JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Coronary Heart Disease Prediction using Machine Learning

[1] Veena Potdar, [2] Lavanya Santhosh, [3] Likith Jadhav M

[1]Associate Professor, Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bangalore, India
[2]Assistant Professor, Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bangalore, India
[3]Student, Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bangalore, India

Abstract

The field of medical analysis is often referred to be a valuable source of rich information. Coronary Heart Disease (CHD) is one of the major causes of death all around the world. Therefore, early detection of CHD can help reduce these rates. The challenge lies in the complexity of the data and correlations when it comes to prediction using conventional techniques. The aim of this research is to use the historical medical data to predict CHD using Machine Learning (ML) technology. The scope of this research is limited to using three supervised learning techniques namely Naive Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT), to discover correlations in CHD data that might help improving the prediction rate. Using the South African Heart Disease dataset of 462 instances, intelligent models are derived by the considered ML techniques using 10-fold cross validation. Empirical results using different performance evaluation measures report that probabilistic models derived by NB are promising in detecting CHD.

Keywords — Diagnosis, Heart Disease, Machine Learning, CNN

I. INTRODUCTION

The medical field is changing at a rapid rate as new diseases emerge on a regular basis, requiring the development of appropriate treatment options. For accurate diagnosis and treatment, an effective modality is required. If the system is automated, it can be extremely useful. However, the problem is that medical practitioners are not efficient enough in each subspecialty, leading to a shortage of human resources. Therefore, implementing an efficient automated medical diagnostic scheme can greatly benefit any stratification involved in the process.

The heart is a muscular organ responsible for blood circulation. The human heart acts as a compressor, regulating blood flow in the circulatory system. The heart pumps deoxygenated blood from other areas of the body into the veins, and oxygenated blood is returned from the lungs to different areas through the arteries. The lungs help with oxygenation. The sinus node's system of electrical impulses located in the heart is used to organize the pumping frequency. It works like a natural pacemaker but is located at the top of the right atrium. The control over contraction and relaxation of the atria and ventricles is provided by messages sent to specific myocardial tissues. As a result, heart disease is the leading cause of death in adults worldwide. According to a recent heart health survey, nearly 1.2 billion people die each year from 4,444 heart diseases. There is no single solution to increase heart disease prediction quickly and automated. There is no single solution to the growing burden of heart disease, given the rapidly changing economic. Prognosis of heart failure has always been an extremely difficult task considering the high-cost rate. The cost of a wide range of modern cardiac imaging and diagnostic methods is very high. The main causes of heart disease are chest tightness, shortness of breath, fatigue, edema, palpation, and fainting, as well as coughing, hemoptysis, and cyanosis.

Coronary artery disease is a type of heart disease in which the arteries of the heart cannot supply enough oxygen-rich blood to the heart. Coronary artery disease affects the large coronary arteries on the surface of the heart.

Another type of heart disease, called coronary microvascular disease, affects the small arteries in the heart muscle. Coronary microvascular disease is more common in women. The cause of coronary heart disease depends on the type. Coronary artery disease is usually caused by cholesterol, a waxy substance that builds up inside the walls of coronary arteries to form plaque. This buildup can partially or completely block the flow of blood through the large arteries of the heart. Coronary microvascular disease occurs when the small blood vessels in the heart do not work properly. For most people, coronary heart disease can be prevented through a heart-healthy lifestyle. Symptoms of coronary heart disease can vary from person to person, even if they have the same type of coronary heart disease. However, many people have no symptoms so they don't know they have coronary heart disease until they experience chest pain, blood flow to the heart is blocked causing a heart attack, or the heart suddenly stops working, also known as cardiac arrest.

In recent times, despite the advances in medical science, heart diseases are diagnosed by ink injection, X-ray, MRI, etc. India has more than a billion people and about one radiologist per 100,000 inhabitants. Therefore, it is difficult to respond immediately to all patients. Our system attempts to combine questions about everyday life and personal health choices, along with MRI image recognition to reduce time and increase detection.

II. LITERATURE REVIEW

Senthil Kumar Mohan et al, [1] proposed Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques in which objective is to finding critical includes by applying Machine Learning bringing about improving the exactness in the expectation of cardiovascular malady. They introduce a technique we call the Hybrid Random Forest with Linear Model (HRFLM). The expectation model is created with various blends of highlights and a few known arrangement strategies. We produce an improved exhibition level with a precision level of 88.7% through the prediction model for heart disease with hybrid random forest with a linear model (HRFLM) they likewise educated about Diverse data mining approaches and expectation techniques, such as, KNN, LR, SVM, NN, and Vote have been fairly famous of late to distinguish and predict heart disease.

Sellappan Palaniappan et al. [2] proposed a model Intelligent Heart Disease Prediction System (IHDPS) built with the assistance of data mining techniques namely, Neural Network, Naïve Bayes and Decision Tree. Results show that each technique has its infrequent strength in realizing the objectives of the defined mining goals. IHDPS is able to answer complex "what if" queries which conventional decision support systems cannot.
The results illustrated the uncouth strength of each of the methodologies in comprehending the goal of the specified mining objectives. IHDPS was capable of responding queries that the traditional decision support systems were not able to. It facilitated the establishment of crucial knowledge such as patterns, relationships amid medical factors connected with heart disease. IHDPS subsists well -being web-based, user-friendly, reliable, scalable and expandable.

Chaitrali S. D et al., [3]. investigated a computation structure for heart syndrome with the help of full amount of input characteristics. A few terms related to medical like blood pressure, sex, cholesterol and 13 more attributes like this were recycled to predict the heart disease to a particular person or patient. He also made use of two different attributes like smoking and obesity. Unlike data mining performances were used like Decision trees, neural networks and Naïve Baye's for analyzing the heart disease database. The concert of these practices depends on the accuracy provided by the system. The accuracy provided by decision tree is 99.62%, neural network is 100% and naïve bayes is 90.74% respectively.
S. Vijayarani, et. al. [4], made use of experimental results carried out using dissimilar classification methods for heart disease dataset. The different classification systems which were used and tested by them are Decision Stump, Random Forest and LMT tree algorithm. WEKA tool was used for comparison.

Niti Guru et al. [5]. introduced the diagnosis of Heart Disease, Blood Pressure and diabetes with the aid of neural networks. Experiments were carried out on a sampled data set of patient's records. The Neural Network is trained and tested with 13 input variables such as Blood Pressure, Age, Angiography's report and the like. The supervised network has been advised for diagnosis of heart diseases. Training was carried out with the help of back propagation algorithm. Whenever unfamiliar data was inserted by the doctor, the system identified the unknown data from comparisons with the trained data and produced a catalog of probable diseases that the patient is vulnerable to.

M. A. Nishara Banu et. Al., [6] published a research paper "Disease Forecasting System Using Data Mining Methods". In this article, the preprocessed data is clustered using clustering algorithms as K-means to gather relevant data in a database. Maximal Frequent Item set Algorithm (MAFIA) is applied for mining maximal frequent model in heart disease database. The regular patterns can be classified into different classes using the

C4.5 algorithm as training algorithm using the concept of information entropy. The result demonstrates that the designed prediction system is capable of predicting the heart attack successfully.

Soodeh Nikan et. al., [7] Machine Learning Application Predict the Risk of Coronary Artery Atherosclerosis: Coronary artery disease is the leading cause of death in the world. In this research, we propose an algorithm based on the machine learning techniques to predict the risk of coronary artery atherosclerosis. A ridge expectation maximization imputation (REMI) technique is proposed to estimate the missing values in the atherosclerosis databases. A conditional likelihood maximization method is used to remove irrelevant attributes and reduce the size of feature space and thus improve the speed of the learning. The STULONG and UCI databases are used to evaluate the proposed algorithm. The performance of heart disease prediction for two classification models is analyzed and compared to previous work. Experimental results show the improved accuracy percentage of risk prediction of our proposed method compared to other works. The effect of missing value imputation on the prediction performance is also evaluated and the proposed REMI approach performs significantly better than conventional techniques.

Lakshmana Rao et al,[8] Machine Learning Techniques for Heart Disease Prediction in which the contributing elements for heart disease are more (circulatory strain, diabetes, current smoker, high cholesterol, etc..). So, it is difficult to distinguish heart disease. Different systems in data mining and neural systems have been utilized to discover the seriousness of heart disease among people. The idea of CHD ailment is bewildering, in addition, in this manner, the disease must be dealt with warily. Not doing early identification, may impact the heart or cause sudden passing. The perspective of therapeutic science furthermore, data burrowing is used for finding various sorts of metabolic machine learning a procedure that causes the framework to gain from past information tests, models without being expressly customized. Machine learning makes rationale dependent on chronicled information.

III. METHODOLOGY

Our system uses dual method to diagnose and determine heart disease. Patient inputs via text form and also image recognition via image of MRI has been used. All the data used to train the models are collected from sources like Kaggle, UCI etc.

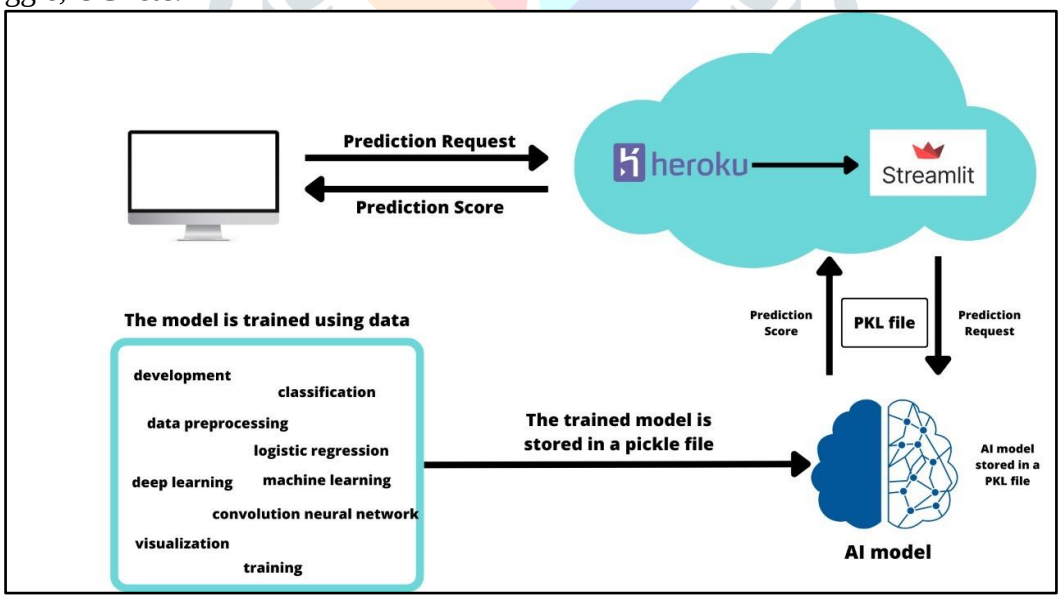

Fig1: Methodology

Indian Heart patient dataset is considered for implementation, which contains images of good health and bad health.

For text-based form dataset required are:

- Race
- Age
- Sex
- BMI
- Sleep timing
- Health Habits

Form Based Detection

A form is created with text and dropdown questions for the patients to fill. These questions are well researched questions which can guide us on their lifestyle, habits and overall general health. Based on the inputs provided the system determines their heart condition is either healthy or weak.

Heroku is a container-based cloud platform as a service (PaaS). Developers use Heroku to deploy, manage, and scale modern applications. Our platform is sleek, flexible, and easy to use, giving developers the easiest path to bringing their apps to market. Heroku is fully managed, giving developers the freedom to focus on their core product without being distracted by maintaining servers, hardware, or infrastructure.
The form is hosted on Heroku and developed using streamlit. The form is trained using logistic regression method. The data collected on phone is processed and analyses itself. Post analysis it determines If the heart is prone to heart disease or not. The form data is processed against the trained data for similarities and based on the similarity the result is determined.

Image Recognition Method

Along with form with patients inputs we also have an option of using MRI scan images. These images are collected and cleaned carefully. The cleaned images are processed for gray scaling. Post the Gray scaling method the images are trained using CNN. During CNN the machine is taught what is a healthy heart and which isn't. The Image dataset used for training in image recognition method is divided into test set and training set. Training set is used to train the AI Model in TensorFlow. The images are cleaned and further converted to grayscale. Post gray scaling the images are used to train the model. The features are extracted and saved in bytes. While testing too the image is uploaded via the option provided in the frontend. This image is converted to grayscale and the features are matched with the trained dataset. If there is a match based on CNN modal then the output received. Result can be healthy or unhealthy based on the match.
This system provides dual validation for heart disease prediction.

Both the models are pre-trained using logistic regression and CNN and uploaded to the pickle. Pickle stores the trained files. This AI model file performs the detection and prediction of heart disease.
The frontend for the users to input the information is developed on streamlit.
Streamlit is a free and open-source framework for quickly building and sharing stunning machine learning and data science web applications. It is a Python-based library specially designed for machine learning engineers

IV. RESULTS

Logistic Regression here is the confusion matrix of the Logistic Regression:

We have used the L2 penalty, the square of the magnitude of coefficients, supported by Logistic Regression to avoid overfitting. The train accuracy is 83.88% and test accuracy is 85.25%. It performs well but not the best for us. The advantage of the Logistic Regression is that it does not need too much computational resources and it is highly interpretable. So, it is easy and sufficient to apply Logistic Regression. However, the limitation of Logistic Regression is that it assumes linearity between the features of the dataset. In the real world, the data is rarely separable, neither as our dataset. That is why we cannot reach a very high accuracy of 90%.

The implementation of both test-based and the image-based detection of heart disease increases the accuracy and provides indebt information for the medical department, researchers and doctors.

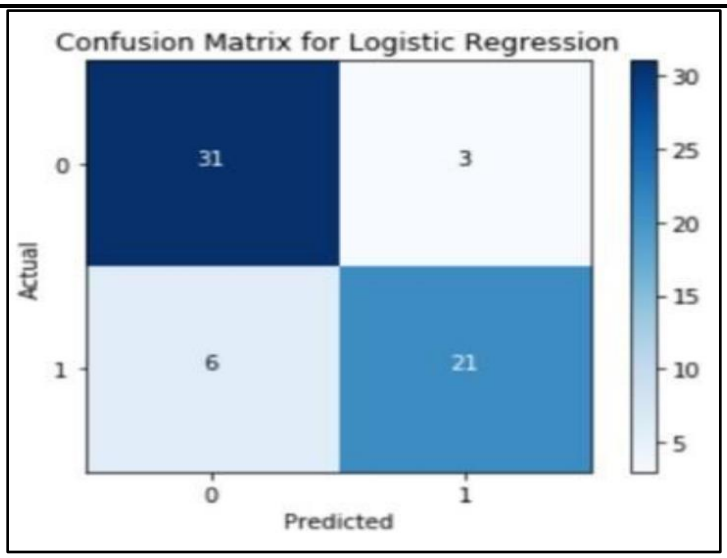

Fig 2: Confusion Matrix for Logistic Regression

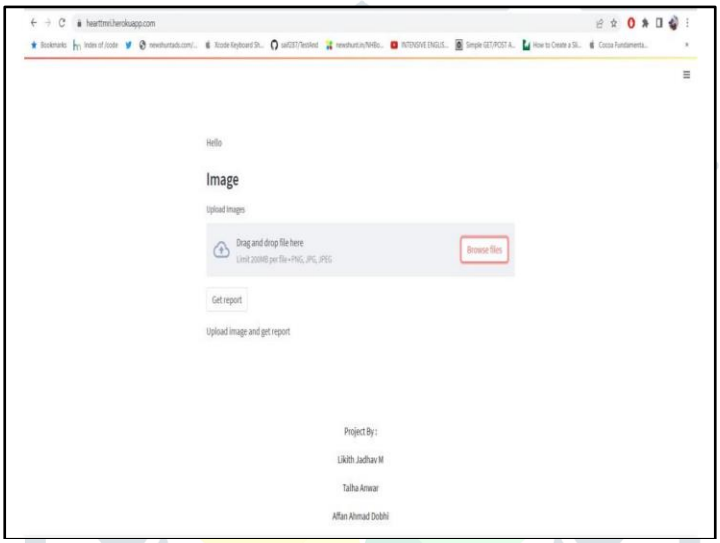

Fig 3: Home page

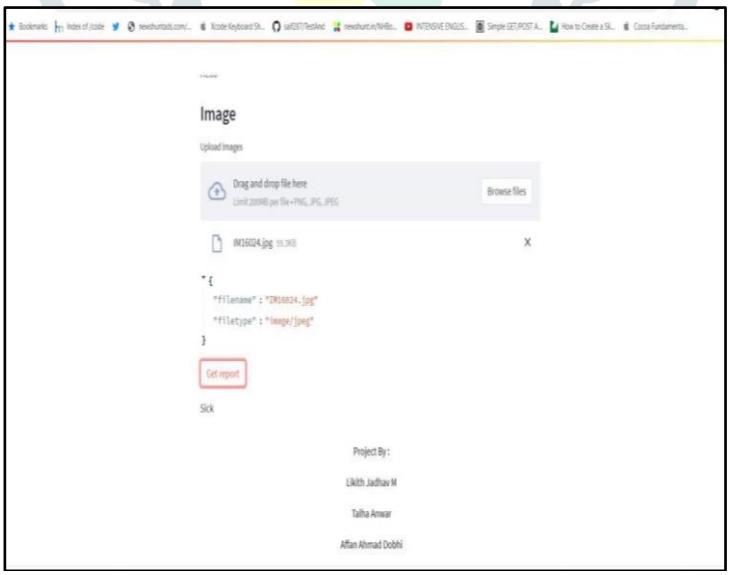

Fig 4: Output as Sick

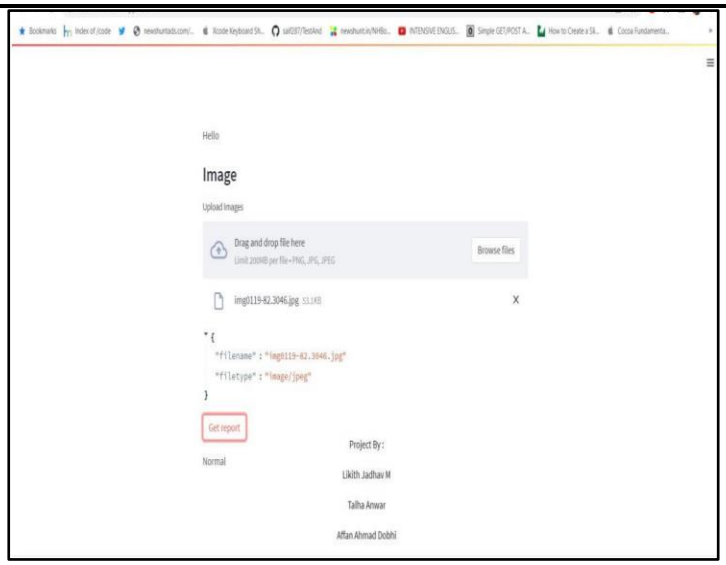

Fig 5: Output as healthy

Fig 6: Form User input

The user interface Fig 2 & Fig. 3 developed for this system is easy and encourages users to easily fill the details required for further diagnoses. The data collected also helped in understanding the patient's routine, regular exercise routine and food habits. This data is invaluable and has helped in better planning and direction for any further treatments, medications and analysis by a medical expert. Further the suggestions provided at the end of the test has added value to the system. This information is like a guide and navigates them to a page or website providing verified details based on diagnoses. If healthy it displays as Fig 4 and if sick as Fig 5. The text input screen is as per Fig 6.

V. Conclusion and future enhancement

This study explores methodologies in chronic heart disease prediction using machine learning and deep learning. Heart disease is especially difficult to diagnose given the subtle nature of its symptoms. Of the 2,626,418 deaths reported in the United States for 2014, chronic heart disease accounted for nearly 38,170 deaths. Prediction by means of computers will continue to grow in importance. This project explored 2 possibilities of prediction model, one is using machine learning models and hybrid model and other using deep learning algorithm. There are many criterions for evaluating the selected feature subset, here this thesis used features such as Total bilirubin, Direct_ bilirubin, Total_protiens, Albumin, A/G ratio, SGPT, SGOT, Alkphos to evaluate the performance of different classification algorithm. In future, we have attempted to classify different feature selection algorithms into four groups: complete search, heuristic search, meta-heuristic methods and methods that use artificial neural network. PSO has been widely used for feature selection to improve heart classification performance. Further, a lot of work is being done using multi-objective PSO for feature selection to improve heart classification performance and to reduce number of features selected as well. Most of the existing multi-objective feature selection based on PSO algorithms use binary tournament selection to select best and uniform and non-uniform mutation. There is a scope to further reduce search space for better heart classification

accuracy if enhanced selection and mutation procedures are being used. The future methodology is used to analyze the liver region into separable compartments i.e., heart etc.

VI. REFERENCES

[1] Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava ―Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, DOI:10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7, 2019 S.P. Bingulac, ―On the Compatibility of Adaptive Controllers, Proc. Fourth Ann. Allerton Conf. Circuits and Systems Theory, pp. 8-16, 1994. (Conference proceedings)

[2] Sellappan Palaniappan and Rafiah Awang, "Intelligent heart disease prediction system using data mining techniques", International Journal of Computer Science and Network Security, vol.8, no.8, pp. 343-350,2008.

[3] Chaitrali S. Dangare, Sulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications (0975 888) Volume 47 No.10, June 2012.

[4] S. Vijayarani and S. Sudha, "An Efficient Classification Tree Technique for Heart Disease Prediction", International Conference on Research Trends in Computer Technologies (ICRTCT - 2013) Proceedings published in International Journal of Computer Applications (IJCA) (0975 – 8887), 2013 (pp 6-9).

[5] Niti Guru, Anil Dahiya, Navin Rajpal, "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Vol. 8, No. I, January - June 2007.

[6] M. A. Nishara Banu, B. Gomathy, "Disease Forecasting System Using Data Mining Methods," International Conference on Intelligent Computing Applications, 2014

[7] Soodeh Nikan, Femida Gwadry-Sridhar, and Michael Bauer "Machine Learning Application to Predict the Risk of Coronary Artery Atherosclerosis", IEEE, August 2016

[8] A. Lakshmanarao, Y. Swathi, P. Sri Sai Sundareswar," Machine Learning Techniques for Heart Disease Prediction", International Journal of Scientific & Technology Research Volume 8, Issue 11, November 2019.