

A Predictive Model for Early Heart Disease Detection using Data Mining Techniques

Sanjay Singla

Department of Computer Science & Engineering, Chandigarh University, Mohali, Punjab, India
dr..ssinglacs@gmail.com

Priti Singla

Department of Computer Science & Engineering, Chandigarh University, Mohali, Punjab, India
pritisingla04@gmail.com

Nandita Manchanda

Department of Computer Science & Engineering, Maharishi Markandeshwar Engineering College (MMEC), Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India
nandita.malik2721@gmail.com

Deepshikha Chhabra

Department of Computer Science & Engineering, Maharishi Markandeshwar Engineering College (MMEC), Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India
114deepshikha.116@gmail.com

Harish Kumar Mittal

Department of Computer Science & Engineering, Chandigarh University, Mohali, Punjab, India
harish_mittal@yahoo.co.in

Narinder Kumar

Department of CSE, University School of Computing, Rayat Bahra University, Mohali, India
nk616319@gmail.com

ABSTRACT

Heart disease is one of the leading causes of deaths throughout the world. Early diagnosis is very essential to cure the disease. Data mining helps in making intelligent diagnostic system through which massive amount of clinical data can be analyzed. In the paper, a better approach for predicting the early heart diseases using data mining algorithm such as SVM, Decision Tree, Random Forest, Naïve Bayes, CN2 Rule Inducer and Majority Voting as base line is suggested. Data cleaning, normalization, feature selection, classifier optimization and cross-validation are all part of the model pipeline. In benchmark heart disease dataset experiments, optimized classifiers improve predicted performance. Naïve Bayes and Random Forest are the most accurate algorithms, with the highest AUC, so they can be used to help doctors make decisions. Healthcare professionals can use the suggested framework to automate, reliably assess risk and add it to practical diagnostic systems that will keep improving patient data and qualitative heart disease risk assessment. An ordinal number from 0 to 4, where 0 indicates that the patient is healthy and 4 suggests that there is a good chance that the patient is ill, is used to describe this probability. If the classifier is able to predict the state of health of the patient in 85% of the cases or if its absolute accuracy exceeds 85%, the results would be impressive.

Keywords: Random Forest, Decision tree, Majority, Support Vector Machine, Naïve Bayes

1. INTRODUCTION

Information withdrawal, which holds a lot of promise, can assist the healthcare sector to create health programs that take into account full usage of data and analytics to detect inefficiencies and optimal methods of improving treatment and saving costs. Most hospitals use hospital information systems to keep track of a lot of patient data. Data mining can yield a plethora of previously unrecognized information that aids physicians in making informed treatment choices [1].

1.1 Data Mining

This technique employs machine learning, database proficiency, and quantitative analysis to uncover concealed configurations within extensive datasets. As the healthcare industry grows, more and more people are submitting

medical data, which makes it important to remove that data from an inquiry. Heart disease wins when you look at deaths around the world [2]. Medical professionals face a compelling challenge when determining if a long-term patient has heart disease, owing to general impediments and specialized knowledge [3] [4].

There have been many iterations of models that use data mining or computational methods. The data extraction method combines math, knowledge of devices, and experience with catalogs to find new configurations in huge datasets [5].

Data mining and machine learning approaches such as neural networks and and neural networks, quantifiable methods and mathematical models are some of the most advanced ways to analyze data. Many schools use data quarrying in their research. In order to get new customers, fight disease, cut costs, and improve benefits, businesses in communications, banking, health, education, and commerce give up data withdrawal. Data Mining is also referred to as information extraction, extraction of facts from statistics, statistical/organizational analysis and statistics investigation, which have different meanings. Statistics removal seeks to understand and predict behavior by finding patterns in statistics. Data mining models have directives, algorithms, or complex "transmission purposes." Their goals divide them into two categories:

- Supervised Prototypes.
- Unsupervised Prototypes.

1.1.1 Supervised Prototypes

Supervised, targeted, directed, or predictive modeling makes predictions about events or estimates continuous numerical attribute parameters. The models have areas or qualities for contributions and areas or goals for production. Input fields are called predictors because they help the model choose a utility for making production forecasts.

1.1.2 Unsupervised Prototypes

Mockups that don't have a verification or objective only have an input field and no production area. No objective factor affects undirected arrangement recognition. These models focus on trends in input field data.

Unsupervised prototypes comprised of:

1. Cluster models.
2. Association and series models.

1.2 Mixture Recommender Structures

The mixture recommender systems employ numerous methods, and they reduce the disadvantages of certain ones by altering the reward structure of others. In order to improve classification optimization and overcome some of the limitations of the clean reference methods, hybrid filtering methods employ multiple recommendation algorithms [6].

1.3 Heart Disease

At the current stage in medicine, heart disease is one of the main issues regarding the health of people throughout the world. There was a significant change in the causes of deaths related to heart diseases in the twenty first century and an impressive increase in the lifespan expectancy. At the current time, it is expected that there will be a decline in the number of deaths by 30%, with the decrease being by 40% for the high-income countries and the small and middle income countries being reduced by 28%.

The ongoing transition has spread throughout the globe at a much faster rate than in the previous century for all groups of people, groups within national and international societies and nations due to economic growth, urbanization, and associated changes in physical activities. There has been an increased incidence of heart failure due to recent changes in life style. According to recent revisions, heart failure symptoms have tripled in the last 25 years. According to recent revision, chronic non-infectious diseases like heart disease have been one of the main causes of death throughout the world. There has been an incredible change in the health status of the general population which has resulted in the increase in cardiac infection [5]. The main cause of death in the world every day is now heart disease. The main reason behind the worldwide increase in cardiac disease is the remarkable change in the health status of the general population around the world. Heart disease has significantly increased during the last 20 years and has become one of the main causes of death in most nations of the world. According to new studies concerning cardiac health, about 1.2 billion people around the world die each year due to heart diseases. Assumed the big differences between communal, national, and economical zones, there isn't any single solution to the increased risk of heart disease. Predicting heart failure before the expensive portion event is normally a difficult task. The cost of new imaging technologies and controlled processes in the diagnosis of heart disease is equally very high [7]. Major signs of heart disease include chest pain, difficulty breathing, tiredness, swelling, palpitations, and fainting; coughing, blood from coughing and blueness are also other examples.

1.4 Machine Learning

In the era of technology, machine learning is being applied with precision in assessments and estimation. Machine learning is done through training of a set of completed tasks to know how to estimate the process. Some of the machine learning includes supervised machine learning, unsupervised machine learning, and reinforcement machine learning [9].

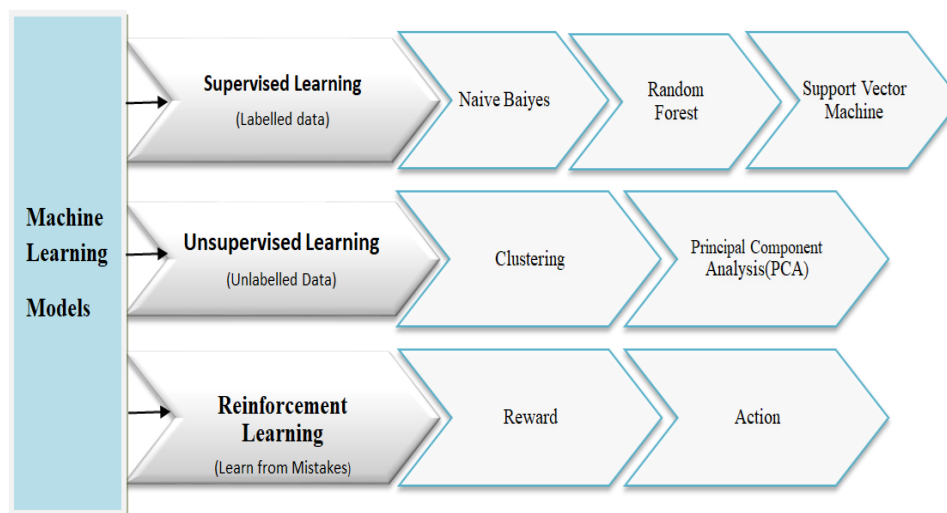


Figure 1 Different Machine Learning Models [11]

1.4.1 Supervised Learning

External support and "factual" or "accurate" definitions for the input dataset are provided by the supervised machine learning algorithms [12]. The training data set and the test data set are both generated from the input data set. The test data are associated with the output variable which must be predicted. The whole procedure depends on some specificities which are drawn from the training data set and are used for the classification and prediction of something in the test data set. There are many machine learning algorithms focused on the optimization of the objective function or the cost function. The difference between the ground truth and the approximation of the algorithm is measured by the cost function. The model training occurs through the minimization of the cost function in such a way that the estimations become sufficiently close to the real values (ground truth). Through the cost function minimization, the gradient descent method has been used [13][14]. In the machine learning approach, several models come together, having different strengths depending upon the situation and application.

Classification

From a machine learning standpoint, classification matters a lot. The two steps are learning, where a classification model is built, and then classification itself, where that model predicts object classes. The methods can be transferred to other sets of data after the processes of data preparation, modeling, feature selection and evaluation [15]. Classification estimates either categorical or sometimes unconditional continuous valued functions. An example of bank loan application system is the database which classifies loan applications as either safe or risky. Supervised machine learning covers a broad set of classification methods. The categorization process is shown in the figure 2. At the start, the system looks at loan decision factors. It uses several variables, such as name, age, and income. After that, the classification model is applied to those inputs. Once the process is finished, it produces an output from the given data, showing clearly whether a person qualifies for a bank loan or does not.

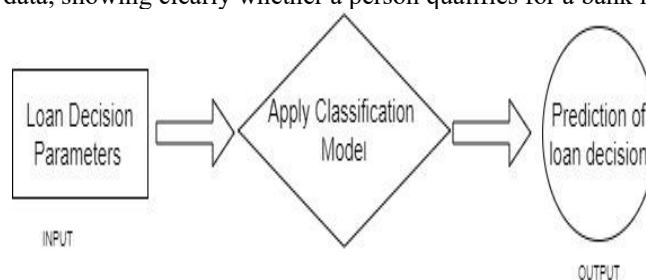


Figure 2 An Example of Classification Technique [15]

Naïve Bayes

Another classification technique based on Bayes' theorem assumes that if a feature is selected for a certain class, then it is likely to be independent of another feature in that class.

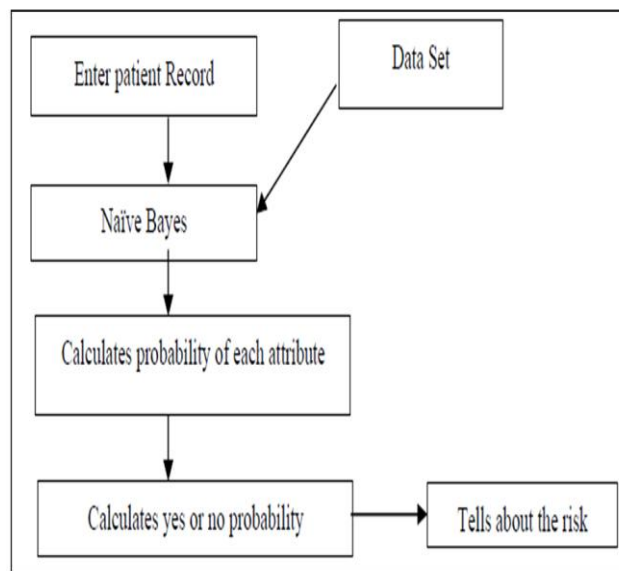


Figure 3 Graphical Representation of Bayesian Classifier [15]

Random Forest

Random Forest is the alternative supervised learning algorithm. It works in both regression analysis and classification, although it is used more often for classification and usually gives suitable results. In practice, it combines several decision trees. Those trees are built from data samples, then each tree gives a predicted value. After that, a vote decides which result wins. Figure 4 displays how a random forest works.

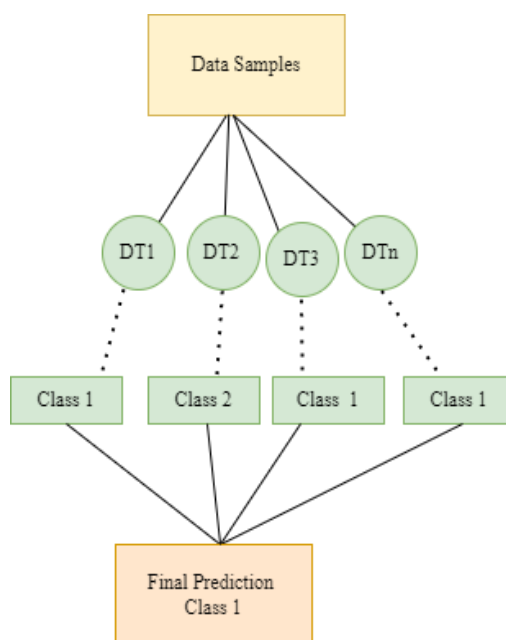


Figure 4 Graphical Representation of Random Forest Classifier [15]

Support Vector Machine (SVM)

SVM handles classification through maximizing . The points nearest to the margin are known as support vectors. It is a type of supervised learning algorithm and can be used for both classification and regression.

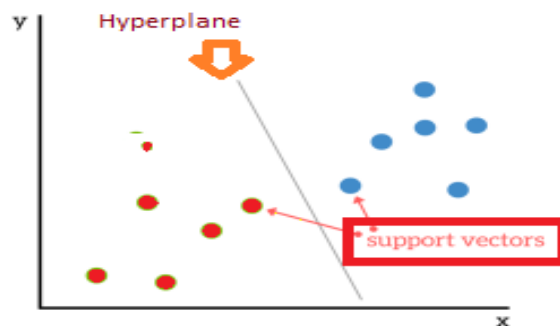


Figure 5 Graphical Representation of Support Vector Machine [15]

Decision Tree

Researchers have examined how the Decision Tree approach can be used in handling heart disease. A decision tree has a simple structure: internal nodes, branches, and leaf nodes. Each branch shows a feature value, while each internal node tests a feature, and the leaf node gives the predicted class or decision output. Starting at the root node, classification moves through the tree based on the value of the predictive attribute. To build pruned decision trees, the process involves data splitting, data classification, decision tree clustering, and the use of pruning to reduce complexity. There are two main types of classification methods: supervised and unsupervised. In supervised grouping methods, Chi-square merging and entropy are usually used, while equal-width and equal-frequency methods belong more to unsupervised ones. Data splitting is tested both with voting and without it, which honestly can change the result more than expected. The study also looks at decision tree types based on Gini Index, Information Gain, and Gain Ratio. In the end, reduced-error pruning gives fewer, tighter decision rules. Figure 6 shows ID3 algorithm implementation on patient data.

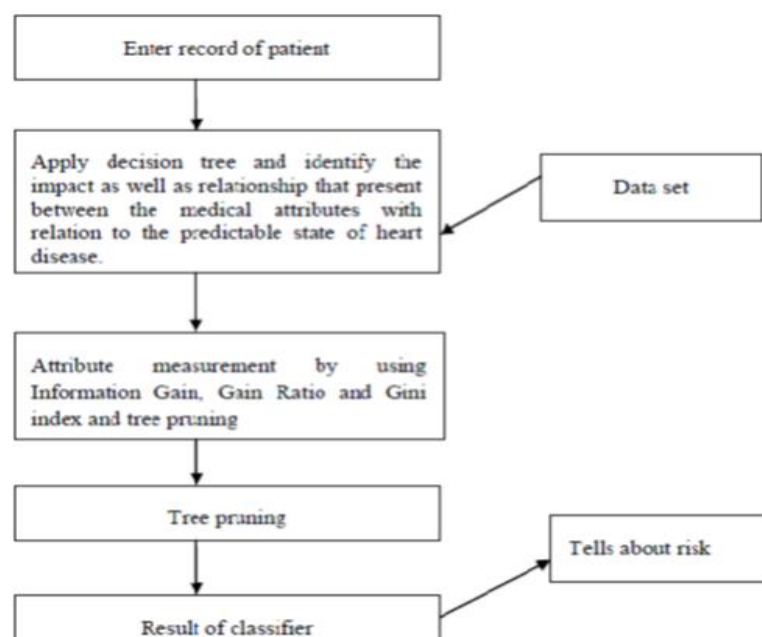


Figure 6 Graphical Representation of Decision Tree

Neural Network

The neural network model is famous due to providing accurate predictions in practical applications. In our case, NN learns from the Heart Diseases dataset, which is a feed-forward neural system model with optimal 10, variable culture degree, and the backpropagation knowledge acquisition process with movement. This system is relatively straightforward; it takes clinical data as input, develops ANN algorithm and produces the prediction outcomes. Computational procedure consists in dividing randomly the clinical data into two equal sets; one is for training, while the other is for testing. Each feature starts with a random weight first. Then those weights get adjusted based on the estimated errors, which is really the core of the learning step. Once the errors meet the stopping condition, the final weight for each feature is fixed. This process is repeated many times. After the training model is built, we can calculate the prediction results from the test data. The neural network procedure on scientific data is shown in Figure 7.

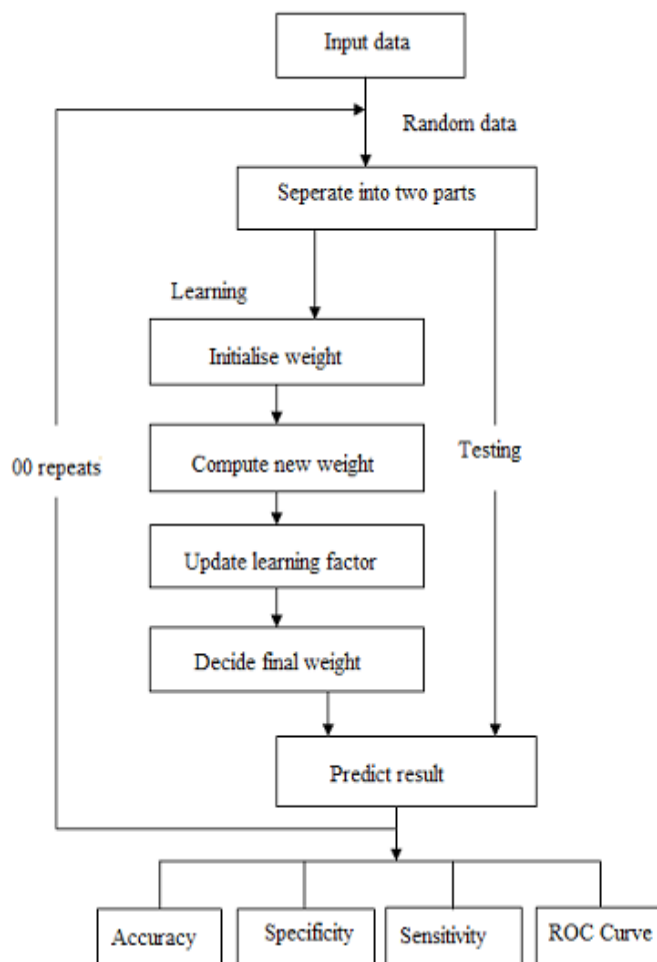


Figure 7 Graphical Representation of Neural Network

1. RELATED WORK

For the category of models, Arsalan Khan et al. [2023] have compared the performance of the recommended machine learning algorithms such as Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) and Support Vector Machine (SVM) under several scenarios [16]. Huang starting k-modes clustering method proposed by Chintan M. Bhatt et al. [2023] enhances the accuracy of classification. The random forest, decision tree classifier, multilayer perceptron and XGBoost model is used. The model parameters are optimized using GridSearchCV hyper-tuned model parameters [17]. A machine learning model for cardiac disease prediction by Daniyal Asif et al. [2023] incorporates several preprocessing, ensemble learning and hyper parameter optimization techniques [18]. K. Butchi Raju et al. [2022] recommended the combination of cloud, fog and edge computing for faster and accurate results. Hardware is used to collect patient data. Removal of sentiment narrative from signs leads to important features [1]. IoT methods and subjects' long suffering past associated microelectronic healthcare statistics saved in the fog are used to extrapolate analytics by Angel Nancy et al. [2022]. The Bi-LSTM smart health system accurately predicts heart disease risk with 98.86% accuracy [2]. Prakash Ramani et al. [2020] compares categorization-based model prediction accuracy. Artificial Neural Network (ANN) outperforms KNN, DT, SVM, and GNB in prediction accuracy. Models are made with ANN [3]. Unsupervised clustering methods examined by Sarangam Kodati et al. [2019] include hierarchical cluster, filtered cluster, basic k-means, OPTICS and furthest first. The display of processes is associated using unsupervised approaches. Building clusters takes time, and each has real positive and negative standards. The main objective is to compare Weka-implemented cluster algorithms to find the best one for the heart illness dataset [4]. The algorithm of MapReduce was proposed by T. Nagamani and S. Logeswari [2019] to predict coronary diseases. Using the MapReduce approach, the authors compared a meta-heuristic technique with a trained fuzzy neural system. Compared to the meta-heuristic approach, the designed system and the conventional fuzzy replication neural system achieved 98.12% accuracy on the 45 cases in the difficult group. Owing to dynamic schema and linear scaling, the designed MapReduce approach provided much more accurate results [5]. Syed Umar et al. [2018] performed neural network testing and learning through backpropagation. A superior procedure called genetic process was used to enhance the NN masses. Thus, this multi-layer neural network had 12 input nodes, 2 hidden nodes, and 2 output nodes. The number of contribution layers influences the possibility of heart disease, and that is quite an important aspect. Moreover, the training function in the network was used to update and track the weight and bias throughout the process. For practical implementation, the designed model was implemented with the help of Matlab R2012a, Universal Optimization Toolbox, and Neural System Toolbox. The accuracy

reached 96.2 for the training group, while the testing set got 89%. In total, 100 patient risk factors were generated [6]. KaanUyar et al. [2017] presented this approach clearly.

The authors then worked with patient data across 297 cases. Of these, 252 went into training and 45 were kept for testing. They compared the ANN-Fuzzy-AHP strategy with the RFNN approach, and on the testing set, the algorithm reached 97.78% accuracy [7].

J. H. and Miao [2016] made use of four distinct information systems namely, the Swiss University Hospital (SUH), Hungarian Association of Cardiology and Cleveland Hospital Foundations to detect heart ailments based on information classification and prediction models. The results achieved by the present research are more accurate than those achieved by previous researches [8]. Another study conducted by Rovina Britto et al. [2016] successfully developed an intelligent medical decision support system based on data mining which is quite efficient and effective. The accuracy rates for the Naive Bayes, Provision Route Engine, and Logistic Lapse were 75%, 80%, and 79% respectively. The authors focused on finding the best classifier, since better information withdrawal methods clearly led to stronger accurateness [9]. Purusothama et al. [2015] used several classification methods to diagnose HD through disease prediction models. The model trained and compared two main setups with two main setups: a single model and a combined, or hybrid, model. Both were trained on the data, then compared to see which one actually performed better. In this study, the attention was kept on classification approaches utilized in the single and joint models. The accuracy rates for each of them were, 86%, 58%, 76%, 69%, and 96% for the choice approach, connotation training, N Bayes, ANN, KNN and the mixture technique respectively. This difference cannot be overlooked. The mixture technique proved to be very successful and, frankly speaking, it appears to be more efficient than the rest. The authors claim that hybrid data mining techniques performed effectively and showed more accurate results in detecting heart disease [10].

2. PROPOSED WORK

This study uses data mining techniques to simplify prediction of the heart disease. Traditional methods for predicting the disease are not applicable to diverse medical datasets because of their high sensitivity to noise and inability to generalize. In order to solve these issues, the proposed approach employs systematic multi-phase data mining pipeline together with comparison and ensemble classification.

Dataset Selection

The study employs secondary data from public databases which are a part of the UCI Heart Disease Dataset. The dataset contains features such as age, resting blood pressure, thalassemia, chest pain type, cholesterol level and other symptoms which are required to determine the stenosis and disease probability.

Data Preprocessing

Taking into account inconsistencies in medical data sets, the following preprocessing will be carried out:

- 1) Data Cleaning: Locate and correct missing values, inconsistencies and duplicates by means of adding and/or deleting. Validation and smoothing reduce the noise level.
- 2) Normalization: Min-Max normalization ensures that the scaling is uniform. It is helpful for those algorithms which are sensitive to the magnitude of attributes.
- 3) Feature Selection: Perform feature selection by means of correlation analysis, information gain and recursive feature elimination. Unnecessary features will be deleted for simplicity and better comprehension.

Model Construction

Several machine learning classifiers are analyzed in terms of their ability to predict. Algorithms that were selected are:

- Naïve Bayes (NB)
- Decision Tree (DT)
- CN2 Rule Inducer
- Support Vector Machine (SVM)
- Random Forest (RF)
- Majority Voting Baseline

This multi-model setup allows recognizing algorithms for predicting the heart disease.

Model Training and Optimization

In this section, model is trained and optimized using the below steps

- 1) Cross-Validation: A 10-fold approach is employed to measure performance and minimize partition bias.
- 2) Optimize hyperparameters: Parameters such as estimators, SVM kernel choice, tree complexity and CN2 rule limits to make the most of ROC curve area under the curve (AUC).
- 3) Implementation of Procedures: The selection of features is followed by training classifiers using 75% of the data and testing on 25%. Classifiers provide predictions about classes based on a particular degree of confidence.

Performance Evaluation

Clinically substantial metrics assess model performance:

- Accuracy
- Area Under ROC Curve (AUC)
- Specificity (True Negative Rate)
- Sensitivity (True Positive Rate)

These metrics assess the classification's dimensions to differentiate between normal (<50% stenosis) and pathological (>50%) representations. The results of the experiment show that Naïve Bayes, Random Forest, and Decision Tree are better than other classifiers, with Naïve Bayes having the highest AUC.

Proposed Predictive Framework

A relative assessment makes the prediction framework effective by combining the most optimal classifiers. Predictions of the final model include binary decisions:

- 0: Stenosis < 50% (normal)
- 1: Stenosis $\geq$ 50% (high risk)

This portrayal helps healthcare providers better understand the clinical diagnostic requirements.

Deployment Considerations

A lightweight decision-support interface makes it easier to use in real life. Doctors fill out an interactive form with patient information. A system then:

- Makes predictions about diagnoses, shows a likelihood score, and keeps anonymous entries for ongoing learning and system improvement.
- This method of updating models over and over again makes them more reliable and makes it easier to connect hospital information systems.

3. RESULTS AND DISCUSSION

Systematic tests on the preprocessed and feature-selected dataset evaluated the suggested framework for predicting heart disease. We show and compare the quantitative results of several machine learning methods in terms of their ability to make predictions. The 10-fold cross-validation approach improved the reliability of the classifier, decreased the sampling bias, and provided the statistical evaluation of the classifier performance.

Accuracy, AUC (Area Under the ROC Curve), Specificity and Sensitivity metrics are clinically meaningful as they demonstrate the ability of the classifier to distinguish individuals with significant heart stenosis (>50%) from normal cases. Detailed analysis of the class label distribution in the dataset revealed the presence of some imbalances which may influence the performance of the classifier.

Algorithms' performance such as Naïve Bayes, SVM, CN2 Rule Inducer, Decision Tree and Random Forest are evaluated. The quality of prediction is influenced by the algorithm structure, dependencies of the attributes, and decision boundaries. This is an issue for categorization process. Graphical illustrations and summary tables help to interpret the classifier accuracy and AUC trends.

This section looks at how models behave, finds the best algorithms for predicting heart disease, and looks into what this means for clinical decision support systems.

Table 1 Classification Results

Output	Analysis/Details
Numeral	Heart Illness Judgement: 0 = No. 1= small possibility. $2 \geq 1$ $3 \geq 2$ 4 = high possibility.

Cross-validation yielded these results:

Table 2 Result of Different Algorithms

Algorithms	Accuracy
Decision tree	0.559
Majority	0.547
Naive bayes	0.545
Random forest	0.588
CN2	0.531

As Figure 8 demonstrates the values of the accuracy scores are very low and are far from meeting the set target of 85%. However, characteristics of the class indicate a significant transformation of the data that can improve the accuracy of the classifier. Indeed, the class not only includes the percentage of major heart vessels stenosis but also its probability. With the increasing value of the integer variable of the class, the percentage increases. The percentage is lower than 50% if the class equals 0. It has been proven that the occlusion greater than 50% creates substantial health problems for the patient with heart disease who needs surgery to open the conduit.

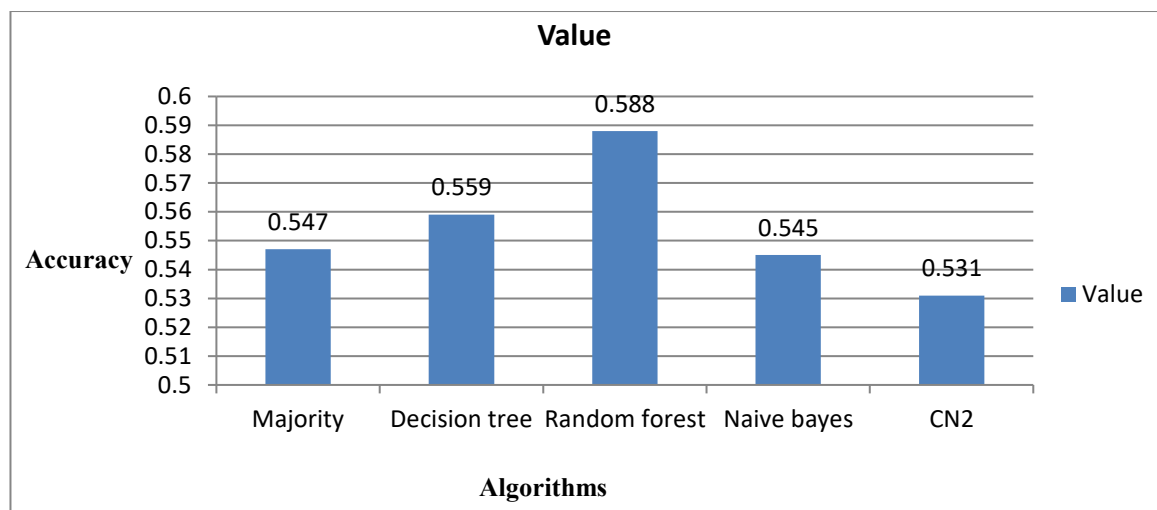


Figure 8 Outcome of Different Algorithms

Table 3 Classification Results

Output	Details
Numeral	Diagnosis: 0: stenosis of major cardiac vessels less than 50%; 1: stenosis of major cardiac vessels greater than 50%.

Class frequencies were evaluated to determine which quality indices to calculate in modeling and evaluation.

Table 4 Classification Results and Frequency Relative

Number	Frequency Relative
0	0.56
1	0.44

These frequencies show that the dataset's patients were not randomly selected from the population. Figure 9 shows that 46% of the population has heart disease, which is unimaginable.

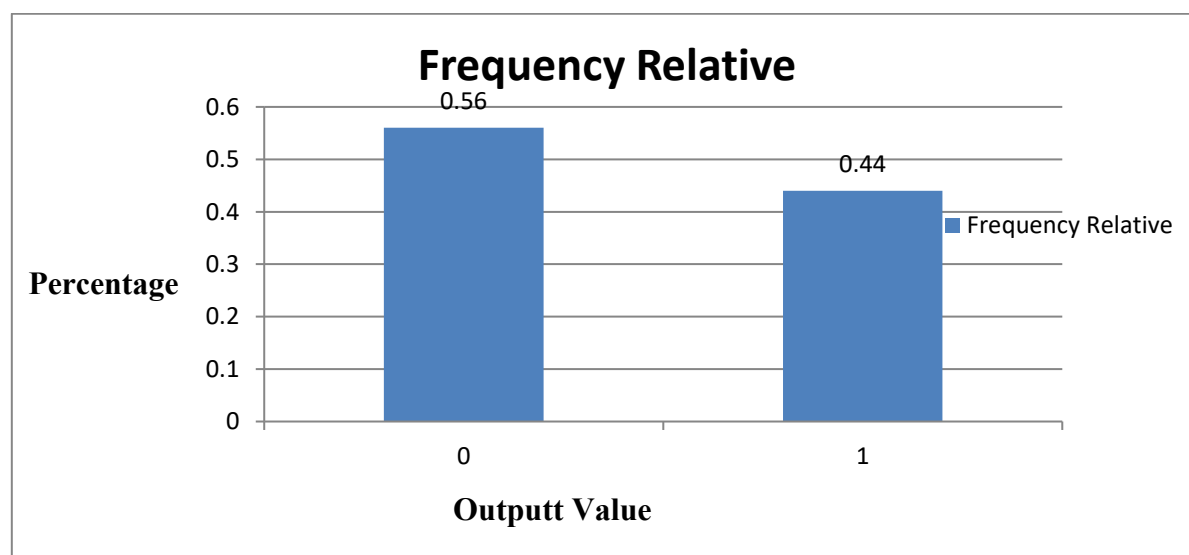


Figure 9 Frequency Relative Results

The remaining 75% of feature selection data is tested. Classifier quality indices were obtained following 10-fold cross validation. Accuracy, AUC, Specificity and Sensitivity were employed.

Random Forest

P is set to the square root of the number of features, thus $9 = 3$. Select the best AUC to stop tree development if there are 8 or less examples.

Table 5 Approximation using AUC

Size Bound	AUC
≤ 6	≤ 0.899
7	0.892
8	0.914
9	0.895
10	0.906
≥ 11	0.895

CN2

Ten is the best length for a rule. It looks like the size of the beam doesn't change the AUC values. So, this dimension was set to its default value of 5. Because the data was always changing, entropy was used as the evaluation metric. Even if you want to use a Laplace estimate, you need to do preventive discretization first. On the other hand, entropy is maximized by discretization. Finally, setting 0.17 as the minimum coverage when looking at the rules' minimum coverage makes AUC better.

Table 6 Minimum Coverage of Rules Evaluation

Cover	Approx
0.12	0.674
0.17	0.698
0.23	0.676
> 0.23	≤ 0.665

Cross-validation yielded these average results:

Table 7 AUC Assessment of Different Methods

Algorithms	Sensitivity	Specificity	AUC
Decision Tree	0.732	0.819	0.871
CN2 Rule Inducer	0.737	0.835	0.844
SVM	0.764	0.823	0.779
Naïve Bayes	0.786	0.862	0.899

The AUC is the evaluation parameter. We start by choosing classifiers based on AUC to find the best one for a given case. Figure 10 shows Nave Bayes because it is the most accurate algorithm for the basic purpose.

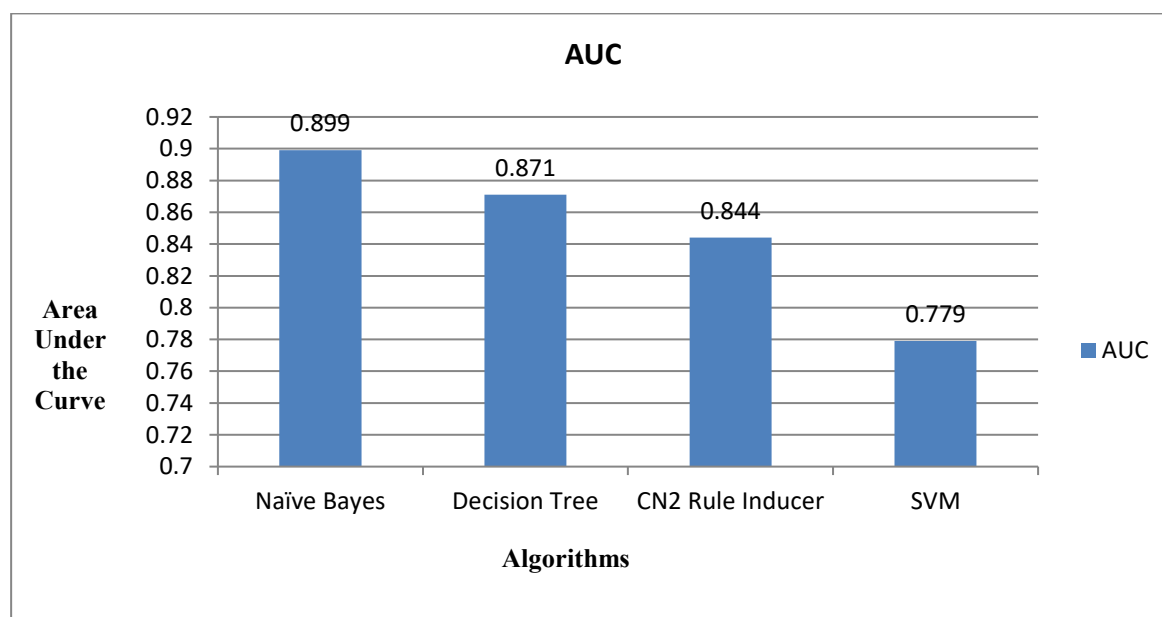


Figure 10: AUC Assessment of Different Methods

CONCLUSION AND FUTURE DIRECTION

This study offers a model for prediction of heart disease based on various methods of data mining and machine learning. In the presented research, such methods as preprocessing, normalization, feature selection, classification adjustment and 10-fold cross validation have been used to build the model, whose results are accurate and clinically feasible. In experiments, Naïve Bayes, Random Forest and Decision Tree classifiers have demonstrated the greatest prediction abilities with respect to AUC, sensitivity and specificity. Naïve Bayes showed the greatest discriminative power of all three classifiers and can be considered as a good decision-making model.

Based on the results obtained, it can be concluded that improved multi-algorithm models predict heart disease better than single classifiers models. Clinical categorization of output (like <50% vs. >50% stenosis) increases the clinical usefulness of the model. The proposed approach can become a highly valuable diagnostic tool for doctors who need to detect high-risk patients. However, the proposed model has numerous possibilities for improvements in performance and practicality despite its promising theoretical appearance:

- The usage of datasets collected from several regions and different institutions improves the chances of successful generalization and reduces the risk of biasing and hence allows using this model on a broad scale.
- The model can be integrated into wearables and hospital systems which are part of the Internet of Things (IoT) and are designed to estimate the risk in real-time mode.

REFERENCES

- [1] K. Butchi Raju, Suresh Dara, "Smart Heart Disease Prediction System with IoT and Fog Computing Sectors Enabled by Cascaded Deep Learning Model", *Hindawi Computational Intelligence and Neuroscience* Volume 2022, Article ID 1070697, 22 pages <https://doi.org/10.1155/2022/1070697>
- [2] A Angel Nancy, Dakshanamoorthy Ravindran, "IoT-Cloud-Based Smart Healthcare Monitoring System for Heart Disease Prediction via Deep Learning", *Electronics* 2022, 11, 2292. <https://doi.org/10.3390/electronics11152292>
- [3] Prakash Ramani, Nitesh Pradhan, "Classification Algorithms to Predict Heart Diseases—A Survey", *Computer Vision and Machine Intelligence in Medical Image Analysis*, pp 65-71, 2020
- [4] Sarangam Kodati, R. Vivekanandam G. Ravi, "Comparative Analysis of Clustering Algorithms with Heart Disease Datasets Using Data Mining Weka Tool", *Soft Computing and Signal Processing*, pp 111-117, 2019
- [5] T. Nagamani, S. Logeswari, B. Gomathy, "Heart Disease Prediction using Data Mining with Mapreduce Algorithm", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Volume-8, Issue-3, 2019
- [6] Syed Umar Amin, Agarwal, K., & Beg, R., "Genetic neural network-based data mining in prediction of heart disease using risk factors", *IEEE Conference in Information & Communication Technologies (ICT)*, 1227-1231, 2018.
- [7] KaanUyar, Ahmetİlhan, "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks", *9th International Conference on Theory and Application of Soft Computing*, Procedia, 2017
- [8] J. H., & Miao, G. J., "Diagnosing Coronary Heart Disease Using Ensemble Machine Learning", *International Journal of Advanced Computer Science and Applications*, 7(10), 30-39, 2016
- [9] Dbritto, R., Srinivasaraghavan, A., & Joseph, V., "Comparative Analysis of Accuracy on Heart Disease Prediction using Classification Methods", *International Journal of Applied Information Systems (IJ AIS)*, 11(2), 2016
- [10] Purusothaman, G., & Krishnakumari, P., "A survey of data mining techniques on risk prediction: Heart disease", *Indian Journal of Science and Technology*, 8(12), 1, 2015
- [11] Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald and Edin Muharemagic, "Deep learning applications and challenges in big data analytics", *Journal of Big Data* (2015) 2:1
- [12] Dane Morgan and Ryan Jacobs, "Opportunities and Challenges for Machine Learning in Materials Science", *Annu. Rev. Mater. Res.*, 2020. 50:71–103
- [13] Neha Sharma, Reecha Sharma, Neeru Jindal, "Machine Learning and Deep Learning Applications-A Vision", *Global Transitions Proceedings 2* (2021) 24–28
- [14] Niklas Kühl, Marc Goutier, Robin Hirt, Gerhard Satzger, "Machine Learning in Artificial Intelligence: Towards a Common Understanding", *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019
- [15] Niklas Kühl, Max Schemmer, Marc Goutier, Gerhard Satzger, "Artificial intelligence and machine learning", *Electronic Markets* (2022) 32:2235–2244, <https://doi.org/10.1007/s12525-022-00598-0>
- [16] Arsalan Khan, Moiz Qureshi, Muhammad Daniyal, and Kassim Tawiah, "A Novel Study on Machine Learning Algorithm-Based Cardiovascular Disease Prediction", *Hindawi Health & Social Care in the Community*, Volume 2023, Article ID 1406060, 10 pages
- [17] Chintan M. Bhatt, Parth Patel, Tarang Ghetia and Pier Luigi Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques", *Algorithms* 2023, 16, 88., <https://www.mdpi.com/journal/algorithms>
- [18] Daniyal Asif, Mairaj Bibi, Muhammad Shoaib Arif and Aiman Mukheimer, "Enhancing Heart Disease Prediction through Ensemble Learning Techniques with Hyper parameter Optimization", *Algorithms* 2023, 16, 308., <https://www.mdpi.com/journal/algorithms>