CARDIO VASCULAR DISEASE PREDICTION USING MACHINE LEARNING

T Suryam [1], S Neelima [2], Y V N S S Gayathri [3], R Harshini [4] and V Meenakshi [5]

[1,2,3,4,5] Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India.

ABSTRACT

A variety of disorders that affect the heart and blood arteries fall under the umbrella term of cardiovascular disease. There is a wealth of medical data in the healthcare sector. As a result, efficient decision-making in the prediction of cardiac illnesses requires the use of machine learning algorithms. For prediction and decision-making at various levels, data pre-processing in the proposed research employs techniques such the removal of noisy data, the removal of missing data, the filling in of default values when appropriate, and the classification of attributes.

INTRODUCTION

The demand for intelligence and accuracy in technology is rising as the Earth travels through a purple technological region. Internet addiction is more common today, but people are less likely to worry about their own health. Humans are surrounded by technology in the twenty-first century since it is an integral part of our daily lives. With this, we are always concentrating on our own health and the health of our earned possessions. Individuals avoid going to the hospital for minor issues that could develop into serious illnesses in the future. Instead of searching through a list of potentially relevant documents on the web, creating question-and-answer forums is increasingly becoming a straight forward method of responding to those inquiries. According to the World Health Organization, cardiovascular diseases account for more than 12 million deaths annually. Individuals have busy, routine schedules, which cause stress and anxiety. Also, a significant increase in the number of people who are overweight, stressed, and addicted to cigarettes. This is a major factor in the development of cardiac problems. At 17.9 million deaths per year, or 31% of all deaths worldwide, cardiovascular diseases (CVDs) are the leading cause of death worldwide. Heart attacks and strokes account for four out of five CVD deaths, and one-third of these deaths happen before the age of 70.

Our main goal is to create a system that can forecast diseases and provide information about them, including their severity and any symptoms that the user enters. The computer will assess the symptoms against the datasets offered in the database. If the symptoms fit the criteria, it will show the probability of cardio disease and risk of the disease. This paper can be used to forecast a potential heart illness using random forest algorithm to predict the disease. Heart failure is a common event brought on by CVDs. Both men and women, regardless of age, are more likely to develop heart disease. A machine learning model can be very helpful in the early detection and management of people with cardiovascular disease or who are at high cardiovascular risk. For instance, in a task to diagnose heart disease, it is essential that the accuracy on healthy individuals be as high as feasible, as a mistake in this category could lead to a healthy patient receiving treatment without a need. The distribution of the system's performance among the various classes could change and depends mainly on the nature of the underlying medical condition and the data gathered. Also, the mortality of patients with various ailments has grown in the majority of the countries due to a shortage of medical specialists. In both urban and rural settings, heart disease has become the leading cause of death in the majority of the world's nations. Machine learning is an art of mastering system without being explicitly computed. They prove to be effective in assisting in making the decision and predictions from the large quality of data. They are used to analyse the analytical arrangement in high dimensional, diverse data

sets like heart diseases. They are used in recognition of the arrangements (patterns) that gives support for forecasting and controlling mechanism for analysis and medication.

LITERATURESURVEY

[1] Abhay Agrahara "Heart disease prediction using Machine Learning algorithms" uses K-Nearest Neighbour, Support Vector Machine, Random forest, Logistics Regression, Decision tree. KNN is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems. In other words, similar things are near to each other. In this paper, it is given an accuracy of 86.09%. The objective of Support Vector Machine algorithm is to find a hyperplane in an N-dimensional space(N-the number of features) that distinctly classifies the data points. Support Vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize themargin of the classifer. These are the points that help us build our SVM, gives anaccuracyof88.04%.Logistic regression is used when the dependent variable(target) is categorical, gives an accuracy of 86.09%. Random forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is allso one of the mostused hyper plane. The loss function that helps maximize the margin is hinge loss,gives an accuracy of 96.82%. It takes as inputa record or object described by a set of attributes and returns a "decision with predicted output value for the input", gives anaccuracyof98.29%.

[2] Jaymin Patel proposed a paper **"**Heart disease prediction" using Machine Learning and Data mining techniques uses J48 algorithm, logistic model tree algorithm, Random Forest Algorithm. In J48 algorithm, it makes decision tree for classification and uses decreased- error p runing. Decision trees are built using information. It is just extension of ID3 algorithm. Coming to the logistic model tree algorithm, it also contains decision trees with logistic regression. It isslower than the other algorithms. Random Forest Algorithm has classifiers consists of many decision trees. The outputs are represented by individual trees for each classifier. It gives T/F values with different logics. The accuracy of J48 is 56.78% and LMT is 55.77%. Among these J48 algorithm has highest accuracy.

[3] Sanjay Kumar Sen proposed a paper "Predicting and Diagnosing of Heart Disease using Machine Learning Algorithms". Like Naïve Bayes Classifier, SVM, Decision tree. This classifier learns from training data theconditionalprobabilityofeachattribute.Itgivesanaccuracyof83%.Support Vector Machine(SVM) exist in different forms, linear and non-linear. But comparatively to Naïve bayes classifier it applied the most, gives an accuracy of 84%. Decision Tree partitions the input space of a dataset into mutually exclusive regions, each of which is assigned a label, a value oran action to characterize its data points. TheDecision tree mechanism is transparent and we can follow a tree structure easily to see how the decision is made, gives an accuracyof77%.

[4] Shashikant proposed a paper "Heart Disease Diagnosis using Support Vector algorithm".In SVM the risk is less and works on the principle of fact error rate and test is bounded by the training error rate. SVM depends on VM Dimensions. The diagnosis of disease is based on dataset like sample data of patient. Here, we use 5-fold and 10-fold cross validation technique on dataset, in which it gives promising results. The accuracy of prediction using SVM is 85%.

PROPOSED SYSTEM

It begins with data gathering and identification of the critical characteristics for system operation. The necessary format is then created by pre-processing the pertinent data. The information is then separated into training and testing data. The model is trained using the algorithms and training data. Testing the system with test data allows for the determination of the system's correctness.

The following list includes the modules that this system implements.

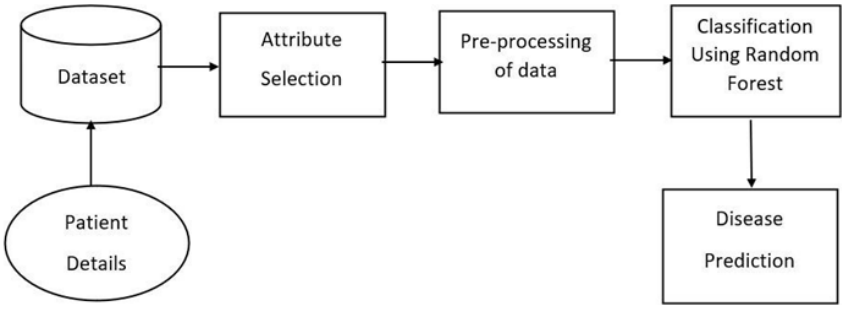

Fig 1: Architecture

Dataset Gathering: To begin, we compile a dataset for our system that anticipates cardiovascular disease. After being gathered, the dataset was split into training and testing data. The training dataset serves as the basis for the prediction model's learning, while the testing dataset serves as the basis for its evaluation. In this case, 30% of the data is utilised for testing, and 70% is used for training.

Attribute or Feature Selection: This step involves choosing the right attributes for the prediction system. This is done to make the system more effective. For the prediction, a number of patient characteristics are used, including gender, the nature of the patient's chest discomfort, fasting blood pressure, serum cholesterol, and exang. The Correlation matrix was mainly used in the selection of the attribute.

Data Pre-Processing: Before building a machine learning model, data pre-processing is a crucial step. If the data is not in the correct format it may have incorrect result. The transforming of the data into the correct format can be done using Pre-processing. It is used to handle the dataset's noise, duplication, and missing values. The set of activities mainly include importing datasets, partitioning datasets, attribute scaling, etc. are all part of data pre-processing. Pre-processing the data is necessary to increase the model's accuracy.

Data Balancing: There are two approaches to balance unbalanced datasets. They both over sample and under sample;

1. **Under Sampling**: In Under Sampling, the size of the large class is reduced in order to balance the dataset. When there is enough data, this process is taken into account.
2. **Over Sampling**: In over sampling, the size of the sparse samples is increased in order to balance the dataset. When there is not enough data, this process is taken into account.

Disease Prediction: Many machine learning methods are used for classification, including SVM, Naive Bayes, Decision Trees, Random Trees, Logistic Regression, Ada-boost, and Xg-boost. The algorithms are contrasted, and the one that predicts cardiovascular disease with the best degree of accuracy is chosen.

RESULTSANDDISCUSSIONS

DATA ENTRY

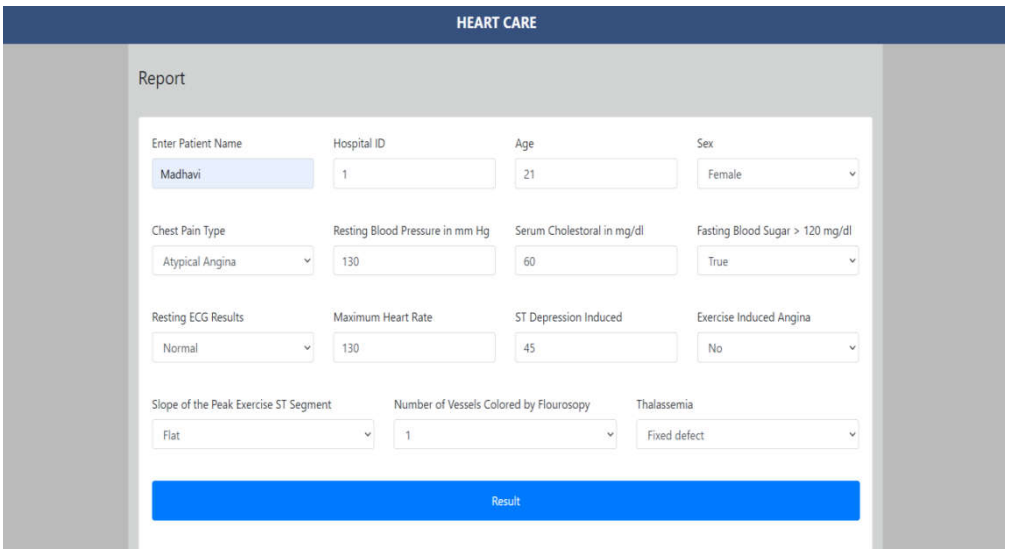

Fig 2: Data Entry

REPORT

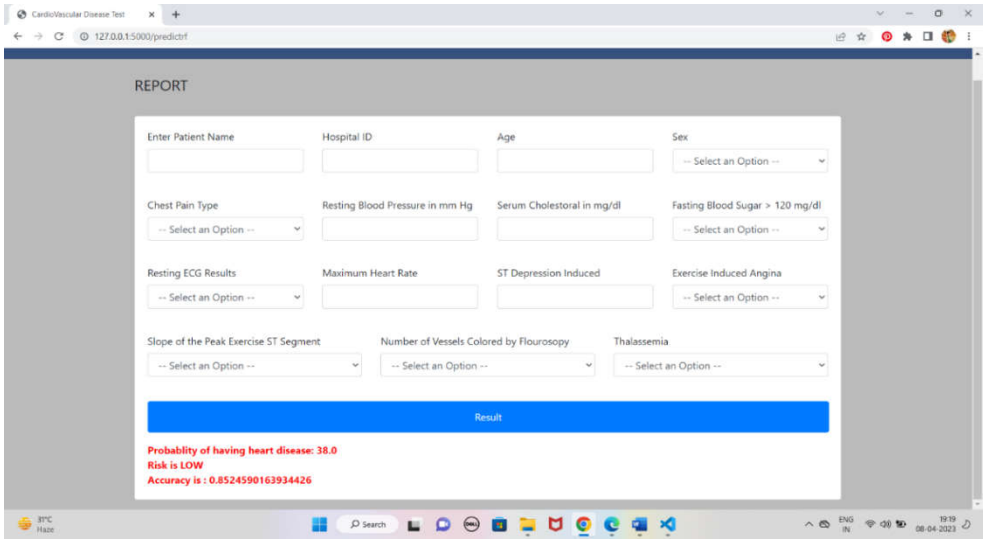

Fig 3: Prediction

CONCLUSION

Utilization of important technology, such as machine learning, to the first prediction of heart problems would have a significant impact because cardio diseases are a leading cause of death in India and around the world. Early detection of cardiac disease can help high-risk patients and make decisions about lifestyle modifications that will lessen problems, which can be a significant advancement in the field of medicine. Each year, more people are diagnosed with cardiac illnesses. The medical field and also patients can both benefit greatly through this use of right technology support. The main goal of this project is to predict the cardio disease with the help of Random Forest algorithm. We take symptoms put down by patient and then predict the probability of the disease. The dataset comprises 14 significant features that are useful to evaluate the system, and among them few related attributes are taken to predict the characteristics that lead to cardio

disease in individuals. Finally, the output screen displays the probability & risk of the disease in particular person with the given details. In this project we used random forest algorithm and achieved the accuracy of 86.1% approximately which is higher than the previous work and also makes this system more reliable than the existing one.

FUTURE SCOPE

Further for its enhancement, we can train on models and predict the types of cardiovascular diseases providing recommendations to the users. The extension of the project can be done in such a way that the user can access the website with ease. We can further use more models for more accuracy. We can further deploy the project in AWS cloud for global accessibility.

REFERENCES

1. Abhay Agrahara, HeartDiseasePredictionusingMachineLearningalgorithmVolume 6, Issue 4 Page Number: 137-149 Publication Issue: July-August-2020

2. Jaymin Patel, Prof.Tejal Upadhyay, Dr.Samir Patel “Heart disease prediction usingMachinelearningandDataMiningTechnique"Volume7.Sept2015 – November 2016.

3. SanjayKumarSen”PredictingandDiagnosingof HeartDiseaseUsingMachineLearning Algorithms”- International Journal of Engineering And Computer ScienceISSN:2319-7242Volume6Issue6June2016.

4. Shashikant “Heart Disease Diagnosis using Support Vector algorithm”- InternationalJournal of Engineering And Computer Science ISSN:2319-7242Volume5 Issue 6 May2016.

5. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and AsmitaMukherjee”HeartDiseaseDiagnosisandPredictionUsingMachineLearningandDataMiningTechniques: AReview”-AdvancesinComputationalSciencesand TechnologyISSN0973-6107,Volume10,Number7(2017).