OPEN ACCESS

Saudi Journal of Engineering and Technology
Abbreviated Key Title: Saudi J Eng Technol
ISSN 2415-6272 (Print) | ISSN 2415-6264 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Original Research Article

Utilizing Machine Learning Techniques to Predict Cardiovascular Diseases and Comparing the Outcomes for Better Accuracy

Shartaj Fatima Mohammed[1], Zia Ul Haq Paracha[1]*

[1]Lecturer, MIS Department, King Saud University, Riyadh, KSA

DOI: https://doi.og/10.36348/sjet.2025.v10i02.002 | **Received:** 02.02.2024 | **Accepted:** 08.03.2024 | **Published:** 27.02.2025

***Corresponding author:** Zia Ul Haq Paracha
Lecturer, MIS Department, King Saud University, Riyadh, KSA

Abstract

With the advancements in technology, several features are now available for diagnosing heart diseases. However, large data sets have some limitations such as limited storage capacity and long access and processing times. Early diagnosis of heart problems is vital for medical treatment. Heart disease is a severe illness that is on the rise in both developed and developing countries, leading to fatalities. This disease causes the heart to not provide enough blood to various parts of the body, hindering its normal functions. Diagnosing this condition early and accurately is crucial to prevent further harm and potentially save lives. Diagnosis for various forms of heart disease can be detected with numerous medical tests, however, predicting heart disease without such tests is very difficult. Many researchers analyzed the risk factors of this disease and proposed machine learning models for the early detection of heart patients. However, these models suffer from the high dimensionality of data and need to be improved to obtain highly accurate results. The proposal was tested using five different standard data sets from the UCI repository. Our proposal consists of two main processes: the first is the data preprocessing process, and the second is the prediction process. In data preprocessing, the data is prepared for the prediction process, and three different feature selection methods (e.g., PCA) are applied to select the most relevant features from the data. In the prediction process, ten different prediction techniques (for example, Random Forest (RF) and Support Vector Classifier (SVC)) were applied to over-employed datasets. The techniques used were evaluated using four evaluation metrics: accuracy, precision, recall, and F1-score. For this research, we collected the dataset from the UCI repository (Kaggle) and used Random Forest Classification algorithm for predicting heart disease. The predictive model achieved an accuracy of 89.4 percent using Random Forest Classifier's default setting to predict heart diseases. Furthermore, the research focuses on the opportunity for training and testing using our model with a larger dataset and modifying different hyper parameters for further improvement. The results show that the LASSO method as a feature selection method with RF as a prediction technique produced the best accuracy (100%). Accuracy (99.57%) was obtained for Decision Tree (DT), Gradient Boosting (GB), AdaBoost (AB), Decision Tree Bagging Method (DTBM), Random Forest Bagging Method (RFBM), K-Nearest Neighbors Bagging Method (KNNBM), AdaBoost Boosting Method (ABBM), and Gradient Boosting Boosting Method (GBBM). The accuracy of SVC, Logistic Regression (LR), Naïve Bayes (NB), and Support Vector Classifier Bagging Method (SVCBM) was very similar to each other (98.73%).

Keywords: Heart Disease Diagnosis, Artificial Intelligence, Machine Learning, Feature Selection, Accuracy, Data Preprocessing.

Copyright © 2025 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)** which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

INTRODUCTION

According to a report by the World Health Organization (WHO), there were 17.9 million deaths caused by cardiovascular diseases (CVDs) in 2019, which accounted for 32% of all global deaths. The annual mortality rate for CVDs was greater than 17.7 million in the same year.

The increasing cost of healthcare has been a major problem for developed countries (Dadgostar, 2019). According to the CDC, an estimated 859,000 people in the US die from cardiovascular diseases, which accounts for one in every three deaths. The healthcare system spends $216 billion on cardiovascular diseases, and $147 billion is lost in productivity (Mayo, 2022). This cost has become a significant concern in the US, emphasizing the importance of early detection. With the advancement of biotechnology and an era of big data generated for healthcare, mainly through EHR (electronic health records) in various structures, it is increasingly important to intelligently use this information to detect abnormalities, predict heart diseases, and make sense of hidden patterns.

Physicians use a patient's medical history, physical examination report, and symptoms analysis to diagnose heart diseases, but this method is not sufficient for detecting all heart disease patients. Additionally, it is costly and computationally challenging to perform.

Artificial intelligence has certainly made computers smarter. Machine learning which is a subset of artificial intelligence plays an important role in mining large datasets and extracting valuable knowledge from them. Training a machine appropriately with proper train data set, the machine's algorithm can learn patterns and therefore detect any abnormalities in the initial stage of a disease which can help patients save overall cost and time.

Machine learning is a data analysis method that automates the construction of an analytic model. It is based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention.

In this research, we will explore the potential benefits of machine learning in the healthcare industry, specifically in relation to heart diseases. By detecting abnormalities early on, medical professionals can minimize healthcare costs and provide better care. We will also examine how data generated by EHR can provide valuable insights for detecting potential chronic diseases.

Moreover, a comparative and analytical study between ten different supervised models was performed to evaluate and validate their accuracy results for the prediction of cardiovascular disease and compare these results with the results of the related work which was published in recent years. These models are tested using five standard datasets from the UCI and evaluated using four evaluation metrics: accuracy, precision, recall, and F1-score. Moreover, the high dimensionality of data is solved by using three different feature selection methods. The results of using the employed models with and without feature selection approaches are compared. The results showed that RF as a classification technique with PCA as a feature selection method achieves an accuracy of 97.05%, SVC can achieve 98.31%, and DT achieves an accuracy of 97.89%.

Research Background

Cardiovascular diseases are a major cause of health issues and economic burden worldwide (Roth *et al*., 2020). These diseases encompass any conditions that affect the heart and blood vessels. When a heart condition occurs, the major blood vessels that supply the heart muscles can become affected. A buildup of cholesterol deposits, known as plaque, can reduce blood flow to various parts of the body and the heart (Heart 3 disease, 2022). Without proper treatment, this can lead to serious conditions like stroke, heart attack, or heart failure. Unfortunately, heart diseases can be difficult to detect and may not show symptoms until they become life-threatening. To diagnose these diseases, doctors often rely on a combination of blood tests, MRIs, CT scans, ECGs, or Holter monitoring. All of this medical data is collected and stored in various databases.

However, this data is only useful when it is integrated and analyzed using advanced techniques like Artificial Intelligence, machine learning, and data mining. By doing so, it is possible to generate accurate diagnostic information that can save lives and reduce costs.

Machine Learning

In today's world, data is growing exponentially and is easily accessible online. Machine learning algorithms can learn and create models without human intervention due to the availability of inexpensive computational power. Machine learning is a subset of artificial intelligence that can collect meaningful knowledge from its training data and improve automatically through exposure without programming. There are four main types of machine learning algorithms: Supervised, Unsupervised, Semi-supervised, and Reinforcement Learning. Supervised Learning has two categories: Classification and Regression, while unsupervised learning has Clustering and Association categories. These approaches use labeled or unlabeled datasets to anticipate outcomes. When applied to medical data, each approach has distinct guidelines for effectively extracting vital knowledge. (Jordan & Mitchell, 2015; Sarker, 2021; Delua, 2021; Gupta *et al*., 2021).

Medical Big Data

Big Data refers to extensive, complex, and diverse datasets that require specialized software to handle. With the advancement of medical technology, various types of medical data have emerged, including audio, lab tests, previous diagnostic reports, clinical records, research, and images. These datasets are valuable in improving healthcare quality. However, analyzing a single dataset may not yield desirable results. Instead, linking various datasets can provide excellent insights. Medical data is sourced from different data sources, such as lab databases, electronic health records (EHR), and electronic medical records (EMR). A medical data warehouse serves as a centralized repository for these datasets, providing better insights than analyzing data from a single database.

Problem Statement

The use of advanced technology tools like AI, machine learning, and data mining can assist healthcare professionals in making informed decisions to reduce deaths caused by heart disease at a minimum cost. By analyzing large databases, machine learning algorithms can detect common patterns that may contribute to heart disease and eventual death.

RESEARCH QUESTIONS

1. Which Machine Learning algorithms are utilized for diagnosing heart disease?
2. In what ways can Machine Learning techniques reduce misdiagnosis, which can lead to costly additional tests and improper treatments for patients?
3. How can Machine Learning aid in the early detection of abnormalities, benefiting both patients and the healthcare system?

OBJECTIVES

1. To explore how Machine Learning algorithms can be used in the diagnosis of heart disease by proposing an optimized model that can be used to predict heart diseases.
2. A novel method is described here to reduce the cost and enhance the accuracy of heart disease prediction in a simple and efficient manner.
3. Appling various feature selection methods (e.g., PCA) to propose an effective framework for the early detection of heart disease patients.

LITERATURE REVIEW

In their study, Neloy *et al*., [2022] introduced a new machine learning model called the Weighted Average Ensemble. This model combines three commonly used machine learning techniques (Random Forest, Decision Tree, and Naive Bayes) and showed superior results. The researchers used a combined dataset from Cleveland, Long Beach VA, Switzerland, Hungarian, and Stat log datasets to evaluate the model's performance using various metrics such as Accuracy, Precision, Recall, F1-score, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The results showed that the average ensemble model's precision, recall, and F1-score were all 0.93, which is a significant improvement. Compared to the other six algorithms, the Weighted Average Ensemble model also performed the best in terms of MAE, MSE, and RMSE, with scores of 0.07, 0.07, and 0.27, respectively. However, it is important to note that the dataset used in the study has a limited number of data points, and the results may not be accurate enough to be used as a reliable model for heart disease prediction.

Ghosh *et al*., [2021] proposed an efficient prediction of cardiovascular disease using machine learning algorithms with Relief and LASSO feature selection methods. The following methods are used which was proposed by a number of researchers: DT, Gradient Boosting (GB), KNN, RF, Decision Tree Bagging Method (DTBM), Random Forest Bagging Method (RFBM), K-Nearest Neighbors Bagging Method (KNNBM), AdaBoost Boosting Method (ABBM), and Gradient Boosting Boosting Method (GBBM). Researchers used a combined dataset from Cleveland, Long Beach VA, Switzerland, Hungarian, and Stat log datasets. Accuracy, Precision, Recall, F1-score, False Positive Rate, False Negative Rate, and Negative predictive value metrics were used to evaluate the employed algorithms. Based on the result analysis, using RFBM and the Relief feature selection method achieved an accuracy above 90%. However, the obtained results are not accurate enough to be a reliable model.

Mienye, *et al*., [2020], the authors proposed an ensemble learning approach for the prediction of heart disease risk using a weighted aging classifier ensemble. The authors compared the performance of ensemble classifiers with machine learning algorithms including (KNN), (LR), linear discriminant analysis (LDA), (SVM), classification and regression tree (CART), gradient boosting, and random forest. Two heart disease datasets are used, the Cleveland dataset obtained from the University of California, Irvine (UCI) repository, and the Framingham dataset obtained from the Kaggle website. There are 303 instances and 14 attributes in the former, while there are 4238 instances and 16 attributes in the latter. The Framingham dataset has missing attributes and has been preprocessed to make its machine-learning compatible. Age, sex, cholesterol level, blood pressure, alcohol consumption, and diabetes are all included in both databases. The models' performance is measured using accuracy, precision, sensitivity, and F1 Score. With 93% accuracy, 96% precision, 91% sensitivity, and a 93% F1 score, the proposed method performed best. This model, however, has a high dimensionality of data.

Li *et al*., [2020] proposed a heart disease identification method using machine learning classification in E Healthcare. Researchers study the impact of using two feature selection methods (i.e., Relief and LASSO) on the performance of six standard machine learning techniques. These techniques are SVM, Artificial Neural Network (ANN), NB, DT, LR, and K-Nearest Neighbors (KNN). The Cleveland heart disease dataset, which was extracted from the UCI machine learning repository and contains 303 instances and 75 attributes, is used in this work. Accuracy, sensitivity, specificity, precision, and Matthews Correlation Coefficient (MCC) metrics are used to evaluate the performance of the employed techniques. The accuracy of SVM with their feature selection algorithm was achieved at 92.37%. However, this model suffers from the small size of the data.

Khan [2020], the authors proposed an IoT framework for improving heart disease prediction based on the Modified Convolution Neural Network (MDCNN) classifier. The authors compared the performance of MDCNN with that of Deep Learning Neural Networks (DLNN) [2020] and Logistic Regression (LR). Data from the UCI machine learning repository, Framingham, Public Health, and Sensor Data were used to train and evaluate the disease. Accuracy, precision, sensitivity, recall, and F1 Score metrics were used to evaluate the performance of the MDCNN and the other employed methods. It was found that the proposed model achieved the best results compared with other methodologies. The MDCNN achieves 98.2% accuracy. In contrast, the existing LR and DLNN have lower accuracy of 88.3% and 81.6%, respectively. However, this model suffers from the high dimensionality of data.

Mohan, *et al*., [2019], the authors proposed a hybrid machine-learning technique for the effective prediction of heart disease. The proposed technique improves the accuracy of the prediction of the cardiovascular disease model for early diagnosis of the disease and protects people's lives. The Naïve Bayes (NB), Generalized Linear Model (GLM), Linear Model (LM), Deep Learning (DL), Decision Tree (DT), Random Forest (RF), Gradient Boosted Trees (GBT), and Support Vector Machine (SVM) methods were implemented and compared. The dataset used in this work was collected from the UCI machine learning repository. There are four databases (i.e., Cleveland, Hungary, Switzerland, and the VA Long Beach). Several standard performance metrics such as accuracy, precision, and classification error have been considered for the computation of the performance efficacy of these techniques. From the obtained results, the authors selected hybrid RF and LM to propose a new hybrid method called Hyper Random Forest Linear Model (HRFLM). The author compared his work using HRFLM with other researchers' methodologies. It was found that the proposed hybrid model achieved better results in the case of using the accuracy, and classification error criterion in the evaluation, but it decreased when using precision, F-Measure, sensitivity, and specificity. HRFLM achieved an accuracy of about 88.4%.

A study by J. & S., [2019] used two supervised classifiers called Naïve Bayes Classifier and Decision Tree Classifiers to predict heart diseases on a dataset. 8 Their Decision Tree model predicted the heart disease patients with an accuracy of 91 percent and the Naïve Bayes Classifier had an accuracy of 87 percent.

Mustafa *et al*., [2018] proposed an ensemble approach for better prediction by combining five classifiers. Their work included SVM, ANN, Naïve Bayes, 10 Regression analysis, and Random Forest. Their goal was to predict and diagnose cardiovascular disease.

Samuel *et al*., [2017] predicted the risk of heart failure using the Artificial Neural Network (ANN). Their work included fuzzy analytic hierarchy (AHP) to calculate the global weights of features depending on individual contributions. Afterward, the feature contributions were applied to train the ANN classifier to predict the patient's risk of heart failure.

Yekkala *et al*., [2017] examined bagging ensemble methods such as Random Forest and Adaboost along with Particle Swarm Optimization (PSO) to predict heart disease. They achieved high accuracy with bagging with PSO.

Dolatabaddi *et al*., [2017] used an optimized Support Vector Machine for their classification model, they extracted HRV signals from ECG in domains, time, and frequency for automated diagnosis of coronary artery disease. The overall accuracy of the research showed the strength of classification.

Table 1 list the discussed related work in summary. It shows the year of publication, algorithms used, employed datasets, and accuracy achieved for each related work. From this table, these works suffer from the small size of the used datasets. Researchers didn't implement all available datasets about cardiovascular disease using available classifiers, and most authors have relied on the use of standard datasets about heart disease. These models suffer from the high dimensionality of data. The achieved accuracy wasn't stable on all classifiers used by the authors and was not accurate enough to be a reliable model.

Table 1: Summary of the related work

Year	Algorithm Used	Dataset	Best Accuracy Achieved	Advantages	Disadvantages
2022	Random Forest, Decision Tree, and Naive Bayes, and Weighted Average Ensemble model	A combined dataset from Cleveland, Long Beach VA, Switzerland, Hungarian, and Stat log datasets are used.	Precision, recall, and F1-score are all 0.93%. MAE, MSE, and RMSE of 0.07, 0.07, and 0.27, respectively	A huge amount of datasets from the UCI Repository are used. In addition, using different evaluation metric to evaluate the performance of their model	The dataset has a limited number of data. And the obtained results are not accurate enough to be a reliable model for heart disease prediction.

Year	Algorithm Used	Dataset	Best Accuracy Achieved	Advantages	Disadvantages
2021	(AB), (DT), (GB), (KNN), (RF), (DTBM), (RFBM), (KNNBM), (ABBM), and (GBBM)	Cleveland, Long Beach VA, Switzerland, Hungarian and Stat lo	Above 90%	Researchers presented a comprehensive test with ten classifiers that include traditional and hybrid classifiers in addition using a dataset Combined from five datasets	The obtained results are not accurate enough to be a reliable model
2020	k-nearest neighbor (KNN), logistic regression (LR), linear discriminant analysis (LDA), support vector machine (SVM), classification and regression tree (CART), gradient boosting, random forest, and ensemble learning model	Cleveland dataset obtained from the University of California, Irvine (UCI) repository, and the Framingham dataset obtained from the Kaggle website	93%, 96%, 91%, and 93% using accuracy, precision, sensitivity, and F1_Score respectively.	A huge amount of data are used and implemented different machine-learning classifiers	The model suffers from high dimensionality of data
2020	Support vector machine, Artificial neural network, Naïve bays, Decision tree, Logistic regression, and K nearest neighbor	Heart disease contains 303 Instances and 75 attributes.	92.37%	Researchers presented a comprehensive test using machine learning classifiers and Artificial neural network	This model Suffers from small size of data
2020	Deep Learning Neural Network (DLNN), Logistic Regression (LR), and Modified Deep Convolutional Neural Network (MDCNN).	UCI machine learning repository, Framingham, Public Health, and Sensor Data	88.3%, 81.6%, And 98.2% when using LR, DLNN, and MDCNN respectively	The MDCNN achieves 98.2% accuracy which is the best results compared with other methodologies	This model suffers from high dimensionality of data
2019	Naïve Bayes, Generalized Linear Model, Logistic Regression, Deep Learning, Decision Tree, Random Forest, Gradient Boosted Trees, Support Vector Machine, VOTE, and HRFLM	Cleveland, Hungary, Switzerland, and the VA Long Beach	88.4%	Accuracy, and classification error criterion achieved good results in the evaluation.	The proposed hybrid model performance decreased when using precision, F-Measure, Sensitivity, and Specificity.

RESEARCH METHODOLOGY

In predicting heart disease, various Machine Learning algorithms are available. These include SVM, Naïve Bayes, Decision Trees, Bagging and Boosting, and RandomForest. Each algorithm has its advantages and disadvantages. However, we recommend using the RandomForest Classifier algorithm to build an optimized model.

Cardiovascular disease prediction is a critical area in medical prediction datasets, and Machine Learning has found vast applications in this field. Identifying high-risk patients early on is crucial as cardiovascular incidents can be fatal. Therefore, timely diagnosis and preventative care are essential goals.

In this section, the overall steps of our proposed model, the most important evaluation metrics used to evaluate the performance of the proposed model, and the machine learning classifiers are discussed.

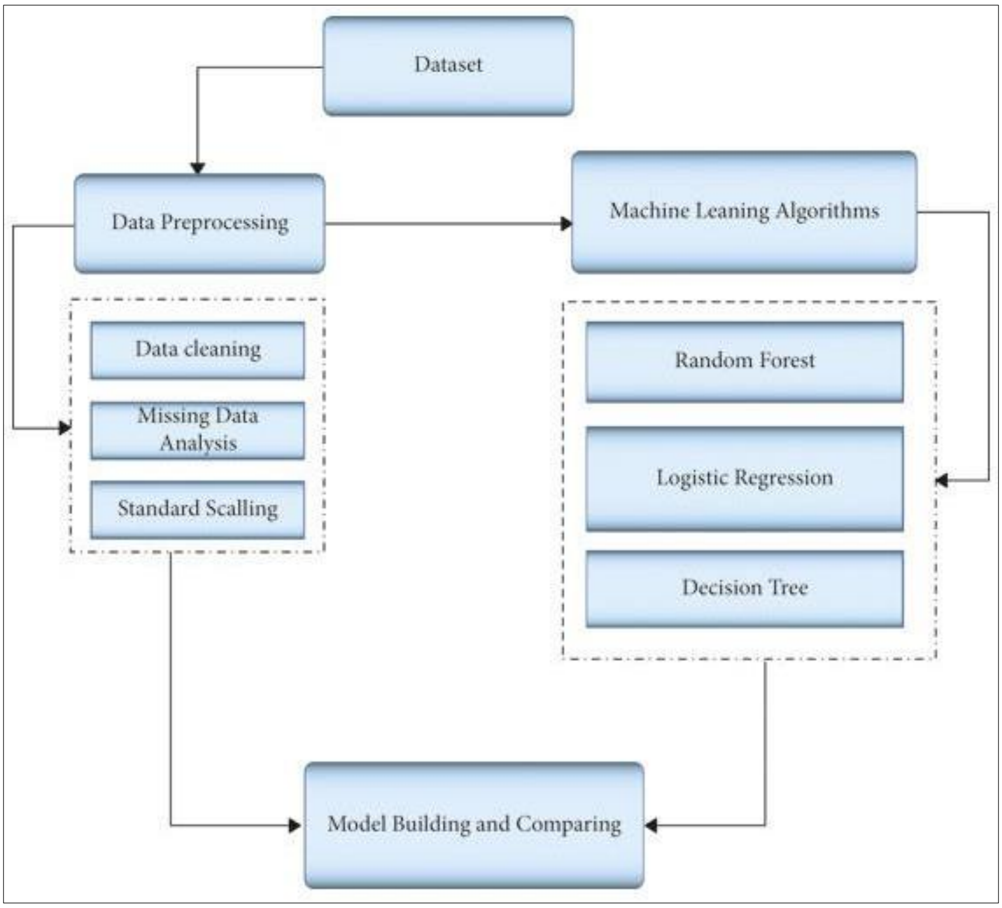

Fig. 1: The system's schematic diagram

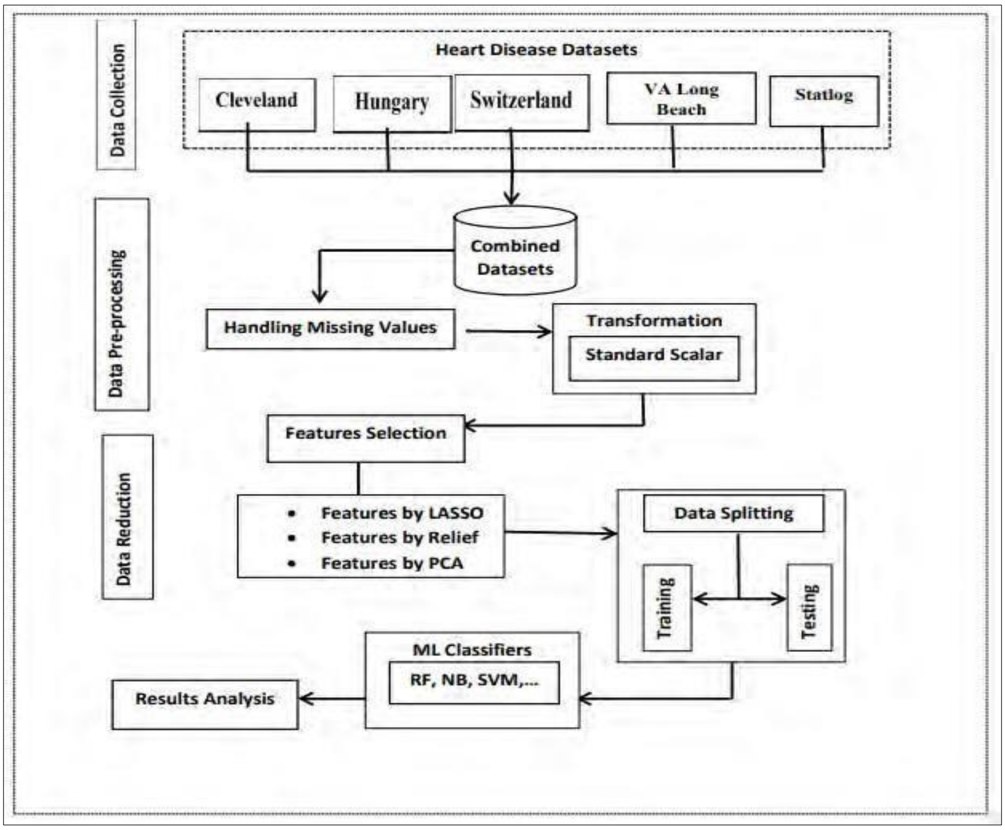

Fig. 2: The architecture of our proposed model

As shown in fig.1 and fig 2. The data used consists of five data sets which are extracted from the UCI repository and then combined for processing. In the data preprocessing stage, the collected data is analyzed to check the value of NaN and replace it with the best value. Missing values can be dealt with using a variety of strategies, including imputation and deletion. This problem is solved in our dataset by replacing all NaN values with the mean value

Performance Measure Indices

The evaluation metrics used for evaluating the employed classifiers are accuracy, precision, recall, and F1 score are discussed in this section. First, some terminologies are discussed:

1. True Positive: - Consider the time when the model's heart disease was accurately recognized.
2. True Negative: - When the model successfully identified the opposing class, such as patients who do not have any heart problems.
3. False Positive: - Refer to when the model incorrectly identified heart disease patients i.e., identifying non- heart disease patients as heart disease patients.
4. False Negative: - When the model wrongly identifies the opposite class, such as heart disease patients as normal patients.

Accuracy refers to the proximity of the measurements to a specified value. The higher the accuracy value, the better the performance of the model used as defined in Eq 1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision as defined in Eq. (2) quantifies the number of positive class predictions that actually belong to the positive class.

$$Precision = \frac{TP}{TP + FP}$$

Recall quantifies the number of positive class predictions made out of all as defined in Eq. (3).

$$Recall = \frac{TP}{TP + FN}$$

In Eq. (4), the F1-score combines precision and recall in relation to a given positive class - The F1 score can be thought of as a weighted average of precision and recall, with 1 being the highest and 0 being the worst.

$$F1\text{-}score = 2*((precision*recall)/(precision+recall)).$$

Overview of the Proposed Algorithms

Machine learning is a type of data analysis that automates the creation of analytical models. It's a subset of artificial intelligence based on the concept that machines can learn from data, recognize patterns, and draw conclusions with little or no involvement from people.

The machine learning algorithms used in our suggested methodology are briefly explained in this section.

1) Random Forest

Random-forest is a supervised machine learning method that can be used for classification and regression. The trees in the random forest run in a straight line. During the tree-building process, there is no interaction between these trees. It operates by training by constructing a huge number of decision trees. Then either the mean prediction (regression) or the category representing the mode of the categories (classification) is output. With certain useful adjustments, it aggregates the results of numerous predictions that aggregate many decision trees. As can be seen in fig.2, by merging decisions from a set of basic models, the random forest can create predictions.

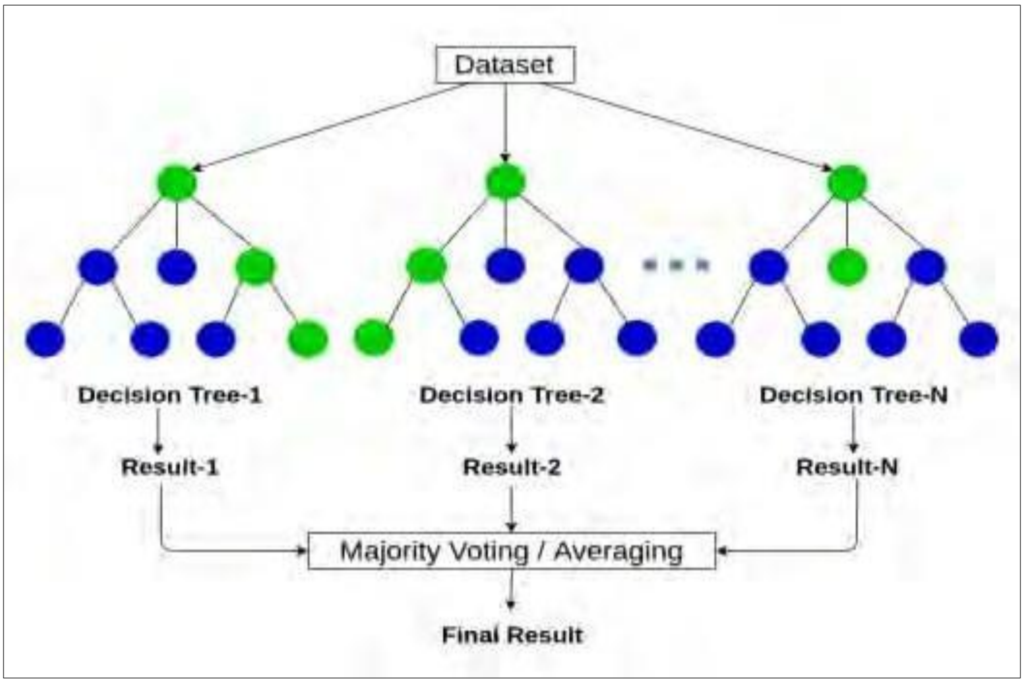

Fig. 3: Flow Chart of Random Forest Algorithm

Decision Trees

One of the most often used forms of supervised learning, a decision tree is a powerful prediction-making/ categorization algorithm that uses previous data to progress from root nodes to decision nodes to leaf nodes. In its most basic form, information is split along branches and ultimately into leaf nodes. The dataset 13 contains independent variables; data pruning without pertinent medical knowledge presents challenges, and a continuous variable decision tree must be employed given the nature of the data. Thus, multiple variables can be considered in this way which is crucial for creating an accurate model.

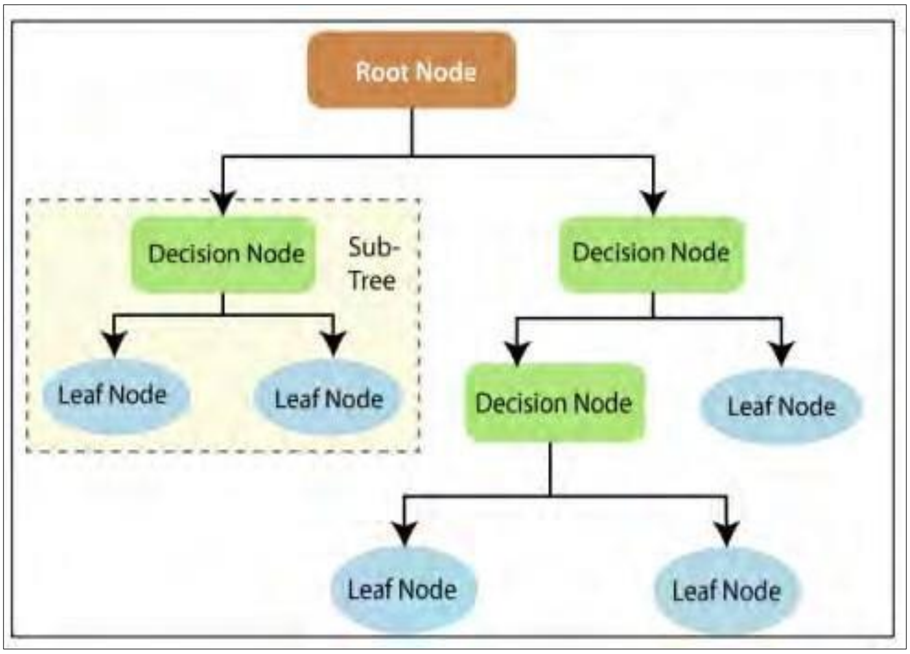

Fig. 4: Flow Chart of Decision Tree Classifier

Support Vector Machine

SVM is a type of supervised machine learning classifier that is commonly used in classification problems. AS shown in fig.5, each data item is represented as a point in n- dimensional space (where n is the number of features), with the value of each feature being the value of a given coordinate. Then, using the information collected from the dataset, a prediction is produced.

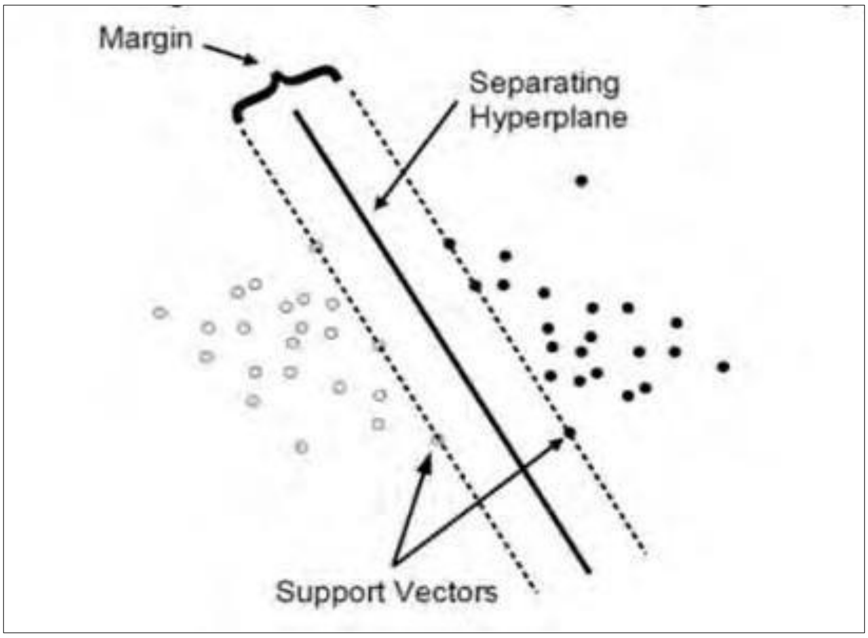

Fig. 5: Example of support vector machine working

Naïve Bayes

A Naive Bayes system is simple to create and does not require entangled iterative parameter estimation, making it particularly effective for large datasets. From P(c), P(x), and P(x|c) (discussed below), Bayes' hypothesis provides a method for determining the returned likelihood, P (c|x) The impact of the estimation of an indicator (x) on a given class (c) is independent of the estimations of multiple predictors as specified in Eq. (5), according to the Naive Bayes classifier.

- $P\ (c|x)$is the opportunity of class (target) given predictor (attribute).
- (c) is the preceding opportunity of class.
- $(x|c)$ is the opportunity of predictor given class.
- (x) is the preceding opportunity of a predictor.
- P (c|x)= P(x|c)P(c)/P(x)

CONCLUSION

This research explored various machine learning approaches for predicting heart disease using feature selection and classification techniques. By utilizing diverse datasets, the study identified a combination of LASSO for feature selection and Random Forest for prediction that achieved a 100% accuracy rate in heart disease prediction. Additionally, other methods like Decision Tree, Gradient Boosting, and AdaBoost achieved near-perfect accuracy, exceeding 99.5%. These findings suggest the potential of machine learning for efficient and accurate heart disease prediction, highlighting the valuable contribution of feature selection in achieving optimal results.

However, limitations exist. The research acknowledges the need for further exploration with larger datasets and hyperparameter tuning to refine the models and enhance generalizability. Additionally, the study acknowledges that these results require further investigation in real-world clinical settings.

Overall, this research demonstrates the promising potential of machine learning techniques, particularly with feature selection, in achieving highly accurate heart disease prediction. Future research can build upon these findings to further refine and validate these models for real-world application in healthcare settings.

REFERENCES

- Bambrick, N. (2022). Support Vector Machines: A simple explanation. KDnuggets. Retrieved November 3, 2022, from https://www.kdnuggets.com/2016/07/support-vector- machines-simple explanation.html
- Camm, J. D., Cochran, J. J., Fry, M. J., & Ohlmann, J. W. (2021). Business analytics: Descriptive, predictive, prescriptive. Cengage
- Chauhan, N. S. (2022, April). Naïve Bayes Algorithm: Everything you need to know. KDnuggets. Retrieved October 29, 2022, from https://www.kdnuggets.com/2020/06/naive-bayes-algorithm everything.html?hss_channel=tw-1318985240
- Dadgostar, P. (2019). Antimicrobial resistance: implications and costs. *Infection and drug resistance*, 3903-3910. https://doi.org/10.2147/idr.s234610.
- Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F. J. M., Ignatious, E., ... & De Boer, F. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access*, *9*, 19304-19326.
- Heart Disease Datasets From UCI Machine Learning

Repository. Accessed: May 31, 2020. [Online]. Available: https://archive.ics.uci. edu/ml/datasets/Heart+Disease

- Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *The Journal of Supercomputing*, *77*, 5198-5219.
- Khan, M. A. (2020). An IoT framework for heart disease prediction based on MDCNN classifier. *IEEE Access*, *8*, 34717-34727.
- Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Heart disease identification method using machine learning classification in e-healthcare. *IEEE access*, *8*, 107562-107582.
- Mienye, I. D., Sun, Y., & Wang, Z. (2020). An improved ensemble learning approach for the prediction of heart disease risk. *Informatics in Medicine Unlocked*, *20*, 100402.
- Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, *7*, 81542-81554.
- Neloy, M. A. I., Nahar, N., Hossain, M. S., & Andersson, K. (2022, February). A weighted average ensemble technique to predict heart disease. In *Proceedings of the Third International Conference on Trends in Computational and Cognitive Engineering: TCCE 2021* (pp. 17-29). Singapore: Springer Nature Singapore.
- Shailaja, K., Seetharamulu, B., & Jabbar, M. A. (2018, March). Machine learning in healthcare: A review. In *2018 Second international conference on electronics, communication and aerospace technology (ICECA)* (pp. 910-914). IEEE. https://doi.org/10.1109/iceca.2018.8474918.
- Tripoliti, E. E., Papadopoulos, T. G., Karanasiou, G. S., Naka, K. K., & Fotiadis, D. I. (2017). Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques. *Computational and structural biotechnology journal*, *15*, 26-47. https://doi.org/10.1016/j.csbj.2016.11.001.
- Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*, *19*(1), 1-16.