IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Heart Disease prediction using Machine Learning

B Mathura Bai,Karne Sreshta, Lingam Naveena , Shaik Reenadh, Suma Manvika
Assoiate Professor, Student, Student, Student, Student
Department of IT,
VNRVJIET, Hyderabad, India

Abstract : In this study, heart disease is anticipated using machine learning. Machine learning plays a crucial role in recognizing motor illnesses like heart diseases if such data is predicted in advance. This gives clinicians the insight they need to treat each patient individually. In our project, we will derive various insights from the dataset that helps us to analyze the weightage of individual features and how they are related to one another. We have used a new algorithm called modified knn (mknn) and compared it with existing algorithms like k-nearest neighbours support vector classifier. The healthcare industry collects a tremendous amount of data, but not all the data finds hidden patterns and can help in making valuable decisions.

I. INTRODUCTION

INTRODUCTION

One of the ailments that has been on the rise recently is heart disease. High blood pressure, smoking, a poor diet, drinking alcohol, and getting too little sleep are the main causes of heart disease. According to estimates, 13% of deaths from CVD are believed to be caused by high blood pressure, 9% by cigarette use, 6% by diabetes, 6% by inactivity, and 5% by obesity. 90% of heart illnesses, according to the estimate, are treatable. Sometimes people discover they have a problem after it is already advanced and there is little chance of a cure. One in three persons worldwide, according to the 2012 World Health Statistics report.

Our application is motivated by the fact that there is least awareness about Heart Disease and to integrate technology and science to make a meaningful and useful solution. People cannot identify the existing heart disease because of mild symptoms which increases the possibility of being in the end stage. The prediction and treatment are expensive which is also a reason why people are ignorant.

Our aim is to make a useful application that can be easily used by any individual who is accessible to technology. In this way, people can get to know about their health condition and can be aware of the seriousness of this disease. We have used MKNN algorithm and compared with existing algorithms like K-Nearest Neighbor algorithm, Random Forest, Decision Tree, Support vector machine, Naïve Bayes, C4.5 and Voting Classifier. We have taken preprocessed dataset and split it into training and testing sets. We have created models by feeding the data and measuring the accuracy of the prediction. We have considered parameters likeage,sex,cholesterol,thalach,slope,oldpeaketc.

NEED OF THE STUDY.

According to one survey, one in every seven American individuals has CKD. In India, about 87% are related to hypertension, 37% to diabetes, 22% to CVD (cardiovascular disease), 6.7% have a history of acute renal injury, and 23% have previously used alternative medications. Treatment and cure of Heart Disease are expensive when they are in renal stages. Hence early-stage prediction is quite necessary. Existing Heart Disease can lead to other diseases and can also affect other internal organs other than the kidneys in the body. A proper diagnosis can lead to sufficient treatment and a healthy lifestyle.

3.1 Effective Prediction of Cardiovascular Disease Using Hybrid Machine Learning Methods

This initiative offers a novel and cutting-edge understanding of heart disease while predicting cardiovascular disease from raw data processing. In this study, K-NN, Naive Bayes, neural networks, decision trees, and neural networks are used to compare the accuracy and precision of six different methods. They employed hybrid machine learning algorithms in order to improve forecasting methods. To improve the accuracy of the dataset, they coupled the linear model with Random Forest to create a new technique called Hybrid Machine Learning. Six patient reports are missing from the dataset, which has 303 patient details, and the remaining records are used to train and test the algorithm. Following instruction on the 297 patient medical records that demonstrate that 137 documents depict the estimation of 1

3.2 A hybrid method for predicting mortality for cardiac patients using ACO HKNN

In this project, the hybrid approach of ACO and HKNN is used, ACO is preferred for feature selection and HKNN is used for classification. HKNN is used to overcome drawbacks of KNN like dependency of K value and execution time. The final outcome

has shown that ACO-HKNN has given the accuracy of 98%, which is higher than other machine learning techniques which exists for cardiovascular disease prediction.

3.3 Utilizing Hybrid Machine Learning Techniques for Effective Heart Disease Prediction

In this project, a distinct methodology is put forth to increase the accuracy and precision of existing heart disease prediction techniques by identifying key features using machine learning algorithms. The predicted model is introduced with combination of many feature selection methods and variety of categorization methods they developed a heart disease prediction model for predicting heart disease. Combining the traits of Random Forest (RF) and Linear Method (LM), the hybrid HRFLM approach was employed, and it was successful in achieving an accuracy level of 88.7%.

RESEARCH METHODOLOGY

The methodology section outline the plan and method that how the study is conducted. This includes Universe of the study, sample of the study,Data and Sources of Data, study's variables and analytical framework. The detailsare as follows;

3.1Population and Sample

In our investigation, the Cleveland preprocessed data is used to remove any missing factors and improve prediction accuracy. The output is produced using the Modified KNN algorithm by categorizing the obtained features. This study compares the outcomes to those of earlier machine learning algorithms with Modified KNN.

3.2 Data and Sources of Data

All the above mentioned algorithms are used to construct, train, and test a model. The dataset, which has 297 items and 13 attributes, was taken from the UCI repository. Algorithms for classifying data have been fed this information. The dataset consists of 13 different features, which include age, sex, resting blood pressure, chest pain type, cholesterol, fasting bp, resting cardio graphic results, thalach, angina, old peak, slope, number of major vessels, thal. It is the first and crucial step in building the model. The data collection should be proper so that the model will be more accurate in predicting. The dataset is taken from the Heart Disease Cleveland UCI It has 297 patient records with 13 attributes

3.3 Objective of the proposed system

- To determine if a person has the chance of getting effected by the heart disease.
- To find the key factors that affect the patient health through the analysis.
- To take precautions before getting more effected by the heart diseases.

3.4 System Design

The workflow of the system can be illustrated here

1. Dataset Collection: The data for training the heart disease prediction model has been taken from the UCI Cleveland heart disease repository. It consists of 13 attributes in text format, with the class labels checked. The feature names include Age Group, Gender, chest pain type, resting blood pressure, cholesterol, fasting blood pressure, resting cardio graphic results, maximum heart rate, angina, old peak, slope and thalach.

2. Model Building: After the dataset is ready, it has been split into training and testing data, and cross validation is performed. Here, 70% of the data is taken for training purpose and 30% of the data has been taken as testing purpose. The different models that were taken into consideration are Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors, Random Forest Classifier, Decision Tree and Modified KNN.

3. Model Training: The training data is now fit into the various supervised learning algorithms and the outputs are predicted. The accuracy metrics are derived based on the outcome of the predictions.

4. Heart Disease Prediction: Now, to test the accuracy of the model, sample data is given fed to the system. The result is given in the form of numerical number, which tells if the person is having heart disease or not.

5. Model Pickling: With the models trained and the outcomes predicted with the accuracy metrics, the models are to incorporate them into a web interface, which works based on the learned data. The web interface is linked to the pickle file by using Flask python library.

6. User Interface: It consists of user login and signup and takes test value as input , shows the presence of heart disease .

3.4.2. Algorithms used

Support Vector Machine: Support-vector networks, sometimes referred to as support-vector machines in machine learning, are supervised learning models for regression and classification. The Support Vector Machine algorithm with a polynomial kernel has been used to try to achieve a higher accuracy and to deal with the enormous number of attributes that contribute to the output.

Decision Tree: This is the widely used supervised machine learning technique that is used in data mining. A decision tree is a diagram that individuals use to illustrate a statistical likelihood or to determine the sequence of events, actions, or outcomes. In this project, we used the attributes of age, gender, education, physical health issues, sleep trouble, anticipatory anxiety, peer pressure, and frequency of suicidal thoughts as the factors to determine whether the student is suffering from mental health problems.

K-Nearest Neighbors: It uses non-parametric machine learning to address problems with regression and classification. The computer predicts future data points based on the training data itself in this sort of instancebased learning. The separations between each new data point and each previous data point in the training set are determined by KNN using the input features. The KNN approach then uses the measured distances to determine the K-nearest neighbors to the new data point. The algorithm then predicts the label or value of the incoming data point using the labels or values of the Knearest neighbors.

Random Forest Classifier: It is a technique for ensemble learning that aids with classification and regression issues. When dealing with enormous datasets, as in this instance, it is a very helpful algorithm. It combines different decision trees to create a more

accurate and trustworthy model. Each decision tree in the random forest is trained using a random subset of the input features and a portion of the training data. The Random Forest Classifier creates numerous decision trees and then combines their predictions to offer a final prediction. Each decision tree in the forest is trained via bagging or bootstrap aggregation using a random subset of the training data.

Modified KNN: In KNN, the output is determined by how long it took to calculate the distance. Assigning weights to each element helps the KNN perform better by reducing processing time, and the k value is chosen automatically. There are six main steps in the Modified KNN. Each element in a class has a distance and a centroid, and the data is separated into classes. In the first stage, the data is separated into classes C0=H1, H2, H3, etc. Each class's centroid is identified. The centroid of each class and the distance between each element in that class are calculated. Centroid = $\frac{1}{v}\sum_j^v$.

The inverse of the distances w(C,H) is determined for each class. The weight value is measured among the class elements and the centre of the class.

$$D(C,H) = \sqrt[2]{\sum_{i=1}^{y}(c_{z,i} - H_{j,i})}$$

$$W(C,H) = \frac{1}{D(C,H)}$$

The test element T is taken. The distance between the test element and the class centre is calculated for each class.

$$\sqrt{\sum_{i=1}^{y}(T,C)}$$

The k element is chosen and the distance is measured from the test element to the k element. To calculate the element strength of the element is the distance between the test element and the k element is multiplied with the weights of that particular class. Based on the strength the rule is framed for deciding the test element belongs to which class.

ST=1/D×W(C,H)

3.4.3 Comparison of the Algorithms

To figure out which algorithm is the best; we have used the following metrics:

Model	Accuracy	Precision	Recall	f1 score
SVM	67.6768	0.725275	0.481752	0.578947
Decision Tree	85.5219	0.856061	0.824818	0.840149
Naive Bayes	84.8485	0.853846	0.810219	0.831461
KNN	75.4209	0.731884	0.737226	0.734545
Hybrid KNN	90	0.76	0.863636	0.808511

From the below graph, it is shown that the Modified KNN(MKNN) has the highest accuracy.

IV. RESULTS AND DISCUSSION

The accuracy metrics of the predicted models were trained and tested based on several factors in the dataset such as Age Group, Gender etc. The metrics show that the Modified KNN is the model with the highest accuracy of 90 percent.

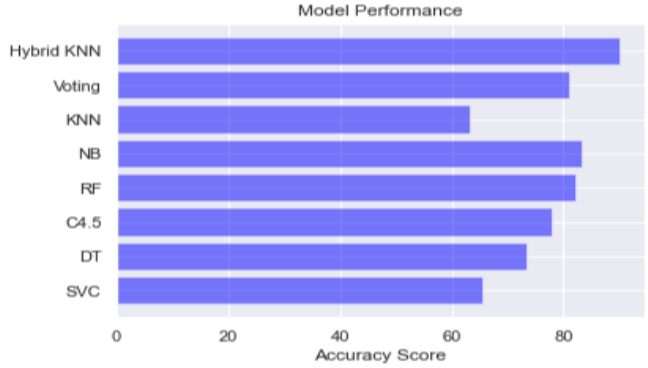


Fig 3: Performance of all the algorithms

User Interface: The web interface interacts with the user and takes the attributes. The input is collected from the user and the model predicts whether the person is affected by the heart disease or not. Fig 4 and Fig 5 shows the interface where the user can give the input to the model and when the input values are given the model will predict the output.

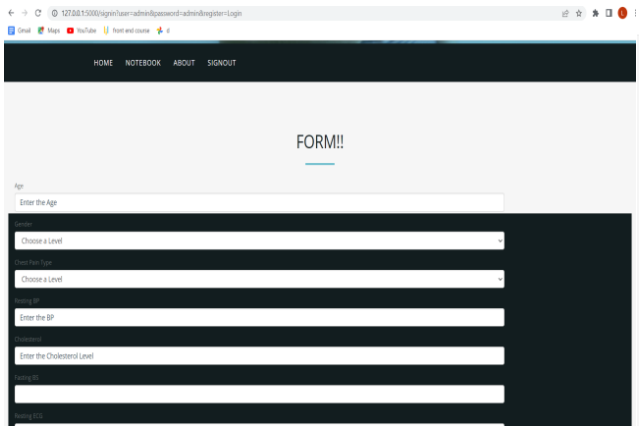


Fig 4 Interface to enter the input values

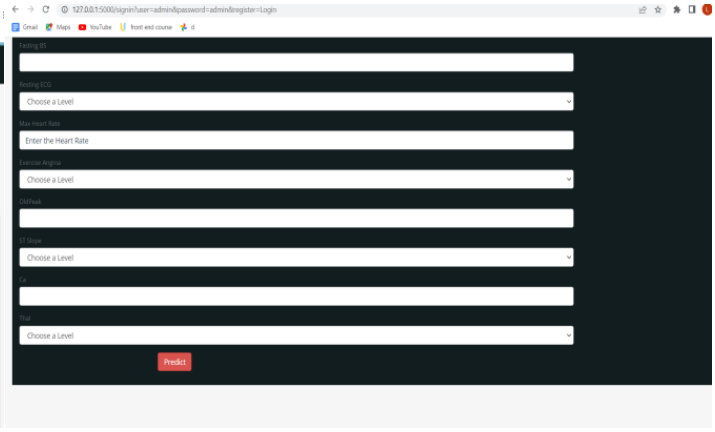
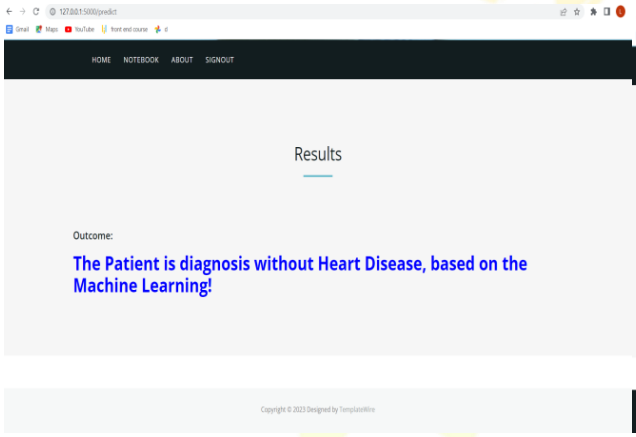


Fig 5:User interface



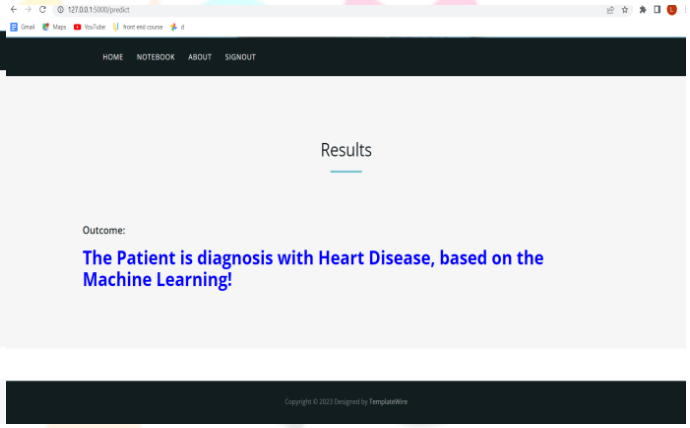


Fig 6: Heart Disease prediction not having heart disease Fig 7: Heart Disease prediction having heart disease

Future Scope:
The study has to be developed further in order to shift the investigations away from theoretical frameworks and simulations and towards actual datasets. A unique feature selection approach must be used to choose the dataset's ideal combination of necessary qualities that enhances prediction performance. This research's future focus can be carried out by combining different machine learning techniques to enhance prediction techniques.

II. ACKNOWLEDGMENT

REFERENCES

[1] Madhavi Veeranki, Jayanag Bayana: Effective Cardiovascular Disease Prediction using Hybrid Machine Learning Techniques vol. 9, no.-4, pp 2249-8958, 2020.
[2] Amin UlHaq, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, and Ruinan Sun: Hybrid Intelligent System Framework for the Prediction of Heart Disease using Machine Learning Algorithms vol. 2018, article id 3860146, https://doi.org/10.1155/2018/3860146.
[3] Navdip Singh, Sonika Jindal: Heart Disease Prediction System using Hybrid Technique of Data.
[4] Mining Algorithms, vol.4, pp 2414-532X, 2018.
[5] Ramkumar P, Thanusha K, Soumya U, Sahana K, Sushma M: Prediction of Cardiovascular disease using Hybrid Machine Learning Algorithm, vol. 7, no. 14 pp. 2394-5125,2020.