International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Enhancing the Prediction of Heart Disease using Machine Learning Algorithms

Rehan K. Nonsola[1*], Chetna G. Chand[2]

[1*]Information Technology Department,

Vishwakarma Government Engineering College, Ahmedabad, Gujarat, India

[2]Information Technology Department,

Vishwakarma Government Engineering College, Ahmedabad, Gujarat, India

[1*]Corresponding author E-mail: rehannonsola@gmail.com

[2]Contributing author E-mail: cgchand@vgecg.ac.in

Abstract

Heart disease represents one of the top causes of death in different parts of the world, indicating a need for better detection as well as more effective risk assessments of this problem at an earlier stage. Here, this paper looks into various machine learning methods to predict a person's risks of heart diseases by considering the seven most imperative algorithms: Logistic Regression, Naive Bayes, Support Vector Machine, K-Nearest Neighbor, Decision Tree, Random Forest, and Adaptive Boosting. By reviewing studies and datasets, we assess how well these algorithms work for heart disease prediction. Our findings reveal that machine learning models have significantly improved the accuracy of predicting heart disease, in comparison to traditional methods. Among all the above-mentioned algorithms, Random Forest has given the best results, with some studies reporting up to 93.44% in identifying potential heart disease cases. The study ends with the finding that machine learning algorithms are able to be really effective at the prediction of heart disease and great potential for actual use, improve patient care and help reduce the global impact of heart disease.

Keywords: Cardiovascular Risk Prediction, Machine Learning Algorithms, Random Forest, Support Vector Machine, Predictive Analytics in Healthcare, Early Disease Detection.

1. Introduction

Worldwide, the majority of people's leading cause of death and illness is cardiovascular disease. [1]. CVD encompasses various heart-related, blood-vessel, or blood circulation ailments [2]. Heart disease has become the condition most likely to kill a great many of victims since it always impairs one's ability to work [3]. Within the last decades, heart disease has become one of the primary causes of death in the world, accounting for approximately 17.7 million deaths annually [4]. Taking a diagnosis by specialized doctors for heart disease is essential for appropriate treatment, although often very subjective and prone to human errors.

It is diagnosed mainly because heart failure has been widely studied and can be complex [5]. Very useful in this area have proved to be computer-aided decision support systems [6]. An example here would be that systems using data mining techniques can help make the predictions concerning the disease faster and more accurate. Heart diseases can arise in various types and may sometimes lead to significant complications, leading to reduced quality of life as well as causing death, majorly in the developing countries [7]. A higher death toll from heart failure is reported from these regions whose healthcare facilities are less developed as compared to developed regions [8]. This justifies the creation of a mechanism that can enable accurate and effective prediction of a patient's possibility of heart failure.

Clinical decision support systems assist physicians in diagnosing diseases and advising on the correct treatment. There's a need to test such systems to ensure they operate safely and soundly as expected. [9] Mobile health technologies can be adapted to integrate decision support systems on mobile devices. These technologies track patients' real-time data, which makes health services more efficient. They assist in checking patients' health without physical visits to a health center. [10,11] Improving the monitoring of heart patients can help lower the death rate. Usually, people see a heart specialist only when the condition is already advanced. [12]

Machine learning is very important in medicine. It helps us diagnose, detect, and predict different diseases. Recently, there has been more interest in using data mining and machine learning to predict the chances of getting certain diseases. There are already some studies that use data mining methods to predict diseases. However, while some studies have tried to predict how a disease might progress in the future, they still haven't been able to get accurate results. [13] In medicine, a large amount of data is created every day using data mining techniques, which helps us discover hidden patterns that can assist in diagnosing diseases. Because of this, data mining is very important in healthcare, as shown by the research done over the past few decades. When predicting heart disease, it’s important to consider factors like high cholesterol, diabetes, high blood pressure, diabetes, and abnormal pulse rates. [14,15]

Improvements in computing power, parallel processing, and quick data storage have all contributed to the evolution of technology. Additionally, it can comprehend the meaning of supplied data and has great predictive capabilities. Tasks like object detection, movement modeling, and data reduction have all benefited from the usage of technology. Deep learning algorithms extract significant features from input with varying levels of detail by using intricate, multi-layered architectures. These algorithms can help find much information in data by focusing on seeking a pattern in it. [16, 17, 18]

To address these challenges, many researchers have developed methods to detect heart disease by considering various factors. Most of these methods use machine learning techniques, which are more effective than traditional statistical methods in analyzing large datasets with complex interactions [19,20,21,22,23]. Some studies have used extensive datasets, relying on historical data collected over time to identify existing diseases. However, until recently, the results often focused on detecting abnormal heart patterns without distinguishing between serious cases and those that are harmless to health.

This study examines seven well-known machine learning methods Decision Tree, Naive Bayes, Support Vector Machine, Random Forest, K-Nearest Neighbor, Logistic Regression, and Adaptive Boosting to see how well they can predict heart disease. By comparing their performance in different studies, the goal is to find the most effective methods for heart disease prediction and explore how they could be used in healthcare settings.

2. Literature Review

The authors Ilias Tougui et al. [24] used data mining and machine learning methods in the study to classify heart disease. The authors tested six data mining tools Orange, Weka, RapidMiner, Knime, Matlab, and Scikit-learn—and six machine learning methods Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Artificial Neural Network, Naïve Bayes, and Random Forest—to determine the models' sensitivity, specificity, and accuracy. The authors used a dataset of 303 cases to test the models. An accuracy of 85.86% was attained by the Artificial Neural Network implementation in Matlab. According to the study's findings, machine learning is an effective tool for medical diagnosis. It also asserts that the user's ability level influences the tools they choose.

Pooja Rani et al [25] created a hybrid decision support system for the prediction of cardiac disease based on machine learning. The system using the Cleveland dataset employs methods like SMOTE to balance classes, a mix of Genetic Algorithm and Recursive Feature Elimination for feature selection, and MICE for handling missing data. The classification algorithms used were Random Forest, Naive Bayes, and AdaBoost. It was found that the highest accuracy was achieved with Random Forest, which was 86.6%. The research shows how this system can be helpful in areas where healthcare facilities are scarce and proposes its potential integration with IoT for real-time data collection and further extension to other chronic diseases.

An AI-driven system that detects heart disease is developed by Victor Chang et al. in [26]. The system employs Python due to its strong features and libraries, such as Pandas, Matplotlib, and SciPy, for analyzing medical data. It tests the efficiency of machine learning methods, including Decision Tree, Random Forest, and K-Nearest Neighbors, achieving 83% accuracy using Random Forest. The study underlines the use of predictive analytics in healthcare and demonstrates how Python could handle and analyze complex datasets, improve diagnostic precision, and offer accessible, scalable solutions to healthcare professionals. The study also covers cybersecurity concerns for protecting patient data.

In the study by Abdul Wahab Ali Almazroi et al. [27], entitled "A Clinical Decision Support System for Heart Disease Prediction Using Deep Learning," a deep learning approach is introduced into the prediction of heart disease. The study employs a Dense Neural Network built through Keras, which has various hidden layers, in analyzing heart disease data from diverse sources. The resulting model was tested for its accuracy, sensitivity, and specificity. Generally, this is supposed to have been done under better performances compared to the traditional methods of machine learning and ensemble. Extensive testing and cross-validation reveal a great promise for the framework with strong and reliable results across many different datasets. The deep model's accuracy reached 83%.

In the paper by Chintan M. Bhatt et al. [28], different machine learning methods were used successfully for the prediction of cardiovascular disease. The researchers point out how early diagnosis significantly decreases death due to heart diseases, which have become the principal cause of deaths worldwide. The research used a large dataset of 70,000 patient records and applied models such as decision trees, random forests, multilayer perceptron, and XGB to improve prediction accuracy. Among them, the best performance was shown by the multilayer perceptron model, with an accuracy of 87.28%. The study also points out the limitations of using a single dataset and calls for further research to explore additional risk factors and improve the model's general applicability.

Md. Imam Hossain et al. [29] has focused on the development of a predictive model for cardiovascular disease by using different artificial intelligence methods. The paper underlines the need for exploratory data analysis to understand the dataset and enhance feature selection, which is very crucial for good prediction. This process includes data gathering and preprocessing, feature extraction, and further application of seven different machine learning algorithms: random forest, decision tree, support vector machine, k-nearest neighbors, naive bayes, logistic regression, and multilayer perceptron. The dataset with selected features was tested by the application of the random forest algorithm, which obtained accuracy at 90%.

Authors in Maria Teresa Garsia-Ordus et al. [30] explain how, with advanced deep learning techniques and feature enhancement methods, they predict heart disease risk. Early detection of heart disease patients greatly impacts survival rates, according to them. They investigate the several determinants of heart health including age, gender, cholesterol level, and heart rate. The study presents a novel approach that performs better than existing best practices by 4.4% while accurately predicting heart disease risk with a value of 90.09%. This paper also discusses the traditional methods applied for heart disease prediction, stating their limitations in handling complex interactions within large datasets.

Mohammed Amine Bouqentar et al. in his study [31] discuss the early detection and diagnosis of CVDs based on machine learning methods. In this research work, he mentions the significance of CVDs, accounting for about 32% of yearly deaths across the globe. So, an efficient prediction tool that can aid in the proper management of such diseases by the medical professionals is emphasized. It makes use of various ML algorithms such as logistic regression, adaptive boosting, decision tree, random forest, and K-nearest neighbors, support vector machine to classify datasets from Cleveland and Statlog with 92% accuracy. The method starts by gathering data and arranging it in a particular order and applying these algorithms for accurate classification. The results have shown that the early detection and diagnosis of CVDs could be significantly enhanced by ML and thus serve as a useful tool for healthcare workers.

These studies show how crucial it is to test and evaluate machine learning algorithms or data mining tools carefully for specific tasks, however some challenges remain despite improvements in ML methods. One of the major challenges is the lack of a standard dataset since each work employed a different dataset, making comparison of results difficult. Furthermore, most research studies use minimal datasets for their models, and this may lower the accuracy of the models. This is due to the fact that ML models require a vast amount of data in order to learn appropriately and generate accurate predictions.

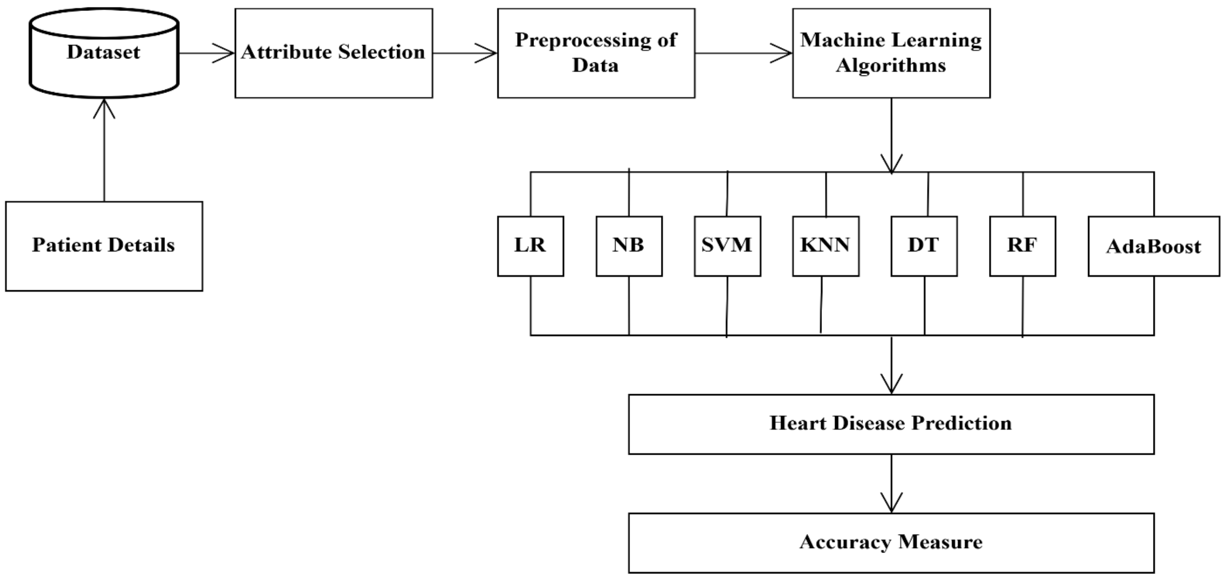
The main issue with these studies is that they primarily compare the effectiveness of various machine learning algorithms rather than trying to optimize the hyperparameters. Hyperparameter optimization is an important step and may seriously affect how the model trains and performs, but it demands considerable time and computing power. In addition, there is little consideration in all these studies about how understandable the models are. If healthcare professionals do not clearly understand how a model arrives at its predictions, they will find it difficult to have confidence in and use such models in applications.

3. Methodology

3.1 The proposed system for predicting heart disease flowchart.

Fig. 1 illustrates the design structure of the proposed heart disease prediction system. It will then be followed by the collection of patient information, which can be inputted via user-friendly interfaces. A pre-trained model is used to predict the chance of a patient being diagnosed with heart diseases. It was designed for easy use in healthcare delivery to aid in quick decision-making.

Fig. 1: Flowchart to predict heart diseases using seven types of machine learning algorithms.



3.2 Cleveland dataset sources

Cleveland dataset which was compiled by David Aha in 1987 is the most common reference source for heart disease prediction research. There are 303 samples and 76 features in this dataset. There are 13 attributes and 1 target variable under the most relevant 14 features we selected according to the given study [31]. It indicates whether or not there is a heart disease in the target variable.

3.3 Data preprocessing

We started the pre-processing stage with duplicate and missing value removals. This involved deleting 16 samples from an original 303 samples in the dataset. After that, we handled outliers as they can seriously affect the performance of statistical and machine learning models. This process entailed identification and management of data points far away from other data points. After addressing the outliers, the next thing is checking the data type of the attributes. If not set up appropriately, the incorrect data types could create problems when doing analysis or modeling. If that's the case, it should be changed to the right data type.

All the steps in preprocessing make sure there is no missing and null values for a dataset. Next step in the pipeline involves finding out if the dataset contains any outlier, outlier here refer to some of the observations or data that varies far and beyond others can badly impact performance in model design. It's easy for it to lead models over fit as well as under fit resulting to highly biased.

The use of box plots for every dataset feature is the most common way to identify outliers. Box plots are used in order to get a visual understanding of the distribution of data, and thus the identification of outliers in the data set is quite easy. Through box plots, outliers can be identified and removed or treated separately.

Table 1: The Cleveland Dataset's attributes

Attribute	Description	Data Type
age	The patient's age in years	Numerical
sex	Patient sex (0 = female, 1 = male)	Categorical
cp	Type of chest pain (0 = typical angina, 1 = atypical angina, 2 = non-anginal discomfort, 3 = asymptomatic)	Categorical
trestbps	Blood pressure at rest (mm Hg)	Numerical
chol	Cholesterol levels in serum (mg/dl)	Numerical
fbs	Blood sugar level during fasting > 120 mg/dl (1 = true, 0 = false)	Categorical
restecg	Results of a resting electrocardiogram (0 = normal, 1 = abnormal ST-T wave, 2 = likely or certain left ventricular hypertrophy)	Categorical
thalach	The highest heart rate attained while exercising	Numerical
exang	Angina brought on by exercise (1 = yes, 0 = no)	Categorical
oldpeak	Exercise-induced ST depression in comparison to rest	Numerical
slope	The exercise ST segment's peak slope (0 = upsloping, 1 = flat, and 2 = down sloping)	Categorical
ca	Number of fluoroscopy-colored major vessels (0–3)	Numerical
thal	Thalassemia is a type of blood disorder (0 = normal, 1 = fixed defect, 2 = reversible defect)	Categorical
target	Heart disease (0 = no, 1 = yes) is present	Categorical

Fig. 2: Cleveland database box plots

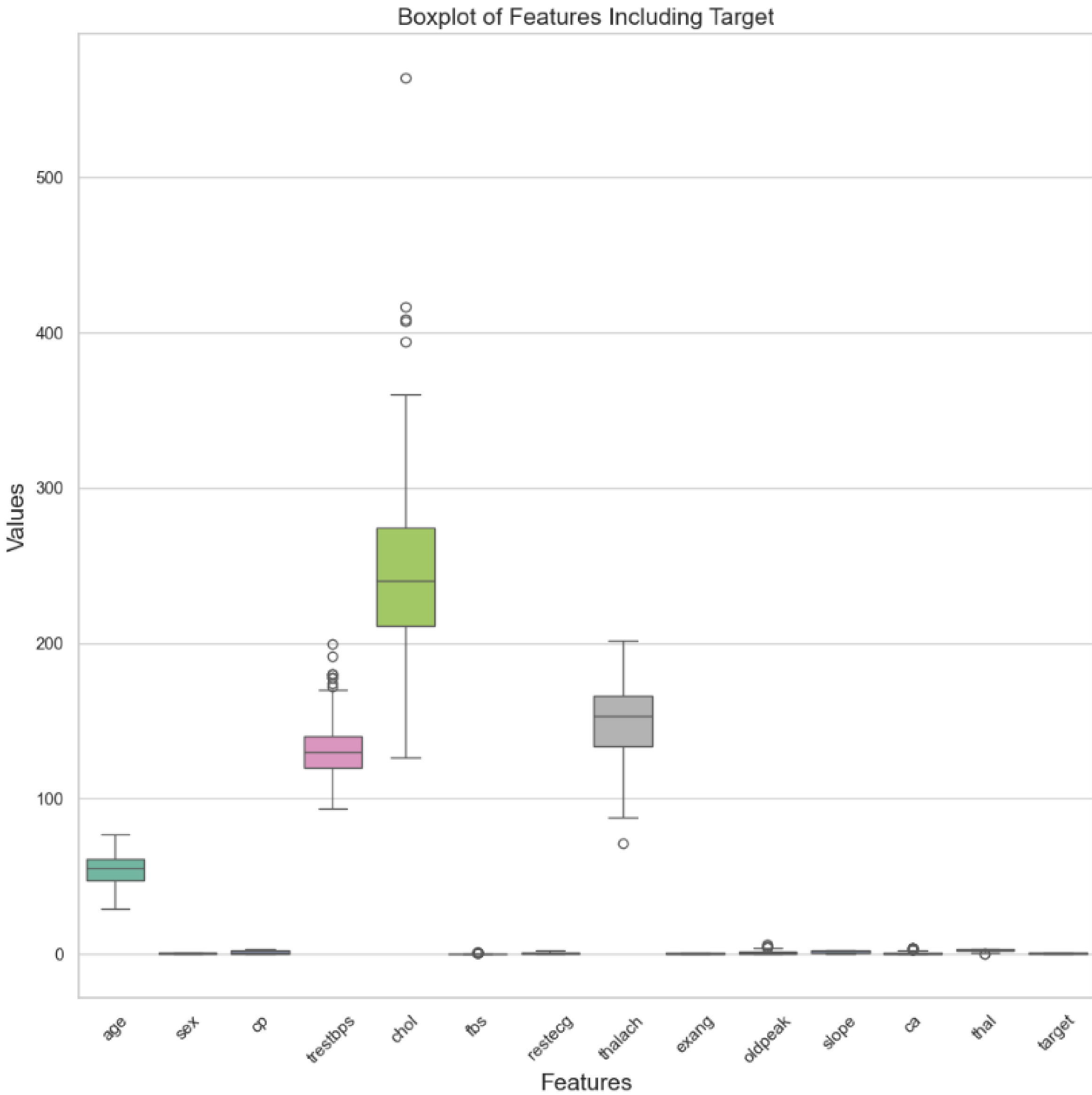


Fig 3: Histogram matrix of features

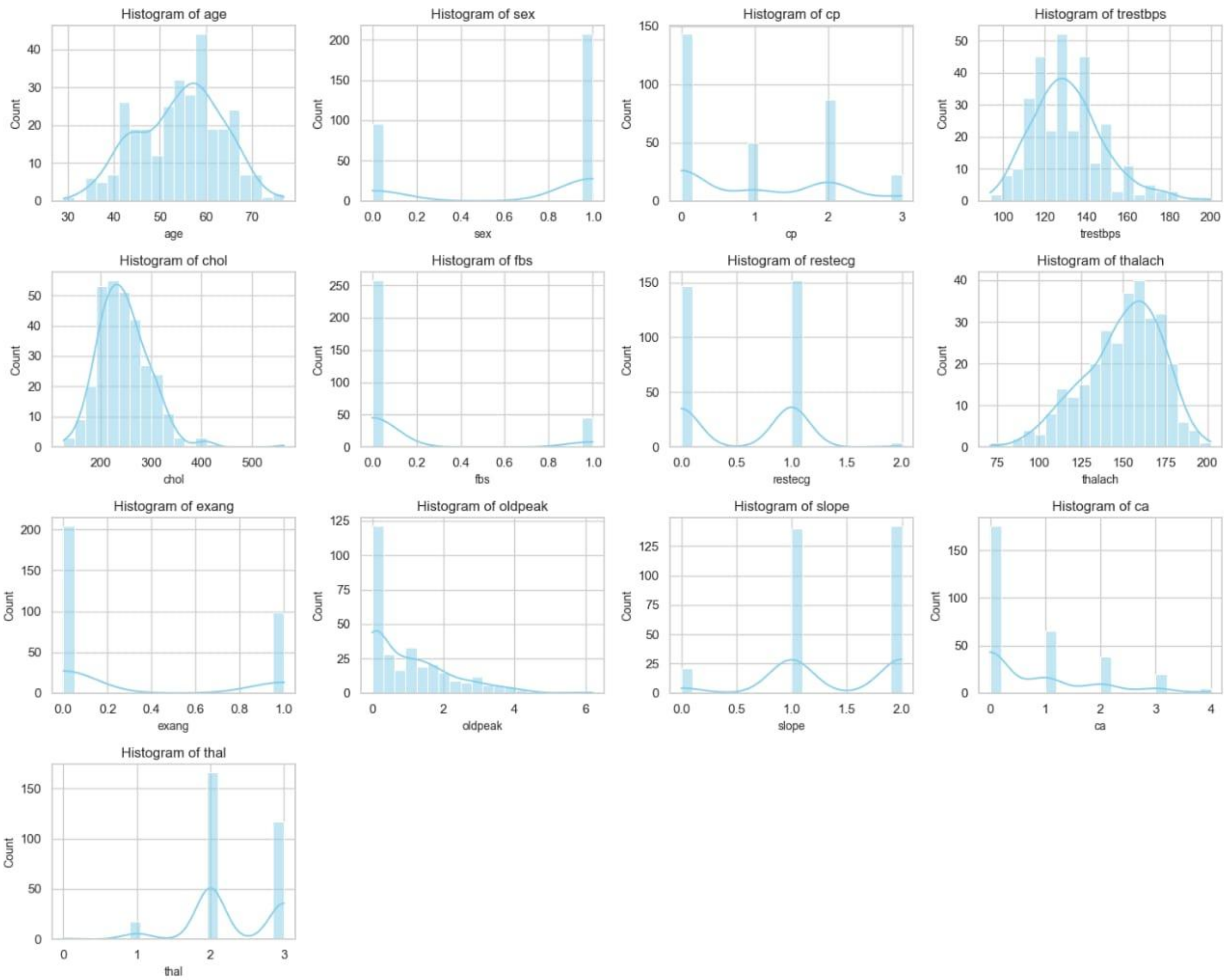
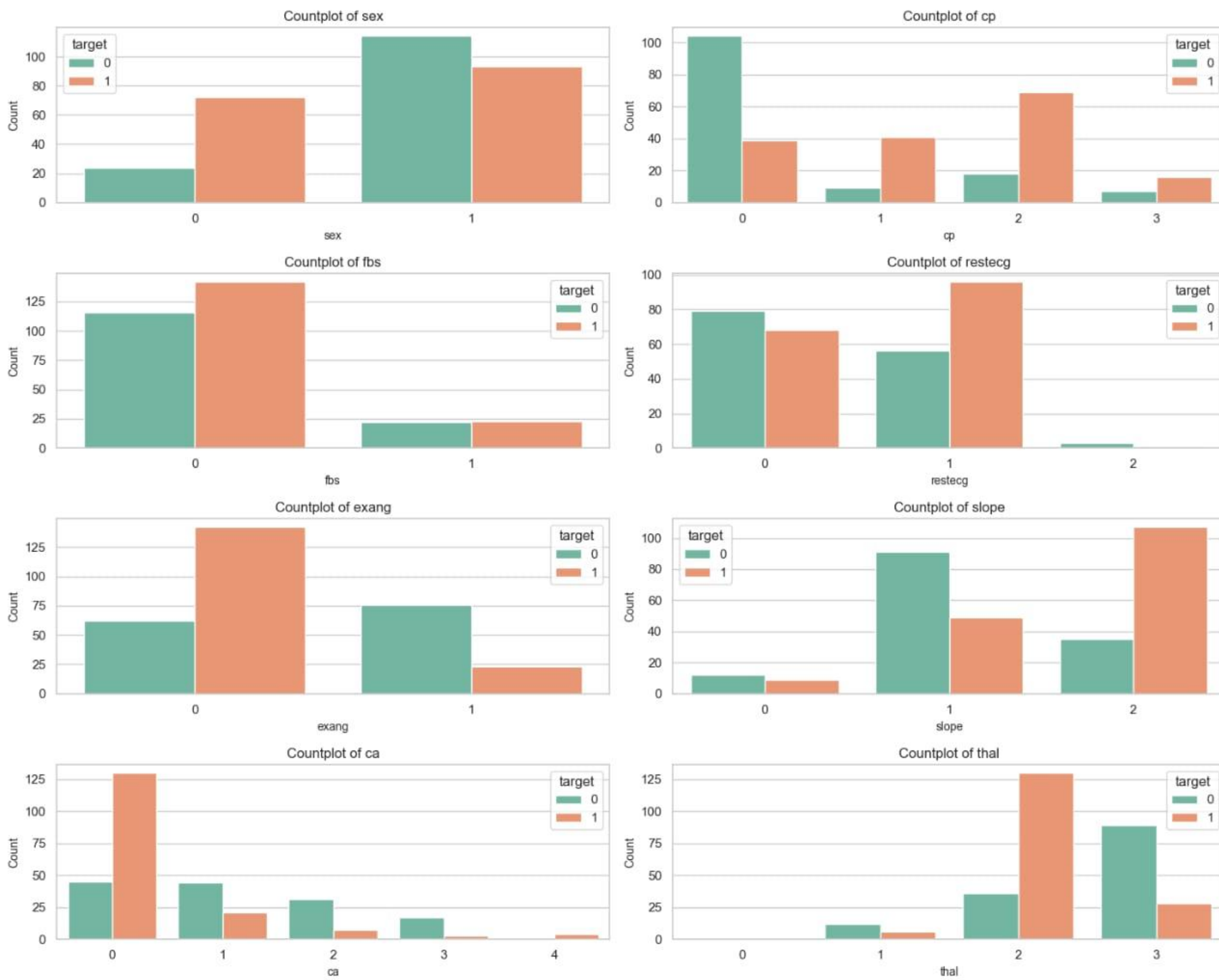


Fig 4: Count plot distribution of instances based on features



3.4 Feature Selection and Reduction

Analyzing feature importance across studies underlined several vital factors that foretell the possibility of heart diseases.

1. Age consistently stood out as a key factor, which aligns with well-known medical insights about cardiovascular risk [38].

2. Chest pain type was often recognized as an important predictor, emphasizing its role in diagnosing heart disease [39].

3. The maximum heart rate reached during exercise testing was frequently ranked as highly important, highlighting the usefulness of stress tests in assessing risk [40].

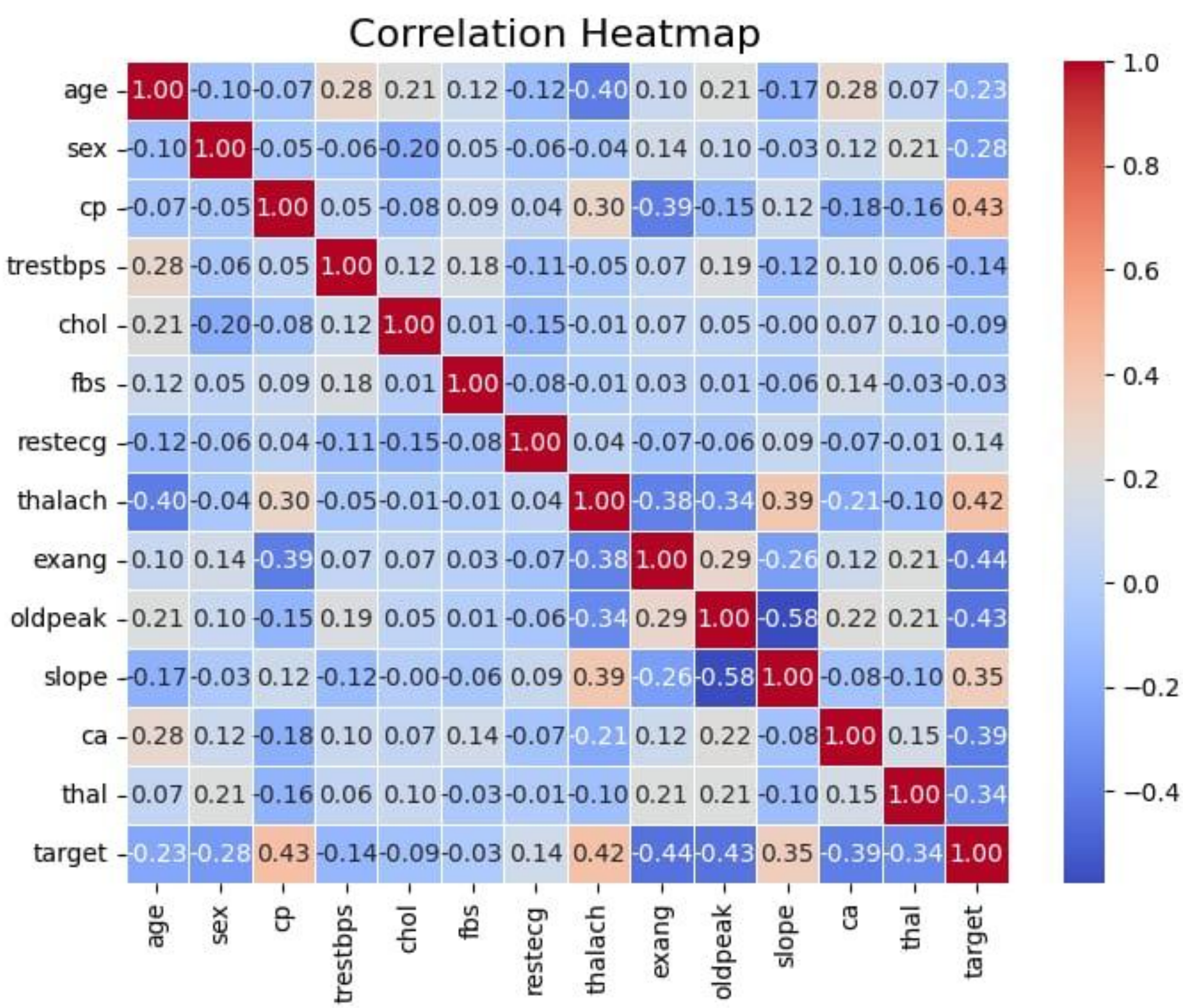
4. The number of major vessels seen in fluoroscopy was consistently important, showing that coronary artery imaging plays a key role in prediction [41].

5. ST depression caused by exercise compared to rest was often emphasized, stressing the importance of ECG changes in identifying heart disease risk [42].

3.5 Correlation table

A correlation table is created to analyze the relationships between different categories. As shown in figure, mean arterial pressure, cholesterol, and age were strongly correlated. This matrix also helps in examining how features depend on each other.

Fig 5: Heatmap-correlation matrix table



3.6 Modeling

The dataset was split into two datasets with 80% for training and 20% for testing. It trained the model on the training data and reported its performances on the test data. Clustering data is used to test the effectiveness of various classifiers using a variety of algorithms, including K-Nearest Neighbor, Support Vector Machine, Adaptive Boosting, Naive Bayes, Random Forest, Decision Tree, and Logistic Regression. Accuracy, precision, recall, and F1 score are then used to assess each classifier's performance.

3.6.1 Logistic Regression

Logistic regression is a classification algorithm that predicts the value of variable Y with two or more possible outcomes: 0 or 1 for binary classification, and more than two for multi-class classification. The probability that input X belongs to class 1 is calculated using the logistic regression equation. [32]

$$\left(\frac{p}{1-p}\right) = b_0 + b_1 x$$

3.6.2 Naive Bayes

It is a Bayes' theorem-based probabilistic classifier. It is effective in many classification tasks: document categorization, spam filtering, and so forth; for instance, it can be used to diagnose diseases. The algorithm uses the assumption that features applied for the prediction are independent of each other. Using Bayes' theorem, it calculates the posterior probability of the class: [33]

$$P(A/B) = \frac{\mathrm{P(B/A)\times P(A)}}{\mathrm{P(B)}}$$

3.6.3 Support Vector Machine

This algorithm classifies data using a hyperplane in such a way that samples from one class lie on one side and samples of the other class lie on the other side. To increase the distance between two classes, it moves the hyperplane. The data points from each class that are nearest to the hyperplane are known as support vectors. The following support vectors are used for the technique.: [32].

$$w_0^T x + b_0 = 1 or w_0^T x + b_0 = -1$$

3.6.4 K-Nearest Neighbor

One technique for classifying a data point based on how close it is to the nearest group of points is K-Nearest Neighbor [34]. The Euclidean distance formula, which is listed below, is used to calculate the distance between the data points.

$$d = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

3.6.5 Decision Tree

The primary uses of a decision tree are data cleaning and pattern recognition. In any dataset, there are some features that are more important than others. Using the Gini Index, irrelevant features or those features that are marginally relevant can be discovered and removed. For the building of a decision tree, every node is chosen through conditional probability calculations. Features with low Gini Index are selected and then these features are used to generate rules. In the field of bioinformatics, decision trees can be used especially in the diagnosis and prediction of diseases. The Gini Index is computed by the formula below. [35].

$$G = \sum_{I=1}^{C} p(i) * \left(1 - p(i)\right)$$

3.6.6 Random Forest

The Random Forest algorithm is constructed on a collection of Decision Trees, as a forest is the aggregate of numerous trees. The Random Forest method employs outcomes from a vast range of trees. Therefore, every tree produces a classification decision, and the forest chooses the most voted tree. Regression analysis employs the following formula to find the average of all trees. [36]

$$Gini = 1 - \sum_{i=1}^{C}(p_i)^2$$

3.6.7 Adaptive Boosting

Adaptive boosting, simply termed as AdaBoost, is one of the boosting algorithms used on weak learners to improve their performance. It begins by training a classifier on the original data set. Subsequent iterations of the classifier follow, each overlearning the previous one and trying to correct its mistakes. Each classifier is trained on a different subset of the data, created by assigning weights to data points. Instances that were misclassified are given higher weights, increasing their chances of being included in the next subset. This is a sequential process, and weak classifiers are combined using a cost function to produce a stronger classifier. The accuracy of the final prediction is improved by giving more weight to better classifiers. With AdaBoost, you can define the weak classifier to be boosted, and by default, the decision tree is used. [37]

3.7 Performance evaluation measures:

The performance of the machine learning models was evaluated by different metrics, such as:

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \quad (1)$$

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

$$Recall = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

4.Results and Discussion

A multiple review of studies and datasets is, therefore, important in emphasizing the efficiency of machine learning algorithms and their potential to predict heart disease. In the section, these key conclusions are underlined, with implications on cardiovascular risk determination.

4.1 Algorithm Performance

To make our system much more effective, we tested it using the Cleveland heart disease dataset and seven different machine learning algorithms. We assessed those using several different performance metrics, namely accuracy, precision, recall, and F1 score. This allowed us to view each model's performance from different angles and get an understanding of how accuracy versus those factors such as false positives or false negatives trade off with one another. We trained and compared all the algorithms on the same dataset with the same evaluation criteria to identify the best-performing model. We carefully selected, implemented, and fine-tuned the algorithms to ensure our results were accurate and trustworthy.

Fig 6: Confusion matrix using Random Forest applied to the Cleveland heart disease data

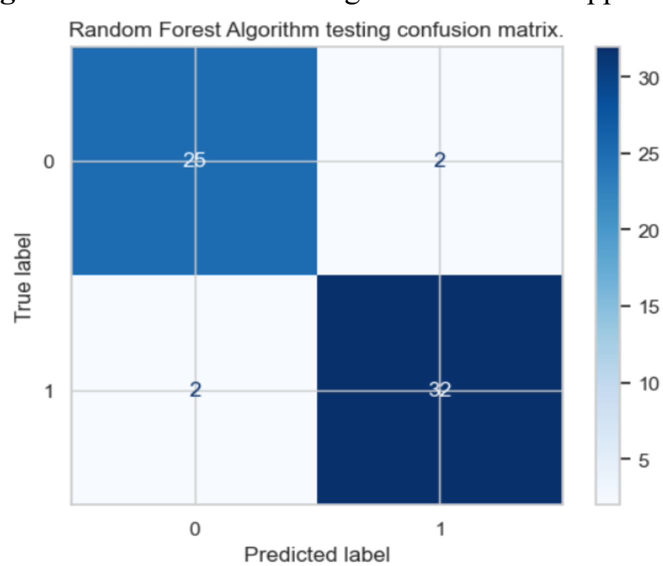


Fig 7: The dataset on Cleveland heart disease was used to test seven distinct machine learning algorithms.

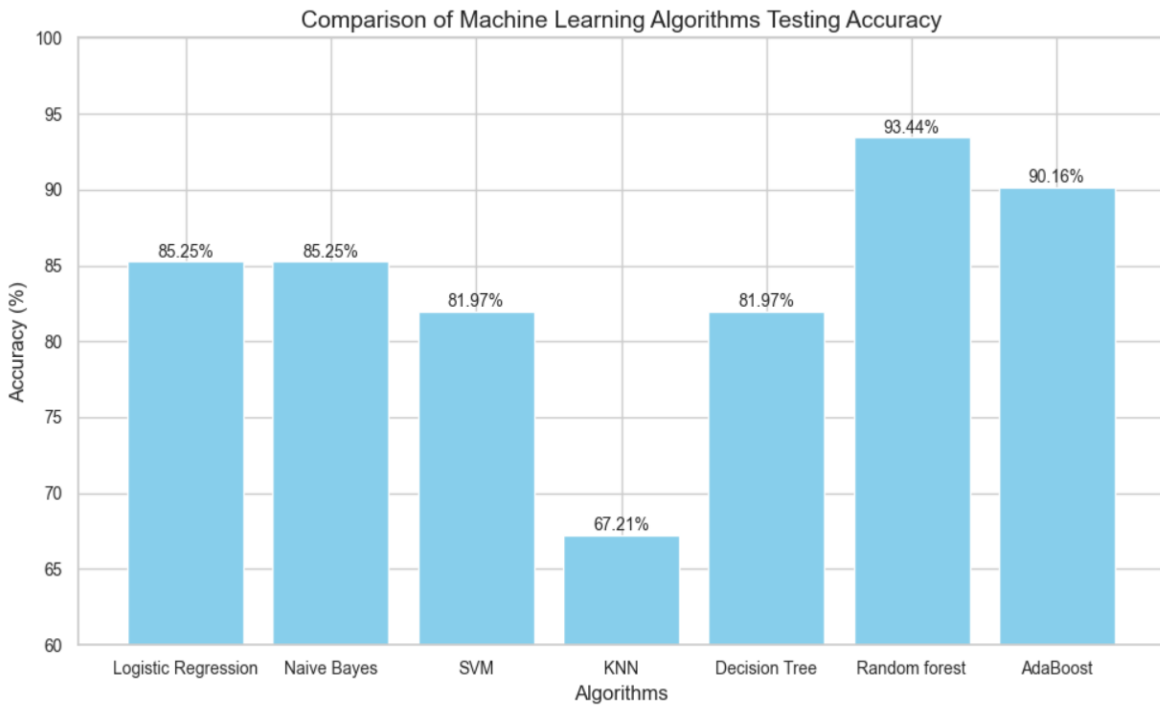


Fig 8: The Random Forest algorithm's maximum accuracy on the Cleveland heart disease dataset.

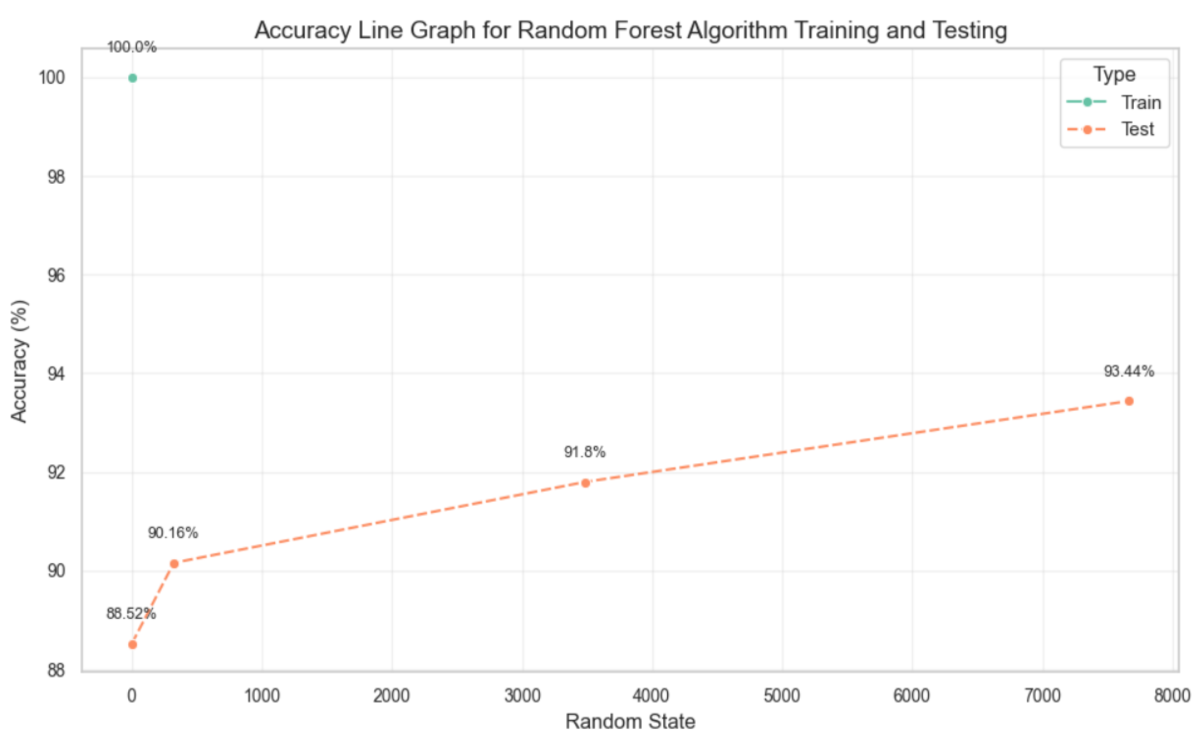


Table 2: Model performance on the training dataset.

Models	Accuracy	Precision	Recall	F1-score
LR	83.47%	76.57%	85.85%	80.24%
NB	83.47%	79.27%	83.80%	80.95%
SVM	84.71%	77.47%	87.75%	81.69%
KNN	72.31%	67.56%	70.75%	68.46%
DT	100%	100%	100%	100%
RF	100%	100%	100%	100%
AdaBoost	92.15%	90.09%	92.59%	90.98%

Table 3: Model performance on the testing dataset.

Models	Accuracy	Precision	Recall	F1-score
LR	85.25%	81.48%	84.61%	82.47%
NB	85.25%	77.77%	87.5%	81.69%
SVM	81.97%	74.07%	83.33%	78.24%
KNN	67.21%	66.66%	62.06%	63.93%
DT	81.97%	81.48%	78.57%	79.47%
RF	93.44%	92.59%	92.59%	92%
AdaBoost	90.16%	92.59%	86.20%	88.89%

Table 4: A comparison between the suggested performance and earlier research

Year	Author	Dataset	Algorithms	Accuracy
2020	Ilias Tougui et al [24]	Cleveland heart disease	ANN, KNN, LR, SVM, NB, RF	85.86%
2021	Pooja Rani et al [25]	Cleveland heart disease	RF, NB, LR, AdaBoost, SVM	86.6%
2022	Victor Chang et al [26]	Cleveland Heart disease	RF, KNN, DT, SVM, LR	83%
2023	Abdul Wahab Ali Almazroi et al [27]	Cleveland Heart disease	DL, KNN, SVM, GP, DT, NB, QDA, AdaBoost, Bagging, Boosting	83%
2023	Maria Teresa Garsia-Ordus et al [30]	11 Clinical features heart disease	RF, XGB, GNB, AdaBoost, DT, KNN, MLP	90.09%

2024	Mohammed Amine Bouqentar et al [31]	Cleveland heart disease	SVM, DT, RF, LR, AdaBoost, KNN	92%
2025	**Proposed Work**	Cleveland heart disease	LR, NB, SVM, KNN, DT, RF, AdaBoost	93.44%

5.Conclusion

This is a dangerous condition that can lead to heart attacks or even death. Using the data on heart disease, this study examined how the application of machine learning techniques aids in the prediction of heart disease. Because imbalanced datasets produce biased results when used in the training of an ML model, there was preparation in dealing with missing values, normalizing the dataset, and balancing the dataset. Various machine learning algorithms were tested. With an accuracy of 93.44%, the best results were attained by the Random Forest. It aimed to seek direct and efficient methods of machine learning that could lead to reliable output on such a dataset. This study is in its early stages about using ML techniques, and the results suggest that it could be a very useful tool for improving patient care.

Future Work

Current machine learning models used to predict heart disease present several challenges. Most are without a diverse dataset, and they are often difficult to interpret. Moreover, their real-time assessments are poor. Current successful models cannot be tested for the long term, and there exists a problem while using the models on various populations. In addition to these, ethical considerations and how to integrate it into the healthcare system pose some difficulties. By solving these issues, it would greatly help make the accuracy, reliability, and acceptance of ML tools for predicting heart disease risk much greater and contribute to better personalized care and improved patient outcomes.

References

[1]. Maini E, Venkateswarlu B, Maini B, Marwaha D, “Machine learning–based heart disease prediction system for Indian population: an exploratory study done in South India”, Med J Armed Forces India 77(3): pp. 302–311. https://doi.org/10.1016/J.MJAFI.2020.10.013, 2021.

[2]. Soni J, Ansari U, Sharma D, Soni S, “Predictive data mining for medical diagnosis: an overview of heart disease prediction”, International Journal of Computer Applications 17(8): pp. 43–48. https://doi.org/10.5120/2237-2860, 2011.

[3]. Diwakar M, Tripathi A, Joshi K, Memoria M, Singh P, Kumar N, “Latest trends on heart disease prediction using machine learning and image fusion”, Mater Today: Proc 37(Part 2): pp. 3213–3218. https://doi.org/10.1016/J.MATPR.2020.09.078, 2021.

[4]. Kondababu A, Siddhartha V, Kumar BB, Penumutchi B, “A comparative study on machine learning based heart disease prediction”, Mater Today: Proc. https://doi.org/10.1016/J.MATPR.2021.01.475, 2021.

[5]. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ et al, "Heart disease and stroke statistics", Circulation, vol 129(3), https://doi.org/10.1161/01.CIR.0000441139. 02102.80, 2014.

[6]. Krishnaiah V, Narsimha G, Chandra NS, "Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review", International Journal of Computer Applications, 136(2): pp. 975–8887, 2016.

[7]. Yang H, Garibaldi JM, "A hybrid model for automatic identification of risk factors for heart disease" J Biomed Inform 58:S171–S182. https://doi.org/10.1016/J.JBI.2015.09.006, 2015.

[8]. Muzammal M, Talat R, Sodhro AH, Pirbhulal S, "A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks", Inf Fusion 53: pp. 155–164. https://doi.org/10.1016/ J. INFFUS.2019.06.021, 2020.

[9]. Cicotti G, "An evidence-based risk-oriented V-model methodology to develop ambient intelligent medical software", Journal of Reliable Intelligent Environments 3(1): pp. 41–53. https://doi.org/10.1007/s4086 0-017-0039-9, 2017.

[10]. Vithanwattana N, Mapp G, George C, "Developing a comprehensive information security framework for mHealth: a detailed analysis", Journal of Reliable Intelligent Environments 3(1): pp. 21–39. https://doi. org/10.1007/s40860-017-0038-x, 2017.

[11]. Jusob FR, George C, Mapp G, "Exploring the need for a suitable privacy framework for mHealth when managing chronic diseases", Journal of Reliable Intelligent Environments 3(4): pp. 243–256. https://doi. org/10.1007/s40860-017-0049-7, 2017.

[12]. Ponikowski P, Anker SD, AlHabib KF, Cowie MR, Force TL, Hu S, Jaarsma T, Krum H, Rastogi V, Rohde LE, Samal UC, Shimokawa H, Siswanto BB, Sliwa K, Filippatos G, "Heart failure: preventing disease and death worldwide", ESC Heart Fail 1(1): pp. 4–25. https://doi.org/10.1002/ehf2.12005, 2014.

[13]. Waigi, R.; Choudhary, S.; Fulzele, P.; Mishra, G, "Predicting the risk of heart disease using advanced machine learning approach", Eur. J. Mol. Clin. Med, 7, pp. 1638–1645, 2020.

[14]. Soni, J.; Ansari, U.; Sharma, D.; Soni, S, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer Applications, 17, pp. 43–48, 2011.

[15]. Mohan, S.; Thirumalai, C.; Srivastava, G, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE Access, 7, pp. 81542–81554, 2019.

[16]. Y. Wang, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction. Neurocomputing", vol. 184, pp. 232–242, 2016.

[17]. M. P. T. Kaandorp, S. Barbieri, R. Klaassen, H. W. M. Laarhoven, H. Crezee, P. T. While, A. J. Nederveen, and O. J. Gurney-Champion, "Improved unsupervised physics-informed deep learning for intravoxel incoherent motion modeling and evaluation in pancreatic cancer patients" Magn. Reson. Med., vol. 86, no. 4, pp. 2250–2265, 2021.

[18]. "The Enterprisers Project. Accessed:" [Online]. Available. https://enterprisersproject.com/article/2019/7/deep-learning-explained-plain, 2022.

[19]. Albert KF, John R, Divyang P, Saleem T, Kevin MT, Carolyn JP et al, (2019) “Machine learning prediction of response to cardiac resynchronization therapy: improvement versus current guidelines” Circ Arrhythmia Electrophysiol, vol 12(7),https://doi.org/10.1161/CIRCEP.119.007316, 2019.

[20]. Khajehali N, Khajehali Z, Tarokh MJ, “The prediction of mortality influential variables in an intensive care unit: a case study”, Personal Ubiquit Comput, https://doi.org/10.1007/s00779-021-01540-5, 2021.

[21]. Kim YJ, Saqlian M, Lee JY, “Deep learning–based prediction model of occurrences of major adverse cardiac events during 1-year follow-up after hospital discharge in patients with AMI using knowledge mining”, Personal Ubiquit Comput 26(2): pp. 259–267. https://doi.org/10.1007/ s00779-019-01248-7, 2022.

[22]. Olsen CR, Mentz RJ, Anstrom KJ, Page D, Patel PA, “Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure”, Am Heart J 229: pp. 1–17. https://doi.org/10. 1016/J.AHJ.2020.07.009, 2020.

[23]. Panahiazar M, Taslimitehrani V, Pereira N, Pathak J, “Using EHRs and machine learning for heart failure survival analysis”, Stud Health Technol Inform 216:40, 2015.

[24]. Ilias Tougui, Abdelilah Jilbab, Jamal El Mhamdi, “Heart disease classification using data mining tools and machine learning techniques”, Health and Technology, vol. 10, pp. 1137-1144, https://doi.org/10.1007/s12553-020-00438-1, 2020.

[25]. Pooja Rani, Rajneesh Kumar, Nada M. O. Sid Ahmed, Anurag Jain, “A decision support system for heart disease prediction based upon machine learning”, Journal of Reliable Intelligent Environments, vol. 7, pp. 263-275, https://doi.org/10.1007/s40860-021-00133-6, 2021.

[26]. Victor Chang, Vallabhanent Rupa Bhavani, Ariel Qianwen Xu, MA Hossain, “An artificial intelligence model for heart disease detection using machine learning algorithms”, Healthcare Analytics, vol. 2, https://doi.org/10.1016/j.health.2022.100016, 2022.

[27]. Abdulwahab Ali Almazroi, Eman A. Aldhahri, Saba Bashir, Sufyan Ashfaq “A Clinical Decision Support System for Heart Disease Prediction Using Deep Learning”, IEEE Access, vol. 11, pp. 61646 – 61659, https://doi.org/10.1109/ACCESS.2023.3285247, 2023.

[28]. Chintan M. Bhatt, Parth Patel, Tarang Ghetia, Pier Luigi Mazzeo, “Effective Heart Disease Prediction Using Machine Learning Techniques. Algorithms”, 16(2), https://doi.org/10.3390/a16020088, 2023.

2[9]. Md. Imam Hossain, Mehadi Hasan Maruf, Md. Ashikur Rahman Khan, Farida Siddiqi Prity, Sharmin Fatema, Md. Sabbir Ejaz, Md. Ahnaf Sad Khan, “Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison”, Iran Journal of Computer Science, vol. 6, pp. 397–417, https://doi.org/10.1007/s42044-023-00148-7, 2023.

[30]. Maria Teresa Garcia-Ordas, Martin Bayon-Gutierrez, Carmen Benavides, Jose Aveleira-Mata, Jose Alberto Benitez-Andrades, “Heart disease risk prediction using deep learning techniques with feature augmentation. Multimedia Tools and Applications” vol. 82, pp. 31759–31773, https://doi.org/10.1007/s11042-023-14817-z, 2023.

[31]. Mohammed Amine Bouqentar, Oumaima Terrada, Soufiane Hamida, Shawki Saleh, Driss Lamrani, Bouchaib Cherradi, Abdelhadi Raihani, “Early heart disease prediction using feature engineering and machine learning algorithms”, Heliyon, Vol. 10(19) https://doi.org/10.1016/j.heliyon.2024.e38731, 2024.

[32]. Rani P, Kumar R, Jain A, Lamba R, “Taxonomy of machine learning algorithms and its applications”, Journal of Computational and Theoretical Nanoscience 17(6), pp. 2509–2514, https://doi.org/10.1166/jctn.2020.8922, 2020.

[33]. Dulhare UN, “Prediction system for heart disease using Naive Bayes and particle swarm optimization”, Biomed Res 29(12): pp. 2646–2649, https://doi.org/10.4066/biomedicalresearch.29-18-620, 2018.

[34]. “World life expectancy, World Health Rankings Live Longer Live Better”. [Online]. Available. http://www.worldlifeexpectancy.com/causeof-death/coronary-heart-disease/by-country/, 2014.

[35]. A. Kumar, P. Kumar, A. Srivastava, V. D. A. Kumar, K. Vengatesan, and A. Singhal, “Comparative analysis of data mining techniques to predict heart disease for diabetic patients. in Proc”, International Conference on Advances in Computing and Data Sciences, Singapore, pp. 507–518, 10.1007/978-981-15-6634-9_46, 2020.

[36]. S. Uddin, A. Khan, M.E. Hossain, M.A. Moni, "Comparing different supervised machine learning algorithms for disease prediction", BMC Medical Informatics and Decision Making, 19, 281, https://doi.org/10.1186/s12911-019-1004-8, 2019.

[37]. Latha CBC, Jeeva SC, “Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques”, Inf Med Unlocked 16:100203. https://doi.org/10.1016/j.imu.2019.100203, 2019.

[38]. North, B. J., & Sinclair, D. A. “The intersection between aging and cardiovascular disease”, Circulation Research, 110(8), pp. 1097-1108, 10.1161/CIRCRESAHA.111.246876, 2012.

[39]. Swap, C. J., & Nagurney, J. T., “Value and limitations of chest pain history in the evaluation of patients with suspected acute coronary syndromes”, JAMA, 294(20), pp. 2623-2629, 10.1001/jama.294.20.2623, 2005.

[40]. Ellestad, M. H., “Chronotropic incompetence: the implications of heart rate response to exercise (compensatory parasympathetic hyperactivity?)”, Circulation, 87(5), pp. 1104-1107, 10.1161/01.cir.93.8.1485, 2003.

[41]. Matthew J Budoff, David Dowe, James G Jollis, Michael Gitter, John Sutherland, Edward Halamert, Markus Scherer, Raye Bellinger, Arthur Martin, Robert Benton, Augustin Delago, James K Min, Am Coll Cardiol, 52(21): pp. 1724-32, 10.1016/j.jacc.2008.07.031, 2008.

[42]. Kligfield, P., & Okin, P. M., “Evolution of the exercise electrocardiogram”, American Journal of Cardiology, 73(15), pp. 1209-1210, 10.1016/0002-9149(94)90183-x, 1994.