International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(18s) 2026 pp. 1059-1073
https://doi.org/10.70917/ijcisim-2026-4941
Article

A Prediction Model for the Early Detection of Angina Pectoris by Using Machine Learning Approaches

Umakant[1], Deepak Dudeja

[1,2]Maharishi Markandeshwar (Deemed To Be University), Ambala, India

Abstract: Angina Pectoris remains one of the leading causes of mortality worldwide, and early detection is essential for effective intervention. Recent advancements in data mining and machine learning have enabled the development of intelligent diagnostic systems capable of analyzing large volumes of clinical data. This study proposes an enhanced prediction model for early detection of Angina Pectoris using multiple data mining approaches, including Decision Tree, Random Forest, Naïve Bayes, CN2 Rule Inducer, Support Vector Machine, and a Majority Voting baseline. The model employs a systematic pipeline comprising data cleaning, normalization, feature selection, classifier optimization, and cross-validation. Experiments conducted on a benchmark Angina Pectoris dataset demonstrate that the combined use of optimized classifiers significantly improves predictive performance. Among the evaluated algorithms, Naïve Bayes and Random Forest achieve the highest accuracy and AUC, indicating their suitability for clinical decision support. The proposed framework can assist healthcare professionals by providing automated, reliable risk assessment and can be integrated into practical diagnostic systems for continuous model improvement. of the patient's data and provide a qualitative assessment of the patient's risk of Angina Pectoris. A natural ordinal number between 0 and 4, where 0 indicates that the patient is healthy and 4 indicates that there is a substantial possibility that he is, is used to express this probability. If the classifier can correctly forecast the patient's state of health in 85% of the cases or if its accuracy in absolute terms surpasses 85%, the findings will be considered remarkable.
Keywords: Majority, Decision tree, Random Forest, Naïve Bayes, Decision Tree, Support Vector Machine

1. INTRODUCTION

By employing information withdrawal, which has significant potential, the healthcare industry may set up health arrangements to totally practice facts and analytics to identify ineffectiveness and the finest performs that increase therapy and save budgets. Hospital information systems are used by most hospitals nowadays to handle huge and gigantic volumes of patient data. These data hold a plethora of unexplored hidden knowledge that might be converted into helpful information that helps medical professionals make informed treatment decisions with the aid of data mining [1].

1.1 Data Mining

This method mines hidden configurations from large records by combining machine learning, databank expertise, and numerical inquiry. Furthermore, the growing volume of frequent submissions in the thriving healthcare industry makes the removal of medical data an essential component of an inquiry. Once all global fatalities have been taken into account, Angina Pectoris becomes the decisive factor [2]. When determining whether a long-term patient may have heart disease, medical professionals face an intriguing debate because it calls for general challenges and the presence of expertise [3] [4].

Much iteration has been made to construct models that use data mining either by itself or in conjunction with computational techniques. The data pullout method, which uses a combinational approach from mathematical study, device knowledge, and catalogue experience to uncover undiscovered configurations, is applied to large datasets [5].

Data mining is one of the most advanced analytical techniques available today; other tools in this toolset include machine knowledge practices (e.g., decision trees or neural systems), measurable methods, and arithmetical models. Data quarrying is one of the most well-liked areas of research that has applications in numerous academic domains. In the fields of communications, banking, health, education, and business markets, data withdrawal is commonly sacrificed in an effort to lure in new customers, combat illness, reduce costs, and increase benefits. Data mining can be referred to by a variety of terms, such as information extraction, facts extraction from statistics, statistics/arrangement investigation, and statistics examination, all of which may have quite distinct meanings. Finding patterns in statistics that may be utilised to comprehend and predict behavior is the aim of statistics removal. Data mining models consist of a series of directives, formulas, or intricate "transmission purposes." They fall into two main categories based on their goals, which are as follows:

- Supervised / Analytical Prototypes.
- Unsupervised Prototypes.

1.1.1 **Supervised / Analytical Prototypes**

The objective of modelling that is supervised, predictive, directed, or targeted is to estimate the parameters of a continuous numerical attribute or forecast an occurrence. These models include both contribution arenas or attributes and production arenas or objective arenas. Input fields are also referred to as predictors since they assist the model in selecting a forecast utility for the production arena.

1.1.2 Unsupervised Prototypes

There is only an input field in unverified or objectiveless mockups; there is no production arena. Undirected arrangement acknowledgement means that it is not influenced by a particular objective aspect. These models seek to identify data trends among the input fields.

Unsupervised prototypes comprise:

1. Cluster mockups.
2. Relationship and series mockups.

1.2 Mixture Recommender Arrangements

Mixture recommender schemes combine many strategies, and they minimize the downsides of some by modifying the rewards of others. In order to progress improved classification optimization and get beyond some of the restrictions and challenges associated with clean reference schemes, hybrid filtering systems pool several recommendation techniques [6].

1.3 Angina Pectoris

In the present-day state of medication, Angina Pectoris illness is a key worldwide wellbeing matter. The twenty-first century saw a prominent change in the origins of Angina Pectoris illness demise global as well as an excellent development in lifespan expectation. Currently, it is assessed that there will be a worldwide drop of about 30%, with the high-income nation longsighted a reduction of around 40% and the little and mid income nations feeling a reduction of around 28%.

This continuing changeover is scattering above the sphere at an even quicker frequency than in the earlier century amongst all participants, national assemblages, and nations due to economic development, suburbanization, and connected everyday bodily variations. The commonness of Angina Pectoris letdown has been extremely amplified by modern variations in life. Agreeing to a fresh revision, Angina Pectoris letdown signs have triplicated over the previous 25 years. Rendering to a current revision, prolonged noninfectious illnesses like Angina Pectoris illness are solitary of the key reasons of demise global. Internationally, there has been a incredible change in publics' wellbeing standing, which has led to an rise in cardiac infection [5].

The foremost root of demise global every day is currently Angina Pectoris. The considerable change in publics' health standing round the world is what sources the worldwide upsurge in cardiac disease. Over the previous 20 years, Angina Pectoris illness has worryingly amplified day by day and has become one of the foremost reasons of demise in the bulk of nations through the world. Allowing to a new reading on cardiac fitness, about 1.2 billion individuals worldwide expire from Angina Pectoris illnesses every year. Assumed the massive variations in communal, national, and financial areas, there is no one resolution to the enlarged threat of Angina Pectoris illness. Forecasting heart letdown the day beforehand a high-cost proportion occasion is characteristically a very problematic responsibility. The price of cutting-edge imaging and controlled approaches for heart illness investigation is moreover great [7]. Key indications of a cardiac disorder comprise chest uneasiness, dyspnea, exhaustion, edema, palpations, and syncope, cough, hemoptysis, and cyanosis are other instances.

1.4 Machine Learning

Machine learning is demonstrating accurate estimates and assessment in the age of technology. A machine learning system uses a training set of finished projects to successfully "learn" how to estimate. Supervised machine learning, unsupervised machine learning, and reinforcement machine learning are a few examples of machine learning approaches [9].

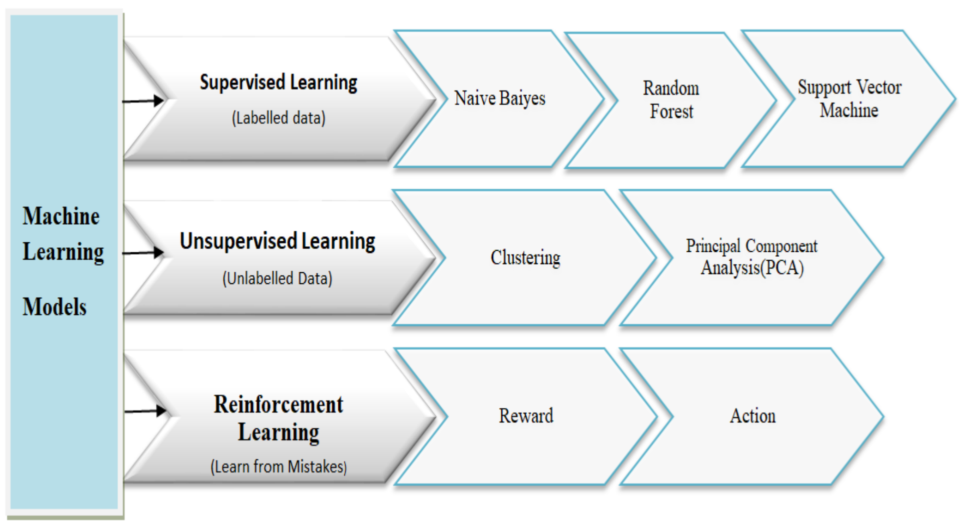

Figure 1: Machine Learning Models: Supervised, Unsupervised and Reinforcement [11]

1.4.1 **Supervised Learning**

An external support and "factual" or "accurate" labels of the input dataset are accessible in supervised machine learning algorithms [12]. A training dataset and a test dataset have been created from this input dataset. Test data has been linked to an output variable that has to be anticipated. All algorithms learn specific patterns from the training dataset and apply them to the test dataset in order to do classification and prediction. Almost all machine learning techniques include optimizing either an objective function or a cost function.

Typically, the error between the ground fact and the algorithm approximations is measured by the cost function. We train our model to generate estimates that are reasonably near to the actual values (ground truth) by minimizing the cost function. The cost function can be minimized by applying the gradient descent technique [13][14]. Within the machine learning technique, various models combine, each with unique capabilities based on the circumstance and application.

Classification

From the perspective of machine learning approaches, classification is crucial. It involves two steps. The two stages are learning, which involves creating a classification model, and classification, which involves predicting object classes using the classification model. Following the procedures of data preparation, model creation, feature selection, and evaluation, these learning methods can be applied to new data [15].

Categorical or unconditional continuous valued functions are predicted by classification. For example, a classification model can be developed based on bank loan applications that are categorized in the database as safe or dangerous. The category of Supervised Machine Learning algorithms includes a wide range of classification techniques. The representation of the categorization technique is shown in the figure. First, consideration is given to the loan decision parameters. It included a number of variables, including name, age, and income. Following that, a classification model will be used. Following completion of this process, an output based on predetermined inputs will be generated that will indicate whether or not an individual is qualified to get a bank loan.

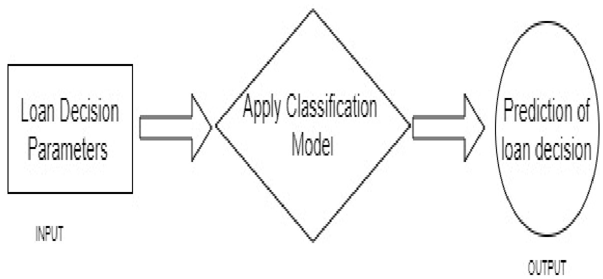

Figure 2: An Instance of Classification Technique [15]

Naïve Bayes

An additional classification method based on the Bayes theorem predicts that a characteristic that is now picked for a class is likely to be independent of another feature in that class.

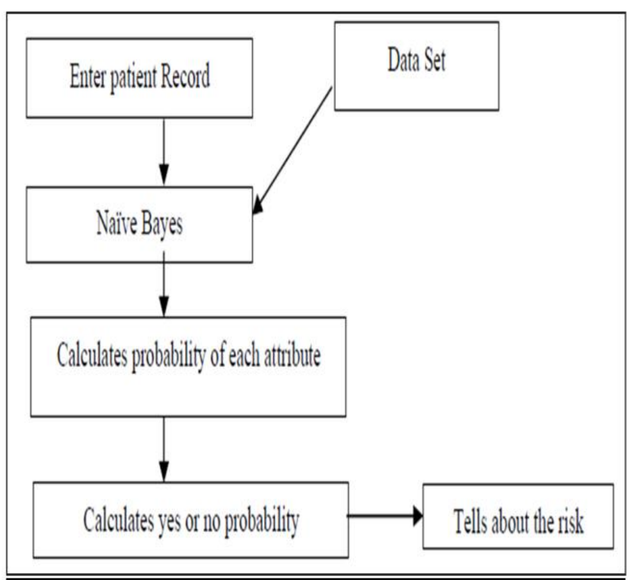

Figure 3: Bayesian Classifier [15]

Random Forest

This alternative kind of supervised learning algorithm is employed in regression analysis and classification. However, it is mostly applied to classification issues and yields suitable outcomes. It is an amalgam of several different decision trees. Decision trees are created using data samples, and each tree's anticipated values are obtained. Finally, a vote will be used to determine which approach is best. The graphic below displays a random forest's operational illustration.

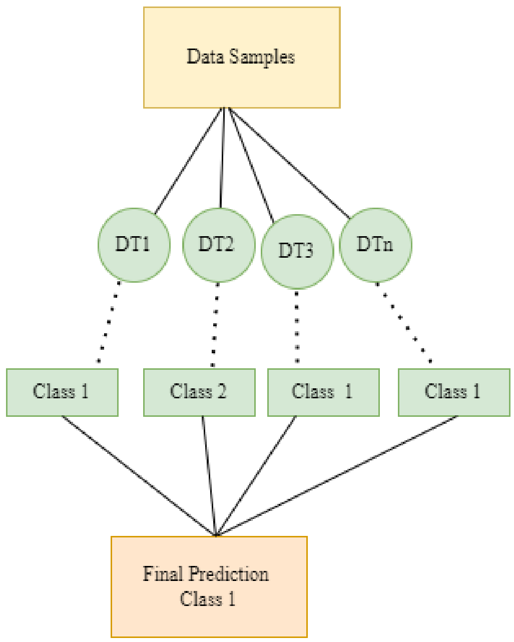

Figure 4: Schematic diagram of Random Forest Classifier [15]

Support Vector Machine (SVM)

By identifying the hyperplane that widens the gap between the two classes, a support vector machine calculates classification. Support vectors are the vectors, or cases, that round the hyperplane. It falls under the category of supervised learning, which is applicable to studies including both regression and classification.

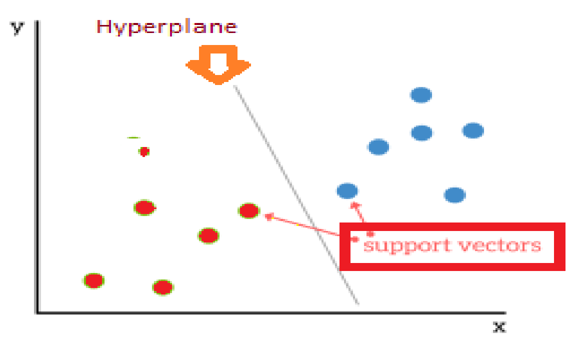

Figure 5: Support Vector Machine w.r.t different Support Vectors and Hyperplane [15]

Decision Tree

Researchers have successfully looked at the demonstration of the Decision Tree approach in the handling of Angina Pectoris. An internal node, a branch, and a leaf node make up a decision tree's tree-like structure. Each branch signifies a characteristic worth, individually inside nodule a trial on a characteristic that is utilized for, and the shoot lump the foretold periods or discussion deliveries. Opening at the source node, the categorization proceeds up the tree depending on the predictive attribute value. In order to construct trimmed decision trees, the process includes information dividing, information categorization, decision tree grouping collection, and the appeal of lessening of accountability adornment. There are dual kinds of cataloging methods: managed and unproven methods. Chi join and entropy are castoff in managed arrangement approaches, whereas similar width and identical frequency are used in unsupervised approaches. Testing with or without voting is done during the data splitting. The Gini Index, Information Improvement, and Gain Ratio Decision Tree kinds are examined. Lastly, fewer closed decision rules can be provided through reduced error cutting. The ID3 algorithm implementation on patient data is shown in the figure 6.

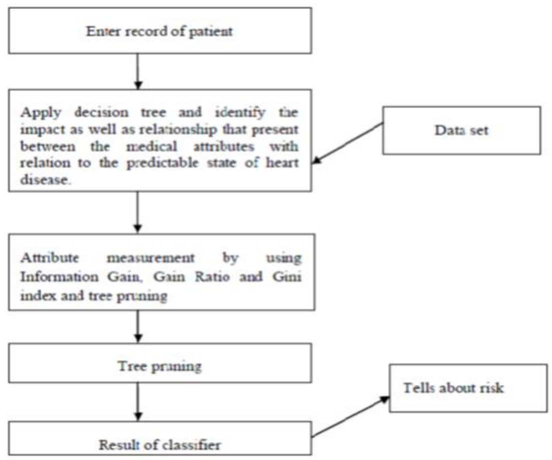

Figure 6: Decision Tree

Neural Network

Neural networks are well renowned for producing incredibly precise outcomes in real-world applications. The NN is trained using the Angina Pectoris database using the feedstuff advancing neural system ideal 10, adjustable culture degree, and back broadcast knowledge process with motion. The model's design is as follows: Clinical data are first entered, and then an ANN algorithm is developed. The prediction results can be produced by the model following training. The categorization of clinical data into two identical shares at random is the first stage in the neural network algorithm's computational process. Both are employed, one for training and the other for testing. Each characteristic is given a starting weight at random. All feature weights are modified using the estimated errors. When the mistakes coincide with the termination requirements, each feature's final weight is determined. Many times are spent repeating the process. We may compute the presentation outcomes from the challenging statistics after building the training models. The neural system procedure implementation on scientific statistics is shown in the figure 7.

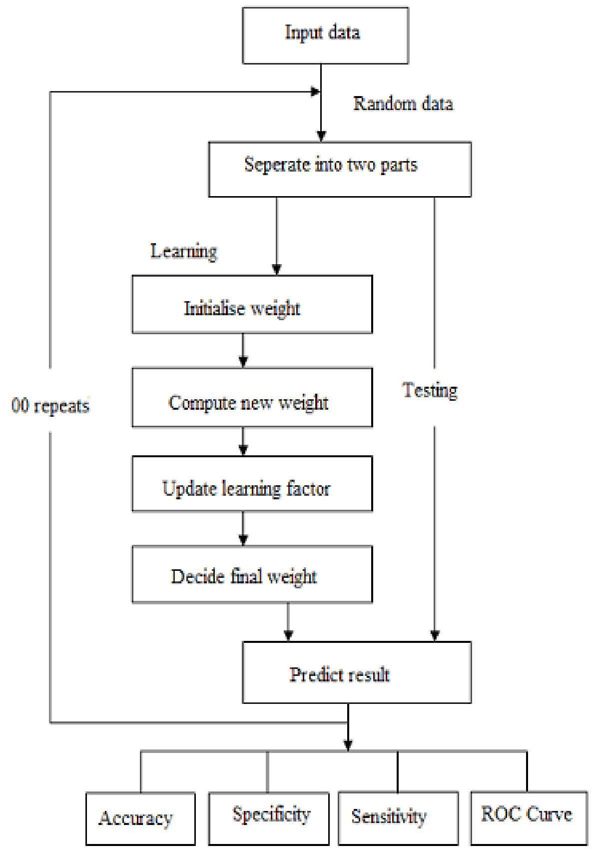

Figure 7: Neural Network

2. RELATED WORK

Arsalan Khan et al. [2023] compared the performance of suggested machine learning algorithms such as decision tree (DT), random forest (RF), logistic regression (LR), Naive Bayes (NB), and support vector machine (SVM) under a variety of scenarios in order to identify the best suited machine learning algorithm in the class of models [16]. Chintan M. Bhatt et al. [2023] presents a k-modes clustering approach using Huang beginning, which can increase classification accuracy. Random forest (RF), decision tree classifier (DT), multilayer perceptron (MP), and XGBoost (XGB) models are employed. GridSearchCV was used to optimize the output by hyper tuning the parameters of the applied model [17]. Daniyal Asif et al. [2023] offer a machine learning model for predicting cardiac illness that uses multiple preprocessing processes, hyper parameter optimization approaches, and ensemble learning algorithms [18]. K. Butchi Raju et al. [2022] suggested a paradigm that combines cloud, fog, and edge computing to provide results quickly and accurately. Hardware elements gather data from various patients. Sentiment story withdrawal from signs is carried out to get significant characteristics [1].

A Angel Nancy et al. [2022] presented a system that gathers information from IoT strategies and subjects' longsuffering past correlated microelectronic healthcare statistics stored in the fog to extrapolative analytics. The Bi-LSTM constructed clever well-being arrangement for tracing and exactly guessing the danger of Angina Pectoris has an accuracy of 98.86% [2]. Prakash Ramani et al. [2020] this study compares the prediction accuracy of models created using various categorization techniques. In comparison to other algorithms like (KNN), (DT), (SVM), and (GNB), Artificial Neural Network (ANN) has the best prediction accuracy. ANN is used for creating models [3]. Sarangam Kodati et al. [2019] examine many unsupervised clustering techniques, including basic k-means approach, filtered cluster hierarchical cluster, OPTICS, and furthest first. In order to associate the presentation of the processes, unsupervised methods are utilized. It takes time to construct the clusters, and each cluster is distinguished by its genuine positive and actual undesirable standards.

The primary goal is to compare cluster algorithms that are examined using the Weka implement and determine which established of procedures may be best suitable for the Angina Pectoris dataset [4]. T. Nagamani and S. Logeswari [2019] recommended using the MapReduce algorithm for coronary disease prediction. In order to predict Angina Pectoris, this study employs the Map reduction method to compare a meta-heuristic methodology with an accomplished tireless ambiguous neural system. When compared to meta-heuristic method, arranged neural system, and traditional ambiguous reproduction neural system, the findings of this revision attained 98.12% accurateness for the 45 occasions of the challenging group. Due to dynamic schema and linear scaling, the suggested MapReduce method produces output accuracy that is noticeably superior [5]. Syed Umar et al. [2018] neural network testing and learning were conducted using the backpropagation technique. A premium process well-known as the genetic procedure was castoff to improve the NN masses.

Hence, there were two output nodes, two hidden nodes, and twelve input nodes in this multi-layered network. The figure of contribution layers disturbs the danger of Angina Pectoris. The network training function helps to update and record Weight and Bias. The Universal Optimization Toolbox, Neural System Toolbox, and Matlab R2012a were exploited in the execution. The correctness outcomes for the preparation group were 96.2, while the accuracy results for the testing set were 89% 100 patient risk components were created [6]. KaanUyar et al. [2017] Angina Pectoris discovery using trained neural networks based on a natural selection approach is presented. The authors used patient data in 297 different situations overall. 45 of them were chosen for testing, and 252 of them were used for training.

The ANN-Fuzzy-AHP strategy and this RFNN approach were contrasted by the authors. By examining the testing set, the aforementioned algorithm achieved a result of 97.78% accuracy [7]. J. H. and Miao [2016] In order to diagnose cardiac illness, four separate information groups—the Cleveland Hospital Groundwork, the Hungarian Organization of Cardiology and the Swiss University Hospital (SUH)—were examined for knowledge cataloguing and forecast mock-ups'. It was determined that the results achieved here were more accurate than those from earlier studies that had been published [8]. Rovina Britto et al. [2016] built a data mining-based intelligent medical decision support system that is effective and efficient.

The accuracy using Nave Bayes, Provision Route Engine, and Logistic Lapse was 75%, 80%, and 79% correspondingly. The writers decided to emphasis on identifying the best classifier that has the ability to deliver improved accurateness by utilizing information withdrawal methods [9]. Purusothama et al. [2015] In order to diagnose HD, a variety of classification techniques for illness prediction models were utilized. The key prototypical, a single model, and the secondary model, a combined model likewise recognized as a hybrid model, were both used to train the data and were then compared. These binary mockups were castoff in the statistics examination. The writers ensure only taken into account the cataloging strategies for together the solo model and the joint model. The findings of comparing algorithms with accuracy of 76%, 58%, 86%, 69%, and 96%, respectively, include choice technique, connotation instruction, KNN, ANN, N Bayes, and mixture methodology. The author suggested that mixture information withdrawal procedures function effectively and that accurate cardiac disease detection findings were reached [10].

3. PROPOSED WORK

The objective of this study is to develop an enhanced prediction model for early detection of Angina Pectoris by integrating multiple data mining techniques. Existing single-algorithm approaches often suffer from limited generalization capability, sensitivity to noise, and reduced performance on heterogeneous medical datasets. To address these limitations, the proposed methodology adopts a systematic, multi-stage data mining pipeline followed by comparative and ensemble-based classification.

Dataset Selection

The study utilizes secondary data obtained from publicly available repositories, specifically the UCI Angina Pectoris Dataset. The dataset contains clinically relevant parameters such as age, cholesterol level, resting blood pressure, chest pain type, thalassemia, and other diagnostic indicators essential for computing stenosis probability and overall disease risk.

Data Preprocessing

Given the typical inconsistencies present in medical datasets, preprocessing is performed as follows:

1) **Data Cleaning:**

Missing values, inconsistent entries, and duplicate records are identified and addressed using appropriate imputation or removal strategies. Noise is handled through smoothing and validation techniques.

2) **Normalization:**

To ensure uniform scaling across features, Min–Max normalization is applied. This is particularly beneficial for algorithms sensitive to attribute magnitude.

3) **Feature Selection:**

Correlation analysis, information gain, and recursive feature elimination are used to identify the most discriminative features. Reducing irrelevant attributes improves model interpretability and prevents overfitting.

Model Construction

Multiple machine learning classifiers are implemented to evaluate their individual and comparative predictive capabilities. The selected algorithms include:

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

This multi-model configuration enables the identification of algorithms best suited for Angina Pectoris predictions and facilitates ensemble-based refinement.

Model Training and Optimization

1) Cross-Validation:

A 10-fold cross-validation strategy is employed to obtain reliable performance estimates and mitigate partition bias.

2) **Hyperparameter Tuning:**

Key parameters—such as maximum tree depth, number of estimators, SVM kernel selection, and CN2 rule constraints—are optimized to maximize the area under the ROC curve (AUC).

3) **Algorithm Execution:**

The models are trained on 75% of the dataset (post feature selection) and evaluated on the remaining 25%. Each classifier generates probability-based predictions for the defined outcome classes.

Performance Evaluation

Model performance is assessed using clinically significant metrics:

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

These metrics provide a comprehensive evaluation of the system's ability to distinguish between normal (<50% stenosis) and abnormal (>50% stenosis) cases. Experimental results indicate that Naïve Bayes, Random Forest, and Decision Tree yield superior performance compared to other classifiers, with Naïve Bayes achieving the highest AUC.

F. Proposed Predictive Framework

Based on comparative evaluation, a refined predictive framework is developed by integrating the best-performing classifiers. The final model provides a binary decision output:

- 0: Stenosis < 50% (normal)
- 1: Stenosis ≥ 50% (high risk)

This representation aligns with clinical diagnostic standards and improves interpretability for healthcare practitioners.

Deployment Considerations

To enhance practical usability, a lightweight decision-support interface is recommended. Clinicians can input patient attributes through an interactive form. The system then:

1. Computes the predicted diagnosis;
2. Displays the associated probability score;
3. Stores anonymized entries in a centralized repository for continuous learning and system improvement.

This iterative update mechanism enhances long-term model reliability and supports seamless integration with existing hospital information systems.

4. RESULTS AND DISCUSSION

The performance of the proposed Angina Pectoris prediction framework was evaluated through a series of systematic experiments conducted on the preprocessed and feature-selected dataset. This section presents the quantitative results obtained from multiple machine learning algorithms, followed by a comparative discussion of their predictive capabilities. A 10-fold cross-validation strategy was employed to ensure robustness, reduce sampling bias, and provide a statistically reliable estimate of classifier performance.

The evaluation focuses on clinically relevant metrics—sensitivity, specificity, accuracy, and AUC (Area Under the ROC Curve)—as these indicators reflect the model's ability to correctly distinguish between patients with significant cardiac stenosis (>50%) and those within normal limits. Additional analysis is provided to examine the distribution of class labels within the dataset, revealing potential imbalances that may influence classifier behavior.

Furthermore, the performance of different algorithms such as Decision Tree, Random Forest, Naïve Bayes, CN2 Rule Inducer, and SVM is compared to identify the most effective models. The results demonstrate varying degrees of predictive strength, highlighting the impact of algorithmic structure, feature dependencies, and decision boundaries on classification outcomes. Graphical representations and tabular summaries are included to facilitate interpretation and illustrate trends in classifier accuracy and AUC values.

Overall, this section provides a detailed examination of model behavior, identifies the optimal algorithms for Angina Pectoris prediction, and discusses the implications of these findings for clinical decision support systems.

Table 1: Classification Results

Output	Details
Number	Angina Pectoris Diagnosis: 0 = No. 1= low probability. 2 ≥ 1 3 ≥ 2 4 = high probability.

The results obtained at the end of the cross-validation procedure are the following:

Table 2: Outcome of Various Algorithms

Algorithms	Value
Majority	0.545
Decision tree	0.557
Random forest	0.587
Naive bayes	0.534
CN2	0.522

It is apparent in figure 8 how low the acquired accuracy numbers are and how poorly the objective (accuracy 85%) is met. The class's characteristics, however, point to a considerable dataset adjustment that could have a significant positive impact on classification accuracy. In fact, the class qualitatively shows both the proportion of stenosis in the major cardiac vessels as well as the qualitative chance of having the condition. A percentage of stenosis less than 50% exists when the class is 0, a percentage of stenosis higher than 50% exists when the class is greater than or equal to 1, and the percentage of stenosis increases as the integer value of the class increases. Furthermore, it is clear from the literature that an occlusion greater than 50% poses major health hazards to the patient, who is diagnosed with Angina Pectoris and requires surgery to reopen the blocked conduit.

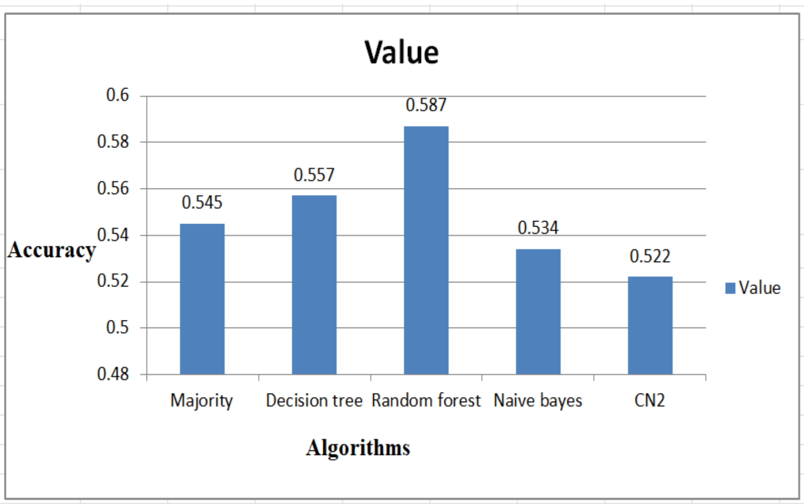

Figure 8: Outcome of Various Algorithms

Table 3: Classification Results

Output	Details
Number	Diagnosis: 0: stenosis of major cardiac vessels < 50%, 1: stenosis of major cardiac vessels >50%.

An evaluation of the frequencies of the class has been carried out, which could indicate which quality indices to calculate in the modeling and evaluation phase.

Table 4: Classification Results

Number	Frequency Relative
0	0.54
1	0.46

These frequencies suggest that the choice of patients within the dataset was not carried out randomly on a sample of subjects belonging to the population. It would be unthinkable, in fact, that 46% of the population is affected by Angina Pectoris as shown in figure 9.

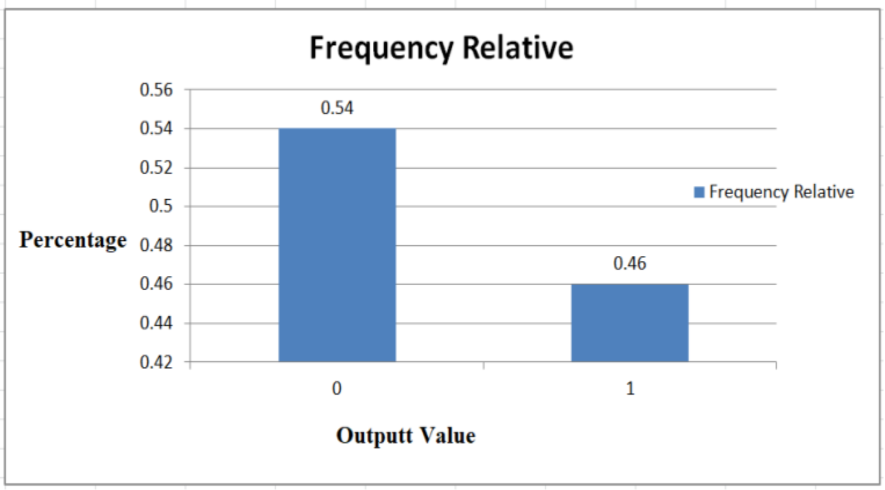

Figure 9: Outcome of Frequency Relative

The testing procedure is performed on the 75% of data remaining from the feature selection. Quality indices of the classifiers obtained after a 10-fold cross validation procedure were obtained. In particular, indices such as sensitivity, specificity, accuracy and AUC were used.

Random Forest

Having 9, p is equal to 3 because P is set to be equal to the square root of the number of characteristics. Additionally, if the number of examples is 8 or less, stop the growth of the trees by selecting the value that produces the best AUC.

Table 5: AUC Estimation

Size Limit	AUC
<=6	≤0.889
7	0.891
8	0.911

9	0.891
10	0.901
>=11	0.897

CN2

The optimum rule length is ten. It would seem that the results in terms of AUC are unaffected by the size of the beam. Consequently, it was decided to set this dimension to 5, which is its default value. Entropy was also selected as the evaluation metric due to the existence of continuous data. In reality, even while a Laplace estimate is preferred; using one would necessitate preventive discretization, whereas for the entropy, discretization is done to maximize it. Finally, it is found that by selecting 0.17 as the minimum coverage while analyzing the rules' minimal coverage that the AUC is improved.

Table 6: Evaluating Minimum Coverage of Rules

Cover	**Approx**
0.12	0.673
0.17	0.697
0.23	0.677
> 0.23	≤0.655

The following average outcomes were achieved following the cross-validation process:

Table 7: AUC Evaluation of Various Algorithms

Algorithms	**Sensitivity**	**Specificity**	**AUC**
Naïve Bayes	0.785	0.863	0.899
Decision Tree	0.731	0.817	0.873
CN2 Rule Inducer	0.736	0.832	0.841
SVM	0.764	0.821	0.778

The evaluation parameter is thought of as the AUC. We begin by selecting them based on the AUC value in order to determine which classifier (or classifiers) should be regarded as the best for such a situation. The algorithms with maximum accuracy are chosen in accordance with the basic goal, so Nave Bayes as shown in figure 10.

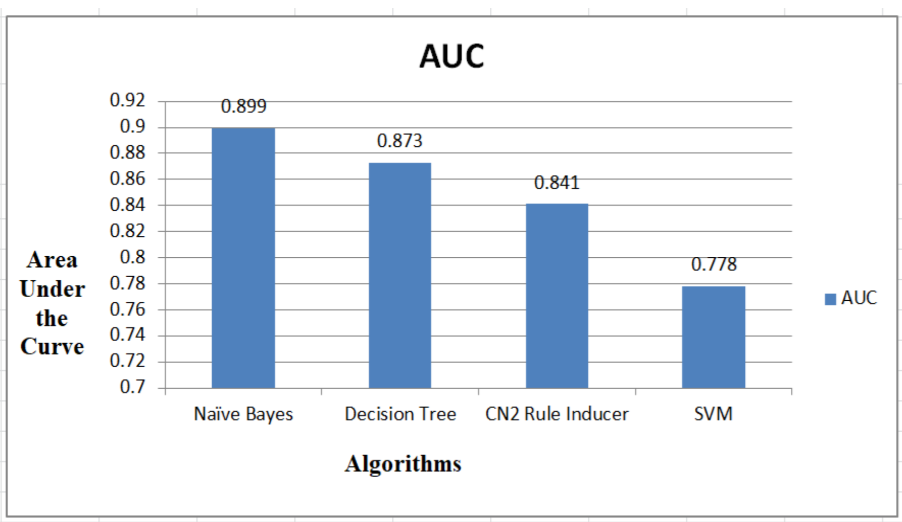

Figure 10: AUC Evaluation of Various Algorithms

5. CONCLUSION AND FUTURE SCOPE

This study presented a data-driven prediction model for early detection of Angina Pectoris using multiple data mining and machine learning techniques. The proposed framework incorporated a complete analytical pipeline—including preprocessing, normalization, feature selection, classifier optimization, and 10-fold cross-validation—to ensure reliable and clinically meaningful results. Experimental evaluation demonstrated that algorithms such as Naïve Bayes, Random Forest, and Decision Tree achieved superior predictive performance, particularly in terms of AUC, sensitivity, and specificity. Among all models, Naïve Bayes exhibited the highest discriminative ability, making it a strong candidate for clinical decision-support applications.

The results confirm that an optimized multi-algorithm approach can significantly improve the accuracy of Angina Pectoris prediction compared to conventional single-classifier models. Furthermore, transforming the output into clinically interpretable categories (<50% vs. >50% stenosis) enhances the usability of the model for medical practitioners. Overall, the proposed system demonstrates strong potential as a supportive diagnostic tool capable of assisting physicians in identifying high-risk patients at an early stage.

Although the proposed model provides promising results, several opportunities remain for advancing its performance and real-world applicability:

- Incorporating multi-regional and multi-institutional datasets can improve generalization and reduce dataset bias, enabling population-wide deployment.
- Implementing the model within IoT-enabled wearable devices or hospital monitoring systems can support continuous, real-time risk assessment.

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 (IJAIS), 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. [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. [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. [18] Daniyal Asif, Mairaj Bibi, Muhammad Shoaib Arif and Aiman Mukheimer, "Enhancing Heart Disease Prediction through Ensemble Learning Techniques with Hyper parameterOptimization", Algorithms 2023, 16, 308., https://www.mdpi.com/journal/algorithms