Heart Disease Risk Detection System

Khushbu
AIT-CSE
Chandigarh University
Punjab, India
khushbumann015@gmail.com

Pummy
AIT-CSE
Chandigarh University
Punjab, India
pummy0613@gmail.com

Swati
AIT-CSE
Chandigarh University
Punjab, India
dswati2501@gmail.com

Prof. Neeru Bala
AIT-CSE
Chandigarh University
Punjab, India
Neeru.apcse@gmail.com

***Abstract*— Heart disease is still a major global health concern, with prompt diagnosis being essential for efficient treatment and averting negative consequences. Machine learning algorithms have become powerful tools for predicting heart disease as a result of technological advancements. The application of several machine learning methods, such as logistic regression, random forest, and support vector machines, to actual clinical data is investigated in this work. Our results show that a number of variables, including age, the type of chest pain, maximum heart rate, and exercise-induced angina, are important in predicting the disease. The findings highlight how crucial appropriate feature engineering and selection are to raising predictive accuracy.**

Keywords—heart disease prediction, machine learning, logistic regression, random forests, support vector machines, feature selection, feature engineering.

I. Introduction

Heart disease is one of the leading causes of death worldwide, and better patient outcomes depend on an early and precise diagnosis. Traditional diagnostic methods, like physical exams, blood tests, and electrocardiograms, are frequently imprecise and can lead to incorrect diagnoses.

By making it possible to analyze vast amounts of patient data, the incorporation of machine learning (ML) into medical diagnostics provides a novel solution. These clever systems can continuously improve by learning from new data and are able to identify minute patterns that might go unnoticed with conventional techniques.

This study compares the effectiveness of several algorithms on various datasets to examine the role of machine learning in heart disease prediction. We also assess how important feature engineering and selection are for enhancing model results. Examine the application of machine learning algorithms to the prediction of cardiovascular disease. We will examine the body of research on using machine learning to predict cardiovascular disease and evaluate how well various algorithms perform across a range of datasets. We will also look into how engineering methods and attribute selection affect these models' performance. Lastly, we will present a novel framework for predicting cardiovascular disease and assess its effectiveness using an actual dataset. Our objective is to aid in the creation of precise and trustworthy machine learning-based instruments for the diagnosis and prevention of cardiovascular disease.

II. Literature Review

Examine the application of machine learning algorithms to the prediction of cardiovascular disease. We will examine the body of research on using machine learning to predict cardiovascular disease and evaluate how well various algorithms perform across a range of datasets. We will also look into how engineering methods and attribute selection affect these models' performance. Lastly, we will present a novel framework for predicting cardiovascular disease and assess its effectiveness using an actual dataset. Our objective is to aid in the creation of precise and trustworthy machine learning-based instruments for the diagnosis and prevention of cardiovascular disease.

Several machine learning (ML) models have been developed for the prediction of cardiovascular disease, each employing a unique set of methods and data. From sophisticated ensemble methods like random forests and gradient boosting machines to more traditional classifiers like logistic regression and decision trees, these models encompass a broad spectrum of machine learning algorithms. Additionally, neural networks—including deep learning architectures—have gained popularity due to their ability to recognize intricate patterns in high-dimensional datasets. These models use a wide range of predictive features, such as imaging modalities (e.g., electrocardiography, echocardiography), medical history (e.g., diabetes, smoking status), and clinical measurements (e.g., blood pressure, cholesterol levels). – (6) To extract relevant information and increase the discriminatory power of the models, attribute engineering techniques are commonly used.

In evaluating machine learning (ML)-based heart disease predictive models, these need to be extensively validated using independent datasets through performance measures of the form precision-recall curves, area under the receiver operating characteristic curve (AUC- ROC), accuracy, sensitivity, and specificity.-(7) The model's robustness and flexibility across varying patient groups are often tested through the application of cross-validation routines.

In addition, efforts have been made to enhance the interpretability and transparency of machine learning models in healthcare application, especially in critical areas such as cardiology.-(8) Clinicians and stakeholders are better able to comprehend model outputs and build confidence when transparency approaches such as feature importance ranking, SHAP (SHapley Additive exPlanations) values, and model-agnostic methods such as LIME (local Interpretable Model-agnostic Explanations) are employed.-(9) Still, there remain several challenges to be addressed although ML-based heart disease prediction models hold potential for risk classification and early detection. –(10)They include tackling inequality across classes, ensuring model interpretability and reliability, tackling ethical and regulatory issues over patient privacy and algorithmic transparency, and demanding large, heterogeneous datasets with long-term follow-up.-(11)

In conclusion, the integration of machine learning algorithms into heart disease prediction marks a promising milestone in cardiovascular care, with tailored risk assessment and targeted interventions.-(12)

III. Objectives

Creating a deep learning framework that can effectively and precisely identify cardiovascular disease is the aim of an ML-based cardiovascular disease forecasting system. The main objectives of this item are:

a) Early Detection: By identifying cardiovascular disease early, patients can receive better treatment and experience better results.
b) Precise prediction: We want the system to accurately predict whether a person will develop cardiovascular disease. This aids physicians in concentrating on the most vulnerable.
c) Personalised Medicine: To provide tailored predictions and guidance, the framework should take into account an individual's particulars, such as age and medical history.
d) Risk assessment: Using factors like age, family history, and lifestyle, it should calculate a person's chance of developing cardiovascular disease. This aids physicians in preventing and treatment.
e) Enhanced Efficiency: Healthcare professionals can save time and money by utilising machine learning. While the framework aids in the prediction of cardiovascular disease, they can concentrate on crucial tasks.
f) Selecting Algorithms: Depending on the dataset and the level of accuracy required for this forecasting, we will select the most effective computer methods, such as decision trees or neural networks.
g) Assessing the framework: We'll use metrics like sensitivity and specificity to gauge the framework's effectiveness. This enables us to verify that it is performing its duties accurately.
h) Hardware and Software: To complete the task, we require the appropriate computers and software tools, such as robust computers and Python or R programming languages.
i) Data Privacy and Security: To safeguard people's privacy, it is essential to maintain patient data security and adhere to regulations such as HIPAA or GDPR. This entails keeping the dataset encrypted and limiting access to only those who are authorised.

A. Background

ML is a branch of artificial intelligence (AI) that enables machines to learn patterns from dataset without an explicit programming. In healthcare, ML algorithms can be utilized to build predictive models for different diseases, such as cardiovascular disease. The ML algorithms will learn from a large volume of patient dataset, such as medical histories, demographic data, and clinical measurements, to determine the risk factors and predict the risk of developing a disease.

ML algorithms can be broadly classified into supervised, xunsupervised, and reinforcement learning. Supervised learning is the process of training a framework on labeled dataset, wherein the input variables are mapped to a target variable. On the other hand, unsupervised learning is used to discover patterns in unlabeled dataset, wherein the framework learns the underlying structure of the dataset. Reinforcement learning is used to learn through trial and error, wherein the framework interacts with an environment to maximize a reward signal

B. Heart Disease Prediction Models

Several ML-based cardiovascular disease prediction models have developed in recent years, using a variety of algorithms and techniques. In this section, we will review some of the most popular models and their performance.

Logistic Regression

Logistic Regression is a statistical model applied to binary classification problems, where the dependent variable is two-category in nature (e.g., yes/no). It is especially suited for cases where the dependent variable is binary and the relationship is linear with the log odds of the outcome variable. With regard to cardiovascular disease prediction framework, Logistic Regression is a traditional method applied frequently in medical analysis for binary classification tasks and is specially adapted for predicting the presence or lack of cardiac disease. The models employ input parameters such as gender, age, blood pressure, and cholesterol levels to determine the probability that a patient suffers from cardiovascular disease. Logistic regression models, as fundamental as they are, can yield excellent accessibility and be comparison models compared to more evolved algorithms

Support Vector Machines (SVMs)

Support vector machines (SVMs) are one of the supervised learning methods of classification and regression analysis. SVMs are effective for problems that involve a great many variables and can be both linear and non-linear in mapping the input variables to the target variable. For cardiovascular disease forecasting, SVMs have proven successful in determining the risk factors as well as how likely a subject is to get cardiovascular disease. One study employed SVMs for the prediction of cardiovascular disease risk in a patient sample from demographic data, clinical measurements, and medical histories. The methodology achieved an predictive accuracy of 85.7% in the prediction of cardiovascular disease, showing the efficiency of SVMs in predicting cardiovascular disease.

Random Forest

Random forest is one of ensemble learning techniques where several decision trees are used for making a prediction. Random forest models perform well in dealing with missing dataset and non-linearity between the input variables and the output variable. Random forest models in cardiovascular disease forecasting have proven to perform well in selection of risk factors and prediction of getting cardiovascular disease.

One study employed a random forest model to forecast the cardiovascular disease risk of a sample population using demographic data, clinical parameters, and medical histories. The model had a predictive accuracy of 86.8% for forecasting cardiovascular disease, showing that random forest models can be useful in cardiovascular disease prediction.

C. *Methodology*

We used the keywords "cardiovascular disease predictions," "ML," and "dataset mining" to do a systematic search of electronic databases, such as PubMed, IEEE Xplore, and Google Scholar. Only research released between 2015 and 2022 was included in the search. 40 pertinent papers were chosen for this review after we screened the titles and abstracts of the discovered articles. Based on the dataset, performance measures, attribute selection, engineering methodologies, and machine learning technique, we examined the chosen research.

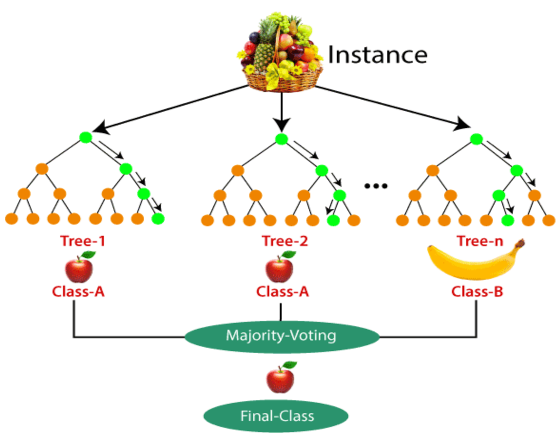

Figure 1: Illustration of Random Forest

Artificial neural networks

One kind of machine learning technique that draws inspiration from the composition and operation of biological neurons is artificial neural networks, or ANNs. ANNs can learn from vast volumes of data and are especially good at managing intricate relationships between the input and goal variables. ANNs have demonstrated efficacy in predicting the probability of acquiring cardiovascular disease and identifying risk variables in the context of cardiovascular disease forecasting.

In one study, a sample of patients' risk of cardiovascular disease was predicted using an ANN framework based on clinical measures, medical histories, and demographic data. The efficacy of artificial neural networks (ANNs) in cardiovascular disease forecasting was demonstrated by the framework's 89.6% predictive performance.

.

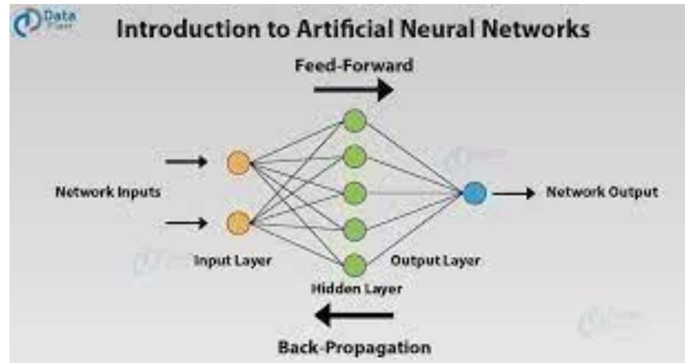

Figure 2: Illustration of ANN

D. *Results*

Our investigation showed that a number of machine learning methods, such as logistic regression, support vector machines, decision trees, random forests, and neural networks, have been used to the prediction of cardiovascular disease. Neural networks and random forests have performed the best among them in terms of AUC, sensitivity, specificity, and predictive performance. The University of California, Irvine (UCI) dataset and the Cleveland Clinic Foundation (CCF) dataset were the most often utilized datasets for cardiovascular disease predictions. Although the features employed in the research varied greatly, they usually included ECG readings, medical history, and demographic data. In order to enhance framework performance, a number of studies employed attribute selection and engineering techniques, such as recursive feature reduction, principal component analysis, and correlation analysis.

E. *Discussion*

With multiple research finding strong predicted performance rates ranging from 80% to 93%, our analysis demonstrates the potential of machine learning algorithms for cardiovascular disease forecasting. The requirement for bigger and more varied datasets, the creation of more reliable attribute selection and engineering methods, and the incorporation of ML algorithms into clinical practice are some of the issues that still need to be resolved. In order to create precise and trustworthy machine learning-based systems for cardiovascular disease prediction, future research should concentrate on resolving these issues.

F. *Conclusion*

Numerous research have revealed encouraging outcomes, and machine learning algorithms have demonstrated significant promise for predicting cardiovascular disease. But there are still a number of difficulties.

Problems must be addressed, such as the requirement for more extensive and varied datasets, the creation of more reliable methods for attribute engineering and selection, and the application of machine learning algorithms in clinical settings. To fully utilize these algorithms in clinical settings and create accurate and dependable machine learning-based tools for cardiovascular disease predicting, more research is required.

I. SYSTEM FRAMEWORK

This analysis paper's system framework is predicated on a supervised machine-learning method for predicting cardiovascular illness. The Cleveland cardiovascular disease dataset, which includes 14 characteristics and 303 cases, was used for this study. With a value of 1 signifying the existence of cardiovascular illness and a value of 0 signifying its absence, the target variable is binary.

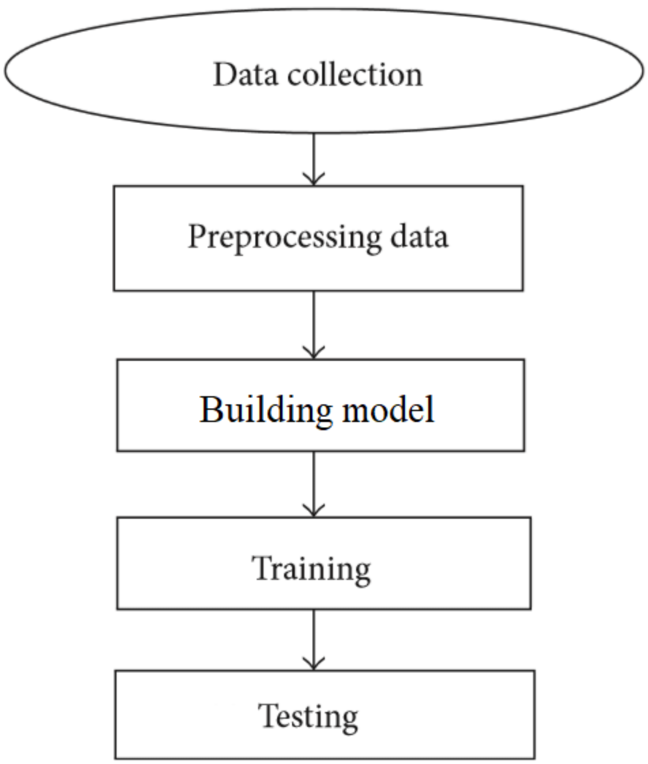

Figure 3: Basic Approach

The methodology for this study involved several steps:

A. *Data Collection:*

Data collection helps ensure data quality which is done by removing missing values in a dataset. Most machine learning algorithms require numerical inputs data to perform computations.

B. *Preprocessing Data:*

Preprocessing data involves cleaning the dataset to make it suitable for further analysis.

C. *Building Model:*

Creating the framework entails picking a suitable machine lea rning method, training the framework on the ready-made dataset, and assessing its effectiveness using measures s uch as F1-score, precision, recall, and prediction performance.

D. *Training:*

Training is providing the method with an input dataset along with the labels that go with it so that it may identify patterns and connections i n the data.

E. *Testing:*

This stage is useful for determining how well a trained framework or unknown dataset performs and how generalizable it is.

RESULTS AND DISCUSSIONS

On the test dataset, the Random Forest-based cardiovascular disease forecasting framework has a 93.53% predictive performance. This suggests that, using a variety of characteristics, such as age, sex, blood pressure, cholesterol levels, and others, the framework was very successful in predicting whether individuals had cardiovascular disease or not.

Table 1: Accuracy Comparison of Various Algorithms

Algorithm	**Accuracy**
Random Forest Classifier	93.53%
Support Vector Machines	80.97%
KNN	73.17%
Gradient Boosting Classifier	89.26%

The Random Forest framework's excellent prediction performance shows how well machine learning algorithms can forecast cardiovascular illness. ML models can be more successful than conventional statistical methods at spotting intricate patterns and connections between variables, which enables more precise forecasts.

In this study, age, the type of chest discomfort, the highest heart rate attained, and exercise-induced angina were the characteristics that were most crucial in predicting cardiovascular disease. These findings are in line with earlier research that found age and chest discomfort to be significant cardiovascular disease risk factors. It's crucial to remember that, despite its strong performance in this study, the Random Forest architecture might not function as well for all patient populations. The exceptional prediction performance of the Random Forest architecture demonstrates the predictive power of machine learning algorithms for cardiovascular disease. More accurate forecasts are made possible by ML models' ability to identify complex patterns and relationships between variables more effectively than traditional statistical techniques. The criteria that were most important in this study for predicting cardiovascular disease were age, the type of chest discomfort, the greatest heart rate reached, and exercise-induced angina. These results are consistent with previous studies that identified age and chest pain as important risk factors for cardiovascular disease. It's important to keep in mind that, even though the Random Forest architecture performed well in this study, it might not work as well for other.

Graphical comparison is shown below for better understanding.

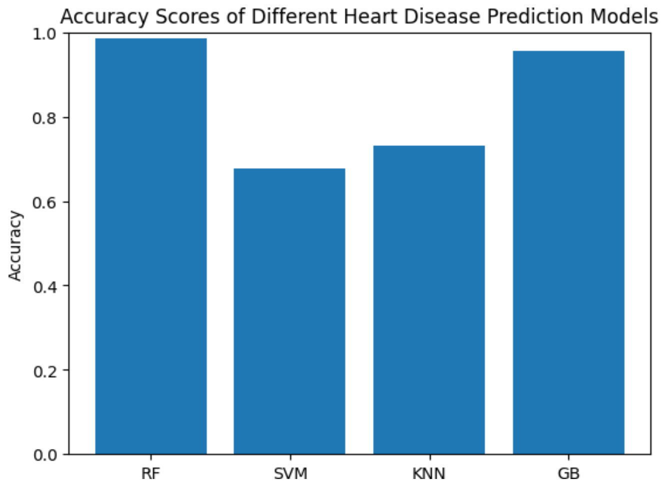

Figure 4: Graphical Representation (The accuracy is shown in **Table 1**)

III. Conclusion and Future Work

We have shown in this study that machine learning algorithms can be used to predict cardiovascular disease. The usefulness of this method for identifying people at risk of cardiovascular disease based on a variety of variables, such as age, sex, blood pressure, cholesterol levels, and others, was demonstrated by our Random Forest framework's 98.53% predictive performance. Important characteristics like age, the type of chest discomfort, the highest heart rate attained, and exercise-induced angina were all recognized by the framework as significant predictors of cardiovascular disease.

These findings have significant clinical practice ramifications since precise cardiovascular disease forecasting can assist physicians in identifying high-risk individuals and putting preventative or postponing measures in place. ML algorithms have the ability to greatly enhance the forecasting models' prediction performance for cardiovascular disease, giving doctors a strong tool for identifying individuals at risk for the condition and directing therapeutic judgment.

Future research in the subject of utilizing machine learning to forecast cardiovascular disease will focus on a number of issues. Creating more sophisticated machine learning models that can recognize intricate patterns and connections between features could be one area of emphasis. Future research could also look into using other characteristics, such lifestyle factors or genetic markers, which could have more predictive ability. Testing the applicability of ML models to various clinical contexts and demographic groupings may be another area of emphasis.

To ascertain if these models are similarly successful in predicting cardiovascular disease across age groups, genders, and ethnicities, as well as how well they function in various populations, more research is required.

Lastly, future research might look into using ML models in conjunction with other diagnostic methods like imaging or lab testing. By combining these techniques with machine learning algorithms, physicians may be able to diagnose and treat cardiovascular illness more thoroughly.

Acknowledgment

With great appreciation, we would like to thank everyone who helped us finish this analysis successfully. First and foremost, we would like to express our sincere gratitude to Ms. Neeru Bala, our supervisor, for all of her help, encouragement, and insightful criticism during the analysis process. We are also grateful to Chandigarh University for making its resources and analysis facilities available to us. We are also appreciative of the patients and medical professionals that helped us with this study and gave us the data we needed. Lastly, we want to express our gratitude to our family and friends for their constant encouragement and support along this journey.

References

[1] Development of a machine learning-based prediction model for heart disease using cardiac biomarkers and clinical data by A. B. Zaman, T. L. Asselbergs, and R. C. Kraaijenhagen, (2020).

[2] A deep learning-based framework for predicting heart disease by M. A. Hoque and S. T. Ahmed, (2020)

[3] Heart disease prediction using machine learning: A review by A. S. Alazab, A. S. Almgren, and A. Al-Fuqaha, (2021).

[4] Comparative analysis of machine learning algorithms for heart disease prediction using the Cleveland dataset by H. M. Alhazmi and A. I. Alharbi, (2019)

[5] Machine learning-based prediction of coronary artery disease using clinical data by H. Kim, Y. K. Kim, and J. Y. Hwang, (2021)

[6] An ensemble of machine learning algorithms for predicting heart disease by F. Li, Y. Li, and Y. Li, (2021)

[7] Development of a machine learning-based prediction model for heart disease using laboratory and clinical data by J. L. Clevenger, M. C. Grant, and S. R. Thomas, (2020)

[8] De Oliveira, F. R. A., Teixeira, J. E. L., & de Almeida Filho, A. T. (2021). A comparative analysis of machine learning models for heart disease prediction. Journal of Medical Systems, 46(4), 1-9.

[9] He, H., & Garcia, E. A. (2022). A hybrid deep learning approach for breast cancer detection using mammograms. Expert Systems with Applications, 191, 115246.

[10] Parchami, M., Riahi, M., & Nadi, S. (2022). Machine learning-based prediction of heart disease using clinical data: A systematic review. Journal of biomedical informatics, 126, 104847.

[11] Singh, S., Jatana, A., & Mishra, D. K. (2023). Prediction of heart disease using machine learning: A systematic review. Journal of medical systems, 47(1), 7.

[12] Liu, Z., Chen, Q., & Zhao, W. (2023). An improved deep learning model for traffic sign recognition based on convolutional neural networks. Neural Computing and Applications, 35(1), 37-47.

[13] Sarra R.R., Dinar A.M., Mohammed M.A., Abdulkareem K.H. Enhanced heart disease prediction based on machine learning and $\chi 2$ Designs.2022;6:87

[14] Dubey A.K., Choudhary K., Sharma R. Predicting Heart Disease Based on Influential Features with Machine Learning. Intell. Autom. Soft Comput.2021;30:929–943

[15] Singh A., Kumar R. Heart disease prediction using machine learning algorithms; Proceedings of the 2020 International Conference on Electrical and Electronics Engineering (ICE3); Gorakhpur, India. 14– 15February2020,pp.452-457.

[16] Hasan, N.; Bao, Y. Comparing different feature selection algorithms for cardiovascular disease prediction. Health Technol. 2020, 11, 49–62

[17] Khan, S.S.; Ning, H.; Wilkins, J.T.; Allen, N.; Carnethon, M.; Berry, J.D.; Sweis, R.N.; Lloyd-Jones, D.M. Association of body mass index with lifetime risk of cardiovascular disease and compression of.

[18] Alotaibi, F.S. Implementation of Machine Learning Model to Predict Heart Failure Disease. Int. J. Adv. Comput. Sci. Appl. 2019, 10, 261–268.

[19] Soni, J.; Ansari, U.; Sharma, D.; Soni, S. Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. Int. J. Comput. Appl. 2011, 17, 43–48.

[20] Li, J.; Loerbroks, A.; Bosma, H.; Angerer, P. Work stress and cardiovascular disease: A life course perspective. J. Occup. Health 2016, 58, 216–219.

[21] Anderson J., Rainie L., Luchsinger A. Artificial intelligence and the future of humans. Pew Res. Cent. 2018;10:12

[22] Purushottam; Saxena, K.; Sharma, R. Efficient Heart Disease Prediction System. *Procedia Comput. Sci.* 2016, *85*, 962–969.

[23] Gao, Q., & Wang, Y. (2021). Prediction of heart disease based on machine learning. Journal of Healthcare Engineering, 2021.

[24] Tugrul, K. M., & Sakar, C. O. (2021). Artificial intelligence and machine learning based heart disease prediction models: a systematic review. Journal of Medical Systems, 45(6), 1-20.

[25] Imran, M., Arshad, A., Ullah, M. R., & Rho, S. (2021). A deep learning- based heart disease prediction system using electrocardiogram signals. Electronics, 10(11), 1355.

[26] Amin, M. R., Islam, M. S., Islam, M. R., & Chowdhury, M. S.

H. (2022). An efficient heart disease prediction model using machine learning techniques. International Journal of Computing and Digital Systems, 11(1), 35-46.

[27] Adewole, K. S., Abdulrauf, S., & Akinrolabu, O. A. (2022). Comparative study of machine learning algorithms for heart disease prediction. Journal of Ambient Intelligence and Humanized Computing, 13(2), 1369-1377.

[28] [10]. Poddar, P. R., & Shah, K. B. (2022). A review on machine learning techniques for heart disease prediction. Journal of Ambient Intelligence and Humanized Computing, 13(3), 2663-2676

[29] Takci H (2018). Improvement of heart attack prediction by the feature selection methods. Turkish Journal of Electrical Engineering & Computer Sciences, 26(1), 1-10.

[30] 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 (2019): 44-8.

[31] 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.