

Health Prediction System For Heart Disease

Kovarthanan K¹, Dr.L.Josephine Mary²

¹Dept of MCA

²Professor, Dept of MCA

^{1, 2}Sri Muthukumaran Institute of Technology, Chennai

Abstract- Big data is difficult to work with using most relational database management systems and desktop statistics and visualization packages. So we can use machine learning. The proposed system shows a machine learning processing model, from the data mining perspective. Using classifiers, we are processing heart percentage and values are showing as a confusion matrix. We proposed a new classification scheme which can effectively improve the classification performance in the situation that training dataset is available. Stent is used for diagnosis of heart disease. We proposed as checking the whole patient heart disease using naive Bayes classification in machine learning. So in that, we will take results of how much percentage patients get disease as a positive information and negative information. Furthermore, the resulting model has a high specificity rate which makes it a handy tool for junior cardiologists to screen out patients who have a high probability of having the disease and transfer those patients to senior cardiologists for further analysis.

Keywords- CardioVascular Diseases (CVD), Coronary artery diseases (CAD), Coronary Heart Disease(CHD), Graphical User Interface (GUI), Content Management System(CMS)

I. INTRODUCTION

Machine learning is one of the most disruptive technologies of this generation. It is a part of data science wherein the computer systems are made to learn from the different data sets on the basis of patterns generated from the datasets. It basically enables a machine to learn on its own based on some useful algorithms specifically developed for it. Machine learning is being heavily used in all the sectors including manufacturing, healthcare, Research and Development etc. In this work we are predicting the heart disease occurrence in a patient based on some important characteristics which are best suited based on our data set that we have collected.

The heart is one of the main organs of the human body. It pumps blood through the blood vessels of the circulatory system. The circulatory system is extremely important because it transports blood, oxygen and other materials to the different organs of the body. Heart plays the

most crucial role in circulatory system. If the heart does not function properly then it will lead to serious health conditions including death.

Types of Cardiovascular Diseases

Heart diseases or cardiovascular diseases (CVD) are a class of diseases that involve the heart and blood vessels. Cardiovascular disease includes coronary artery diseases (CAD) like angina and myocardial infarction (commonly known as a heart attack). There is another heart disease, called coronary heart disease(CHD), in which a waxy substance called plaque develops inside the coronary arteries. These are the arteries which supply oxygen-rich blood to heart muscle. When plaque begins to build up in these arteries, the condition is called atherosclerosis. The development of plaque occurs over many years. With the passage of time, this plaque can harden or rupture (break open). Hardened plaque eventually narrows the coronary arteries which in turn reduces the flow of oxygen-rich blood to the heart. If this plaque ruptures, a blood clot can form on its surface.

A large blood clot can most of the time completely block blood flow through a coronary artery. Over time, the ruptured plaque also hardens and narrows the coronary arteries. If the stopped blood flow isn't restored quickly, the section of heart muscle begins to die. Without quick treatment, a heart attack can lead to serious health problems and even death. Heart attack is a common cause of death worldwide. Some of the common symptoms of heart attack are as follows.

Chest pain

It is the most common symptom of heart attack. If someone has a blocked artery or is having a heart attack, he may feel pain, tightness or pressure in the chest.

Nausea, Indigestion, Heartburn and Stomach Pain

These are some of the often overlooked symptoms of heart attack. Women tend to show these symptoms more than men.

Pain in the Arms

The pain often starts in the chest and then moves towards the arms, especially in the left side.

Feeling Dizzy and Light Headed

Things that lead to the loss of balance.

Fatigue

Simple chores which begin to set a feeling of tiredness should not be ignored.

Python Language

Python is an object-oriented programming language created by Guido Rossum in 1989. It is ideally designed for rapid prototyping of complex applications. It has interfaces to many OS system calls and libraries and is extensible to C or C++. Many large companies use the Python programming language include NASA, Google, YouTube, BitTorrent, etc. Python programming is widely used in Artificial Intelligence, Natural Language Generation, Neural Networks and other advanced fields of Computer Science. Python had deep focus on code readability & this class will teach you python from basics.

Web Applications

You can create scalable Web Apps using frameworks and CMS (Content Management System) that are built on Python. Some of the popular platforms for creating Web Apps are: Django, Flask, Pyramid, Plone, Django CMS. Sites like Mozilla, Reddit, Instagram and PBS are written in Python.

Scientific and Numeric Computing

There are numerous libraries available in Python for scientific and numeric computing. There are libraries like: SciPy and NumPy that are used in general purpose computing. And, there are specific libraries like: EarthPy for earth science, AstroPy for Astronomy and so on. Also, the language is heavily used in machine learning, data mining and deep learning.

Creating software Prototypes

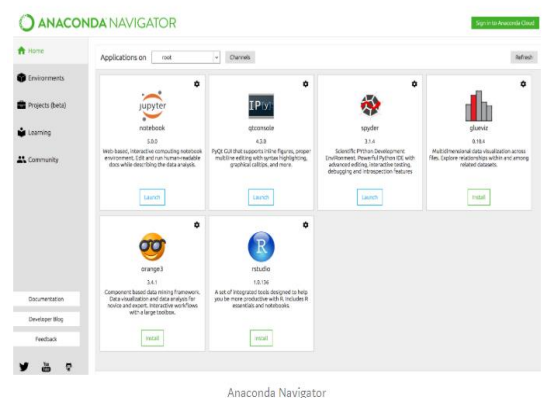
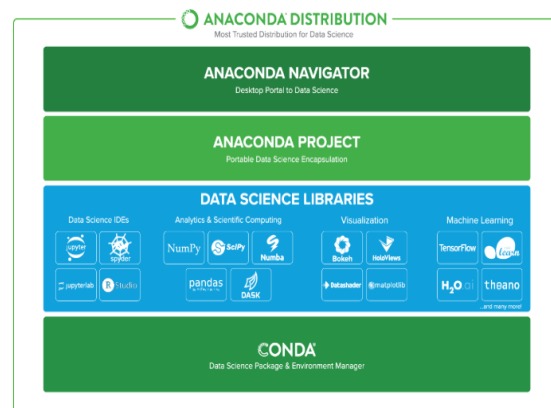
Python is slow compared to compiled languages like C++ and Java. It might not be a good choice if resources are limited and efficiency is a must. However, Python is a great language for creating prototypes. For example: You can use Pygame (library for creating games) to create your game's

prototype first. If you like the prototype, you can use language like C++ to create the actual game.

Anaconda

Anaconda is a free and open source, easy to install distribution of Python and R programming languages. Anaconda provides a working environment which is used for scientific computing, data science, statistical analysis and machine learning. The latest distribution of Anaconda is Anaconda 5.3 and is released in October, 2018. It has the conda package, environment manager and a collection of 1000+ open source packages long with free community support.

Anaconda Navigator is a desktop graphical user interface (GUI) included in the Anaconda distribution.



It allows us to launch applications provided in the Anaconda distribution and easily manage conda packages, environments and channels without the use of command-line commands. It is available for Windows, macOS and Linux. The paper is organized in 4 sections. Section 2 presents Literature Survey. Section 3 describes the implementation of

proposed application design and model. Further, section 4 discusses conclusion and future enhancement.

II. LITERATURE SURVEY

With the rampant increase in the heart stroke rates at juvenile ages, we need to put a system in place to be able to detect the symptoms of a heart stroke at an early stage and thus prevent it. It is impractical for a common man to frequently undergo costly tests like the ECG and thus there needs to be a system in place which is handy and at the same time reliable, in predicting the chances of a heart disease [1].

In modern society, Heart disease is the noteworthy reason for short life. Large population of people depends on the healthcare system so that they can get accurate result in less time. Large amount of data is produced and collected by the healthcare organization on the daily basis. To get intriguing knowledge, data innovation permits to extract the data through automatization of processes. Weighted Association Rule is a type of data mining technique used to eliminate the manual task which also helps in extracting the data directly from the electronic records. This will help in decreasing the cost of services and also helps in saving lives. In this paper, we will find the rule to predict patient's risk of having coronary disease. Test results have shown that vast majority of the rules helps in the best prediction of coronary illness [2].

The heart disease prediction modelled using partially observable markov decision process (POMDP) is proposed. In emergency, the patient is alerted through the doctor by fog computing. Ambulance sent to the location of patient at critical situations. The doctor gets the data through fog computing iFogSim. Fog computing in healthcare is a new area, which gains more attraction in research community. Many researches focus on cardiovascular disease i.e. heart disease. The important risk factor for cardiovascular disease is increase in blood viscosity. The highly viscous nature of blood does not allow the blood to flow creating a resistance in the blood flow. Heart disease risk factors are high blood pressure, obesity, diabetes, increased blood viscosity, etc. The POMDP model for heart disease prediction computes the policy approximation using states and timeslots. Rewards are tabulated using policy approximations over different iterations[3].

Heart disease indicates the type of condition which leads to heart malfunction. Not all people with coronary artery disease have chest pain as a symptom. There are other factors such as R-Blood Pressure, S-Cholesterol, F-Blood Sugar, R-ECG, Thalach, Ex-Ang, Number of major Vessels blocked, Thallium Scan which also results in heart disease. At an initial

stage the prediction of heart disease can save human lives. Data mining techniques such as Naive Bayes, Decision Tree, K-nearest neighbor, etc., are used in the Heart Disease prediction based on the parameters / factors. It also explores about the existing techniques and models used for prediction. This work concludes with the key challenges which are not yet explored[4].

The significant advances in biotechnology and health science have led to produce large amount of data. As the heart disease causing the major problem because it is very difficult to identify disease based on symptoms. For that we need a lot of experience and knowledge. Finding the exact cause of disease can help to patient cure at early stage of risk level. Ischemic heart Disease is one of the most common causes of death nowadays. In this research work, we have applied data mining classification techniques on the stroke dataset and proposed a recommendation system in which a person can come to know their risk level of IHD. By taking proper cure and treatment can help to survive more years if they have high risk of IHD. We have used five data mining algorithms Logistic Regression, Decision Tree, K nearest Neighbor, Naïve Bayes and SVM on Ischemic Stroke Dataset and got the highest accuracy 97.91% using Support Vector Machine algorithm[5].

Over the last few decades, heart disease is the most common cause of global death. So early detection of heart disease and continuous monitoring can reduce the mortality rate. The exponential growth of data from different sources such as wearable sensor devices used in Internet of Things health monitoring, streaming system and others have been generating an enormous amount of data on a continuous basis. The combination of streaming big data analytics and machine learning is a breakthrough technology that can have a significant impact in healthcare field especially early detection of heart disease. This technology can be more powerful and less expensive. A real-time heart disease prediction system based on apache Spark which stand as a strong large scale distributed computing platform that can be used successfully for streaming data event against machine learning through in-memory computations. The system consists of two main sub parts, namely streaming processing and data storage and visualization. The first uses Spark MLlib with Spark streaming and applies classification model on data events to predict heart disease. The seconds uses Apache Cassandra for storing the large volume of generated data[6].

Cardiovascular disease is the leading cause of mortality for both sexes in worldwide. Heart disease is increasing at a rapid rate in both older and younger generation of males and females now days. So in need demand of right

strategies, development and implementation of effective health monitoring policies should be emphasized to combat the epidemic of heart related diseases. So early detection and treatment with the use of both conventional and innovative technique must be preferred In this paper we have used the UCI machine learning repository Cleveland heart disease database having 303 instance and 76 attributes. For the proposed method we have used the Information gain concept for selection of best attribute and processes the selected features using weka and python. The various techniques, processes which have used to train the model of heart datasets such as feature selection, numpy, pandas library, decision tree classifier, KNN classifier, entropy, gini-index, confusion matrix. The result shows that decision tree classifier is most effective and appropriate for prediction of UCI repository Cleveland heart dataset[7].

Cardiovascular illness is the most widely recognized infection and many individuals are experiencing this malady from an extremely youthful age. It is the enormous purpose behind mortality. Accordingly it is imperative to analyze the issue and treat it on time before it turns out to be deadly. A computerized framework will help in indicating how lethal the issue is and how soon it should be dealt with. This will likewise decrease the endeavors of a specialist to assemble all the moment subtle elements of the individual to look for the issue inside and out[8].

Linear Discriminant Analysis (LDA) is an easy and efficient method for pattern classification, while it is also broadly utilized for the initial discovery of diseases using Electronic Health Records (EHR) data. The time complexity will be reduced in this algorithm by optimal scoring analysis of LDA and will be comprehensive to execute sparse discrimination through the combination of Gaussians if limits between classes are nonlinear or if subgroups are available inside every class [9]. To estimate the heart disease status, we apply a generalized linear model with regularization paths via coordinate descentX[10].

There are many approaches within the literature for detecting the arteria coronaries illness. Within the existing work, decision tree classification algorithm has been to assess the events associated with cardiac disease. In decision tree ID3, C4.5 algorithms are used to perform CAD analysis. Under classification stream, and fuzzy logics are used.

We proposed system acts as a system and will prove to be an aid for the physicians with the diagnosis. To design a Health Prediction System for medical data classification and early disease prediction by using supervised learning. The system effectively identifies the disease and its sub types, the

sub type which is referred as the level of class such as normal and mild or extreme. Using combinatorial methods from data mining decision making has been simplified and our proposed algorithm achieved highest accuracy.

III. IMPLEMENTATION OF HEALTH PREDICTION SYSTEM FOR HEART DISEASE

The System Architecture of Health Prediction System for heart disease is shown in Fig. The following modules and its functions which are used in health prediction system for heart disease are given below.

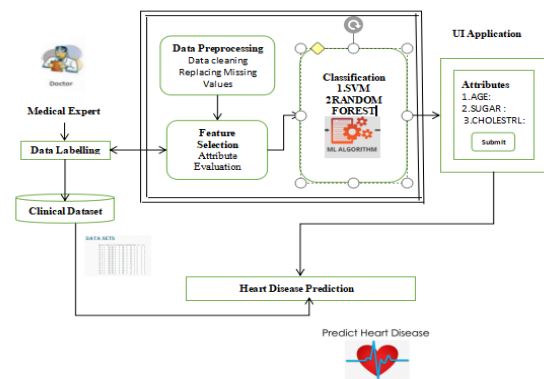


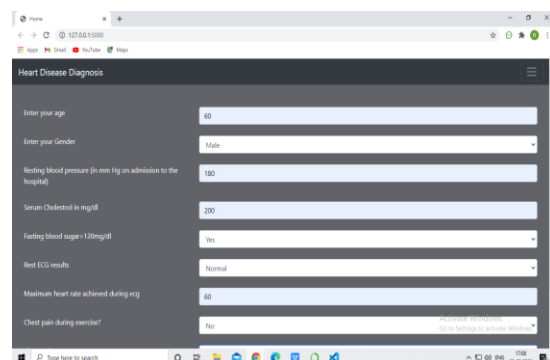
Fig. System Architecture Of Health Prediction System for Heart Disease Data

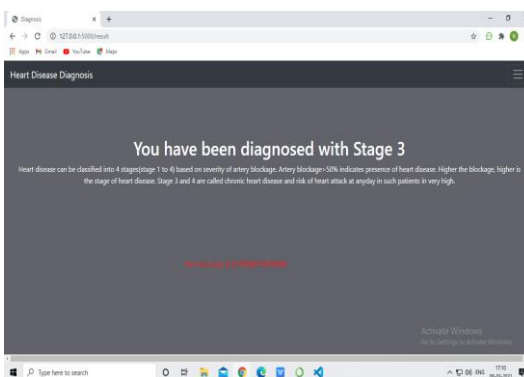
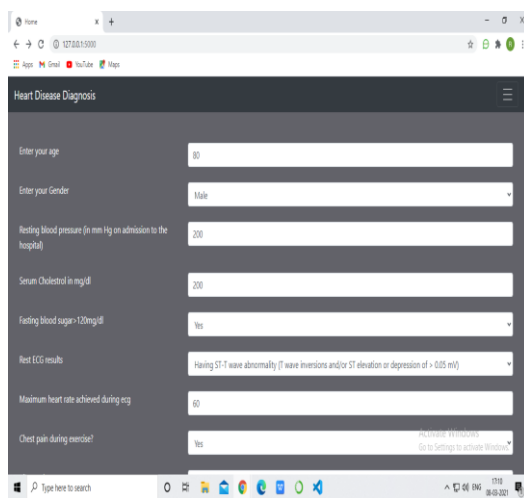
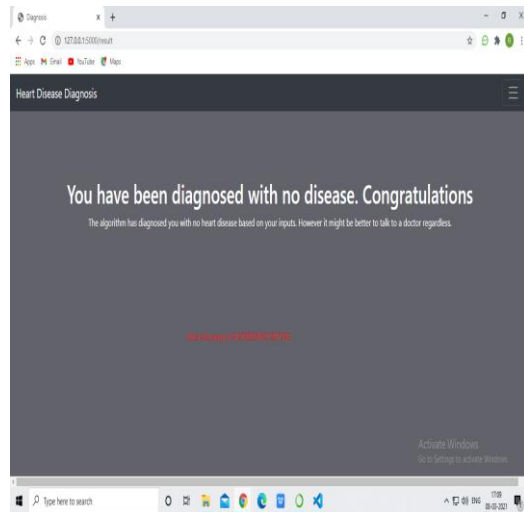
3.1. Data Collection

In this phase the data is collected to analyze the best proposed method. UNSW-NB15 dataset is used in our research for identifying the performance of proposed methods.

3.2. Data Preprocessing

It is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.





The collected data is preprocessed in this phase. It consists of Data Transformation, Data Normalization and Feature Selection.

3.3. Data Integration:

Data with different representations are put together and conflicts within the data are resolved.

3.4. **Data Transformation:** Data is normalized, aggregated and generalized.

3.5. **Data Reduction:** This step aims to present a reduced representation of the data in a data warehouse.

Here the raw data in the heart disease data is cleaned and the metadata is appending to it by removing the things which are converted to the integer. So, the data is easy to train. In this pre-processing, we first load the metadata into this and then this metadata will be attached to the data and replace the converted data with metadata. Then this data will be moved further and remove the unwanted data in the list and it will divide the data into the train and the test data. For this splitting of the data into train and test we need to import `train_test_split` which in the scikit-learn this will help the pre-processed data to split the data into train and test according to the given weight given in the code.

The division of the test and train is done in 0.2 and 0.8 that is 20 and 80 percent respectively. Here SVM is used for these process.

IV. CONCLUSION AND FUTRUE ENHANCEMENT

Heart diseases are complicated and take away lots of lives every year .When the early symptoms of heart diseases are ignored, the patient might end up with drastic consequences in a short span of time. Sedentary lifestyle and excessive stress in today's world have worsened the situation. If the disease is detected early then it can be kept under control. However, it is always advisable to exercise daily and discard unhealthy habits at the earliest.In the above paper we have studied various classification algorithms that can be used for classification of heart disease databases also we have seen different techniques that can be used for classification and the accuracy obtained by them. This investigation tells us about dissimilar technologies that are used in dissimilar papers with dissimilar count of attributes with different accuracies depending on the tools designed for execution.

REFERENCES

- [1] Prediction of Heart Disease Using Machine Learning, Aditi Gayhane ;Gouthami Kokkula ;Isha Pandya ;Prof. Kailas Devadkar,2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA).
- [2] Heart Disease Prediction using EvolutionaryRule Learning,Aakash Chauhan ;Aditya Jain ;Purushottam Sharma ;Vikas Deep, 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT).

- [3] Blood Viscosity based Heart Disease Risk Prediction Model in Edge/Fog Computing, R Latha ; P Vetrivelan, 2019 11th International Conference on Communication Systems & Networks (COMSNETS).
- [4] An Investigation of Heart Disease Prediction Techniques, G. Shanmugasundaram ; V. Malar Selvam ; R. Saravanan ; S. Balaji, 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN).
- [5] Newfangled Approach for Early Detection and Prevention of Ischemic Heart Disease using Data Mining, Dhara B. Mehta ; Nirali C. Varnagar, 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI).
- [6] Real-time machine learning for early detection of heart disease using big data approach Abderrahmane Ed-Daoudy ; Khalil Maalmi, 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS).
- [7] Predictive Analysis of Rapid Spread of Heart Disease with Data Mining, Radhanath Patra ; Bonomali Khuntia, 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT).
- [8] Heart Disease Prediction System using CART-C, Priyal Chotwani ; Asmita Tiwari ; Vikas Deep ; Purushottam Sharma, 2018 International Conference on Computer Communication and Informatics (ICCCI).
- [9] Automated Prediction of Heart Disease Patients using Sparse Discriminant Analysis, K. M. Zubair Hasan ; Shourob Datta ; Md Zahid Hasan ; Nusrat Zahan, 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE).
- [10] Non-Laboratory-Based Risk Factors for Automated Heart Disease Detection, H. Mai ; T. T. Pham ; D. N. Nguyen ; E. Dutkiewicz, 2018 12th International Symposium on Medical Information and Communication Technology (ISMICT).