HEART FAILURE PREDICTION USING MACHINE LEARNING

Abhishek Pinjarkar[1], Rajeshwari Hande[2], Priyanka Malode[3], Sakshi Nagdive[4], Dr.V.K.Taksande[5]

B.E Students[1,2,3,4], Professor[5]
Department of Electronics & Telecommunication Engineering
Priyadarshini College of Engineering, Nagpur, Maharashtra, India

ABSTRACT

In this modern era people are very Busy and working hard in order to satisfying their materialistic needs and not able to spend time for themselves which leads to physical stress and mental disorder. There are also reports that heart suffer because of global pandemic corona virus. Inflammation of the heart muscle can be caused by Corona virus. Thus heart disease is very common now a day's particularly in urban areas because of excess mental stress due to corona virus. As a result heart disease has become one of the most important factors for death of men and women in the so called material world. It has emerged as the top killer that has affected both urban and rural population. CAD (Coronary artery disease) is one of the most common types of heart disease. In the medical field predicting the heart disease has become a very an complicated and challenging task, requires patient previous health records and in some cases they even need genetic information as well. So, in this contemporary life style there is an urgent need of a system which will predict accurately the possibility getting heart disease. Predicting a heart disease in early stage will save many people's life. There were many heart disease prediction systems available at present, the authors have been researched well and proposed different classification and prediction algorithms but each one has its own limitations. The main objective of this paper is to overcome the limitations and to design a robust system which works efficiently and will able to predict the possibility of heart failure accurately.

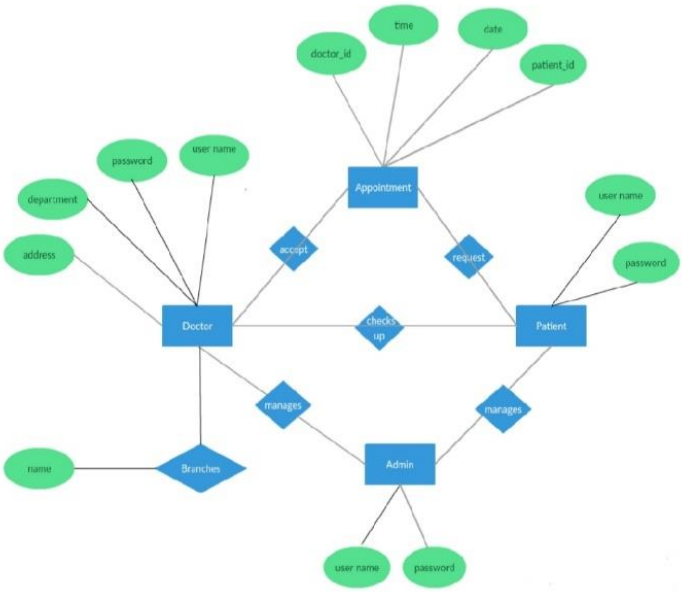
I. INTRODUCTION

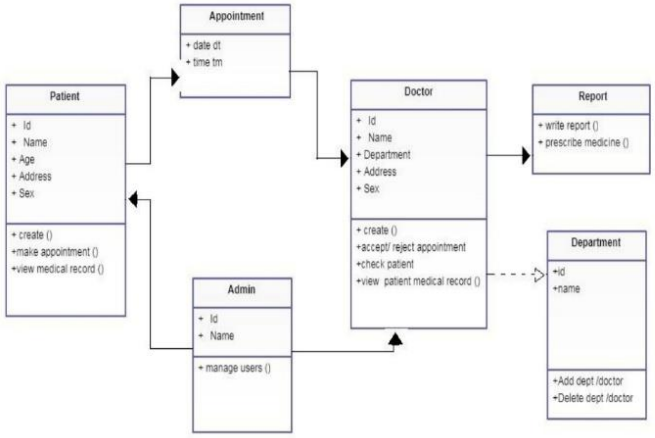
Heart is the essential organ of the human body. Human life is completely depends on the efficient working of heart. The heart pumps blood over blood vessels to the different body parts of the body, with enough oxygen and other essential nutritional components that are required for smooth functioning of the body. Healthy heart leads a healthy life. But, in today's world heart disease has become vital cause of death for both men and women in the world. Corona virus causes inflammation of the heart muscle leading to heart failure. Experimental evidences suggest that 1 in every five patient having heart injury due to Corona virus irrespective of respiratory symptoms. The diagnostic method for heart failure is generally based on the patient's medical and family histories, a physical examination, and test results. Identification of heart disease is difficult because of various risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate and many other factors. Heart failure is a serious manifestation or late stage of various heart diseases. Generally, heart Failure would result in insufficient cardiac ejection. Since heart disease is the most common, hence there is an urgent need of very accurate and early detection of heart disease which can save the life of many patients. There are many scanners to detect the heart disease but predicting a heart disease before it attacks can save many people. Machine learning holds great potential for the healthcare industry to enable health systems to use the Data to analyze and predict the disease (identify) by using machine learning. We can analyze the data and predict the disease and this reduces the cost, and saves the time.

II. METHODOLOGY

This paper shows the analysis of various machine learning algorithms, the algorithms that are used in this paper are K nearest neighbors (KNN), Logistic Regression and Random Forest Classifiers which can be helpful for practitioners or medical analysts for accurately diagnose heart disease. This paperwork includes examining the journals, published paper and the data of cardiovascular disease of the recent times. Methodology gives a framework for the proposed model. The methodology is a process which includes steps that transform given data into recognized data patterns for the knowledge of the users. The proposed methodology includes steps, where first step is referred as the collection of the data than in second stage it extracts significant values than the 3rd is the pre-processing stage where we explore the data. Data pre-processing deals with the missing values, cleaning of data and normalization depending on Algorithms used [15]. After pre-processing of data, classifier is used to classify the pre-processed data the classifier used in the proposed model are KNN, Logistic Regression, Random Forest classifier. Finally, the proposed model is undertaken, where we evaluated our model on the basis of accuracy and performance using various performance metrics.

III. MODELING AND ANALYSIS





The class diagram is chosen to explain the design phase of the system. A class diagram describes classes of the system, attributes, and operations and relationships of the classes in a better way. We can also say that class diagrams are used to justify the structure or behaviour of use cases of the system. Class diagrams best explain the conceptual model of the system in terms of entities and their relationships. The class diagram looks like a shape of a rectangle, comprising three compartments stacked vertically. The first top box comprises the class name, the second middle box contains the attributes of the class and third the last box contains the methods or functions performed by that class. The first compartment /box of the name is compulsory while rest of the two can be omitted to simplify the diagram. So, in any class diagram first compartment must be drawn while the second two compartments are optional. The class "patient" contains multiple parameters (such as id, name, age, address), which depict the information of all the registered patients. The user class also contains the methods performed by these users such as get appointment, view/ create own medical record etc. In the same way, the class "doctor" has the parameters id, name, department, address possessing all the required information of the users registered as a doctor on to the system. Methods include accept/reject the appointment, check the patient, view a medical record of any patient etc. These methods are the functions performed by the users registered as a doctor on the system. The class "appointment" has the parameters of date and time, explaining what time or day patient user has requested for the appointment to the doctor. The class "department" has the parameters id and name and methods include add/ delete doctor and add or delete department. Every doctor user must belong to any department class. The "report" is another class containing methods like write report or prescribes medicine. In the end, class "admin" contains the parameters like id and name and methods of this class include manage users.

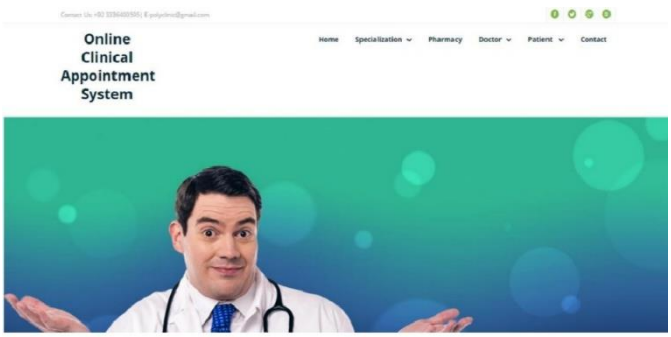
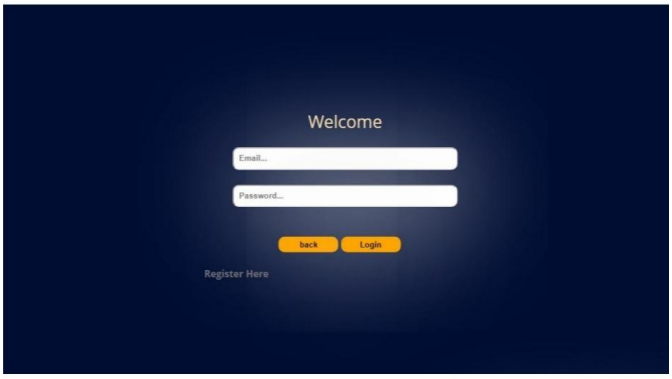
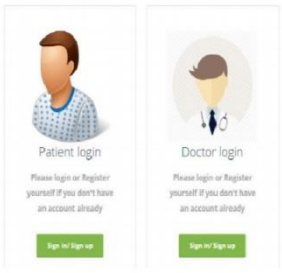


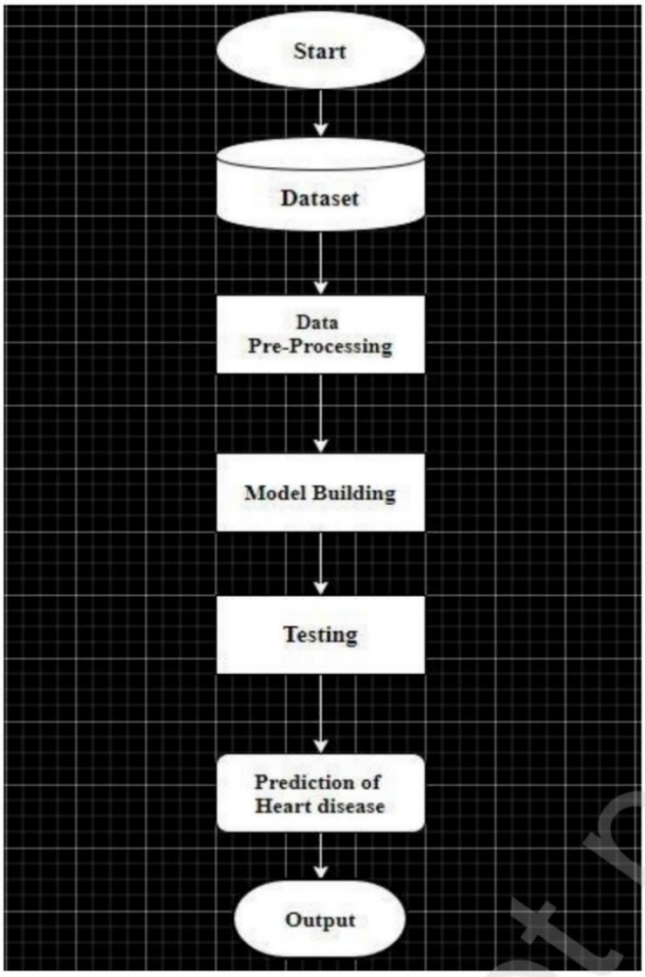
Figure 7: Layout part 1



h 1:User Sign in/ up page



Architectural model of proposed system



Implementation

In this portion of our report we are going to discuss how we prepared or designed the whole System

Dataset

We found our data set that has been used in our research from kaggle (https://www.kaggle.com/Ronitf/heart-disease uci/version/1). The dataset that we used in our thesis has in total 14 columns and 303 rows. First 13 of those columns are the features that we will be using later on in order to predict the final column 'diagnosis' which will tell us if the patient is going to be affected by heart disease or not. The 303 rows represents data of 303 patients that we found from the dataset.

Preprocessing

Before we start let us give a brief information about what data pre-processing actually is. Data Pre-processing may be a data processing technique that involves re-modeling data into a lucid format. Real-world data is often incomplete, inconsistent and lacking in certain behaviours or Trends and is likely to contain many errors .Data pre-processing may be a tried technique of partitioning such problems. Data pre-processing prepares raw data for further processing. Data pre-processing is used in database-driven applications such as customer relationship management and rule-based applications. For our thesis we are using standard scaler from the sklearn library for pre-processing our data. We choose this one over the many other ones because it suits very well with our system.

Load data

We created an array called col names and put down all our columns on that array. Then we read the csv file also known as the dataset file.

Analyze features

In this section we are going to distribute the target value is vital for choosing appropriate accuracy metrics and consequently properly assess different machine learning models. First of all we are going to count values of explained variable otherwise known as the determining variable which is going to give us the prediction of a patient being affected by heart disease or not. Second of all we are going to separate numeric features from categorical features. Then we are going to show the relation between the categorical features in various plots and try to figure out or rather observe the influence of those categorical features in the actual determining variable "diagnosis".

Modeling and predicting with machine learning

The main goal of the entire project is to predict heart disease occurrence with the highest accuracy. In order to achieve this we will test several classification algorithms. This section includes all results obtained from the study and introduces the best performer according to accuracy metric. I have chosen several algorithms typical for solving supervised learning problems throughout classification methods. First of all, let's equip ourselves with a handy tool that benefits from the cohesion of SciKit learn library and formulate a general function for training our models. The reason for displaying accuracy on both, train and test sets, is to allow us to evaluate whether the model over fits or under fits the data (so-called bias/variance trade off). Then we are going to split the data then test and train them in the ratio of 70:30. Then we are going to create a model where we are going to run all our algorithms.

Finding the result

At the end we are going to create a summery table where we are going to show the different accuracy percentage of different algorithms. It is worth to emphasize that in most cases hyper parameter tuning is essential to achieve robust results out of these techniques. By producing decent results, simpler methods proved to be useful as well. Machine learning has absolutely bright future in medical field. Just imagine a place where heart disease experts are not available. With just basic information about a certain patient's medical history, we may quite accurately predict whether a disease will occur or not .

IV. RESULTS AND DISCUSSION

This research work use the data set consists of various clinical attributes like patient Age, Sex, Chest pain, Fbs and etc . Next step is to divide the data set into two sets, training set is 70% and testing set is 30%. Training set is used to construct the model and testing to evaluate the correctness of the model. By using this method this research work could able to predict whether a patient is suffering from heart disease or not (0 is absence and 1 is presence). By using machine learning we could able to predict the heart disease.

V. CONCLUSION

Heart is the most essential organ of the human body and day by day the loss of human life is increasing exponentially due to heart failure. It was experimentally found that the global pandemic Corona Virus causes heart injury among a lot of patients. Hence there is an urgent need for research to focus into the causes for heart failure and to design a robust prediction system to detect at early stage so that loss of life can be avoided. Even though there were many heart diseases prediction systems available at present but each one has its own limitations. The main objective of this research work is to overcome the difficulty faced by other researchers and to build a robust system which works efficiently and will able to predict accurately the possibility of heart attack at very early stage. This research work could able to design a very robust and accurate model to predict the possibility of heart Failure in the current scenario. By using the machine learning we could able to predict the heart disease at earlier stage.

ACKNOWLEDGEMENT

Priyadarshini College of Engineering is a well-established & renowned institute and follows a goal of creating technocrats and brings it into reality. We wish to avail this opportunity to express our sincere Thanks to our guide DR.V.K. TAKSANDE who continuously supervised our work with utmost care and zeal. He has always guided us in our endeavor to present our Project on “HEART FAILURE PREDICTION USING MACHINE LEARNING ”.

VI REFERENCES

[1]. B. Jin, C. Che, Z. Liu, S. Zhang, X. Yin and X. Wei, “Predicting the Risk of Heart Failure with HER Sequential Data Modeling,” in IEEE Access, vol. 6, pp. 9256-9261,2018.

[2]. S. Mohan, C. Thirumalai and G. Srivastava, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques,” in IEEE Access, Vol. 7, pp. 81542-81554, 2019.

[3]. B. Wang et al., “A Multi-Task Neural Network Architecture for Renal Dysfunction Prediction in Heart Failure Patients With Electronic Health

Records," in IEEE Access, vol. 7, pp. 178392-178400,2019.

[4]. S.Adithya Varun, G.Mounika, Dr. P.K. Sahoo, K. Eswaran, "Efficient System for Heart disease prediction by applying Logistic regression. Ijcst Vol 10, issue 1, march 2019.

[5]. Rajalakshmi, S., & Madhav, K. V., A Collaborative Prediction of Presence of Arrhythmia in Human Heart with Electrocardiogram Data Using Machine Learning Algorithms with Analytics. 278 287. Doi:10.3844/jcssp.2019.278.287, 2019.

[6]. Purushottam, Saxena, K., & Sharma, R, "Efficient Heart Disease Prediction System". Procedia Computer Science Published by Elsevier 85, 962–969. Doi:10.1016/j.procs.2016.05.288.

[7]. A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor and R. Nour, "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection," in IEEE Access, vol. 7, pp. 180235-180243, 2019.

[8]. H. Benjamin Fredrick David and S. Antony Belcy, "heart disease Prediction using data mining techniques". ICTACT JOURNAL, oct 2018, Volume: 09, issue: 01, DOI: 10.21917/ijsc.2018.0253.

[9]. L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed and J. A. Khan, "An Automated Diagnostic System for Heart Disease Prediction Based on ${\chi^{2}}$ Statistical Model and Optimally Configured Deep Neural Network," in IEEE Access, vol. 7, pp. 34938-34945, 2019.