Global Journal of Medical Research: F
Diseases
Volume 23 Issue 4 Version 1.0 Year 2023
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-4618 & Print ISSN: 0975-5888

Heart Disease Detection using Machine Learning

By Rashmi S K

Alvas Institute of Engineering and Technology

Abstract- Every person's health is impacted by a confirmation of circumstances, and certain diseases are fatal and have serious side effects. One such serious condition that affects people of all ages is heart disease. this paper suggests a preprocessing strategy to improve the categorization precision of ECG data We are suggesting an ECG sensor-based healthcare monitoring system. Since the values are so crucial, ECG sensors are necessary for patient remote monitoring. Elements from the ECG wave are extracted using a verification of extraction techniques to be able to accurately predict cardiac disease The patient's ECG is continuously monitored using a mobile app. The different algorithms used in data mining eliminate the extra time and work required to perform multiple tests to identify diseases. Data collection employs ECG sensors. The acquired data is stored on a storage device before data Mining techniques are used to it. These equations indicate the patient's potential for cardiac disease. Doctors may utilise the outcomes for diagnostic purposes. The technology will predict cardiac illness by utilizing machine learning methods.

Keywords: heart disease prediction, UCI dataset, machine learning.

GJMR-F Classification: LCC Code: RC685-688

Strictly as per the compliance and regulations of:

Heart Disease Detection using Machine Learning

Rashmi S K

Abstract- Every person's health is impacted by a confirmation of circumstances, and certain diseases are fatal and have serious side effects. One such serious condition that affects people of all ages is heart disease. this paper suggests a preprocessing strategy to improve the categorization precision of ECG data We are suggesting an ECG sensor-based healthcare monitoring system. Since the values are so crucial, ECG sensors are necessary for patient remote monitoring. Elements from the ECG wave are extracted using a verification of extraction techniques to be able to accurately predict cardiac disease The patient's ECG is continuously monitored using a mobile app. The different algorithms used in data mining eliminate the extra time and work required to perform multiple tests to identify diseases. Data collection employs ECG sensors. The acquired data is stored on a storage device before data Mining techniques are used to it. These equations indicate the patient's potential for cardiac disease. Doctors may utilise the outcomes for diagnostic purposes. The technology will predict cardiac illness by utilizing machine learning methods.

Keywords: heart disease prediction, UCI dataset, machine learning.

I. Introduction

The healthcare sector has been utilizing new technologies to deliver better and more sophisticated healthcare facilities. including commerce, healthcare, social networking, and education. To do so, provide better and more advanced healthcare facilities, the healthcare The sector has been embracing new technology. Remote and real-time patient monitoring is now possible Thanks to the IoT (Internet of Things), continuous health is now a possibility. permits clinicians to provide immediate advise or therapy through monitoring. Early identification of Heart disease is critical to saving lives and maintaining a healthy lifestyle among people. early detection of heart disease is crucial because it affects a bigger population than previously thought. The development of various IoT capabilities and tools to follow patients' health problems on a regular basis has greatly improved healthcare monitoring. Additionally, Patients and clinicians can communicate more readily, which enhances the quality of care and reduces hospital stays and medical expenses. Setting The main objective of utilising IoT in healthcare systems is to provide a fully automated environment for patient monitoring and real-time help and treatment. It is increasingly vital to have portable technology that patients may use at home to measure their ECG patterns and quickly identify their ailment. Therefore, a thorough assessment of the accessible technologies for tracking heart-related disorders is done in this work. The research shows that the raw data was gathered using noise and useless information. These are inaccurate and unrelated information that is useless for formulating a diagnosis. since the noise and extreme data volatility Reduced categorization sensitivity, accuracy, and precision. Therefore, in this study, a unique pre-processing technique is used to get rid of noise and unimportant data from ECG signals. Relevant features are discovered using the correlation technique to increase data efficiency. ECG signals are categorised according to waveforms using classifier method for machine learning algorithms like KNN, naive Bayes, and Decision trees. The ECG waveform variation can be diagnosed and the kind of abnormality and disorder can be determined using the classifier with the best metrics. Around the world, machine learning is applied in many different fields. The healthcare industry is not an exception. Machine learning can be very helpful in identifying whether disorders like heart problems, mobility issues, and other challenges if it will even exist. Such information, if anticipated long in ahead, can clinicians with benificial insights, allowing them to then tailor their diagnosis and course of treatment for each patient.

Problem Statement: One of the most exciting and difficult topics is using machine learning to forecast heart illness. A quick and effective detection system must be created owing to the lack of professionals and the high rate of cases that are incorrectly diagnosed. This work's primary aims to employ a classifier model to extract the important patterns or characteristics from the medical data. You can see the quality that are more significant for diagnosing heart disease. This will make it easier for medical professionals to fully comprehend the underlying causes of disease.

II. Objectives

- The goal of our research is to identify the heart disease dignosis more precisely using fewer variables

Author: Alvas Institute of Engineering and Technology.
e-mail: skrashmikanth@gmail.com

- The stages of the disease are determined using fourteen heart
- Disease prediction criteria
- We are sending the doctor's report via Telegram Web.

III. Literature Survey

A method that predominantly prioritized algorithm accuracy was put out by Archana Singh and Rakesh Kumar et al. [1]. It was one of the parameters taken into consideration by the authors when analyzing the algorithmic performance. The dataset utilized for testing and training purposes determines how accurate machine learning algorithms will be. KNN was exhibited as the best algorithm when they compared the algorithms using a dataset with parameters like age, type of chest discomfort, cholesterol, and resting time. They employ the following methodologies: K-nearest Neighbour, support vector machine and leniar regression and Decision Tree One drawback is that it requires numerous data sets because it uses so many attributes.

A method with a strong emphasis on feature selection and the prediction was put out by Rahul Kataria and P. Srinivas et al. [2]. For every automated system, these two are necessary. We can forecast heart disease more accurately by making wise feature choices. A few techniques, such as the random search algorithm and the hybrid grid search algorithm, were summarised by the authors as being helpful for choosing the features. They chose some key traits that are commonly utilized to predict cardiac disease, such as gender, age, resting blood pressure, ECG readings, heart rate, etc. They employ the following methodologies: Support vector machines (SVM), Decision trees (DT), Artificial neural networks (ANN), and Random forests (RF).

An effective machine learning-based diagnosis method for the diagnosis a diagnosis of heart disease has been made according to Jian Ping Li et al. [3]. Creating a system for heart disease diagnosis identification is the studys originality. A small improvement in prediction accuracy made by Jian Ping Li et al. has a significant impact on the identification of serious diseases. Four well-known feature selection technique are used in this case, and one additional feature selection approach is provided. Both the LOSO CV technique and performance evaluation measures are employed. The information for Cleveland heart disease is utilized for testing. The dataset has been pre-processed using methods including Standard Scalar (SS), Min-Max Scalar, and deleting attribute missing values. For preparation and testing.

B. Keerthi Samhitha, et al. [4] contributed to the UCI machine learning archive's Cleveland heart dataset. Additionally, they put forth a cutting-edge strategy that increases the reliability of cardiovascular disease prognosis and emphasises literature survey using machine learning approaches, Rahul Kataria and P. Srinivas identified the major highlights in their publication of machine learning techniques. Different combinations of highlights and a few well known grouping strategies accustomed to display the forecast model. They were able to produce an improved display level with a precision level of 88:7% using the halfbreed irregular woods and a straight model.

Machine learning's "Application for the Diagnosis of Coronary Heart Disease" [5] was put up by Sameer S. Yadav et al. after extensive research and analysis of An technique with greater precision was developed as a result of numerous machine learning algorithms. The problem is determining which research will help with cardiovascular identification and which will be insufficient. The Cleveland dataset, which has high statistical precision, is utilised for local development programmes.

They have combined cross-validation and the test train division approach to ascertain the ideal set of the parameters for analysis. For example, Naive Bayes, K-Nearest Neighbors, Logistic Regression, and Neural Network algorithms are employed.

B. Keerthi Samhitha, et al. [6] proposed a novel approach that targets finding critical highlights by applying machine learning procedures, improving the precision in the forecast of cardiovascular illness. They used the Cleveland Heart dataset based on the data from the UCI machine learning archive for preparation and testing purposes. The forecast model is displayed with different highlights mixtures and a few well-known grouping techniques. Utilizing expectation model for heart illness with the half-breed irregular woods and a straight model, they were ability to accomplish an improved exhibition level with a precision level of 88:7%.

Geetha S, Santhana Krishnan J[7]. The author attempted to concentrate on male patients in this paper and took into account a amount of variables that may contribute to heart disease, such as risk factors and risk categories. One of the most often used data mining technique was used by the author, WEKA, and KNN algorithm was employed for the prediction portion. While the author utilised two algorithms in order to get analysis, other algorithms and prediction approaches can be employed, and while the elements examined during the prediction are rather few, we can uncover other aspects that are affecting cardiac condition and work on those.

Mohan et al. [8]. The accuracy of cardiovascular disease prediction can be enhanced by utilising hybrid machine learning approaches, with the objective of identifying crucial by utilising machine learning. The expectation model is created using a few well-known arranging procedures. which is composed of different highlighted mixtures. Several data mining techniques and methods, including KNearest Neighbours, LR, SVM,

neural networks, and vote, have been successfully utilized to forecast heart disease.

Chandra P, Shekar K, et al. [9], to lower the risk of heart disease, which is a major health concern for millions of people worldwide. This study tried to examine a dataset on heart illness using several sorts of data mining methodologies in order to produce a 100% right model based on data mining algorithms. The findings might be critical for gaining insights from the data, forecasting the prognosis of new patients with heart disease, and establishing effective techniques for improving the accuracy, efficacy, and calibre of heart disease treatment processes. This research developed a basic SVM model for predicting heart disease. The appropriate healthcare data is created through the usage of the methodical framework provided in this study for cardiac disorders.

Monther Tarawneh, Osama Embarak [10], The Nave Bayes method is used in this paper's DSS (Decision Support System). In the proposed system, users' information on their sex, smoking habits, blood sugar, type of chest discomfort, etc. are first collected. For supervised learning, they next utilised a Nave Bayes classifier. As input, it receives an independent variable. The Advanced Encryption Standard algorithm is utilised to encrypt the patient's data because it is less time-consuming and

When compared to other methods, it provides superior accuracy. Despite reducing the characteristics, the dominant technique exceeds the Naive Bayes in terms of accuracy, yielding an accuracy of 89.77%.High accuracy is attained here by using only a few criteria for prediction.

Ravinder Kumar and Kanika Pahwa [11], Divide the process into three steps for this article. Raw data is subjected to preprocessing in the first stage. Age, sex, Cp (associated to chest discomfort), cholesterol, etc. are examples of raw data. In this stage, data transformation is completed. This problem is transformed into a binary class problem using data transformation. The output of this second stage, which applies feature selection to the pre-processed data as input, is a relevant feature. Unimportant data are eliminated during feature selection. The Gain Ratio is used to compute the score of an attribute. At the very end, the classification algorithm is used. They employed the Naive Bayes approach and the Random Forest algorithm to anticipate cardiac attacks. They made advantage of a database that is open to the public. The ROC curve and area under the curve for the confusion matrix are assessed. Two algorithms' accuracy is compared.

Shubhankar Mayank, Ritika Chadha, and others [28], KNN, SVM, and ANN are utilised in this article to predict cardiac disorders. Compare the outcomes of these three algorithms and the ensemble classifier as well. They classified using binary and multiclass systems. Cross-validation and percentage split are the two methods of evolution employed. Multiclass classification is inferior to binary classification. The percentage split result is superior to cross-validation. Data are first collected in this method. from a dataset and then, using feature selection, data are chosen and used as model input.

IV. Proposed Work

A system architecture is a conceptual framework for organizing a system. the model outlines the structure, behavior, and other features of a system. An architecture description is a formal description and representation of a system that is designed to make it easier to analyze its structures and behaviors. System architecture may be made up of system elements and developed subsystems that work together to complete the entire system. The architecture description languages (ADLs) collectively refer to efforts to formalize languages that describe system architecture. Various organizations define systems architecture in a variety of ways, including the fundamental organization of a system as seen in its parts, how they interact with one another and the environment, and the ideas that have guided its development. A system graphic that shows how users interact with various components, how functionality is assigned to hardware and software components, and how software architecture is mapped onto hardware architecture. An allocation of physical elements that provides a design solution for a consumer product or life cycle process that tries to meet the objectives of functional architectures and the requirements baseline. The most important, ubiquitous top-level strategic innovations, choices, and related arguments about the overall structure (i.e., the core elements and their relationship), along with related traits and behavior, make up an architecture. A description of the design and elements of a computer system. If it is documented, it might include information like a comprehensive summary of the current networking, software, and hardware capabilities. a comprehensive component-level design of the system to guide development or a formal description of the system.

The combination of a product's life cycle processes and design architectures. The organization of components, how they relate to one another, and the rules and ideas that have guided their development over time. A system architecture can be viewed as a collection of representations of a current (or upcoming) system. These representations start off by describing a broad, high-level functional organization and then gradually get more specific and concrete.

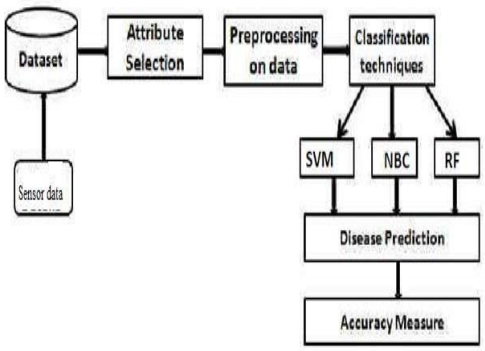


Fig. 1: Proposed system architecture

Algorithms Used

KNN: A well-known machine learning approach. K-Nearest Neighbours (KNN) is a classification and regression algorithm.

When it comes to predicting heart disease, KNN may be used to separate people into two groups based on whether they have heart disease or not. The approach works by determining the K data points in the training set that are closest to a new data point—a patient—and categorizing the patient based on the class with the most members among the K closest neighbors.

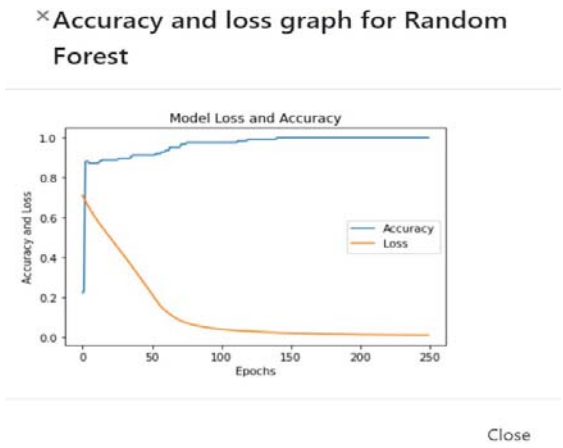
Support Vector Machine: To handle classification and regression problems, powerful machine learning techniques called Support Vector Machines (SVM) are used. In the context of heart disease prediction, SVM may be used to categorise individuals into two groups based on whether they have or do not have heart disease. The method searches for the best hyperplane to partition the two classes in a high-dimensional space.

Naïve Bayes: A well-known machine learning algorithm for classification problems, Naive Bayes is utilised. It is based on the Bayes theorem, which states that when a hypothesis (such as a patient having heart disease) is known to exist, the likelihood of the evidence (such as the patient's symptoms) is inversely associated. When it comes to predicting heart disease, Naive Bayes may be used to classify people into two groups according on whether they have heart disease or not.

Logistic Regression: Is a popular machine learning strategy for dealing with binary classification problems. It is used to evaluate the chance of an event occurring based on specified independent criteria. In the context of heart disease prediction, logistic regression may be used to classify patients as having or not having heart disease.

Random Forest: For classification and regression problems, a powerful machine learning approach is used. It employs an ensemble learning approach that blends various decision trees to provide more exact predictions. When it comes to predicting heart disease, Random Forest may be used to separate people into two groups based on whether they have heart disease or not. For classification and regression problems, a powerful machine learning approach is used. It employs an ensemble learning approach that blends various decision trees to provide more exact predictions. When it comes to predicting heart disease, Random Forest may be used to separate people into two groups based on whether they have heart disease or not.

Accuracy Results



Accuracy Graph

The accuracy and loss graphs for Random Forest models might not look like the typical graphs for nural networks. Instead, they might show a plot of the metric used to evaluate the models performance over time, such as the MSE or R-squared value.

V. Conclusion

Heart attacks are a major public health issue in modern culture. Initially, we attempted to incorporate several techniques in this project using machine learning algorithms and the Internet of Things. Finding the missed predictions in the confused matrix provides us with the error rate, which we can then remove to determine the classifier's accuracy. We built a dataset from the patient's family history and compared it to the patient's record to see if there was a chance they may have a heart attack. We tried to apply it by using the machine learning algorithm.

References Références Referencias

1. Rakesh Kumar and Archana Singh, "Heart Disease Prediction Using Machine Learning Algorithms" 2020 International Conference on Electrical and Electronics Engineering (ICE3), 2020, pp. 452-457, doi: 10.1109/ICE348803.2020.9122958.
2. Jian Ping Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, Heart Disease Identification Method Using

3. Machine Learning Classification in EHealthcare, IEEE Access, vol. 8, no. 8, 2020, pp.107562–107582, doi 10.1109/ACCESS.2020.3001149.
4. B. Keerthi Samhitha, M. R. Sarika Priya, C. Sanjana, S. C. Mana. Rahul Katarya and P. Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp.302 305, doi:10.1109/ICESC48915.2020.9155586.
5. Sameer S. Yadav, Shivaji M. Jadhav, S. Nagrale, doi:10.1109/ICIMIA48430.2020.9074954.
6. "Improving the Accuracy in Prediction of Heart Disease Using Machine Learning Algorithms," 2020 International Conference on Communication and Signal Processing (ICCSP), 2020, pp. 1326–1330, doi: 10.1109/ICCSP48568.2020.9182303.
7. "Prediction of Heart Disease Using Machine Learning Algorithms.," in 2019's first International Conference on Innovations in Information and Communication Technology (ICIICT), pp. 1–5, doi: 10.1109/ICIICT1.2019.8741465.
8. Chandrasegar and Senthilkumar Mohan Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, Gautam Srivastava Thirumalai, Digital Object Identifier10.1109/ ACCESS.2019.2923707, IEEE Access, Volume 7, 2019, 2019. S.P. Bingulac, "On the Compatibility of Adaptive Controllers," in Proc. Fourth Ann. Allerton Conf. Circuits and Systems Theory, pp. 8-16, 1994. (Transactions from a conference)
9. Chandra P., Venugopala Rao K., and Chandra Shekar K. (2019) An Ensemble Classifier for the Prediction of Heart Disease Using Genetic Algorithm and Decision Tree. As published in Saini H., Sayal R., Govardhan A., and Buyya R. Computer science advances and Engineering. Volume 74 of Lecture Notes on Networks and Systems. Singapore's Springer. https://doi.org/10.1007/978-981-13-7082-3_2
10. Ossama Embarak and Monther Tarawneh (2019) Using a Hybrid Approach, Data Mining Techniques are used to predict heart disease. Advances in Internet, Data, and Web Technologies, edited by Barolli L., Xhafa F., Khan Z., and Odhabi H. EIDWT 2019. 29th volume of Lecture Notes on Data Engineering and Communications Technologies. https://doi.org/10.1007/978-3-030-12839-5_41 Springer, Cham
11. KanikaPahwa and Ravinder Kumar, "Prediction of heart disease using hybrid technique for selecting features," 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), 2017, pp. 500-504, doi: 10.1109/UPCON.2017.8251100.
12. Shubhankar Mayank, T. Pradhan, A. Vardhan, and Ritika Chadha (2016)., (), -. doi: 10.1007/978-81-322-2553-9_38. Application of Data Mining Techniques on Heart Disease Prediction: A Survey.