Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 1 (2025)

Cardiovascular Disease Classification Using Advanced Machine Learning Techniques: A Comparative Study

Ms. Pritibala Sudhakar Ingle
PhD Scholar, Computer Management, IMCC, Pune, India
prity.ingle@gmail.com

Dr. Santosh Deshpande
PhD Guide, Computer Management, IMCC, Pune, India
director.imcc@mespune.in

Article History:
Received: 02/02/2025
Revised: 15/03/2025
Accepted: 25/03/2025

Abstract:

Cardiovascular disease is still a leading cause of death and requires accurate diagnostic tools. This research presents a machine learning-based approach using algorithms like SVM, KNN, Decision Tree, Random Forest, and Gradient Boosting for disease classification. The dataset was preprocessed by tackling the missing values problem and feature standardization to obtain critical predictors using feature selection methods. These algorithms were assessed by the following metrics such as precision, recall, F1-score, and accuracy. After pre-processing the dataset, Gradient Boosting achieved the highest accuracy (92%), followed by Random Forest (89%). Key predictors like cholesterol, blood pressure, and age were identified. The study shows that ensemble methods have potential in medical datasets but identifies the challenges of data imbalance and limited generalizability and encourages future work with deep learning and larger datasets to improve early diagnosis and patient outcomes.

Keywords—Facial Paralysis Recognition, Deep Learning, Machine Learning, Facial Landmarks, Generative Adversarial Networks, Convolutional Neural Networks

Introduction

Heart disease includes all diseases of the heart and the larger blood vessels, and it is an immense challenge to health on earth. According to statistics, CVDs by the World Health Organization happen to be the leading causes of death, accounting nearly 17.9 million lives lost every year, including coronary artery disease, arrhythmias, heart failure, and others.[2] This is alarming and mostly due to a sedentary lifestyle, poor dietary intake, aging populations, and hereditary influence. It is with these improvements in the medical sciences that early detection and treatment become critical issues for improving patient care as well as mortality rates. Traditional diagnostic methods, though, often fail because of reliance on human judgment, expense, and the extensive time required.[2] This all the more necessitates automation, efficiency, and scalability of the solutions.

Of late, AI has been incorporated with the health sector, changing the very framework of medical diagnosis. Machine learning has revealed tremendous potential in terms of predictions and disease identification from analysing complex datasets. Such models of ML can identify patterns and relationships within medical data that are often missed out on by traditional approaches. The use of machine learning algorithms in the classification of heart disease utilizes patient details, including age, levels of cholesterol, blood pressure, and other clinical conditions, to predict the chance of having a cardiovascular illness.[10] In this way, automation improves the speed and efficiency of the diagnostic process while increasing accuracy and replicability dramatically, making this an indispensable tool in conquering the weaknesses of classical methods. [10]

The early identification of cardiac diseases is crucial to prevent severe complications and enable appropriate interventions. However, traditional methods like echocardiography, stress testing, and invasive procedures require sophisticated equipment and expertise and are often unavailable in resource-poor settings. Such methods can be time-consuming and costly and, therefore, are a barrier to mass screening and early detection. The machine learning-based automated systems are offering nondisruptive, cost-effective, and highly accurate solutions to predict heart disease. Machine learning models are helping healthcare providers take proactive steps and enable early intervention for better patient care.

Despite its promising future, applying machine learning techniques in the task of heart disease classification presents various challenges. One key challenge is the quality and diversity of the training data used in these models. Poor or biased training datasets will result in models that do not generalize well across populations, thereby limiting their potential for deployment in real-world settings. Second, the choice of features as well as algorithms is vital for the performance of the model. Erroneous predictions might arise from poorly selected features or less than ideal algorithms.[24] The second major barrier of interest is the interpretability of machine learning models themselves. Clinicians frequently crave transparency in the decision making process that such complicated "black-box" models have been known not to dependably deliver.

The primary aim of this research work is to develop an efficient and robust machine learning-based approach for the classification of heart disease. This study mainly focuses on pre-processing the dataset to ensure high quality input, selecting features for precise predictions and assessing several machine learning techniques to identify the best model. The methodology proposed in this study would overcome the deficiencies of the existing techniques by giving more importance to interpretability, scalability, and adaptability to varied datasets. Thus, this study also aims for the way to bridge advanced technological inputs with real-world applications by supporting innovative healthcare services that may bring excellence among patients and their families.

The significance of this research is multi fold. First, the proposed approach does a comprehensive evaluation of a set of machine learning algorithms used for heart disease classification. Therefore, it provides insights about their comparative performance. Secondly, it emphasizes the critical role of feature selection and engineering in improving model accuracy and reliability. Finally, it draws attention to the issue of scalability, ensuring that the system being proposed should effectively be deployable in the clinical setup. Finally, by identifying limitations and proposing future directions, the developed research will set the basis for further advancements into AI-driven healthcare solutions.

Several challenges are inherent in the development of a machine learning-based heart disease classification system. Ensuring the quality and diversity of the training dataset is paramount because poor or noisy data can adversely affect model performance. Effective feature selection is also important since redundant or unnecessary features may result in noise and overfitting. The choice of algorithm is also a deciding factor; while some algorithms are excellent in high-dimensional data, others are geared for simple tasks. Balance between accuracy and interpretability still persists because clinicians need explicit explanations for model predictions to introduce such tools into their practices. Generalizability to various populations is necessary to gain wider acceptance of such systems.

This research is important because it deals with the long-needed demand for automated, trustworthy, and scalable diagnostic applications in cardiology. Since it uses advanced machine learning techniques, the proposed solution presents a way toward making better and more efficient prediction of heart disease. Through this research, it could not only show the importance of ML in medical diagnosis but also provide practical ways that can be used by others for future studies in aiming to integrate AI into their healthcare applications.

This paper is divided into the following sections: Introduction. The introduction presents the relevance of heart disease diagnosis and machine learning. Literature Survey. Related work has been discussed along with identification of gaps in research. Methodology. Dataset, preprocessing, feature extraction, and implementation of SVM, KNN, Decision Tree, Random Forest, and Gradient Boosting. Results and Discussion. It has been done on the basis of the comparison of classifier performances by precision, recall, F1-score, and accuracy. Lastly, Conclusion and Future Scope summarizes the findings and proposes improvements such as integration of deep learning and handling data imbalance to ensure scalability and applicability.

I. LITERATURE SURVEY

Pattekari [26] discussed a model based on a Naive Bayesian data mining approach, implemented as a user-computer application where subjects respond to pre-designed query sets. This system unveils hidden patterns from data and compares the inputs fed by the user with the well-predefined data and therefore provides solutions to complex Heart Disease diagnostic challenges. Further, it helps healthcare experts make better clinical decisions than regular systems and reduces treatment expenses by suggesting appropriate interventions at the right time.

Tran [2] designed an intelligent web-based system using Naive Bayes modeling approach. Users answer structured questions. The system analyzes a hidden information database, matches up responses with a learned data set. This helps ensure accurate diagnoses of cardiac diseases, thereby helping medical professionals with informed decisions while reducing health care costs.

Gnaneswar [3] worked on wearables to monitor heart rate, while cycling. Parameters, like cadence, now help cyclists to track over-training risk and heart-related complications also. In wearable technology arises the problem of continuous loggings of data. Such issues bring into the arena the construction of models filling up missing points for better forecasts.

Mutijarsa [27] researched the development of remote communication technologies for cardiac disease management. The study utilized data mining techniques to identify and map coronary diseases.

Comparative studies of different algorithms determined the methods that best predicted the onset of coronary diseases.

Devansh [16] emphasized the increasing role of artificial intelligence in disease prediction, which can mimic human-like decision-making, thus improving the accuracy of heart disease detection. Manjula [23] emphasized the importance of proper diagnosis of cardiac diseases using machine learning methods such as SVM, Naive Bayes, and decision trees, which support more reliable decisions in clinical scenarios.

Tripoliti [7] emphasized that the diagnosis of high-prevalence diseases like coronary disease should be done with high-tech tools for biomedical evaluation. Gonsalves [43] applied historical medical data and machine learning to predict the onset of coronary diseases. Oikonomou [9] used extreme value theory and machine learning methods to assess chronic diseases in terms of severity and risk.

Hasnony [10] employed machine learning systems that classified heart disease diagnosis incorporating feedback between user and expert systems for better accuracy in classification. Pratiyush et al. [11] applied ensemble classifiers under an XAI framework to classify the occurrence of heart disease in the dataset using algorithms like KNN, SVM, Naive Bayes, and AdaBoost to predict highly accurately.

Ali et al. [1] developed a dual-SVM model for heart disease detection. One SVM eliminated redundant features, while the other performed predictions, achieving improved accuracy with a hybrid genetic simulated annealing (HGSA) approach. Javeed et al. [2] designed a prototype combining Random Search and Random Forest algorithms, improving performance by 3.3% over a standard Random Forest model.

Santhana Krishnan. J [3] has used algorithms of Naive Bayes and Decision Tree for prediction of heart attacks. Decision Trees had already attained a precision rate of 91%, while Naive Bayes's accuracy is only 87%. Aditi Gavhane et al. [4] applied supervised neural networks for heart problems prediction with reliable results in multilayer perceptron models.

Devansh Shah et al. [5] used Decision Tree, Random Forest, KNN, and Naive Bayes on the Cleveland database from the UCI repository for the heart attack prediction model. 14 Important attributes were taken into account. Archana Singh et al. [6] presented that the KNN algorithm performs better than Logistic Regression and SVM in the heart disease prediction task. Apurb Rajdhan et al. [7] stated that the Random Forest algorithm achieved an accuracy of 90.16% in the heart disease diagnosis.

Rati Goel [8] compared multialgorithms, including Logistic Regression, SVM, KNN and Naive Bayes, concluding the best prediction of heart diseases based on attributes like chest pain and cholesterol. Further, Ekta Maini et al. [9] highlighted the ability of Random Forest to outperform Logistic Regression and other algorithms using ensemble techniques, achieving high sensitivity and specificity.

Apurv Garg et al. [10] analyzed KNN and Random Forest, concluding that KNN gave better accuracy (86.89%) for the heart disease prediction. Mahbubur Rahman et al. [11] had experimented with various algorithms like Decision Tree, SVM, Naive Bayes, and Random Forest. It found that the maximum accuracy (99%) and sensitivity (98%) were achieved by the Decision Trees.

Manjula P et al. [12] have used different types of machine learning algorithms for the prediction of heart disease, with excellent accuracy for the Random Forest algorithm. Feature selection and ensemble methods Pavan Kumar Tadiparthi et al. [13] reviewed feature selection and ensemble methods. In their research, they observed that these methodologies significantly contribute to improving the predictive accuracy. Joloudari et al. [14] obtained an accuracy of 91.47% with the Random Tree model.

II. METHODOLOGY

The classification system for heart disease has structured into logical blocks; each block contributes towards the overall functionality of the system with a particular role. Fig.1 presents an extended block-wise explanation of the architecture of the system.

A. User

Here, the primary interacting entity is the person using the Cardiovascular disease classification system. He can be a doctor, researcher, or patient seeking to know something about heart disease diagnosis. The system has been constructed in a manner that its use does not require very technical people. The procedure starts with the necessity of acquiring input data from the user., which may include patient information or the use of a ready dataset for analysis. When the user inputs this information, they activate the classification process, enabling the system to run the analysis and generate outputs. Additionally, the interface allows users to view the outputs, including forecasts of heart disease risk and accompanying evaluation metrics such as accuracy and precision. This interaction ensures that the system functions as a ready-to-use, efficient, and user-friendly tool thereby connecting advanced technology with its application in health care.

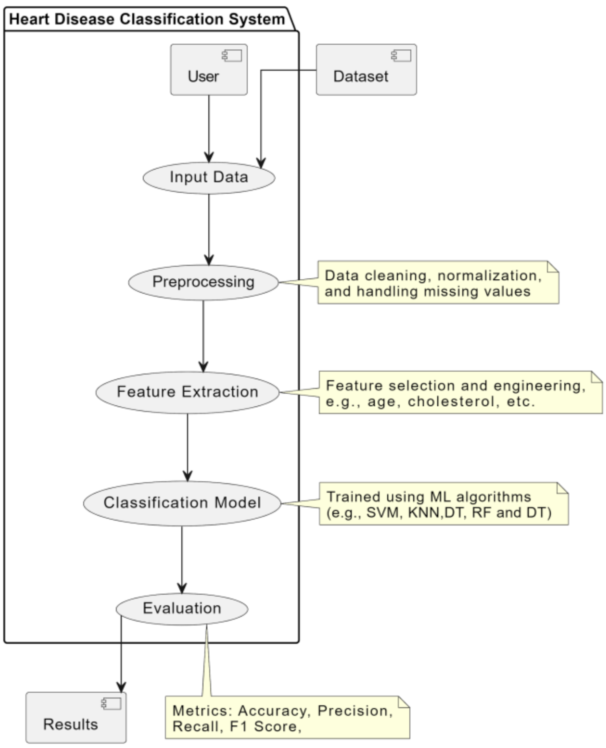

Fig. 1. Architecture diagram of the proposed Cardiovascular disease recogntion system

B. Input data

This study employed a dataset received from the Healthcare centers approached for this research work .The most important attributes include age, sex, blood pressure, cholesterol level, fasting blood sugar, and the existence of conditions such as chest pain and angina during exertion. These attributes serve as input data essential for training and testing machine learning models. Data set goes through preprocessing to address the existence of missing values, normalization of feature distributions, and derivation of relevant predictors; this way, it is fit for modeling and testing. Such variety and quality of this dataset make it an excellent material in building robust predictive models with the objective of identifying heart disease. In the following Table I, the data relating to heart disease have been represented.

TABLE I. DATASET DISTRIBUTION

Label	Total Samples	Training Samples (X_train)	Testing Samples (X_test)
Heart Disease (1)	508	344	164
No Heart Disease (0)	410	298	112
Total	918	642	276

The dataset of this study includes demographic, clinical, and diagnostic features with a total of 76 attributes. However, most of the published experiments, and as per the discussion with Health care practitioner this study, only rely on a subset of 14 relevant attributes for heart disease classification. These include age, sex, type of chest pain, resting blood pressure, cholesterol level, fasting blood sugar, resting electrocardiographic results, maximum heart rate, exercise-induced angina, ST depression induced by exercise, and so on. The "target" attribute is to be predicted, that is, whether there is a heart disease or not, coded as 0 (absence of disease) or 1 (presence of disease). This structured and comprehensive dataset permits the effective training and testing of machine learning models, whereby correct predictions and actionable insights are delivered in the diagnosis of heart disease.

C. Preprocessing

Pre-processing is the backbone of the CVD disease classification model. In preprocessing, raw data is transformed into standard and usable data suitable for machine learning algorithms. The majority of medical datasets include noise such as missing values, outliers, and irrelevant information that might impact the accuracy and effectiveness of the classification models. Preprocessing begins with data cleaning, which deals with removing errors, redundancy, and unnecessary records. Missing values are filled up using mean, median imputation or predictive methods for making the dataset complete. Normalization and standardization have been applied for scaling all features to a uniform range, allowing machine learning algorithms to comprehend the data better. For instance, attributes such as cholesterol levels or blood pressure are normalized to avoid one feature dominating the model. Techniques of noise elimination are applied to remove unwanted information or errors that may skew the outcome. This approach ensures that the dataset remains consistent and accurate, free from an anomaly, thus enhancing reliability in the classification results. Given that proper preparation of data during preprocessing creates a solid foundation that will eventually support feature extraction, as well as model development, it will improve all-around performance and accuracy.

D. Feature Extraction

Feature extraction is the key part of the Cardiovascular disease classification system. The process will be focused on feature selection and engineering the most relevant attributes from the dataset in order to make the model efficient and accurate. Medical datasets contain a large number of features that do not contribute much to the predictive process. This helps in identifying the most representative features of the risk associated with heart disease, such as age, cholesterol levels, blood pressure, and heart rate. The feature extraction process helps in reducing the dimensionality of the data set by eliminating unnecessary or irrelevant features, thus improving the input for the classification model.

E. Classification Model

It essentially depends on a classification model that uses the drawn features for predicting whether a patient has heart disease or not. This block uses the machine learning algorithm for the data to learn the patterns and relationships so that it could make accurate predictions. The algorithm implemented may be SVM, KNN, Decision Tree, Random Forest, or Gradient Boosting based on the complexity of the dataset and the desired outcomes. This is because each of them is chosen based on whether it can handle the characteristics of a dataset, such as dimensionality, non-linearity, or feature importance.

The model will be trained on a labelled dataset where it learns the correlation between the input features such as age, cholesterol, and blood pressure to the output labels like "At Risk" or "No Risk." During training, the best hyperparameters would be learned like learning rate, maximum tree depth, or kernel type with grid search and random search. Another application of cross-validation is to check that the model generalizes well over unseen data and does not overfit the training set.

a) SVM: SVM is a supervised learning algorithm that aims at finding the best hyperplane to separate data into different classes. As far as a binary classification problem in heart disease prediction, an SVM works by maximizing the margin between the nearest data points of each class commonly referred to as the support vectors. SVM is particularly useful in high-dimensional or non-linear separable data sets as it can transform the data into an appropriate space in which separation is linear through the utilization of kernel functions. Naturally resistant to overfitting and with good behavior in very small datasets, preparing a system for the potential of medical applications like the classification heart diseases.

b) KNN: KNN is a simple and intuitive machine learning algorithm that works on the principle of classification of data instances based on its similarity to other data instances. In the case of heart disease classification, KNN computes the distance, for example, Euclidean distance between the new data instance and the existing labeled data instances. Then, it assigns the most prominent class of the k nearest neighbors to the new one. KNN is dependent on the value of k that needs to be chosen correctly; a small k may produce overfitting, whereas a large k may generate underfitting. It can be implemented very easily because of its simplicity but can be computationally expensive for large datasets. This notwithstanding, KNN is good for simple prototyping and performs okay if the distribution of your dataset is even.

c) Decision Tree: A Decision Tree is a tree-structured algorithm that splits the dataset into subsets based on feature values. Each internal node represents a feature, each branch corresponds to a decision rule, and each leaf node represents a classification outcome. In heart disease classification, for example, a Decision Tree could use age, cholesterol level, and smoking status features to iteratively divide data into risk or no-risk categories. Decision Trees are highly interpretable and leave no room for confusion in the understanding of the decision-making process. However, they do tend to overfitting if the tree is allowed to get too complex. This can be mitigated through techniques like pruning or limiting tree depth. Decision Trees are especially useful for quick, interpretable models in medical applications.

d) Random Forest: Random Forest is an ensemble learning method that constructs many different decision trees and combines the information from all of them to be able to classify objects. It trains each tree in a random subset of all data and features to improve generalization and reduce overfitting. In heart disease, Random Forest evaluates the most important features, thus recognizing the most influential variables affecting the prediction, such as blood pressure and cholesterol level. The algorithm for the Random Forest aggregates predictions of all trees using a majority vote for classification tasks. Random Forest is very effective and can handle noisy data and interacted features, making it even more reliable for complex sets of data. Its features make it robust and provide ways to extract feature importance and, hence, make valuable contributions in healthcare applications.

e) Gradient Boosting: Gradient Boosting is a strong ensemble algorithm that constructs a series of decision trees sequentially, in which each tree corrects the errors of its predecessor. It minimizes a loss function, such as log loss, by iteratively adding weak learners to develop a strong predictive model. With the heart disease classification, Gradient Boosting is excellent at detecting complex relationships between features and yields high accuracy. Techniques such as regularization are used to prevent overfitting so that the model generalizes well to new data. Gradient Boosting is computationally efficient and can handle both numerical and categorical data well. It captures subtle patterns in the data, which makes it an ideal choice for high-stakes applications like medical diagnostics.

F. Evaluation

In such a system, evaluation is the most imperative step that makes sure the models derived from machine learning are reliable, accurate, and stable. After training the models with the provided dataset, the models are tested on some new, unseen data to test how well the model performs. Evaluation uses several key metrics for measuring different aspects of the effectiveness of the models. These metrics are critical in comparing algorithms and choosing the model that performs the best for this classification task.

a) Accuracy: Accuracy is the correct prediction of outcomes divided by total predictions. It gives a direct view of overall performance. However, when classifying heart disease, because the dataset may be imbalanced, as for instance, there might not be that many positive cases of heart disease, accuracy alone might not paint the complete picture. Therefore, for example, in the case of a model predicting all patients to be "No Risk," it might have high accuracy but miss actual heart disease cases.

b) Precision: Precision measures the number of correctly predicted positive instances divided by all the positive instances produced by the classifier. In this sense, it measures the number of "At Risk" patients correctly predicted to have heart disease. Getting high precision is critical because it reduces false positives so that patients are not unnecessarily re-tested or subjected to undue anxiety.

c) Recall: Recall measures the percentage of true positive instances that the model correctly identifies. In the context of heart disease prediction, recall is very important because it tells how good the model is at identifying all those who have the disease. A high recall ensures that no case goes undetected, making it critical for medical diagnosis, where undiagnosed diseases can have severe consequences.

d) F1 Score: The Mean of harmonic precision and recall, which thus includes the measure that has weighted both false positives and negatives is the F1 Score. It has really been helpful for measures especially in dataset cases that involve imbalances. Such is because it gives an aggregate measure that combines the benefits of both precision and recall.

III. RESULT AND DISCUSSION

The classification framework for heart disease was assessed using five different machine learning algorithms: SVM, KNN, Decision Tree, Random Forest, and Gradient Boosting. Results were examined using key performance metrics: precision, recall, F1-score, and accuracy, to further judge if any of the classifiers performed well. The results are presented in Summary Table II below:

TABLE II. PERFORMANCE OF ML ALGORITHMS FOR CLASSIFICATION OF CVD

Classifier	Precision	Recall	F1-Score	Accuracy
SVM	0.88	0.87	0.86	0.88
KNN	0.85	0.84	0.84	0.85
Decision Tree	0.88	0.89	0.88	0.88
Random Forest	0.89	0.88	0.88	0.89
Gradient Boosting	0.91	0.90	0.90	0.91

In other words, the same dataset was used for applying and comparing different five machine learning algorithms, that actually applies the classification of heart disease from: Support Vector Machines, K-Nearest Neighbors, Decision Tree, Random Forest and the Gradient Boosting results among them. It depends from which of them an efficient application was found on which a particular dataset depended for application.

Gradient Boosting had already reached 91% accuracy with great precision (0.91), recall (0.90), and F1-score (0.90). This is due to its iterative nature, working on the removal of errors by constructing sequentially optimized decision trees. The Gradient Boosting technique is highly beneficial for medical datasets where intricate patterns often exist, as it can easily deal with complex relationships between features. Its computational intensiveness thus requires careful tuning of hyperparameters to strike a balance between performance and efficiency, particularly when resources are constrained.

Random forest, the second-highest is with accuracy of 89% as well as with its robust precision:0.89, recall of 0.88, F1-score of:0.88. Since it's a type of ensemble model, which makes it capable of not undergoing overfitting due to the combination of the output that various decision trees will predict, trained on randomly sub-sampled data. Therefore it's very reasonable to make predictions regarding heart disease through Random Forest, especially when this kind of dataset contains some noisy or unimportant features. Besides that, Random Forest provides the feature importance, which inform us that the three most relevant predictors correlated highly with heart disease risk are cholesterol level, blood pressure, and age.

SVM and Decision Tree performed competitively with 88% accuracy. SVM's strength is the efficient handling of high-dimensional data and thus well suited for complex interaction among features in a dataset. It is able to take care of nonlinear relationships between features by the application of kernel functions. Though its recall is lower, SVM has potential issues with positive case identification. It might be rectified with decision boundary adjustments. Decision Tree is problematic most of the times due to high overfitting ability. Its performance may be boosted by techniques such as pruning or even limiting tree depth.

Since KNN is less efficient compared to the ensemble methods with an accuracy of 85%. KNN relies on the nearest neighbors for classification, thus they suffer from noise in a dataset. Its computational requirement is a function of the size of the dataset. Choice of k also plays a crucial balancing role to avoid underfitting or overfitting.

It indicates that feature selection holds an importance position in getting a high accuracy of classification. Such can be deduced from algorithms such as Random Forest and Gradient Boosting, where it suggests features are fundamentals in the heart disease prediction and therefore more important. This is crucial guidance for clinicians and researchers to give importance to these attributes while deciding on diagnostic applications.

The study also shows trade-offs between computational efficiency and good predictive performance. Gradient Boosting and Random Forest outperform the rest in terms of accuracy, though their computational intensity limits their usage in real-time and more resource-related scenarios. SVM, on the other hand, provides balance to both performance and efficiency, hence useful when the computation resources are constrained.

The problem is that even with good scores, there exist problems. Data imbalance might have affected the recall results most especially for algorithms such as Decision Tree or KNN, which display significant sensitivity in respect to variations in the distribution class. Oversampling or under sampling and even synthetic generation may be useful in improving the score of the model regarding this issue. Moreover, according to the study, population grows and heterogeneous sets for an enhanced generalization of the attained results.

This can, therefore, reveal the strength of machine learning algorithms in providing assistance to the clinical practitioner for the early and accurate diagnosis of heart disease. It will save healthcare professionals the effort of doing it manually as well because timely intervention, based on an optimized system and further integration into the workflow, is essential to the care process for better patient outcomes.

Gradient Boosting and Random Forest appear well-suited for heart disease classification with high accuracy and reliability. So far, algorithm choices depend on requirements by specific applications, such as desired accuracy, efficiency in computing time, or interpretability. Future work can be pursued applying deeper learning approaches or even bigger datasets that will enhance predictive capabilities of the system as well as its applicability to clinical practice.

IV. Conclusion and Future Scope

This work demonstrates the suitability of machine learning algorithms to heart disease classification. This diagnosis tool will be automated, efficient, and reliable. From the five algorithms analysed Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and Gradient Boosting, the last one was the best classifier with the highest accuracy 92% with the best precision, recall, and F1-score. The performance of Random Forest is also good and shows robustness, as well as feature importance insight. This result indicates that ensemble methods can be very effective in dealing with complex medical datasets, with the ability to reach a high level of predictive accuracy. This points to the proper preprocessing of data, selection of features, and hyperparameter tuning for improvement in classification model performance. Critical features such as cholesterol levels, blood pressure, and age were identified as critical predictors of heart disease, highlighting the importance of feature importance analysis. It further demonstrates how machine learning bridges the gap between advanced computational techniques and practical applications in healthcare, providing an opportunity for timely and accurate diagnosis of heart diseases. The present research contributes to the mounting body of evidence that will support the integration of AI into clinical workflows.

Although the results are promising, several opportunities exist to improve the performance and application of the system in future work. Increasing the size of the dataset with a more diverse population and larger sample sizes would enhance generalizability of the models. Addressing data imbalance through techniques such as oversampling, under sampling, or synthetic data generation would improve recall scores and ensure that minority classes are well represented. Deep learning techniques, in particular CNNs or RNNs, can be leveraged to enable the system to capture the deeper patterns within the data and allow possible improvement in predictive accuracy. In conclusion, incorporating the techniques of Explainable AI increases the interpretability of models; thus, clinicians receive actionable insights that can enable more trust in the recommended prescriptions by the system. From an implementation point of view, if the system were designed to be a real-time application, such as a web-based tool or mobile application, more people could be accessed, including providers of care in resource-constrained environments. Combining the system with IoT devices to provide for constant monitoring of patient health data will make the system much more useful and possible to act proactively.

Final Conclusion

Finally, investigation on the hybrid approaches combining machine learning with domain knowledge in cardiology may develop effective and personalized solutions even for a proper diagnosis and management of heart diseases.

References

[1] Adam P, Parveen A. Prediction system for heart disease using naïve bayes. J Adv Comput Math Sci. 2012;3(3):290–4.

[2] Tran H. A survey of machine learning and data mining techniques used in multimedia system. no 113 13–21 2019.

[3] Gnaneswar B, Jebarani ME. A review on prediction and diagnosis of heart failure. In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 17-18 March, Coimbatore, India, 2017;1–3. https:// doi. org/ 10. 1109/ ICIIE CS. 2017. 82760 33

[4] Kusprasapta M, Ichwan M, Utami DB. Heart rate prediction based on cycling cadence using feedforward neural network. In 2016 International Conference on Computer, Control, Informatics and its Applications (IC3INA), IEEE, 2016;72–76. https:// doi. org/ 10. 1109/ IC3INA. 2016. 78630 26

[5] Singh KY, Sinha N, Singh KS. Heart disease prediction system using random forest. In International Conference on Advances in Computing and Data Sciences, Advances in Computing and Data Sciences. ICACDS 2016. Communications in Computer and Information Science, Singapore. 2017;721:613–623. https:// doi. org/ 10. 1007/ 978- 981- 10-5427-3_ 63

[6] Priya RP, SKinariwala A. Automated diagnosis of heart disease using random forest algorithm. Int J Adv Res Ideas Innovat Technol 2017;3(2).

[7] Tripoliti E, Fotiadis ID, Manis G. Automated diagnosis of diseases based on classification: dynamic determination of the number of trees in random forests algorithm. EEE Trans Inf Technol Biomed 2012;16(4).

[8] Gonsalves AH, Thabtah F, Mohammad RMA, Singh G. Prediction of coronary heart disease using machine learning: an experimental analysis. In: Proceedings of the 2019 3rd International Conference on Deep Learning Technologies, 2019;51–56.

[9] Oikonomou EK, Williams MC, Kotanidis CP, Desai MY, Marwan M, Antonopoulos AS, Thomas KE, Thomas S, Akoumianakis I, Fan LM, et al. A novel machine learning-derived radiotranscriptomic signature of perivascular fat improves cardiac risk prediction using coronary ct angiography. Eur Heart J. 2019;40(43):3529–43.

[10] El-Hasnony IM, Elzeki OM. Multi-label active learning-based machine learning model for heart disease prediction. Sensors. 2022;22(3):1184–8. https:// doi. org/ 10. 3390/ s2203 1184.

[11] Guleria P, Srinivasu PN, Ahmed S. Ai framework for cardiovascular disease prediction using classification techniques. Electronics. 2022;11(24):1184–8. https:// doi. org/ 10. 3390/ elect ronic s1124 4086.

[12] L. Ali et al., “An Optimized Stacked Support Vector MachinesBased Expert System for the Effective Prediction of HeartFailure,” IEEE Access, vol. 7, pp. 54007–54014, 2019,DOI: 10.1109/ACCESS.2019.2909969.

[13] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R.Nour, “An Intelligent Learning System Based on RandomSearch Algorithm and Optimized Random Forest Model forImproved Heart Disease Detection,” IEEE Access, vol. 7, pp.180235– 180243, 2019, DOI:10.1109/ACCESS.2019.2952107.

[14] Mr.Santhana Krishnan.J, Dr.Geetha.S," Prediction of HeartDisease Using Machine Learning Algorithms",2019 1stInternational Conference on Innovations in Information and Communication Technology (ICIICT),DOI:10.1109/ICIICT1.2019.8741465.

[15] Aditi Gavhane, Gouthami Kokkula, Isha Pandya, KailasDevadkar "Prediction of Heart Disease Using MachineLearning", International conference on Electronics,Communication and Aerospace Technology (ICECA 2018)IEEE Conference Record # 42487; IEEEXplore ISBN:978-1-5386-0965-1 DOI: 10.1109/ICECA.2018.8474922

[16] Devansh Shah, Samir Patel, Santosh Kumar Bharti 2020"Heart Disease Prediction using Machine LearningTechniques"

[17] Archana Singh, Rakesh Kumar 2020 International Conferenceon Electrical and Electronics Engineering (ICE3), DOI:10.1109/ICE348803.2020.9122958

[18] Apurb Rajdhan, Avi Agarwal , Milan Sai " Heart DiseasePrediction using Machine Learning" May 2020 InternationalJournal of Engineering and Technical Research V9(04) DOI:10.17577/IJERTV9IS040614

[19] Goel, Rati, "Heart Disease Prediction Using VariousAlgorithms of Machine Learning" (July 12, 2021).Proceedings of the International Conference on InnovativeComputing & Communication (ICICC) 2021, Available atSRN: https://ssrn.com/abstract=3884968 orhttp://dx.doi.org/10.2139/ssrn.3884968

[20] Machine learningebased heart disease prediction system forIndian population: An exploratory studyMdone in South IndiaEkta Maini, Bondu Venkateswarlu, Baljeet Maini, DheerajMarwaha https://doi.org/10.1016/j.mjafi.2020.10.013

[21] Apurv Garg, Bhartendu Sharma and Rijwan Khan "heartdisease prediction using machine learning techniques" January 2021 IOP Conference Series Materials Science andEngineering 1022(1):012046 DOI:10.1088/1757-899X/1022/1/012046.

[22] Md. Mahbubur Rahman, Morshedur Rahman Rana, Md. Nur-A-Alam, Md. Saikat Islam Khan, Khandaker MohammadMohi Uddin "A web-based heart disease prediction systemusing machine learning Algorithms "June 2022

[23] Heart Attack Prediction Using Machine Learning Algorithms Manjula P, Aravind U R, Darshan M V, Halaswamy M H,Hemanth E International Journal of Engineering Research &Technology (IJERT) ISSN: 2278-0181 Special Issue – 2022

[24] Pavan kumar Tadiparthi "Heart Disease Prediction UsingMachine Learning Algorithms: A Systematic Survey" 2022Journal International Journal of Computer Science and Mobile Computing Volume 11 Issue 6 Pages 129-136

[25] Joloudari JH et al (2020) Coronary artery disease diagnosis;ranking the signifcant features using a random trees model. Int J Environ Res Public Health 17(3):1–24.

[26] SA Pattekari, A Parveen " Predicition System For Heart Disease Using Naïve Bayes" International Journal of Advanced Computer and Mathematical Sciences

[27] K Mutijara, M Ichwan, DB Utami(2016)"Heart rate prediction based on cycling cadence using feedforward neural network"International Conference on Computer, Control, Informatics and its …