International Journal of Research in Engineering, Science and Management
Volume 6, Issue 5, May 2023
https://www.ijresm.com | ISSN (Online): 2581-5792

Early and Accurate Detection of Cardiovascular Diseases

Radhika Singh[1*], Anirudh Sharma[2], Jyoti Chauhan[3]
[1,2,3]*Department of Computer Science and Information Technology, Meerut Institute of Engineering and Technology, Meerut, India*

***Abstract*: A heart is integral organ for all living beings. Any disturbance in the heart causes disturbances in the whole body. We live in the modern era, and each day brings about significant shifts that have an impact on our lives in one way or another. One of the world's five leading causes of death is heart disease. As a result, the disease's prognosis is crucial because it enables the right decisions to be made at the right time. Machine learning and data mining use huge amounts of data to find and improve useful information. The process of defining a database and mining for hidden patterns and useful information is fundamental. When dealing with complicated non-linear problems and optimization algorithms' adaptability are helpful. In medical science, machine learning methods are used to solve real-world health issues like disease prediction and treatment. The decision tree algorithm outperforms all other classifiers and tests accuracy with 97.29 percent.**

***Keywords*: Detection of heart diseases, python, machine learning.**

1. Introduction

One of the most dangerous diseases that can strike the human body at any time is heart disease. As a result of this and other diseases that are similar, hundreds of people die every day. As per a concentrate by the Indian Heart Affiliation, in excess of 17 lakh individuals pass on from coronary illness consistently, and this figure is supposed to increment to 23 million by 2030. A wide range of conditions that impair heart function are referred to as heart disease. Atherosclerotic coronary illness influences the courses of the heart and is one of a few sorts of coronary illness.

A condition known as valvular heart disease affects how the heart's valves direct blood flow in and out of the body. Cardiomyopathy influences how the muscle of heart are contracts. The heart's electrical conduction is disrupted by arrhythmia. Before birth, there is possibility of both infections and structural issues with the heart. high levels of compounds lipids in the human blood cause cholesterol plaque, these plaques can narrow or completely block the arteries and cause problems throughout the body, resulting in coronary artery disease. These coronary arteries help in supplying blood to the heart. A heart attack, occurs when a cholesterol plaque breach and forms a clot in an artery and prevents blood flow, this is the most common type of cardiovascular disease. The person's professional behaviour and personal health are completely dependent on the human heart's health. It is necessary to develop a model that can easily predict the heart disease with accuracy and includes symptoms and dietary habits due to the rapid rise in the disease. The following characteristics may be exhibited by people at risk for heart disease: increased levels of cholesterol. Smoking causes hypertension. a lot of lipids in there. being overweight and being obese. Heart disease history in the family Various characteristics like age, gender, cholesterol, resting blood pressure, and so on can be used to predict, diagnose, and analyse heart diseases in a way that enables experts to make knowledge-based decisions that are better and more accurate. Machine learning is a new field that enables knowledge to be extracted from massive amounts of data because of the rapid increase in large amounts of data. In healthcare, data mining is used to diagnose, treat, and solve real-world health issues. Using a variety of tools and methods, this research paper aims to collate various ML algorithms and their accurate tests and results. Additionally, it emphasizes the potential for predictive models in healthcare in the future through comprehensive data mining and machine learning algorithm analysis.

2. Literature Survey

Using data mining, deep learning, and machine learning tools, numerous studies have been conducted to predict heart disease. The researchers used a variety of datasets, algorithms, and approaches, and they observed results. More research is being done to find effective medical diagnostic strategies for heart disease. Researchers have been developing predictive models for the diagnosis of cardiovascular disease (CVD) for the past ten years. An important real-world medical issue is automatic CVD detection. Better treatment depends on early heart disease detection. This is only possible with AI such as deep learning and ML, the two fundamental components of predictive analytics that makes predictions about future outcomes using historical data combined with statistical modeling, data mining techniques and machine learning. Heart disease has been detected and predicted in its early stages through a variety of approaches developed by researchers worldwide.

The following are some of the most common methods: Jaimin Patel and co-workers 2015): This study aims to create a

*Corresponding author: radhikasingh172001@gmail.com

model with some success for predicting cardiovascular diseases. The experiment was carried out using the Cleve Land UCI dataset in WEKA, and the data underwent the following procedures during training: visualization, classification, association, regression and clustering prior to processing the classification tools are tested in WEKA Explorer mode. With 10-fold cross-validation and low error pruning, decision tree classifiers like J48, logistic regression, decision tree algorithm, and random forest algorithm were utilized for the purpose of analysis. With low error pruning, j48 achieved the highest level of accuracy. Various discretization methods, multiple classifiers, voting methods, and other decision tree algorithms (gain ratio, Gini index) can improve accuracy. In order to achieve greater precision in terms of sensitivity, specificity, and accuracy, there is a insufficiency of appropriate amalgamation and complex models. In the future, a method for detecting imbalances with other data mining models could be developed using the aforementioned algorithms in this study.

Sonam Nikhar's 2016 work: By removing irrelevant and unwanted features from the dataset and selecting only informative ones, the goal of this study is to enhance the Naive Bayes classifier's performance. There were 303 records and 76 features in the Cleave Land Heart Disease Database, but only 19 were taken into consideration for analysis. C4.5 trees are created using the Selective Naive Bayes classifier. in order to accomplish this. In this review, the use of Gullible Bayes and Choice Tree with data gain computations gives improved results than different classifiers. The greedy algorithm is used to incorporate the decision tree, and it is found that the decision tree is more accurate than Naive Bayes.

(2017) Syedahmin Pouriyeh: On a limited dataset, the primary objective of this paper is to compare various machine learning methods. The Cleave Land Database provided the study's dataset, which consisted of 303 examples with a total of 76 features, of which a maximum of 14 were utilized. The data part has been validated with a 10-fold cross-validation because it has less variance than other estimators like the single-fold approach. Decision trees, naive bayes, a multi-layer perceptron, k nearest neighbours, a single conjunctive rule, radial basis functions, and support vector machines have all been utilized in machine learning classifiers.

In addition to the aforementioned machine learning strategies, boosting, stacking, and bashing were utilized. Precision, recall, the F-measure, and the ROC are all examples of metrics. This study's primary performance indicators. There were two phases to the experiment. Machine learning algorithms were utilized with 10-fold cross-validation in the first step, which utilized the entire dataset. SVM exhibited the highest accuracy, at 89.12%, according to the findings. In the second instance, bagging, boosting, and stacking were used in an experiment. Packing expanded the DT precision from 77.55% to 78.54%. The DT accuracy improved by boosting from 77.55% to 82.17%. However, the highest accuracy of 84.15 percent was achieved when SVM and stacking were combined.

(2018): Youness Khourdifi Cardiovascular disease was predicted making use of machine learning algorithms that were improved through the use of particle swarm and ant colony optimization. A technique known as quick connection based highlight determination (FCFS) was utilized to work on the nature of coronary illness grouping by staying away from extra elements. Particle swarm optimization and ant-colony optimization (ACO) are two methods by which classification is carried out. Other methods include KNN, SVM, NB, RF, and MLP (ANN). The UCI Machine Learning Repository provided the data. The binary type classification problem is prepared for with the training dataset, and WEKA is used to classify new features. The dataset is processed using the proposed hybrid method. With a precision of 99.65 percent, the FCBF, PSO, and ACO optimized model was the most accurate. As a result, the hybrid model mentioned earlier performs better than the other classification methods mentioned. The creator's information base, the gear, and how much time accessible are the essential constraints of this review.

Alex Mamta (2019): This study aims to develop and systematize rapid detection methods for timely disease detection and prediction. The shortcomings of data mining algorithms are highlighted in this paper., such as the fact that diagnosis takes longer and predictions are less accurate. Face-to-face interactions with patients and the details of heart disease patients' discharge are used to collect real-time data from "Jubilee Mission College &Research Thrissur." A sum of 2200 records were gathered with 20 credits, then, at that point, arranged and organized in a progressive way. SVM, RF, KNN, and ANN were used to process the features, with KNN achieving the highest accuracy of 92.21 percent.

3. Methods and Materials

Logistic Regression algorithm: This algorithm is used to predict or calculate the probability of an event occurring binary i.e., yes or no.

The following data driven equation: $Y = ß_0 + ß_1x_1 + ß_2x_2 + ... + ß_n x_n$

Decision Tree: It is a supervised non-parametric algorithm, which is used for both tasks such as classification and regression. It has a tree like structure in hierarchical form, which consists root node, internal nodes and leaf nodes.

Artificial Neural Networks (ANN): It is a deep learning method and based on brain function that used for complicated patterns and forecast issues. This method arose from the concept of biological neural network.

Naïve Bayes: This is a supervised machine learning algorithm, which is used for classification tasks, like text classification.

The algorithm follows the following steps:

Step 1: Distinct by Class.

Step 2: Consolidate the Dataset.

Step 3: Sort the data into categories.

Step 4: Gaussian Likelihood Thickness Capability.

Step 5: Probabilities by Class.

Random Forest Algorithm: This algorithm helps in consolidating the result of numerous choice trees to arrive at a

solitary outcome. it handles both order and relapse issues.

LSTM: LSTMs are used to learning, processing, and classification of sequential data because these networks can learn long-term dependencies between time steps of data.

Support Vector Machines (SVM): A supervised learning algorithm known as a support vector machine divides data into two groups.

4. Materials Required

Hardware Requirements:

- Data acquisition
- The microcontroller communication unit
- Alarm unit

Software Requirements:

- Arduino IDE
- Apache Cordova
- Python libraries

5. Proposed System

- The extracted feature space is cleaned out of irrelevant and noisy data using a variety of machine learning classification algorithms.
- The properties or characteristics used in traditional methods are not suitable. Therefore, for feature selection purposes, the right features for the prediction process are chosen using the Random Forest algorithm.
- In this project, boost model is used to accomplish higher rate of accuracy than traditional model of early detection of heart diseases.
- In both developed and developing nations, cardiovascular diseases have emerged as the leading cause of death over the past few decades. Early cardiac intervention can reduce mortality rates. disease detection and ongoing clinician supervision. An effective machine learning method that was derived from a unique analysis needs to be used in order to accurately detect heart diseases.
- We investigated the performance of machine learning models that have been trained to use longitudinal electronic health records to find early diagnostic heart failure in primary care patients. The length of the prediction window (the time before a clinical diagnosis), the length of the observation window (the time before the prediction window), and the number of distinct data areas (diversity of data) were used to evaluate the model's performance number of patients. recordings in the training dataset (number of data) and frequency of patient encounters (data density).
- Early and timely diagnosis of this problem is very important to prevent further harm to patients and save their lives. Angiography is the most well-known of the traditional invasive methods for diagnosing heart issues, but it has some drawbacks.

6. System Architecture

There are various stages involved in the process of detecting heart diseases. First, we need to train the datasets and then check each dataset for accuracy in order to have a deep understanding of specialized computers. After that, we need to gather databases and figure out what the attributes are. Furthermore, for advancement of approaches we need to assess the entire dataset by directing strategic relapse. Thirdly, we must evaluate, synthesize, and have the capability of decoding and sending packets to search. Finally, we use the cyber security application developed by LSTM and the scan network.

When the patient's data is loaded and stored which is then used by the proposed system to analyse whether the patient will be affected by a heart disease or not. Data such as patient's age, gender, weight, cholesterol, glucose, whether the person smokes, whether the person is alcoholic and various other fields are taken into account by the system to perform calculations and then give accurate results based on the analysis.

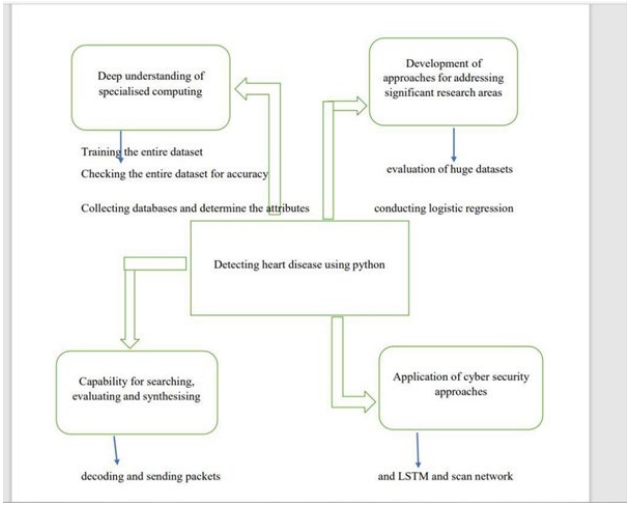

Fig. 1. System architecture

7. Objective

As cardiovascular disease (CVD) is the main reason for morbidity and mortality therefore early diagnosis is essential. The main objective of this study is to accurately predict whether the patient has heart disease. The patient's health report's input values are entered by the healthcare professional. A model that predicts the likelihood of heart disease is fed the data.

Our project's objective is to figure out if the patient will get a heart disease diagnosis or not; this is a binary outcome. A positive result of one indicates that the patient will receive a cardiovascular disease diagnosis, while a negative result of zero indicates that the patient will not receive one.

The contribution of our current work is to detect cardiovascular disease in an early stage with accurate results. Our goal is to provide an accurate system to society so that this chronic disease may not lead to large number of deaths.

8. Conclusion

The concentrate fundamentally centered around, or endeavored to zero in on, beginning phase forecast of cardiovascular disease(s), and hence, an examination of different ML calculations was intended to meet this objective. Pre-processing methods dealt with contaminated and missing

values in order to enhance the quality of the data set. A block diagram, on the other hand, was suggested by us for creating a machine learning model. These models are able to classify the predisposition of a possible patient (to cardiovascular disease) based on the influential parameters of the data set. The disease was predicted using six distinct machine learning algorithms, and the outcomes were compared to a variety of statistical measures. Experiments have shown that random forests are suitable for our purpose. dataset that was considered the best among all classifiers (100 percent of the training set and 97.29 percent of the test set). The reliability of different machine learning algorithms (LR, NB, SVM DT, RF, ANN) was evaluated using the 10-fold cross-validation method.

The RF model performed well with other statistical parameters like precision, recall, the F1-score, specificity, the rate of false positives and false negatives, and the negative predicted value. One believes that larger data sets will need to be used to train and test machine learning classifiers in order to improve disease prediction. in the future. For improving the quality of this study's research, additional research—from data collection to visualization of results—is still needed. The efficiency, dependability, and validity of current research hybridization or consolidation can be improved for earlier human life preservation.

References

[1] https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1
[2] V. Shorewala, "Early detection of coronary heart disease using ensemble techniques," in Informatics in Medicine Unlocked, vol. 26, 2021.
[3] Nashif, S., Raihan, Md.R., Islam, Md.R. and Imam, M.H. (2018) Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System. World Journal of Engineering and Technology, 6, 854-873.