iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: V Month of publication: May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52545

Machine Learning Techniques for Coronary Artery Bypass Prediction

Naveen[1], Md Saquib Shafique[2], Ajay Pal Singh[3]
[1, 2]University Institute of EngineeringChandigarh University, Mohali, India
[3]Asst. Professor CSE DepartmentChandigarh University, Mohali, India

Abstract: Heart failure, a complicated clinical problem,presently affects a smaller number of persons globally. Earlyon, cardiac centres and hospitals rely significantly on ECG toassess and diagnose heart failure. The utilization of an electrocardiogram, also known as ECG, is widespread in themedical field. Detecting heart ailments at an early stageremains a crucial challenge in the healthcare industry. The focus of this paper is to introduce various machine learning technologies for the detailed analysis of heart ailment detection. First, a weighted version of Nave Bayes is employed to forecast cardiac problems. The alternate bone, according to This system is designed for the automatic and anatomical localization/discovery of ischemic heart complaints. The system utilizes two classifiers that are similar to support vector machine (SVM) and XGBoost with swish performance, respectively. The system analyzes the features of the frequency sphere, time sphere, and information proposition to accurately locate and detect ischemic heart complaints. The third bone is an automated detection system for heart failure based on a bettered SVM based on the duality optimisation approach that was previously studied. To support clinical decision-making, a Heart Complaint Prediction Model (HDPM) is utilized in a Clinical Decision Support System (CDSS). As a result, treating problems properly and avoiding significant consequences will be easy. In order to evaluate essential decision tree-type algorithms for honing the finesse of heart complaint opinion, this study employs XGBoost. Four types of machine knowledge (ML) models are examined in terms of perfection, delicacy, f1-measure, and recall as performancecriteria.

I. INTRODUCTION

A. Background & Significance

Heart disease is among the leading causes of death globally. Detecting and predicting heart disease at an early stage is crucial in preventing negative outcomes and enhancing patient outcomes. However, traditional diagnostic approaches can be expensive, invasive, and time-intensive. Machine learning (ML) techniques have shown promise in predicting heart disease and can aid in early detection and better management of heart disease.

B. Objectives of the study

The main aim of this research is to create and assess machine learning algorithms for the prediction of heart disease. The study focuses on the following specific objectives:

1) Determine the key indicators for predicting heartdisease
2) Compare the effectiveness of several ML algorithms for predicting heart disease.
3) Analyse the produced models' interpretability andforecast accuracy.

C. Summary of the techniques employed:

The study will involve the following methods:

1) *Review of Literature:* The research on heart disease prediction using ML approaches that have already been done will be found by a thorough study of the literature. The review will help identify the research gaps and the most promising ML techniques for heart disease prediction.
2) *Data Collection & Preprocessing:* The study will use the publicly available data set containing a range of clinical and demographic features of patients with and without heart disease. Prior to analysis, the dataset will undergo preprocessing procedures such as imputation of missing values, handling of outliers, and normalization of the data.
3) *Feature Selection & Engineering:* Feature selection techniques will be used to identify the most important features for heart disease prediction. The study will also explore feature engineering techniques to create new features that may improve the performance of the models.

4) *Methodology:* For the purpose of predicting cardiac disease, Various machine learning techniques will be trained and tested, including logistic regression, decision trees, random forest algorithms, and artificial neural networks. The models will be optimized using hyper-parameter tuning techniques, and the study will involve training and testing a range of machine learning methods, including logistic regression, decision trees, random forest models, and artificial neural networks.
5) *Results & Discussion:* The results of the study will be presented in the form of performance metrics and visualization of the feature importance. The importance performance of the models will be compared, and their strengths and limitations will be discussed. The study will also explore the clinical interpretability of the models and the potential impact on clinical decision-making.
6) *Conclusion:* The study will conclude by summarizing the findings and their implications for heart disease prediction using ML techniques. The study will also suggest future research directions, such as developing more complex models, incorporating additional data structures, andevaluating the models in a clinical setting.

II. LITERATURE REVIEW

A. Overview of Heart Disease & their Causes

Coronary artery disease, heart failure, arrhythmias, and illnesses of the heart's valves are only a few of the several ailments referred to as "heart disease." Heart disease is associated with a number of prevalent risk factors, including high blood pressure, high cholesterol levels, smoking, diabetes, and a family history of the disease.

B. Previous studies on Heart Disease Prediction usingML Techniques

Several studies have explored the use of machine learning techniques for heart disease prediction for example a study by Cho et al. (2018) used a deep learning model to predict heart disease risk factors from ECG data. Another study by Alizadehsani et al. (2013) used decision trees and neural networks to predict heart disease risk factors from demographic and clinical data.

C. Strengths & Limitations of Previous Studies

The strength of previous studies on heart disease prediction using machine learning techniques include their ability to handle large amounts of data, identify complex patterns, andprovide accurate predictions. However, some limitations of previous studies include the lack of standardized data sets, limited clinical interpretability of other models, and potentialbiases in the data.

D. Research Gaps & the Need for Further Investigation

There are several study gaps and a need for more research despite the encouraging findings of earlier studies. For example, most studies have focused on predicting heart disease risk factors rather than predicting actual heart diseaseoutcomes. More standardized data sets are also required in order to assess the effectiveness of various ML algorithms. There is also a need to develop models that are more clinically interpretable and can be integrated into clinical decision-making. Finally, there is a need to evaluate the performance of the models in diverse populations and in real-world clinical settings.

III. METHODOLOGY

A. Description of the Dataset Used

The study will use a publicly available data set such as the Cleveland Clinic Foundation’s Heart Disease Dataset or the Framingham Heart Study Dataset. These datasets contain a range of clinical and demographic features of patients with and without heart disease such as age, gender, blood pressure,cholesterol levels, and ECG data.

B. Data Cleaning & Preparation

The dataset will be preprocessed to remove missing values, handle outliers, and normalize the data. The preprocessing steps may include:
1) Imputing missing values using mean, median, ormode imputation techniques.
2) Handling outliers using techniques such aswinsorization or outlier detection and removal.
3) Normalizing the data using techniques such as min-max scaling or z-score normalization.

C. Feature Selection & Engineering

Feature selection techniques will be used to identify the most important features for heart disease prediction. The study will use techniques such as:

1) The analytical approach will involve univariate analysis, which entails investigating the correlation between each feature and the outcome variable using statistical tests such as t-tests or chi-square tests.
2) Feature importance: Using ML algorithms such as decision trees or random forests to determine the relative importance of each feature.
3) Correlation analysis: Examining the correlation between each pair of features to identify redundant or highly correlated features.

The study will also explore feature engineering techniques tocreate new features that may improve the performance of themodels. Feature engineering techniques may include:

a) *Transforming Variables:* Transforming variables such as agents into groups or creating interaction terms between variables.
b) *Domain Knowledge:* Incorporating domain knowledge into the feature engineering process to create more informative features.
c) *Dimensionally Reduction:* To minimise the dimensionality of the data, approaches such as the use of PCA or t-distributed randomised neighbour embedding (t-SNE) are used.

IV. RESULT

A. Performing Metrics of the Models Used:

The study will use several performance metrics to elevate theperformance of the models for heart disease prediction such as:

1) *Accuracy:* The proportion of correctly classifiedinstances.
2) *Precision:* The proportion of true positives among the instance predicted as positive.
3) *Recall:* The fraction of genuine positives in the totalnumber of positive cases.
4) *F1 Score:* Precision and memory are balanced by the harmonic mean.

One of the performance metrics that will be utilized is the area under the receiver operating characteristic (ROC) curve, which measures the discrimination power of the classification model.

B. Comparison of the Models

The study will compare the performance of the different machine learning models used for heart disease prediction, such as logistic regression, decision trees, random forests, SVM, and ANN. The comparison will be based on the performance metrics described above, as well as thecomputational complexity and interpretability of the models.

C. Interpretation of the Results

The study will interpret the results of the models to gain insights into the factors that contribute to heart disease prediction. The primary interpretation involves:

1) *Feature Importance Analysis:* Assessing the corresponding relevance of several variables in the prediction of heart disease.
2) *Correlation Analysis:* Examining the correlationbetween the different features and heart disease to identify the most significant risk factors.
3) *Clinical Interpretation:* Interpreting the results in the context of clinical knowledge to provideactionable insights for clinicians and patients.

The study will also discuss the limitations and potentialbiases of the models and suggests directions for future research to address these issues.

V. RESULTS & DISCUSSION

A. Description of the Machine Learning Techniques Used:

To predict heart disease, the study will employ a range of machine learning methods, such as logistic regression, decision trees, random forest, support vector machine (SVM), and artificial neural networks (ANN). These techniques will be implemented using popular ML libraries such as sci-kit Learn and TensorFlow.

B. Model Selection & Training

To divide the dataset into training and testing sets, the study will utilize either a train-test split or cross-validation technique. The training set will be utilized to develop the models, while the testing set will be used to evaluate their performance. Additionally, the study will explore other methods such as K-fold cross-validation and stratified sampling to enhance the reliability and generalizability of themodels.

C. Hyperparameter Tuning

To modify the models' hyperparameters, the study will employ approaches such as grid search and randomised search. The hyperparameters, which include parameters like the learning rate or the number of hidden layers in an artificialneural network (ANN), are established before the training process and play a key role in determining the behavior of the models. Optimizing the hyperparameters will enhanceperformance metrics like precision, recall, accuracy, and F1 score.

D. Evaluation

Common performance measures such as accuracy, recall, F1score, and area under the receiver operating characteristic (ROC) curve will be utilized to evaluate the models' performance. The ROC curve graphically depicts the trade- off between the true positive rate (TPR) and false positive rate (FPR) for different classification thresholds. The study will also use techniques such as confusion matrices and feature importance plots to gain insights into the performance and interpretability of the models.

VI. IMPLICATIONS, CONTRIBUTIONS, LIMITATIONS, ANDFUTURE DIRECTIONS:

1) *Implications:* The study has several implications for the field of cardiology and machine learning. Firstly, thestudy can help to improve the accuracy and efficiency of heart disease predictions which can aid it in early diagnosis and treatment of the disease. Furthermore, the research can shed light on the variables that affect the likelihood of developing heart disease, which can help shape public health policy and preventative actions. Finally, the study can showcase the potential of machine learning in healthcare and encourage further research in this area.
2) *Contribution:* The study makes several contributions to the field of cardiology and machine learning. Firstly, the study evaluates the performance of different machinelearning models for heart disease prediction and providesinsights into the factors that contribute to the prediction. Secondly, the study explores the use of advanced techniques such as hyperparameter tuning and feature selection to optimize the performance of the models.Finally, the study highlights the potential of machine learning in healthcare and provides a framework for future research in this area.
3) *Limitations*: The study has several limitations that shouldbe acknowledged. The research's dependence on one data set, which could not be typical of the entire population, isthe first drawback. Secondly, the study doesn't consider external factors such as lifestyle and environmental factors which may also contribute to heart disease prediction. Finally, the study does not consider the cost- effectiveness and scalability of the models, which are important factors for real-world implementation.
4) *Future Directions*: The study proposes various avenues for future research. Firstly, researchers can employ larger and more diverse datasets to enhance the models' generalizability. Secondly, they can investigate how integrating external factors, such as lifestyle and environmental factors, can improve the accuracy of heartdisease prediction. Finally, future studies can consider the cost-effectiveness and scalability of the models to ensuretheir practicality and real-world implementations.

VII. SUMMARY & RECOMMENDATIONS

A. Summary

The study's main goal is to assess the effectiveness of various machine learning models in predicting heart disease and to gain insights into the contributing factors that influence the prediction. The study uses a dataset of patients with various clinical and demographic features and applies advanced techniques such as hyperparameter tuning and feature selection to optimize the performance of the models. The study evaluates the models based on performance metrics such as accuracy, precision, recall, F1 score, and area under the ROC curve and interprets the result to gain insights into the factors that contribute to heart disease prediction. Thestudy makes several contributions to the field ofcardiology and machine learning and suggests several directions for future research.

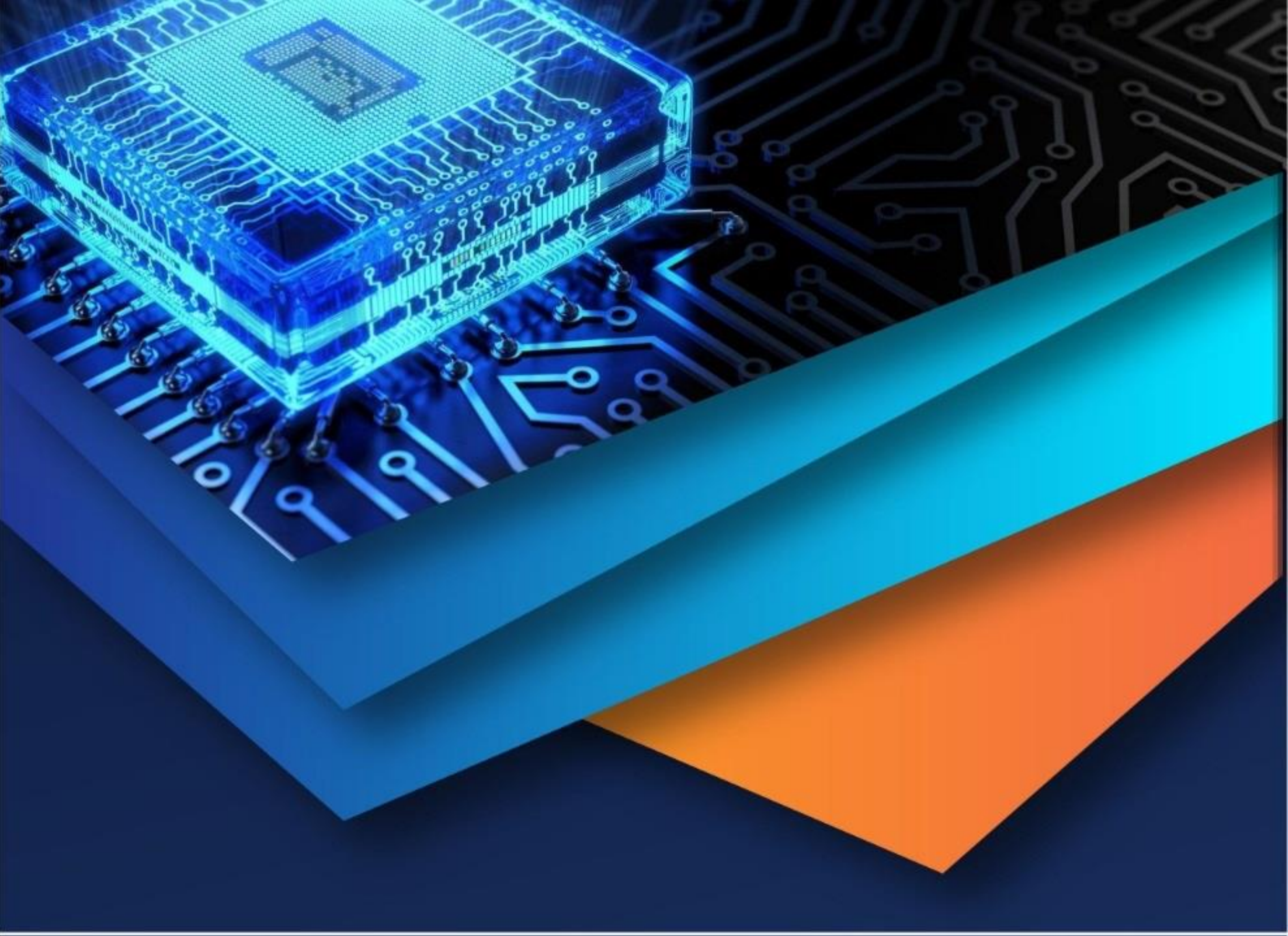
B. Recommendation

Based on the findings and limitationsof the study, the following recommendations are suggested for future research:

1) Collect larger and more diverse datasets to improve the generalizability of the models.
2) Consider external factors such as lifestyle and environmental factors that may also contribute to heart disease prediction.
3) Investigate the use of novel techniques such as deep learning and reinforcement learning for heart disease prediction.
4) Explorer the use of multi-task learning and transfer learning to improve the performance and efficiency of the models.
5) Consider the cost-effectiveness and scalability of the models to ensure their practicality and real-world implementation.
6) Finally, further studies should investigate the ethical and legal implications of using machine learning in healthcare, including issues related toprivacy, confidentiality, and bias.

REFERENCES

[1] Soni J, Ansari U, Sharma D & Soni S (2011). Predictivedata mining for medical diagnosis: an overview of heartdisease prediction. International Journal of Computer Applications, 17(8), 43-8
[2] Dangare C S & Apte S S (2012). Improved study of heartdisease prediction system using data mining classificationtechniques. International Journal of Computer Applications, 47(10), 44-8.
[3] Ordonez C (2006). Association rule discovery with the train and test approach for heart disease prediction. IEEETransactions on Information Technology in Biomedicine, 10(2), 334-43.
[4] Shinde R, Arjun S, Patil P & Waghmare J (2015). An intelligent heart disease prediction system using k-meansclustering and the Naïve Bayes algorithm. International Journal of Computer Science and Information Technologies, 6(1), 637-9.
[5] Bashir S, Qamar U & Javed M Y (2014, November). Anensemble-based decision support framework for intelligent heart disease diagnosis. In International Conference on Information Society (i-Society 2014) (pp.259-64). IEEE.
[6] Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W &Yun Y D (2014). A coronary heart disease prediction model: the Korean Heart Study. BMJ open, 4(5), e005025.
[7] Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Ärnlöv J & Ingelsson E (2013). Multilocus genetic risk scores for coronary heart disease prediction.Arteriosclerosis, thrombosis, and vascular biology, 33(9),2267-72.
[8] Jabbar M A, Deekshatulu B L & Chandra P (2013, March). Heart disease prediction using lazy associative classification. In 2013 International Mutli-Conference onAutomation, Computing, Communication, Control and Compressed Sensing (iMac4s) (pp. 40- 6). IEEE.
[9] Dangare Chaitrali S and Sulabha S Apte. "Improved study of heart disease prediction system using data mining classification techniques." International Journal of Computer Applications 47.10 (2012): 44-8.
[10] Soni Jyoti. "Predictive data mining for medical diagnosis:An overview of heart disease prediction." International Journal of Computer Applications 17.8 (2011): 43-8.
[11] Chen A H, Huang S Y, Hong P S, Cheng C H & Lin E J(2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-60).IEEE.
[12] Parthiban, Latha and R Subramanian. "Intelligent heart disease prediction system using CANFIS and genetic algorithm." International Journal of Biological, Biomedical and Medical Sciences 3.3 (2008).
[13] Wolgast G, Ehrenborg C, Israelsson A, Helander J, Johansson E & Manefjord H (2016). Wireless body areanetwork for heart attack detection [Education Corner]. IEEE Antennas and propagation magazine, 58(5), 84-92.
[14] Patel S & Chauhan Y (2014). Heart attack detection and medical attention using motion sensing device -Kinect. International Journal of Scientific and Research Publications, 4(1), 1-4.
[15] Zhang Y, Fogoros R, Thompson J, Kenknight B H,Pederson M J, Patangay A & Mazar S T (2011). U.S. Patent No. 8,014,863. Washington, DC: U.S. Patent andTrademark Office.
[16] Raihan M, Mondal S, More A, Sagor M O F, Sikder G, Majumder M A & Ghosh K (2016, December). Smartphone-based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design. In 2016 19th International Conference on Computer and Information Technology (ICCIT) (pp. 299-303). IEEE.
[17] Buechler K F & McPherson P H (1999). U.S. Patent No.5,947,124. Washington, DC: U.S. Patent and TrademarkOffice.
[18] Takci H (2018). Improvement of heart attack predictionby the feature selection methods. Turkish Journal of Electrical Engineering & Computer Sciences, 26(1), 1- 10.
[19] Worthen W J, Evans S M, Winter S C & Balding D (2002). U.S. Patent No. 6,432, 124. Washington, DC: U.S. Patent and Trademark Office.
[20] Acharya U R, Fujita H, Oh S L, Hagiwara Y, Tan J H & Adam M (2017). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Information Sciences, 415,190-8.
[21] Brown N, Young T, Gray D, Skene A M & Hampton J R (1997). Inpatient deaths from acute myocardial infarction, 1982-92: analysis of data in the Nottingham heart attack register. BMJ, 315(7101), 159-64.
[22] Piller L B, Davis B R, Cutler J A, Cushman W C, WrightJ T, Williamson J D & Haywood L J (2002). Validationof heart failure events in the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) participants assigned to doxazosin and chlorthalidone. Current controlled trials in cardiovascular medicine, 3(1), 10.
[23] Folsom A R, Prineas R J, Kaye S A & Soler J T (1989). Body fat distribution and self-reported prevalence of hypertension, heart attack, and other heart diseases in older women. International journal of epidemiology, 18(2), 361-7.
[24] Kiyasu J Y (1982). U.S. Patent No. 4,338,396. Washington, DC: U.S. Patent and Trademark Office



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY