HEART DISEASE DETECTION USING MACHINE LEARNING

S Spandana
Assistant professor
CSE dept
Koneru Lakshmaiah Education Foundation, Bowrampet, Hyderabad
Telangana
500043
spandana@klh.edu.in

N. Vijayasri
Asst Professor Department Of Computer Science & Information Technology MLR Institute of Technology Hyderabad, India
vijayasri@mlrinstitutions.ac.in

Gaddam Ravindra Babu
CSE dept
Koneru Lakshmaiah Education Foundation, Bowrampet, Hyderabad
Telangana
500043
ravindragaddam@gmail.com

T. Gandhi
Asst Professor Department Of Computer Science & Information Technology MLR Institute of Technology Hyderabad, India
tellagandhi@gmail.com

B. Madhavi
Asst Professor Department Of Computer Science & Information Technology MLR Institute of Technology Hyderabad, India
madhavi.guntha4@gmail.com

Abstract- Heart diseases has now became the most effecting and leading cause to death of most people. Early detection of this heart diseases can be more effective way for treatments and also, we can save most of the people’s life only the early detection. However, monitoring a patient for 24 hours is not possible for the doctor’s and also through the existing system, Also, this will also not give the best results. Now here we are creating a heart disease detection model which helps us to detect and analyze the disease in the early stages only. This is done through the machine learning algorithms. We have a effective machine learning algorithm and also however there are so many algorithms which gives us the best results. The performance of this algorithm is done by using 10-fold cross validation method. When we used the svm algorithm the accuracy level was 97.53%. The sensitivity and specificity values are 97.50% and 94.94%. Additionally, a real time parameters like body temperature, heart rate, body pressure, and humidity were calculated round-the-clock which means the patient will be monitored by doctor with or without his presence. This will make the disease to be cured more fastly and in more effective way. Our server will receive the data for every 10 seconds and update the data to the doctor which is the best advantage of the system. So the doctor can see the patient and check his vitals frequently and give him medicines if required.

KEYWORDS: – heart disease detection and prediction using data mining, machine learning, and patient monitoring systems

I. INTRODUCTION

Cardiovascular illnesses have been the leading cause of mortality globally in the last decade. According to the World Health Organisation, it is estimated that more than 17.9 million fatalities occur annually due to cardiovascular disorders. Of these deaths, 80% are caused by coronary artery disease and cerebral stroke [1]. Heart disease can be attributed to various causes, including personal and professional behaviours as well as hereditary predisposition. Cardiac disease is generally determined by a combination of risk factors, including smoking, excessive alcohol and caffeine use, stress, physical inactivity, obesity, hypertension, high blood cholesterol, and pre-existing cardiac diseases. An expedient, precise, and timely medical diagnosis of heart disease is crucial in implementing preventative actions to mitigate the problems associated with such conditions. The primary obstacle encountered in the field of medical sciences presently is the delivery of high-quality service and the attainment of efficient and precise prediction. The subsequent issue can be resolved with the implementation of automation utilising Data Mining and Machine Learning techniques. Data mining is the systematic procedure employed to extract valuable information from a vast collection of unprocessed data. It involves the examination of patterns in extensive sets of data through the utilisation of diverse tools. Additionally, it entails efficient gathering and storage of data, combined with computerised data processing. Machine Learning (ML) is a subset of data mining that focuses on efficiently handling large-scale, well-formatted datasets. Machine learning has applications in the medical industry for the purposes of diagnosing, detecting, and predicting a wide range of disorders. The accuracy of different Machine Learning methods, including Logistic Regression, Naïve Bayes, Support Vector Machine, KNearest Neighbour, Decision Tree, Random Forest, and the ensemble technique of XGBoost, is compared to determine the most precise model. The cardiac disease dataset from the UCI repository is utilised in this context. This research involves a thorough examination and comparison of the current classification techniques. The study also discusses the potential for future research and several opportunities for advancement.

The most important part of the human body is heart. The circulatory system will help the blood flow to circulate through the body[4]. As the heart only sends the blood and oxygen and other important elements to the body’s others organs so the circulatory system is very crucial for survival [5]. The heart is the most important part of the human system. Serious health issues including the heart attack issues will be caused by the heart only [6]. So, the heart disease detection came into the picture where a doctor can monitor the patient’s vitals for every one hour and can give the appropriate medicines [7]. In our day-to-day life heart disease are increasing very rapidly so in order to prevent people from this we came with an idea that is heart disease
detection using machine learning algorithms [8]. This is very important and useful application that we have made. We used different approaches and choose the best approach. Several approaches and discoveries mentioned about the benefits of machine learning based heart disease prediction [9]. The main aim of the project is to detect the heart disease from the early stage itself without the help of doctors. Here we will use 13 parameters as input. Our result will be in the form of a bar graph [1-3].

II. LITERATURE SURVEY

As a result of this, the doctors will be able to provide the medications from the very beginning of the illness detection process, which will be of greater assistance to the patient in managing his or her health [16]. The primary objective of the project to detect heart disease is to detect the disease in its early stages. The approach that has been proposed is also useful in the healthcare industry; however, the experts do not possess the necessary knowledge and skills [11].

Data mining techniques such as Naive Bayes and decision trees were utilised by Gomathi et al. in order to make predictions for a variety of sickness categories. Their primary concentration was on the forecasting of diseases related to the cardiovascular system, diabetes, and breast cancer. The findings were achieved by using the confusion matrix as your source. • Miranda et al. presented a Naive Bayes classifier method for the purpose of predicting cardiovascular diseases. For the purpose of establishing whether or not heart disease is present, the authors have taken into consideration a number of major risk factors. Researchers led by Avinash Goleande and his colleagues carried out a study that investigated a variety of machine learning techniques that have the potential to be applied to the categorization of heart illness. For the aim of classification, a study was carried out to investigate the Decision Tree, KNN, and K-Means algorithms, and the accuracy of each of these algorithms was compared. Based on the outcomes of the research, it appears that the Decision Tree algorithm succeeded in achieving the highest level of accuracy. In addition, it was discovered that the effectiveness of the algorithm can be improved by combining it with other strategies and by fine-tuning its settings. • Fahd Saleh Alotaibi has created a machine learning model that evaluates and draws comparisons between five different algorithms. When compared to the Matlab and Weka tools, the Rapid Miner tool produced results that were significantly more accurate. An investigation on the accuracy of classification algorithms such as Decision Tree, Logistic Regression, Random Forest, Naive Bayes, and Support Vector Machines was the purpose of this study. The algorithm that attained the highest level of accuracy was the decision tree algorithm.

There are a number of medical factors that are utilised by our heart disease detection system in order to ascertain whether or not an individual is suffering from heart disease [12]. These characteristics include heart rate, blood sugar levels, blood pressure, age, gender, and so on. In order to compute dataset analysis, the WEKA programme is utilised. Additionally, a neural network that uses an EHR sequential data model is used to provide predictions on the likelihood of heart failure [17]. The predicted heart disease utilising electronic health records (EHR) is the subject of this study, which is currently being suggested and tested. The real-world data and the congestive data were the sources of the data sets and the training data that we utilised [13]. The one-time encryption model was the one that we utilised most frequently in order to diagnose the events that occurred and the ailment that was brought on by the patient. We are also likely to shed light on the significance of the model described above. Once all of the data sets have been featured, the data set will be divided into leaf nodes [10]. [14] The separation of the datasets is the primary and initial algorithm under consideration. In addition to this, the system adheres to the principles of the second algorithm and determines whether or not everything is in order [15]. In the following procedure, a less error classifier is utilised, which contributes to the reduction of errors. A classification of the algorithm into portions is only possible based on the minimum and maximum rates. Table 1 provides an explanation of the dataset description.

| age | trestbps | chol | thalach | oldpeak | target | sex_0 | sex_1 | cp_0 | cp_1 |
|-----|----------|------|---------|---------|--------|--------|--------|------|------|
| 0   | 0.06     | 0.76 | 0.26    | 0.02    | 1.09   | 1      | 0      | 1    | 0    |
| 1   | -1.02    | -0.09| 0.07    | 1.63    | 2.12   | 1      | 0      | 1    | 0    |
| 2   | -1.47    | -0.09| -0.02   | 0.98    | 0.31   | 1      | 0      | 1    | 0    |
| 3   | 0.18     | -0.66| -0.20   | 1.24    | -0.21  | 1      | 0      | 1    | 0    |
| 4   | 0.09     | -0.06| 2.06    | 0.58    | -0.38  | 1      | 1      | 0    | 1    |

Fig 1. Dataset Description

III. PROPOSED SYSTEM

The overview of the proposed system is shown as below also with examples and development methods and various tools are also shown. The proper software solution is required to run a massive dataset and also make differentiate between various algorithms and make the understandable model to the doctors. We first take the details of the patient and then make a dataset with those details then we will send them to the attribute section and followed by preprocessing data then the most important step to classify the techniques and then we will get the result along with accuracy measures. Figure 2 shows proposed system architecture of proposed method.
IV. EXPERIMENTAL RESULTS
Here we will be getting the accuracy for each algorithm and we will also know if the patient is affected with the heart disease or not. Also accuracy of the algorithms are clearly known. Figure 3 shows about proposed system dataflow of proposed method.

![Fig 3. Proposed System Dataflow](image)

![Fig 4. Maximum KNN Score](image)

Figure 4 & 5 shows about maximum KNN score & comparison of testing accuracy and training accuracy

V. CONCLUSION
In conclusion we would like to conclude that this is the best model for the heart disease detection in early stages. By detecting them in the early stage we can make treatment accordingly. When heart conditions are not good people get completely out of control. So inorder to make them live we made this model which will help the patients to know about their disease in the early stage.

REFERENCES
[1]. M. Sen, K. Sharma, S. Mishra, A. Alkhayyat and V. Sharma, “Designing a Smart and Intelligent Ecosystem for Autistic Children,” 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2023, pp. 1-5, doi: 10.1109/ICIEM59379.2023.10166057.

[2]. Chaudhury, P., & Tripathy, H. K. (2017). An empirical study on attribute selection of student performance prediction model. International Journal of Learning Technology, 12(3), 241-252.

[3]. A. Srivastava, S. Samanta, S. Mishra, A. Alkhayyat, D. Gupta and V. Sharma, "Medi-Assist: A Decision Tree based Chronic Diseases Detection Model," 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2023, pp. 1-7, doi: 10.1109/ICIEM59379.2023.10167400.

[4]. Yeshvendra K. Singh, Nikhil Sinha, and Sanjay K. Singh, "Heart Disease Prediction System Using Random Forest," International Conference on Advances in Computing and Data Sciences, 2004. Singapore: Springer, 2016.

[5]. Prerana T H M1, Shivaprakash N C2 and Swetha N3 "Prediction of Heart Disease Using Machine Learning, Algorithms-Nave Bayes, Introduction to PAC Algorithm, Comparison of Algorithms and HDPS" International Journal of Science and Engineering Volume 3, Number 2 - 2015 PP: 90-99.

[6]. B.I. Deekshatulu Priti Chandra "K-Nearest Neighbor and Genetic Algorithm Classification of Heart Disease" Modeling at the International Conference on Computational Intelligence Techniques and Applications (CIIMTA) 2015.

[7]. Michael W. Berry and colleagues, Lecture notes in data mining, World Scientific (2006) [8]. "Feature selection for medical diagnosis: Evaluation of cardiovascular diseases," S. Shilaskar and A. Ghatol, Expert Syst. Appl., vol. 40, no. 10, pp. 4146-4153, Aug. 2013.

[8]. "Applying decision tree and neural network to improve dermatologic diagnosis," C.-L. Chang and C.-H. Chen, Expert Syst. Appl., vol. 36, no. 2, Part 2, pp. 4035-4041, Mar. 2009.
