Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.199

IJCSMC, Vol. 8, Issue. 5, May 2019, pg.119 – 125

PREDICTING THE PRESENCE OF HEART DISEASE USING MACHINE LEARNING

Akshay Jayraj Suvarna[1]; Arvind Kumar M[2]; Ajay Billav[3]; Muthamma K M[4]; Asst. Prof. Gadug Sudhamsu[5]

[1]Department of CSE, School of Engineering and Technology - Jain University, Bangalore, India
[2]Department of CSE, School of Engineering and Technology - Jain University, Bangalore, India
[3]Department of CSE, School of Engineering and Technology - Jain University, Bangalore, India
[4]Department of CSE, School of Engineering and Technology - Jain University, Bangalore, India
[5]Department of CSE, School of Engineering and Technology - Jain University, Bangalore, India
[1] akshaysuvarna7338@gmail.com, [2] arvindkumar771377@gmail.com, [3] ajaybillav95@gmail.com, [4] ankithamuthamma14@gmail.com, [5] g.sudhamsu@gmail.com

Abstract: Earlier techniques in predicting the heart disease have not been much accurate in terms of prediction and analysis. The advancement in Computer Science has brought vast opportunities in different areas. One of it is machine learning which is widely utilized in different domains. Machine learning techniques are used to predict the medical condition at an early stage of human life. We will use Heart Disease related dataset to carry out our experiments. In this system the most known ensemble learning algorithms Random Forest and Support Vector Machine will be applied to create a classifier model which will predict disease with higher performance and accuracy.

Keywords– Machine Learning, Random Forest, Support Vector Machine, Dataset, Heart diseases.

I. INTRODUCTION

Heart is one of the most important organs in the human body and life is dependent on component functioning of heart. On the off chance that there is some issue in the siphoning activity of the heart, the fundamental organs of the body like the brain and kidneys are antagonistically influenced. On the off chance that the working of the heart stops, at that point the demise of the individual happens inside couple of minutes. In this day and age, huge number

of populace is experiencing various sorts of heart sicknesses and the tally of patients experiencing and dying from such diseases in bolstering consistently.

Heart disease is currently the leading cause of death in the world. The survey says 70% mortality rate is due to heart related problems. The term heart diseases implies various issues that influence the ordinary working of circulatory framework, which comprises of heart and veins. There are various classifications of heart ailments like cardiovascular infection in which the heart and veins are influenced and because of which the blood isn't siphoned and coursed appropriately all through the body. In the event that the coronary illness is identified at beginning time and the patient is given proper and sufficient treatment, at that point it tends to be relieved totally and furthermore the expense of the treatment can be decreased essentially. So there is a need to build up an expectation framework to identify the nearness or nonattendance of heart diseases in the patient with higher exactness.

Machine learning algorithms can be used for heart disease prediction systems. Applying machine learning is a key approach to utilize large volumes of available Heart-related data. Machine learning is of great concern when it comes to diagnosis, management and other related clinical administration aspects. Various machine learning techniques include ensemble classifiers can be used in improving prediction accuracy. Machine learning techniques helps in identifying the data and automatically makes the predictions. Machine learning algorithms like Support Vector Machine and Random Forest will be used in the proposed system. Hence in the framework of this study, efforts are made to predict the presence of heart diseases using random forest and support vector machine algorithm.

II. FORMER WORK

Many people have given their ideas and imposed new systems, each overcoming the drawbacks of the previous one. In [1] they have proposed a system in which the dataset is taken from data mining repository of the university of California. In this dataset a total of 14 attributes are such as age, sex, blood sugar, cholesterol, etc. are used to diagnose the presence of heart disease. This methodology uses PSO and CPSO for prediction of presence of heart disease within a person. PSO stands for particle Swarm Optimization. This method is used by group of birds to find food or for interaction with other birds. In this case birds move around in separate groups in search for food and the group which finds best location for food is termed as optimum search. In this algorithm when the partial needs to decide where to move next, own experience based on past positions are taken. This implementation methodology contains 297 datasets out of which 50% is used for training and another 50% for testing dataset.

In the system proposed in [2], they have used data mining techniques for diagnosis and treatment of heart diseases. This paper recognizes breaks in the investigation on heart diseases examination and treatment and recommends a model to efficiently close those ruptures to decide whether spreading information mining methods to coronary illness treatment information can give as reliable execution as that accomplished in diagnosing heart diseases. In [3], information order depends on directed machine learning calculations which result in exactness, time taken to assemble the calculation. Tanagra instrument is utilized to arrange the information and the information is assessed utilizing 10-overlap cross approval and the outcomes are compared.

In the system proposed in [4], For data pre-processing and active decision making One Dependency Augmented Naïve Bayes classifier (ODANB) and naive credal classifier 2 (NCC2) are used. Utilizing medical profiles, for example, age, sex, pulse and glucose it can anticipate the probability of patients getting a coronary illness. It empowers critical learning, for example designs, connections between medicinal components identified with coronary illness, to be set up.

III. METHODOLOGY

In the proposed system, methodology comprises of two segments i.e. Admin's side and User's side. Both has their different flow of work, each of them is explained below.

A. Admin's side

The admin is the one who develops the entire methodology and train the system for future automation of prediction of heart disease. The system here is trained and tested using two different algorithms i.e. Support Vector Machine and Random Forest. Both of these algorithms have different approach for training and testing the system and both of them will give us results with different accuracy.

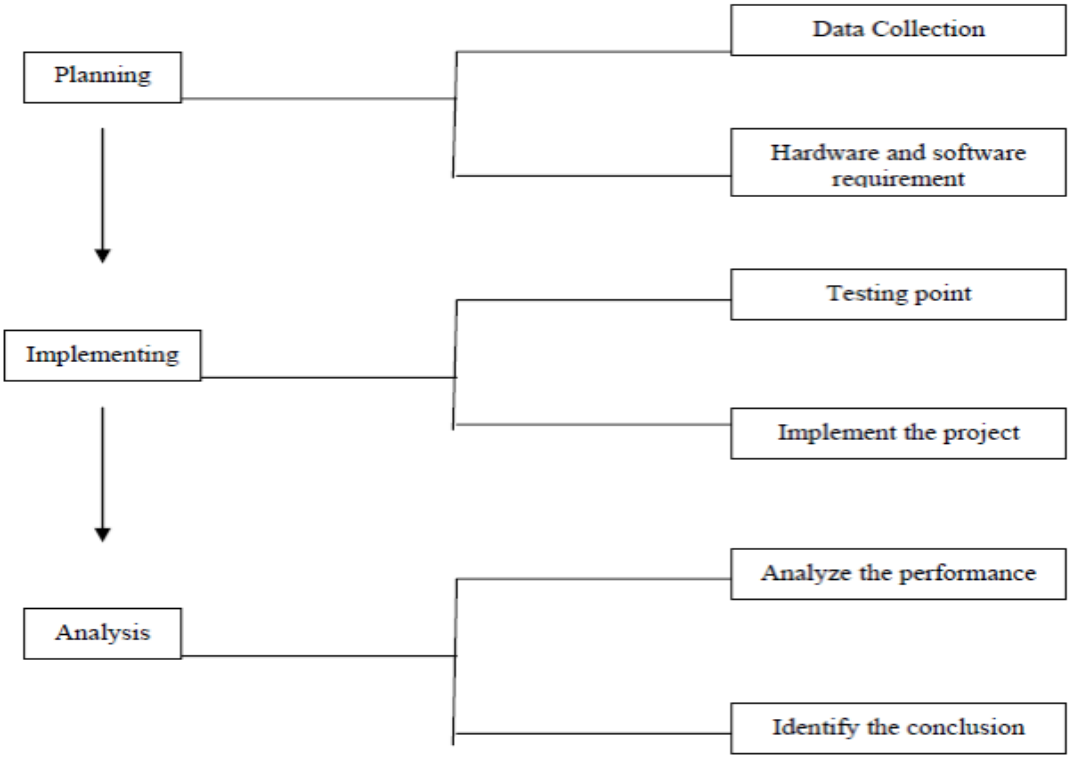

Figure 1. Steps Of Methodology

1) RANDOM FOREST:

The random forest is a model settled on up of numerous choice trees. As opposed to just averaging the expectation of trees (which we could call a "woodland"), this model uses two key concepts that gives it the name random:

- When building branched trees, random sampling of data points.
- While splitting nodes, random subsets of the considered attributes.

2) SUPPPORT VECTOR MACHINE (SVM):

A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be utilized for both classification and regression purposes. SVMs are all the more generally utilized in classification problems and in that capacity, this is the thing that we will concentrate on in this post. SVMs evaluates a hyperplane that can divide the dataset into two categories as shown in the Fig. 2.

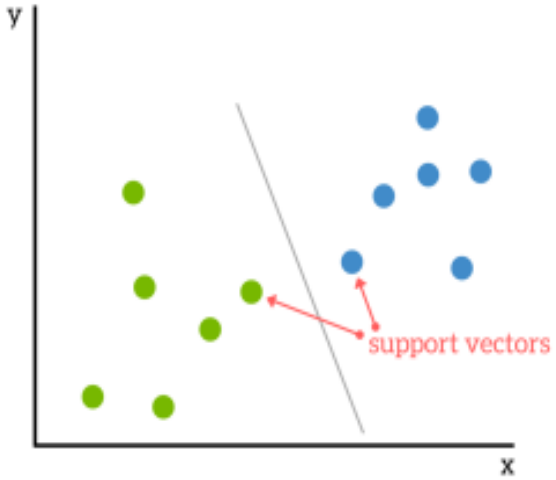

Figure 2. SVM Hyperplane

B. User's side

The user is one who will be using the system after it is trained by the admin. After the system is trained, the user will be asked to sign up if he/she is using the model for the first time or login to the system if he/she is an existing user. After successful completion of this step, the user will be asked to give details of the attributes which will be needed to predict whether the user has heart disease or not. On the basis of values given by the user for all the attributes both the algorithms use their unique approach for the prediction of heart disease. The algorithm giving better accuracy for prediction is taken into consideration.

IV. SYSTEM IMPLEMENTATION

The proposed system is implemented in two segments – internal and external. The system architecture is shown below.

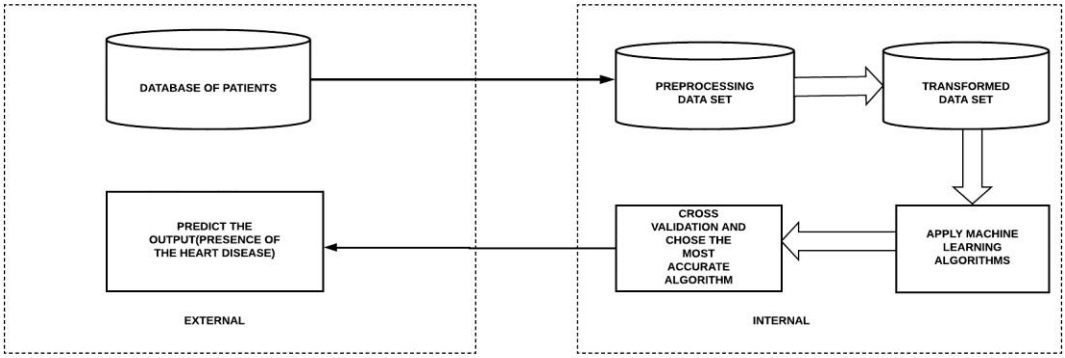

Figure 3. System Architecture

A. INTERNAL:

The internal architecture includes 4 major steps:

1) Pre-processing data set:

The process in which the raw data sets are converted into a clean data sets is called pre-processing. The data which has been collected from the user might not be in an accurate format. It standardizes the data features with feature scaling and analyses datasets and its transformations applied to our data before feeding it to the algorithm. Pre-processing of data is important as the data gathered from different sources might be in unstructured format so conversion of data for further analyses becomes important.

2) Transformed dataset:

The main purpose of Transformed dataset is to enhance the dataset which is been fed to the system so that it increases the likelihood of the algorithm will be able to make the prediction meaningful with accuracy. However, the structure of the data and values with additional features (feature extraction) may be changed or added.

The following types of data can be transformed:

a. Training Data

b. Unlabeled Data

c. Test Data

d. Validation Data.

3) Apply machine learning algorithms.

- Random Forest: Random Forest, also known random decision forest is associated with the predictive models for both the classification and regression problems. In Random Forest the model creates an entire forest of random uncorrelated decision and are trained with bagging method. Bagging method is a combination of learning models to increase the overall result.

- Support Vector Machine (SVM): In machine learning, support vector machine are supervised models are associated with machine learning algorithms that analyze data used for classification and regression analysis. A support vector machine constructs a hyperplane or set of hyperplanes which are suitable for classification and regression.

4) Cross validation and choose the most accurate algorithm:

It is the statistical method to evaluate the accuracy of machine learning models.

It is generally used in machine learning to compare and select a model because it is easy to understand and implement.

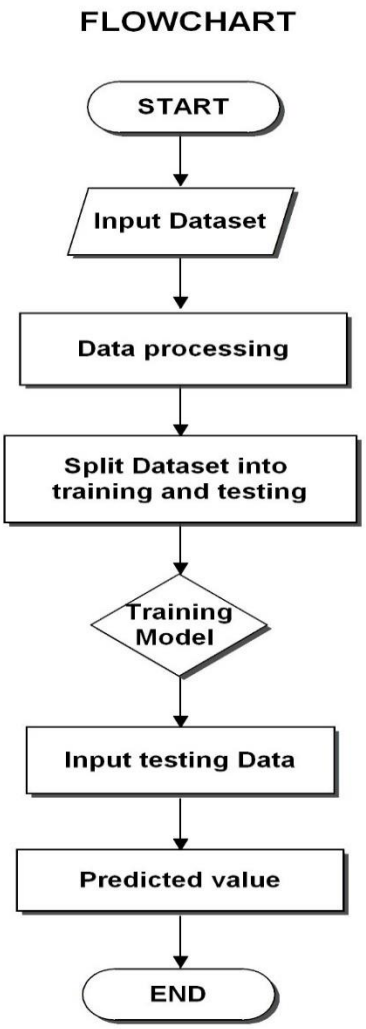

Figure 4. Flowchart

B. EXTERNAL:

External architecture holds the user input data base allowing system to work efficiently. It also holds the output of the data which has been processed. Initially the data which has been collected from the patients is taken in the database and then it is being forwarded to the internal system. In the external system it also provides the interpreted data after it has been gone through several data processing, machine algorithms and mathematical calculations.

V. CONCLUSION

Heart diseases are a critical issue, if taken lightly can result into fatal outcomes. With current technology in medical sector, it is possible to cure them with appropriate treatments. However, if it is diagnosed late, then even the high-tech medical equipment cannot help. Hence, it is necessary to identify roots of any such diseases when it is still in basic stage. Thus, we are using Machine Larning to identify such diseases and its chances to occur in the coming future. When the disease is identified, appropriate treatments can be prescribed to the patient in order to get a cure to it.

REFERENCES

[1] Chaitanya Suvarna, Abhishek Sali, Sakina Salmani "Efficient Heart Disease Prediction Using optimization Technique" in Institute of Electrical and Electronic Engineering, 2017.

[2] Mai Shouman, Tim Turner, Rob Stocker "Using Data Mining Techniques in Heart Disease Diagnosis and Treatment" in Institute of Electrical and Electronic Engineering, 2012.

[3] Asha Rajkumar, Mrs. G Sophia Reena "Diagnosis of Heart Disease Using Datamining Algorithm" in Global Journal of Computer Science and Technology, 2009.

[4] K. Srinivas, B. Kavihta Rani, Dr, A. Govrdhan "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks" in International Journal on Computer Science and Engineering, 2010.

