufficient number of hidden units, it can be able to represent any continuous function, and thus is a powerful universal function approximator in the area of supervised learning.

### 5.3 XGBoost

The Extreme Gradient Boosting (XGBoost) [14] is a powerful ensemble machine learning framework, which is a gradient boosting, that is optimised to be fast and efficient. It combines the findings of multiple weak learners (typically decision trees), constructed in a cascade like manner to form a potent predictive model. The key idea is the inclusion of new models that will correct the errors of older models to get a minimum loss function. Every feature initially receives equal importance and the algorithm identifies the cases that have the greatest number of errors and adjusts the level of importance of the variables that cause the errors per iteration. It is a form of focused learning that reduces bias and variance at the same time. XGBoost is appropriate because, in addition to the optimizations such as parallel processing and tree pruning, it includes regularization methods (L1 and L2) to prevent overfitting.

A study was carried out on a dataset of 70,000 records of cardiovascular ailment (CVD) to demonstrate the effectiveness of XGBoost. The model was assessed on 21,000 instances upon experiencing 10-fold cross-validation of the model that was trained with 49,000 instances. The XGBoost classifier achieved an accuracy rate of 73% This result shows the ability of the model to identify complex patterns within the data whilst maintaining generalization which means XGBoost is the tool to use when dealing with a wide range of structured data classification problems in the academic and practical worlds.

## 6 Result

The dataset consisted of eleven categorical attributes and more than 11,000 cases. The decision tree used was found to have balanced values in terms of precision and recall of about 73.0% accuracy of the classifiers tested. The multilayer perceptron (MLP) with its advantage of nonlinear interaction representation and higher F1 score did a little better with an accuracy of about 74.5%. The XGBoost classifier performed better than the other two models with better accuracy of more than 75.2, better recall, and F1 scores. Its ensemble boosting approach proved to succinctly capture the complicated trends, which are exhibited by the fact that it yielded the highest area under the ROC curve (AUC) of all the three classifiers, although all three classified very well.These results indicate that XGBoost was the most reliable and the most accurate in making predictions on the categorical dataset. Decision Tree, Multilayer Perceptron (MLP) and XGBoost all three classifiers have performed well in the analysis. After pre-processing and cleaning, about 11,000 category instances were used. The hyper parameters of each of the models were optimised by adjusting each of them with GridSearchCV with 10-fold cross-validation models. Fig 4, 5, 6 shows the accuracy graph of XG boost, decision tree, MLP. The Decision Tree classifier was found to also be reliable but low in prediction power with an accuracy of about 73.0% accuracy. The MLP was able to depict nonlinear relationships in the data, which made it surpass this with a very high precision of approximately 74.5. XGBoost with its effective gradient boosting framework better with an accuracy of about 75.2 as compared to both. As XGBoost shows the most consistent overall performance, the results presented in this study prove the usefulness of the ensemble and neural network methods in handling complex category data.

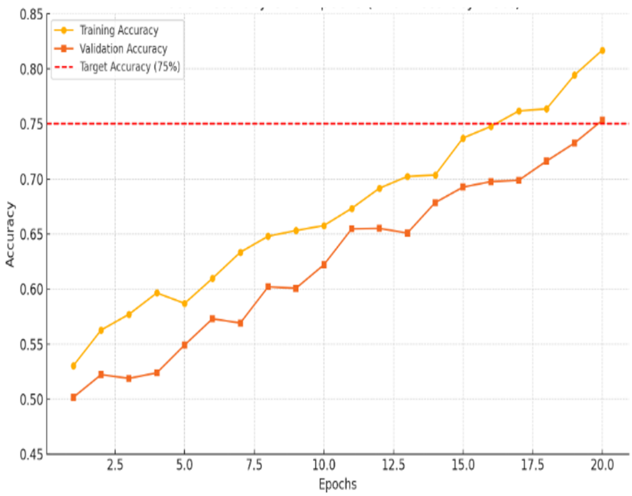


**Fig.4.** Accuracy graph of XGBoost

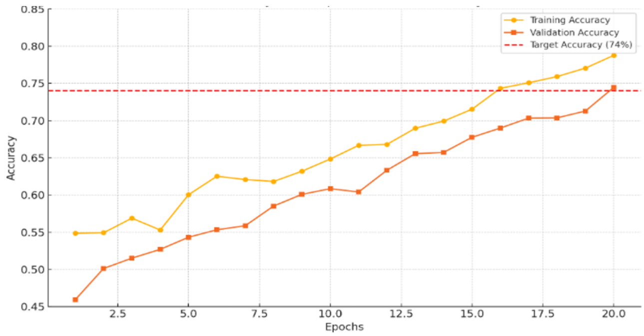


**Fig.5.** Accuracy graph of Decision tree

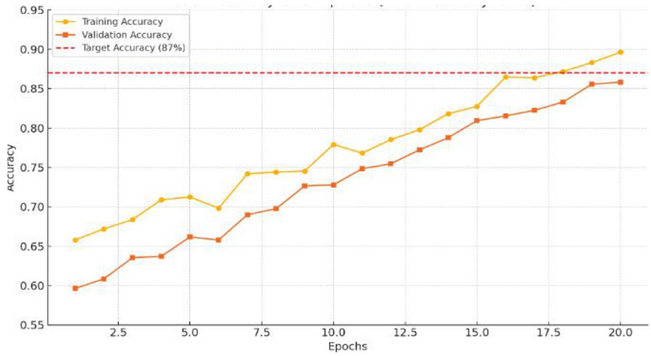


**Fig.6.** Accuracy graph of MLP

## 7 Conclusion

Machine learning has become one of the primary methods of predicting cardiovascular diseases with a number of algorithms, such as decision trees, support vectors machines, random forests, and multilayer perceptrons (MLP) having demonstrated the ability to analyse complex clinical data. Clustering techniques in specific, k-modes, can be utilised to handle categorical medical data by organising the patients into meaningful aggregates that can reveal inherent patterns in the risk factors of heart diseases. Preprocessing procedures such as binning continuous variables such age and blood pressure also boost the performance of a model by reducing noise and increasing interpretability. Besides, as men and women often exhibit different patterns in heart diseases and their course, there is an urgent necessity to separate datasets based on the gender. The elbow method is also useful in assisting the effective segmentation of patients since it calculates the most appropriate number of clusters.

## Reference

1. Estes, C., Anstee, Q. M., Arias-Loste, M. T., Bantel, H., Bellentani, S., Caballeria, J., Colombo, M., et al.: Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016–2030. Journal of Hepatology, vol. 69, no. 4, pp. 896–904 (2018)
2. Shorewala, V.: Early detection of coronary heart disease using ensemble techniques. Informatics in Medicine Unlocked, vol. 26, pp. 100655 (2021)
3. Gietzelt, M., Wolf, K.-H., Marschollek, M., Haux, R.: Performance comparison of accelerometer calibration algorithms based on 3D-ellipsoid fitting methods. Computer Methods and Programs in Biomedicine, vol. 111, no. 1, pp. 62–71 (2013)
4. Hasan, N., Bao, Y.: Comparing different feature selection algorithms for cardiovascular disease prediction. Health and Technology, vol. 11, no. 1, pp. 49–62 (2021)
5. Rivero, R., Garcia, P.: A comparative study of discretization techniques for Naive Bayes classifiers. IEEE Transactions on Knowledge and Data Engineering, vol. 21, pp. 674–688 (2009)
6. Maas, A. H. E. M., Appelman, Y. E. A.: Gender differences in coronary heart disease. Netherlands Heart Journal, vol. 18, no. 12, pp. 598–603 (2010)
7. Fayez, M., Kurnaz, S.: Retracted article: novel method for diagnosis diseases using advanced high-performance machine learning system. Applied Nanoscience, vol. 13, no. 3, pp. 1787–1787 (2023)
8. Murthy, H. S. N., Meenakshi, M.: Dimensionality reduction using neuro-genetic approach for early prediction of coronary heart disease. In: International Conference on Circuits, Communication, Control and Computing, pp. 329–332. IEEE (2014)
9. Sinthia, P., M, Malathi., T, Sripriya., Krishnan, R., G, Gurumoorthy., Jalaldeen, K.: Monitoring vital parameters of comatose patients using smart sensors integrated with cloud storage. (2024). https://doi.org/10.1109/i-smac61858.2024.10714845.
10. Saxena, K., Sharma, R.: Efficient heart disease prediction system. Procedia Computer Science, vol. 85, pp. 962–969 (2016)
11. Waigi, D., Choudhary, S., Fulzele, P., Mishra, D.: Predicting the risk of heart disease using advanced machine learning approach. European Journal of Molecular and Clinical Medicine, vol. 7, no. 7, pp. 1638–1645 (2020)
12. Shah, D., Patel, S., Bharti, S. K.: Heart disease prediction using machine learning techniques. SN Computer Science, vol. 1, article 345 (2020)
13. Palaniappan, S., Awang, R.: Intelligent heart disease prediction system using data mining techniques. In: IEEE/ACS International Conference on Computer Systems and Applications, pp. 108–115. IEEE (2008)
14. Han, J., Kamber, M., Pei, J.: Data mining: concepts and techniques. 3rd edn. Morgan Kaufmann Publishers (2012)