A Review on Human Heart Prediction System Using Machine Learning

[1]Sangya Ware, [2]Mrs. Shanu .K. Rakesh

[1]M.Tech Scholar, [2]Assistant Professor,
CSE Department
Chouksey Engineering College, Bilaspur, Chhattisgarh, India

***Abstract*: Heart disease is a deadly disease that large population of people around the world suffers from. When considering death rates and large number of people who suffers from heart disease, it is revealed how important early diagnosis of heart disease. Traditional way of diagnosis is not sufficient for such an illness. Developing a medical diagnosis system based on machine learning for prediction of heart disease provides more accurate diagnosis than traditional way. In this, a heart disease prediction system which uses SVM algorithm is proposed. 13 clinical features were used as input for the SVM and then the SVM was trained to predict absence or presence of heart disease with accuracy of 95%.**

***Keywords*: SVM, Heart Diseases Prediction System, Data mining**

I INTRODUCTION

Heart disease has created a lot of serious concerns among researches; one of the major challenges in heart disease is correct detection and finding presence of it inside a human. Early techniques have not been so much efficient in diagnosis [2].There are various medical instruments available in the market for predicting heart disease but there are two are very much expensive and secondly, they are not efficient enough to be able to calculate the chance of heart diseases. According to latest survey conducted by WHO, the medical professionals are able to correctly predict only 67% of heart diseases [3]. So, there is a need to find better and efficient approach to diagnose heart diseases at early stage. With advancement of computer science in different research areas including medical sciences, this has been made possible. As application areas of computer science varies from meteorology to ocean engineering and medical sciences. In last decade, artificial intelligence has gain momentum because of the improved technologies and machine learning algorithms. Machine learning implementations are applicable to vast research areas including depression predictions, image and speech recognition, medical sciences, genomics and natural language processing etc. A machine-learning system is trained rather than explicitly programmed. It is presented with many examples relevant to a task, and it finds statistical structure in these examples that eventually allows the system to come up with rules for automating the task [4]. Machine learning could be a better choice for achieving high accuracy for detection of heart diseases. This survey paper is dedicated for wide scope survey in the field of machine learning technique in prediction of heart disease. Later part of this survey paper will discuss about various machine learning algorithms and their relative comparison based on various performance metrics like F1 Score, specificity, accuracy etc

II LITERATURE REVIEW

Different researchers have contributed for the development of this field. Predication of heart disease based on machine learning algorithm is always curious case for researches recently there is a wave of papers and research material on this area. Our goal in this chapter is to bring out all state of art work by different authors and researchers.

Marjia Sultana, Afrin Haider and Mohammad Shorif Uddin [5] have illustrated about how the datasets available for heart disease are generally a raw in nature which is highly redundant and inconsistent. There is a need of pre-processing of these datasets; in this phase high dimensional data set is reduced to low data set. Different methods have their own merits and demerits in work done by M.A. Jabbar, B.L Deekshatulu , Priti Chndra [6], an optimisation of feature has been done to achieve higher classification efficiency in Decision Tree .It is an approach for early detection of heart disease by utilizing variety of feature.These kind of approach can also be utilize for other sphere of research. Other than decision tree various other approach where adopt for achieving the goal of perfect detection of heart disease in human Yogeswaran Mohan et.al [7] have collected raw data form EEG device and used to train neural network for pattern classification . Here input output are depressive and non depressive categories in the hidden layer scaled conjugate gradient algorithm is used for training to achieve efficient result. authors have got efficiency up to 95% with help of trained neural network watching the success of neural network researches working in the domain of SVM have used this technique to classify and achieve more better result in case where the feature vector are multi dimensional and non linear these method defeated all other existing quantum contemporary techniques because it has capability to work under dataset of high dimensionality.

RELATED WORK SECTION

There are some existing studies that have addressed the problem of heart disease prediction [8][9]. The method proposed by [8] integrates web-based system with machine learning and they have methods like Decision trees, Naïve Bayes and neural networks. One of the limitations of their work is the limited size of dataset and also, they have used just three data mining techniques. While,

in our work we plan to use SVM and Random forest. Also, our dataset is more comprehensive and our work could be extended to multiple datasets available in our set. The method proposed by [9] implements genetic algorithm to do feature selection related to heart disease and they have implemented Decision tree, Naïve Bayes and classification using clustering. We are planning to use more sophisticated learning algorithms**.**

SUMMARY

Machine Learning comes under the umbrella of artificial intelligence (AI). It enables the computers to learn without programming it explicitly. Machine learning aims at developing computer programs that can change whenever exposed to new sets of data. The machine learning algorithms are classified as Supervised or Unsuperv

III PROBLEM IDENTIFICATION

Existing work There are some existing studies that have addressed the problem of heart disease prediction [8][9]. The method proposed by [8] integrates web-based system with machine learning approach and they have methods like Decision trees, Naïve Bayes and neural networks. One of the limitations of their work is the limited size of dataset and also, they have used just three data mining techniques. Some disadvantages of decision tree are Data may be over fitted or over classified. And only one attribute at a time is tested for making decision. Disadvantage of Naive Bayes Classifier is that there is a loss of accuracy.

Limitation of Existing work

Many hospital information systems are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. They can answer simple queries like "What is the average age of patients who have heart disease?","How many surgeries had resulted in hospital stays longer than 10 days?", "Identify the female patients who are single, above 30 years old, and who have been treated for cancer." However, they cannot answer, complex queries like "Identify the important predictors that increase the length of hospital stay", "Given patient records on cancer, should treatment include chemotherapy alone, and "Given patient records, predict the probability of patients getting a heart disease."

IV PROPOSED METHODOLOGY

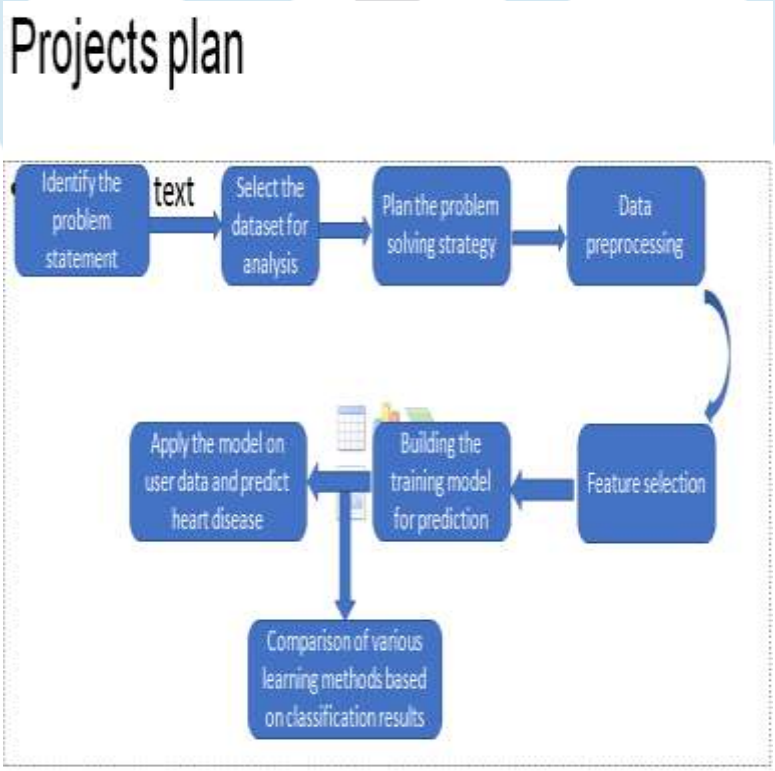

Figure 4.1. Project plan

Figure 1 shows the project plan where first step is to identify the problem statement, where the problem is identified weather the person has heart disease or not. The second step is to select the dataset for analysis, where we are going to use two data set. First one is Hungarian dataset and second one is Switzerland data set. The third step is plan the problem solving strategy, for this we are using two classifiers one is SVM i.e. support vector machine and Random forest for comparison. The next step is Data preprocessing. Data preprocessing is a process of removing noisy and missing data from the data set. The next step is feature selection where we choose 4-5 relevant features. Suppose there are 14 features if we use all the 14 features there is a possibility of

poor result. But if we use only 4-5 relevant features it will give us good result. The next step is building a training model for prediction in this we will build our model by publicly available dataset such as sex, age etc. The next step is Apply the model on user data and predict heart disease. The last step is comparison of various learning methods based on classification results. In the last step it will compare both the classifiers SVM and Random forest for better choice.

DATA SET USED

The heart disease dataset is a very well studied dataset by researchers in machine learning and is freely available at the UCI machine learning dataset repository https://archive.ics.uci.edu/ml/datasets/Heart+Disease.Though there are 4 datasets in this, I have used the Cleveland dataset. The dataset has 76 attributes and 303 records. However, only 13 attributes are used for this study & testing as shown in Table 1.

Table 1: Selected Heart Disease Attributes

Name	Type	Description
Age	Continuous	Age Age in years
Sex	Discrete	0 = female 1 = male
Cp	Discrete	Chest pain type: 1 = typical angina, 2 = atypical angina, 3 = non-anginal pain 4 =asymptom
Trestbps	Continuous	Resting blood pressure (in mm Hg)
Chol	Continuous	Serum cholesterol in mg/dl
Fbs	Discrete	Fasting blood sugar>120 mg/dl: 1-true 0=False
Exang Continuous Maximum heart rate achieved	Discrete	Exercise induced angina: 1 = Yes 0 = No
Thalach	Continuous	Maximum heart rate achieved
Old peak ST	Continuous	Depression induced by exercise relative to rest
Slope	Discrete	The slope of the peak exercise segment : 1 = up sloping 2 = flat 3 = down sloping
Ca	Continuous	Number of major vessels colored by fluoroscopy that ranged between 0 and 3.
Thal	Discrete	3 = normal 6 = fixed defect 7= reversible defect
Class	Discrete	Diagnosis classes: 0 = No Presence 1=Least likely to have heart disease 2= >1 3= >2 4=More likely have heart disease

We are planning to use supervised machine learning called SVM. We will build a model by training on publicly available dataset for heart disease detection. We will optimize the analysis on standard metrics (F1 score, accuracy). The optimized model will then be tested on the user data (features) to predict if the user has a heart disease or not.In addition, we also plan to compare the results from SVM with the results from another classifier- Random forest. This will help us get an insight of which algorithm is a better choice.As a part of data preprocessing, we are planning to remove noisy and missing data to improve data consistency and quality.

SCREEN SHOT

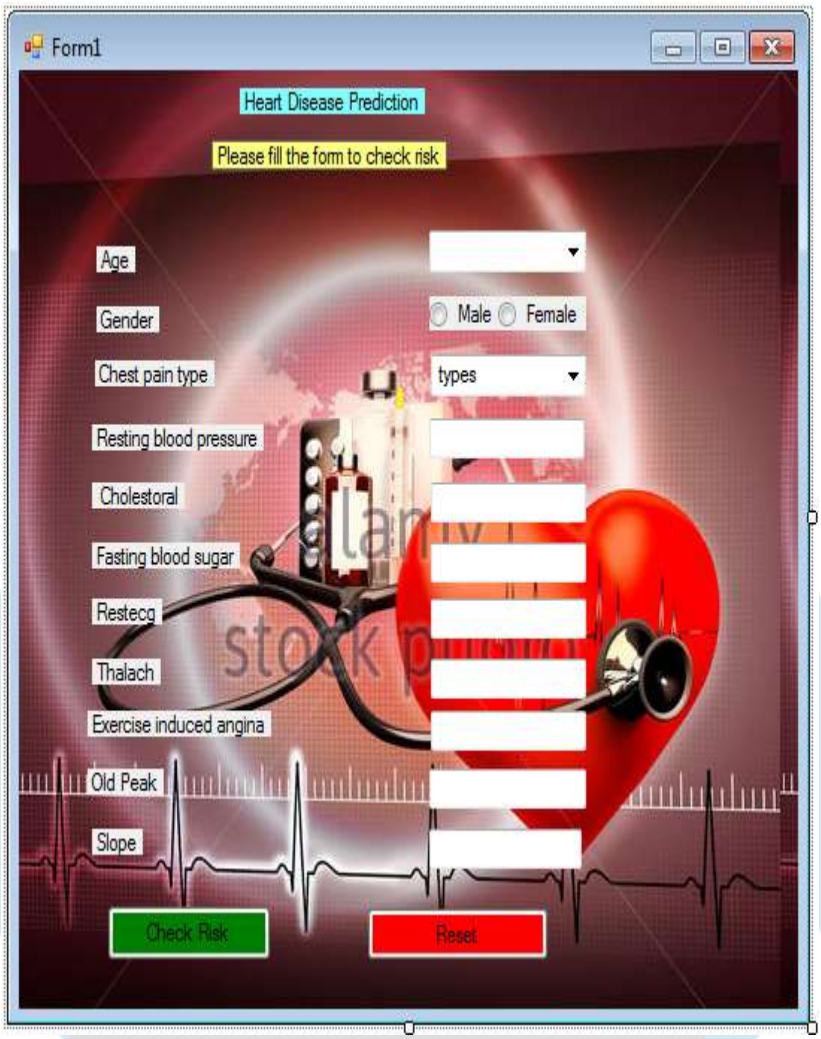

Figure 4.2 Screen shot of heart disease prediction

Figure 5 shows the screenshot of heart disease prediction. User will input the details like Age, gender, chest pain type, Resting blood pressure, cholesterol, fasting blood sugar, Restecg, thalach, Exercise induced angina, Old peak, slope. User will fill the form to check whether they are suffering from heart disease or not. There are two buttons one is check risk, by clicking check risk user can know whether they are suffering from heart disease or not and another one is reset.

V CONCLUSION AND FUTURE SCOPE

CONCLUSION

Heart disease is the number one killer according to World Health Organization (WHO) statistics. Millions of people die every year because of heart disease and large population of people suffers from heart disease. Prediction of heart disease early plays a crucial role for the treatment. If heart disease could be predicted before, lots of patient deaths would be prevented and also a more accurate and efficient treatment way could be provided. A need to develop such a medical diagnosis system arises day by day. The important key points of such medical diagnosis systems are reducing cost and obtaining more accurate rate efficiently. Developing a medical diagnosis system based on machine learning for prediction of heart disease provides more accurate diagnosis than traditional way and reduces cost of treatment.

The dataset needed tremendous efforts for cleaning and had a lot of noisy and missing data. This would certainly improve the data quality and hence would improve the classification accuracy. Based on the feature ranking from the feature selection method, it would be better to extract only relevant features and consider them for further analysis. Feature selection strategy has improved results base on the existing literature. We have proposed Supervised Learning Algorithm for finding the risk of heart disease of a patient using the profiles collected from the patients. We can detect heart related problems by using the model trained from a publicly available dataset. We believe only a marginal success is achieved in the creation of predictive model for heart disease patients and hence there is a need for combinational and more complex models to increase the accuracy of predicting the early onset of heart disease.

EXPECTED OUTCOME

Results

Feature Rank	Feature Name
1	cholestrol
2	age
3	thalach
4	Resting blood pressure
5	Chest pain
6	Old peak
7	slope
8	Gender
9	Exercise induced angina
10	Rest ecg
11	Fasting blood sugar

Figure 5.1 Expected outcome

Figure 6 shows the expected output where we have selected some of the relevant features from the attributes for better results.

FUTURE SCOPE

In future, we plan to extend our research by identifying and including more features. We also plan to use more classification methods like deep learning etc.

REFERENCES

[1] William Carroll; G. Edward Miller, "Disease among Elderly Americans: Estimates for the US civilian non institutionalized population, 2010," Med. Expend. Panel Surv., no. June, pp. 1–8, 2013.

[2] M. A. Jabbar, P. Chandra, and B. L. Deekshatulu, "Prediction of risk score for heart disease using associative classification and hybrid feature subset selection," Int. Conf. Intel. Syst. Des. Appl. ISDA, pp. 628–634, 2012.

[3] V. Kirubha and S. M. Priya, "Survey on Data Mining Algorithms in Disease Prediction," vol. 38, no. 3, pp. 124–128, 2016.

[4] Chollet, Francois. Deep learning with python. Manning Publications Co., 2017.

[5] M. Sultana, A. Haider, and M. S. Uddin, "Analysis of data mining techniques for heart disease prediction," 2016 3rd Int. Conf. Electr. Eng.Inf. Commun. Technol.iCEEiCT 2016, 2017.

[6] M. Akhil, B. L. Deekshatulu, and P. Chandra,"Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm," Procedia Technol., vol. 10, pp.85–94, 2013.

[7] S.Kumra, R. Saxena, and S. Mehta, "An Extensive Review on Swarm Robotics," pp. 140–145, 2009

[8] Palaniappan, Sellappan, and Rafiah Awang. "Intelligent heart disease prediction system using data mining techniques." Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on. IEEE, 2008.

[9] Anbarasi, M., E. Anupriya, and N. C. S. N. Iyengar. "Enhanced prediction of heart disease with feature subset selection using genetic algorithm." International Journal of Engineering Science and Technology 2.10 (2010): 5370-5376.

[10] Dua, D.and karra Taniskidou,E.(2017), Hungarian Institute