IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HEART DISEASE DETECTION USING MACHINE LEARNING

D.Saritha, NazeraTabassum, Adarsh, Dr.SN.Chandrashekhar

B.E ECE student, B.E ECE student, B.E ECE Student, Internal guide, Assistant professor,
Department of Electronics and communication Engineering
Sreenidhi Institute of science and Technology, Telangana, India.

Abstract: Heart disease is a severe health issue that might impact millions of individuals worldwide (HD). Early diagnosis of a cardiac condition will thus benefit in therapy. Because the frequency of heart disease is rapidly growing, it is critical to create technology that can identify the condition more promptly. The presence or absence of a condition is determined by the patient's smoking history. The heart disease system may be able to detect high-risk individuals, explain the essential characteristics of cardiovascular patients, and develop a model that allows them to be distinguished. Machine learning algorithms are applied and assessed based on characteristics such as age, heartbeat, cholesterol, sex, blood pressure, and chest pain. The primary goal of this effort is to create a simple machine learning model.

Index Terms **– Cardiovasular, promptly, diagnosis.**

I. INTRODUCTION

Because cardiac illness is difficult to forecast, the technique must be computerized in order to reduce risks and provide patients with early warning. This study makes use of the heart disease dataset from the repository for machine learning at UCI. The proposed study utilizes a number of data mining approaches, including Naive Bayes, Decision Trees, Rep tree, and Random Forest, to assess the likelihood of heart disease and identify patient risk levels. This study provides a comparison analysis by analyzing the efficacy of several algorithms used in machine learning. As opposed to the other Learning algorithms tested, the Random Forest method has the best accuracy (90.16%), judging by the trial results.

Acute myocardial infarction, sometimes known as a coronary, is one of the worst cardiovascular disorders (AMI). It happens when the blood flow to the heart's muscle is cut off, causing the heart's muscle to be injured. Establishing a heart illness diagnosis is an extra critical task. To determine the diagnosis, the manifestations of cardiac disease, a physical assessment, and knowledge of the various symptoms must all be employed. Many variables, including cholesterol and heredity, are involved.

A variety of factors can contribute to heart disease, including smoking, hypertension, physical exercise, fat, and heart disease. The major cause of heart attacks is a blockage of blood flow to the coronary arteries. As blood flow is reduced, red blood cells (RBC) begin to diminish; as a result, the body stops obtaining oxygen and the person loses consciousness. If the prediction is precise enough, utilizing symptoms and indicators to make an early diagnosis may help patients prevent heart attacks. Figure 1 depicts several heart attack symptoms. The work being presented has 13 traits or qualities with numerical values. It has been stated that adopting minor lifestyle adjustments, such as stopping smoking, can help.

Heart attacks can be avoided by avoiding alcohol, nicotine, and smoking, eating a healthy diet, and exercising regularly. Everyone who lives a healthy lifestyle and seeks treatment as soon as they are diagnosed can dramatically enhance their outcomes. Nevertheless, determining who is at high risk of getting heart disease can be difficult when additional disorders such as diabetes, high blood pressure, and cholesterol problems are present. In these cases, ML can contribute in the early diagnosis of illness.
In order to highlight the relevance of the proposed work, this paper discusses some of the existing machine learning-based diagnosis tools. Researchers have suggested several machine learning-based diagnostic approaches to diagnose HD. Detrano and colleagues [11] created the HD classification algorithm, which had a 77% accuracy rate, using machine learning classification approaches. The Cleveland dataset was used with the features selection approach and the global evolutionary method. Gudadhe and co.

The categorization method has an accuracy rating of 87.3 percent. Result and co. [19] reached a precision of 89.01 percent, sensitivity of 80.09 percent, and specificity of well-being percent by combining the statistical gauging system enterprise miner (5.2) with an ANN ensemble-based HD diagnostic system. Akhil and colleagues [24] devised an ML-based HD diagnostic method. Both the ANN-DBP and the FS algorithms performed admirably. Palaniappan and colleagues [17] proposed an expert medical diagnosis approach for diagnosing HD. During system development, predictive machine learning models such as Navy Bays (NB), Rep tree (DT), and neural network models were used. The NB classifier was 86.2 percent accurate, the ANN classifier was 88.1 percent accurate, and the DT classifier was 80.4% accurate. Olaiynvi et al. [18] developed a three-stage strategy based on the artificial brain network technique for HD prediction in angina that achieved 88.89% accuracy.

II. LITERATURE SURVEY

According to the European Institute of Cardiology [6,] around 26 million cases of HD are detected and diagnosed each year. The majority of people in the United States have heart disease [7]. A clinician typically diagnoses HD after analysing a patient's history, physical exam findings, and any worrying symptoms. The results of this diagnostic method, however, do not accurately identify HD patients. Its analysis [8] is also costly and computationally difficult. To address these issues, a new diagnostic approach based on machine-learning algorithms must be developed.

The choice of parameters and data balance are critical for improving model performance. Several diagnostic procedures have been proposed by many researchers in the field, however they do not accurately diagnose HD. Data standardization is critical for models developed using machine learning to grow increasingly predictive. Min-Max Scalar, Standard Rates are reported (SS), and other preprocessing methods are used to streamline operations, improve learning outcomes, and provide clear and intelligible data. Selecting the finest characteristics from big data is a difficult task due to its complexity and plethora of aspects.

There are further concerns about scalability and stability, as well as identifying characteristics for organized, diverse, and streaming data. For large-scale data analytics, feature selection issues must be handled. Topic hyper chart hashing, an autonomous hashing technique, was developed in [15] to describe the boundaries. By wrapping supporting sentences around images, topic hypergraph hashing effectively addresses the semantic deficit of hashing codes. In terms of performance, the proposed Topic mega graph hashing is more suited for mobile image representation and also can outperform various cutting-edge algorithms.

The selection of features is more influenced by global relevance. On the other hand, successful outcomes necessitate a good machine learning model. The performance of a competent machine preparing mechanism on real-world data is obvious.

Furthermore, it is essential to use efficient cross validation techniques and performance evaluation standards when a model is trained and evaluated on a dataset. In this research, we proposed a diagnosis method for HD based on machine learni

III.SYSTEM ANALYSIS

When assessing performance, the program's output is taken into consideration. One of the most important stages in the study of a system is specifying the requirements. A system may only be developed to function in the defined environment when the relevant specifications are accurately provided. Since they are the ones who will really utilize the system, users of the current system should provide the necessary needs. This is due to the fact that without early knowledge of the requirements, the system cannot be developed to satisfy them. Once a system is designed, it is extremely difficult to change it; nevertheless, designing a system that does not meet the demands of the user is of little utility.

Each system's need definition can be broadly stated as follows:

The system ought to be able to communicate with the current system.

The system ought to be precise.

The system ought to outperform the current system.

The user is entirely responsible for all tasks in the current system.

HARDWARE REQUIREMENTS

Processor: Pentium IV from Intel

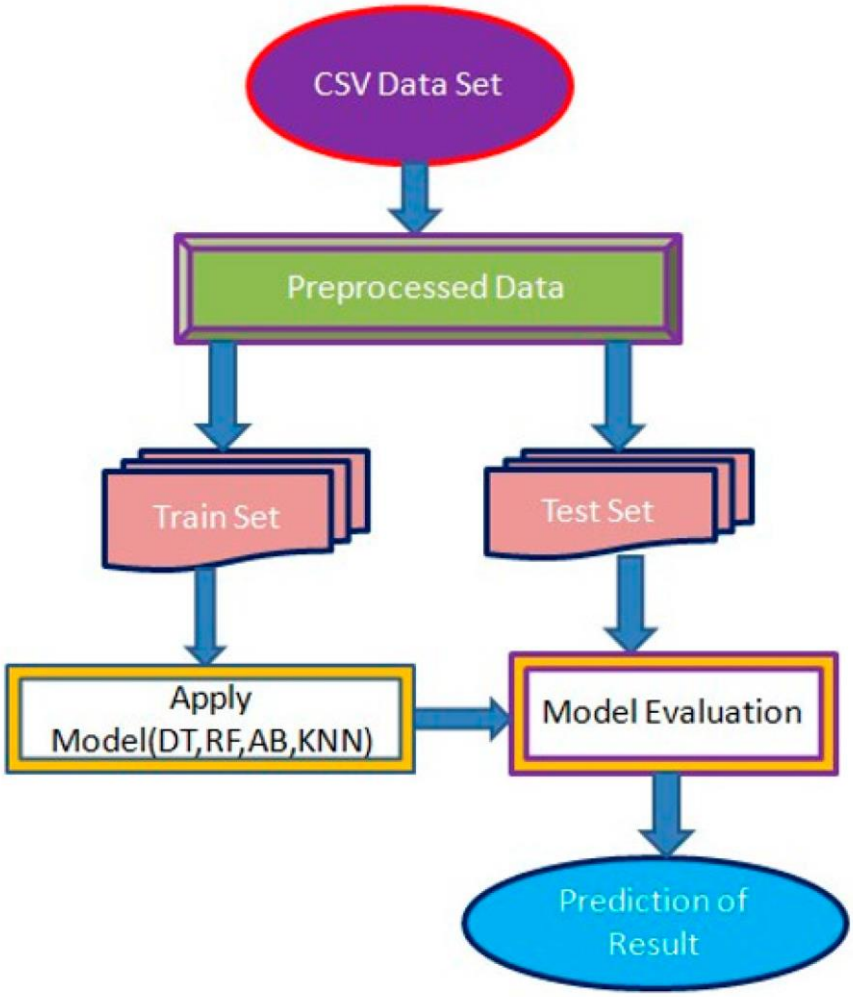
RAM: 4 GB or more

Hard Drive: 120 GB Processor: 1.6 GHz or higher

Display: 1Devices for Input: Keyboard and Mouse

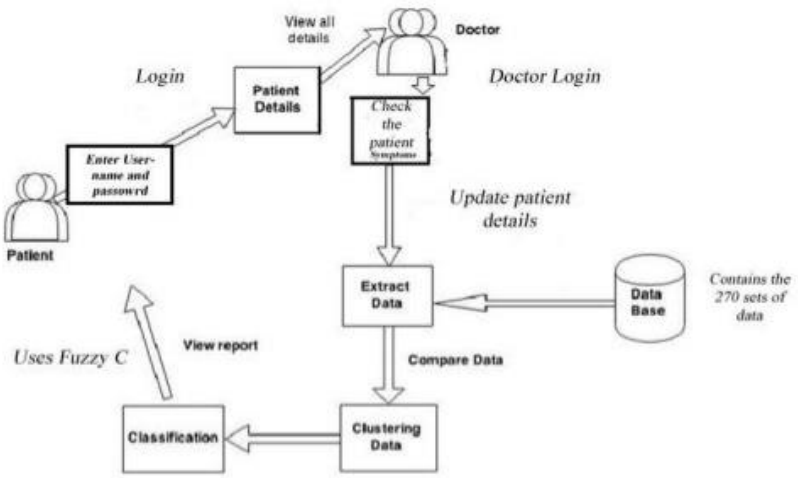
SOFTWARE REQUIREMNTS

The OS is Windows 8. The Languages are Python, HTML, and CSS. Operating system: Google Chrome 50 and above jupyter, Notebooks, and Visual Studio Code are special tools. Also, we used Google Collab in this project to run a software.



IV. SYSTEM DESIGN

Under the suggested method, a developer is in charge of leveraging the hospital's dataset to build the machine learning model. The heartbeat sensor.Based on the idea that each pulse's blood flow via the finger was influenced by light modulation, the heartbeat sensor was created. As blood also absorbs light, changes in the intensity of light passing through a vascular region can be anticipated using the rate of heartbeat pulses. Hence, these signal pulses are the same as heartbeat pulses. When a finger is placed on the structure, it produces an o/p of the heartbeat. For the purpose of measuring Beats per minute, that digital o/p can also be directly fixed to Arduino board.The proposed system's overview is shown in this section, along with examples of each of the system's components, development methods, and tools. An effective software solution is required to train massive datasets. le models is contained.



V.Discussion and results

The testing revealed that backward elimination outperformed the other techniques for feature selection. Sequential Backward Deletion with Classification Model and Retrograde Cancellation without or with KFold were two of the strategies investigated. They ranged in accuracy from 85% to 85.5%, with 85.5% become the greatest. Both RFEV and Backward Elimination exhibited equal reliability, but Reverse Elimination had more Clear Negative misclassifications and higher volatility in exactness than RFEV. The RFEV & backwards elimination accuracy rates are 84% or 86%, respectfully. Also, the recall are 0.99 and 1. accuracy & recollection are both 0.99.

RFECV has less discrepancies than Backward Elimination, as demonstrated. Evaluation Criteria Backward Elimination RFECV accuracy is 83%. Precision is 0.99, recall is 0.99, and precision is 0.84, and recall is 0.86. Table 3: Following development and evaluation, a regression analysis model was used to identify feature selection models.

2.1 Related work

Many research and testing in the fields of machine learning and medicine have been done in recent years, culminating in the release of significant articles. To predict heart disease, the paper [1] suggests utilizing WEKA computer with K Star, Javanica, SMO, and Bayes Net. When k-fold cross validation is used, SMO (89% accuracy)

Linear Regression (87% accuracy) surpass K star, Naive Bayes, and J48 methods. The accuracy performance of these algorithms continues to fall short of expectations.

As a result, improving accuracy performance will allow for more exact illness diagnosis. [2] For cardiac illnesses, 303 cases from the Cleveland Clinic database were investigated utilizing 10-fold Cross Validation, 13 characteristics, and four different techniques. The researchers discovered that Gaussian Nave Bayes and Random Forest were effective.

3.1 conclusion

Three ML categorization simulation models were used to create a coronary heart disease detection model.

This study predicts cardiovascular illness by pulling the individual's medical history that leads to deadly myocardial infarction from a file containing patients' health records such as chest discomfort, insulin levels, hypertension, and so on.

This Myocardial Infarction Warning System aids a patient by using laboratory information from a prior heart disease diagnosis. Regression model, Wild Rainforest Separator, and KNN [22] are the methods used to create the provided model.

Our model has an accuracy of 87.5%. A greater amount of training data increases the model's odds of effectively predicting whether a specific person has high cholesterol or not.

3.2 Acknowledgment

We wish to express our deep sense of gratitude to our Internal guide Dr.SN.Chandrashekhar and project coordinator. Dr.SN.Chandrashekhar, Assistant Professor (ECE) who have taken great care in the project work undertaken, by devoting her valuable time in advising and guiding us at each phase, leading to successful completion of our project work. It has been educative and enlightening experience working with him. We are greatly in debited to prof. Dr.S.P.V SUBBA RAO, Head of the department of Electronics and Communication for providing valuable guidance at every stage of this project work. We are thankful to our principal Dr.T.CH. SHIVA REDDY for providing valuable guidance and support at every stage of this project work. We are profoundly grateful to our director Dr.C.V.Tomy and college authorities for helping us in completing our project work. We would also like to thank all our other faculty members and staff who supported us completing the project work.

4. References

1. The International Journal of Recent Technology and Engineering, vol. 7, no. 6S, pp. 1009 15, S. I. Ansarullah and P. Kumar, "A Systematic Literature Review on Cardiovascular Disorder Identification Using Knowledge Mining and Machine Learning Method," Volume 4, 2016 19

2.Design and Implementation of Heart Disease Prediction Using Naive Bayesian, International Conference on Trends in Electronics and Information, Anjan Nikhil Repaka Sai Deepak Ravika, and Ramya G. Franklin (ICOEI 2019).

3. [4] a world map for managing and preventing cardiovascular disease. id Health Organization, Geneva

4.[5] 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication, Research of Machine Learning Algorithms for Specific Disease Prediction Using Principal of Component Analysis Mahale Kishor M, Dhomse Kanchan B, and others.

5. [7] Cardiovascular Risk Prediction Technique Using the 2017 Random Forest Classification Framework and CFS Subset Evaluation.

6.[8] Developing a Cardiovascular Disease Prediction Model by Vijay K. Mago and Levi Monteiro Martins.