

Estimation of Cardiovascular Disease using Linear Support Vector Machine

Dr.SARANYA AKS

Associate Professor/Department of ECE
Rajiv Gandhi Institute of Technology
Bangalore

Abstract—Cardiovascular disease is one of the major cause of morbidity and mortality world wide. About 60 percentage of the death in the world is due to heart disease. The major cause for the cardiac disease are sedentary lifestyle, unhealthy diet, obesity, diabetes, hypertension, smoking and consumption of alcohol, stress, hereditary factory etc. The early diagnosis of heart failure is an essential factor to protect the life of the human being. The objective of the research work is to estimate the heart rate precisely in the early stage of the cardiac failure itself. The proposed system uses Support Vector Machine method for prediction of cardiovascular disease, compare with other methods the testing sample is classified exactly based on the training sample. The SVM classifier the classification error is reduced and has low computational cost. The dataset is increased to improve the prediction of the cardiac disease. The proposed system will analyze the cardiac disease based on the parameters and diagnose the disease exactly. The aim of the proposed method is even technically unknown person can predict the condition of the cardiac person using Support Vector Machine Classifier.

Keywords—*prediction, cardio vascular diseases, SVM, machine learning, feature selection*

I. INTRODUCTION

Heart disease is a leading cause of death worldwide. The common risk factors associated are identified as age, family history, Stress, high cholesterol, Heart rate, smoking, alcohol intake, overweight, physical inactivity, chest pain type and poor diet. In existing method uses patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited [4]. Mining is a way of exploring large amount of data to take out patterns that are hidden and formerly unknown relationships and expertise detection to assist the better expertise of medical records to save you heart disease[7]. The mining method is an unsupervised learning the training data has to learn without any knowledge on the testing sample and the classification accuracy gets affected. The other disadvantage is number of parameters taken for the prediction of cardiovascular disease is low.

The exact classification of heart beat is essential for the physician to diagnose the heart disease with high precision, since the life of the patient can be saved by the early diagnosis.

A person gets beaten by means of cardiovascular disorder from aspects including family tree, age, hypertension; sports including smoking and alcohol etc. For automatic classification of heart beat effective system is needed. At the right time curing and treating of those diseases might be hard in order to truly result in massive variety of deaths.

In cardiovascular disease, the data provided to the medical field is vast, multi-dimensional and critical, which makes it challenging and cumbersome to understand such extensive data. The major symptoms of heart disease is chest pain, shortness of breath, radiating pain in the arms, left shoulders and elbows, discomfort while walking, high blood pressure, dizziness, nausea, fatigue, etc. while the risk factors causing CVD are cholesterol, hypertension, genetic or heredity, obesity, diabetes, dietary habits, aging, etc[8].

Life fashion adjustments with normal aerobic pleasant sporting events followed by means of a coronary heart-wholesome food regimen, quitting smoking and alcohol consumption, utilization of tobacco, salt, sugar and fats consumption can save you CVD. As per the data from the world wide data CVD affects even young men and women, so prediction is required in the initial stage itself[11]. To envisage the detection of CVD at the incipient stage and intervene appropriately, Support Vector Machine (SVM) is used for classifying the patients depending on their symptoms and risk factors. By using these machine learning algorithms, it becomes easy to understand the nature and type of heart disease in all aspects. The important aspect of this method is to train the machine to analyze a massive set of data with the known inputs and outputs, i.e., Supervised machine learning is applied. The main objective of the work is to improve the accuracy of prediction of cardiovascular disease, which can help for preventive treatments and reduce the cases having CVD. The existing method suffers from less sample set and prediction of cardiovascular disease is affected. Using Support vector machine tries to find the optimal separating hyper plane such that the expected classification error for unseen patterns is minimized. The identification and prediction of such diseases at early stages are helpful to reduce the extremity of it compared to existing method the classification error is low.

II. PROPOSED METHOD

Patient with irregular heartbeat doesn't mean having a heart attack. But if it's a new symptom, or if you have chest pains or problems breathing, may be the preliminary symptom for heart attack. The proposed system can predict cardiovascular disease based on the parameters used for prediction of cardiac disease. The main purpose of the proposed method is to predict the occurrence of heart disease for early detection of the disease in a short time. In our approach, we are using Linear Support Vector Machine (Linear SVC) to predict the heart disease based on some health parameters. In Support Vector Machine the training samples are trained based on the supervised learning. Another feature is the even consider the minute factor for the classification. The testing sample is compared with large set of training sample. In this technique the hyperplane separate the sample of data and the classification is carry out with each sample set and misclassification of testing sample is minimized. In this technique input variable is taken from the dataset and the output is the presence or absence of heart beat prediction. The data set consist of two attributes X and Y. input variable is X and the target output is Y. The input data testing data sample is given as input to the computer. The dataset is divided into two parts; the training data and testing data. The system is trained with the expected output, so in testing sample the desired output is obtained. The device is educated with the schooling statistics, consequently the gadget learns by way of itself. Once the system is trained, the trying out dataset is given to the device and the expected output is accomplished.

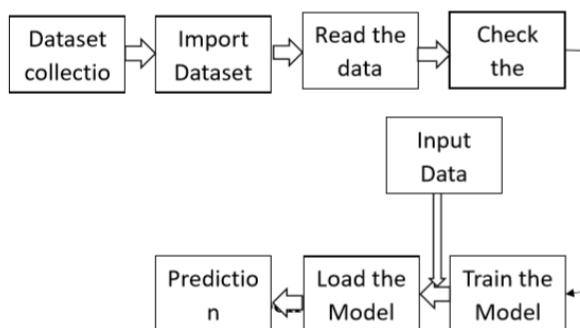


Fig.Block Diagram of Proposed Method

DATASET COLLECTION

The data set consist of 143 records and 10 attributes is collected from UCI. The dataset is divided into 80:20 ratios of training and testing data. The 10 attributes which is been collected are Age, Gender, Smoke, Cholestrol, Consumption of alchcohol, habit, Oldpeak, Slope. The data collected is used to predict the cardiovascular disease of the patient using the parameters the above parameters and comparing the testing and training sample of the patient.

IMPORT DATASET

In Python, we use the import keyword to make code in one module available in another. Transferring data in Python are important for structuring the code effectively. Thus data is imported from the dataset to pandas .Using imports properly

will make more productive, allowing to reuse code while keeping projects maintainable.

READ THE DATA

Identify or study on heart disease patient data and then decides which factors are common in same heart disease patient. Experiments were conducted with Weka. 3.6.0 tool. Data set of 1000 records with 8 attributes is used.

CHECK THE MISSING

Check the missing factors. The training sample will take even the small feature for classification to avoid misclassification.

TRAIN THE MODEL

After collecting the dataset from the above attributes, it can be classified into 80% of training data and 20% data of testing data. The data is gathered in an excel sheet and loaded. The training data contains the information's of patients for each attributes which is been collected.

INPUT DATA

The input data is given to the computer system that is testing sample to be classified is given as input

LOAD THE MODEL

In the previous modules the data set is collected, trained and then the algorithm is implemented. On this module check information is loaded and it exams the classifier the usage of the take a look at dataset in which we can are expecting the class label for check information. The output is given with the aid of evaluating the given schooling and checking out records. The output is given by comparing the given training and testing data.

PREDICTION

The prediction is done for the given input. The output is in the form of whether the patient is having cardiovascular disease or not. The classification is carried out using SVM classifier and the result is executed as the person is affected by cardiac disease or not affected by cardiac disease. SVM is a supervised machine learning algorithm which can be used for both linear and nonlinear data for classification from the field called computational learning theory been widely used in classification problems. On this algorithm, each records item is plotted in a factor as n-dimensional space (wherein n is the range of capabilities) with the value of every function in the testing sample. Then, classification is done by finding the hyper-plane which differentiates the two classes well. The testing sample is categorized based on age, blood pressure, Sugar level and classified using Support Vector Machine Classifier. In an SVM classifier, the separation hyper plane is chosen to minimize the anticipated generalization error of the unobservable sample data. SVM is a sophisticated classifier that can tell the difference between two classes. The test image is assigned to the group with the largest separation to the training's closest point by SVM. The SVM training procedure develops a strategy that can identify whether an input image belongs to this class or not. To find an accurate decision border, SVM requires a considerable amount of training data,

which increases the computational cost. The SVM is a supervised learning algorithm for categorization of testing sample. It aims to find the optimum separation hyperplane for unobserved sequences with the lowest possible predicted classification error. In this system using supervised method for classification labeled data to understand the datasets and learn about each data, once the training and processing are done then we test the model by providing a sample data to check whether it is predicting the exact output or not and it is tested in the testing sample.

The proposed system description is the data is collected from UCI dataset. The data collected is classified based on certain parameters such as age, gender and blood pressure level etc. Next step is importing the data set using the import in python. Pandas are used in python. Pandas is fast, simple and more expressive than other tools. Python changed into successful for information education, but it best supplied confined assist for information evaluation. So, Pandas came into the photograph and better the abilities of data analysis. It is able to perform five giant steps required for processing and analysis of facts no matter the beginning of the statistics, i.e., load, manipulate, put together, version, and examine. After importing of data set next is the reading of data. In this step the patient is classified based on the factors that is same type of disease is consider under the same category. The next step is the checking of missing factor, all the training samples are trained based on the supervised learning that is each training sample is trained and check with the expected output. In order to avoid the missing sample even the small feature is also consider for testing sample to avoid the misclassification.

The next step is the training of sample eighty percentage of sample is consider under training sample. The sample is trained based on patients parameter such as blood pressure, glucose level. The next step is the testing sample is given as the input. The next step is the input testing sample is loaded to the dataset. Supervised learning method is used to predict the output. Supervised learning is to map input information with the output records. The supervised mastering is primarily based on supervision, and it's miles the same as when a student learns matters inside the supervision of the trainer. Support Vector Machine is used for the classification of testing sample because, Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The intention of the SVM set of rules is to create the excellent line or decision boundary which could segregate n-dimensional area into instructions so that we can without problems put the new records point in the proper category within the future. This satisfactory decision boundary is known as a hyperplane. SVM chooses the acute factors/vectors that help in creating the hyperplane. These extreme instances are known as as aid vectors, and consequently set of rules is named as guide Vector system. In creating the hyperplane. In our proposed system using Linear as Support Vector Machine, Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

Thus the proposed system uses Support vector Machine classifier, with the comparison of testing sample with the training sample come to know that the person is affected by the cardiac disease or not. The classifier also measures the systolic blood pressure, Diastolic blood pressure, Glucose level, whether the person is a smoking, alcoholic consumer Based on all these factors Support vector Machine will predict the person is having cardiovascular disease. Thus comparing with the existing method the proposed method have high accuracy in predicting the output of the patient, since the early diagnosis of the disease is an essential factor to diagnose the disease in the initial stage itself and the life of patient is saved.

III. RESULT AND DISCUSSION

The proposed method is 98 percent more accurate than other methods. The results of comparing the proposed feature selection approach to different methods. The Nearest neighbour (KNN) with Genetic Algorithm(GA) is used for effective classification. Using the Random Forest algorithm, the proposed method utilized stronger and smarter features for classification. As a result, classification accuracy has improved with fewer characteristics, resulting in lower test costs and greater test accuracy in identifying ischemic heart diseases.

Fig. Cardiovascular disease prediction of Proposed system

The cardiovascular disease prediction has obtained high accuracy comparing with the existing method. The output used to predict the various factors such as gender, systolic and diastolic pressure. Alcohol consumption and smoking habits.

The Random Tree(RT) model is proven to be poor in the classification process, with a sensitivity and specificity of just 74.41 percent and 71.11 percent, respectively. The existing method of heart beat evaluation using RPA learning shows the accuracy of 97 percentage.

Sensitivity is the ratio of sum of true positive to the sum of true positive and sum of true negative.

Specificity is the ratio of sum of true negative to the sum of true negative and false positive.

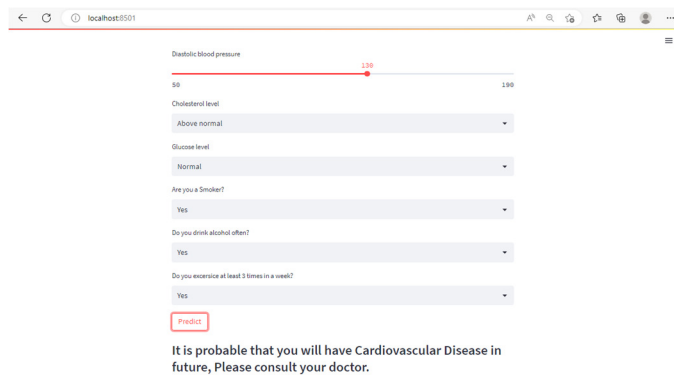
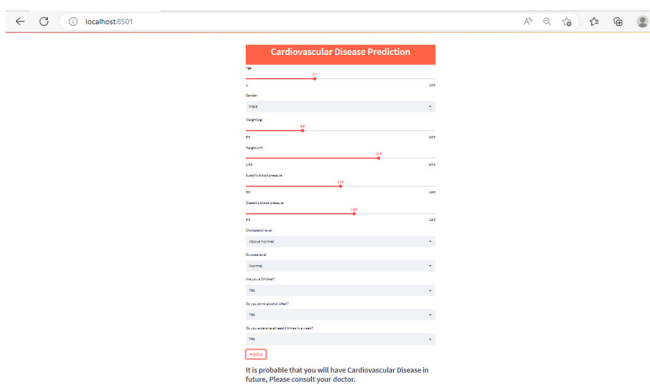


Fig. Cardiovascular disease prediction using Random Forest

The NB Tree model has attempted to outperform previous models by achieving sensitivity and specificity of 80.24 percent and 77.94 percent, respectively. Thus with the existing methods the determination of cardiovascular disease include the factor such as gender, blood pressure and sugar level. The proposed system shows 98 percentage accuracy.



There are so many machine learning techniques to detection and prediction of the heart diseases. In this paper use the support vector machine to predict and identify the heart diseases of patients and compare the result of the support vector machine algorithm with the other machine algorithms. The SVM algorithm gives the better accuracy, specificity and sensitivity when compare to the other machine learning algorithms.

IV. CONCLUSION

In this modern era, heart disease is very decisive health problem in recent times in our society. This work follows advanced techniques for predication of this heart disease. Machine learning is a developing area of artificial intelligence which has showed good result in all fields of medical diagnosis with more accuracy. This system have worked by reducing the attributes of dataset to increase the accuracy of prediction. Linear Support Vector Machine (Linear SVC) classifier can predict that the patients getting a heart disease or not. In

addition classification accuracy, sensitivity, and specificity of the classifier have been found to be high thus making it superior alternative for the diagnosis work continues doing analysis on the data from which we are getting at which age it mostly occur or which region influenced by that disease. So precaution can be taken to avoid the death due to the heart disease. Future work is extended to diagnose the covid disease thus early detection can save the life of the patient.

V. REFERENCES

- [1] M.Durairaj, K.Meena, A Hybrid Prediction System Using Rough Sets and Artificial Neural Networks], International Journal Of Innovative Technology & Creative Engineering (ISSN: 2045-8711) VOL.1 NO.7 JULY 2011.
- [2] K. Srinivas, B. Kavitha Rani and Dr. A. Govrdhan, Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks, International Journal on Computer Science and Engineering (2010).
- [3] Kala John Kappiarukudil and ManeeshaVinodini Ramesh "Real-Time Monitoring and Detection of „Heart Attack“ Using Wireless Sensor Networks", 2010 Fourth International Conference on Sensor Technologies and Applications ISSN:978-0-7695-40962/10IEEE,2010.
- [4] Mai Shouman, Tim Turner, and Rob Stocker "Using Data Mining Technique in Heart Disease Diagnosis and Treatment", 2012 JapanEgypt Conference on Electronics, Communications and Computers ISSN:978-14673-0484-9/ IEEE,pp.173-177,2012. 3
- [5] Vikas Chaurasia and Saurabh Pal "Data Mining Approach to Detect Heart Diseses", International Journal of Advanced Computer Science and Information Technology (IJACSIT) Vol. 2, No. 4, 2013, Page: 56-66, ISSN: 2296-1739.
- [6] J Peter and K. Somasundaram, "An Empirical Study on Prediction of Heart Disease Using Classification DataMining Techniques", Proceedings of IEEE International Conference on Advances In Engineering, Science And Management (ICAESM), (2012), pp. 514-518.
- [7] Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. Online: 25 March 2017 DOI: 10.1007/s10462-01.
- [8] Prerana T H M, Shivaprakash N C et al "Prediction of Heart Disease Using Machine Learning Algorithms- Naïve Bayes,Introduction to PAC Algorithm, Comparison of Algorithms and HDPS",Vol 3, PP: 90-99@IJSE, 2015.
- [9] Salam Ismaeel, Ali Miri et al "Using the Extreme Learning Machine(ELM) technique for heart disease diagnosis",IEEE Canada International Humanitarian Technology Conference, DOI:10.1109/IHTC.2015.7238043, 03 September 2015.
- [10] Salha M. Alzahani, Afnan Althopity, Ashwag Alghamdi, -"An Overview of Data Mining Techniques Applied for Heart Disease Diagnosis and Prediction" Lecture Notes on Information Theory, Volume: 02, No. 04.
- [11] H K Shifali, Dr. B. Srinivasu, Rajashekar Shastry, B N Ranga Swamy"Mining of Medical Data to Identify Risk Factors of Heart Disease Using Frequent Itemset", International Research Journal of Engineering and Technology (IRJET),Volume: 04 Issue: 01.
- [12] Prof. Mamta Sharma 1, Farheen Khan2, Vishnupriya Ravichandran "Comparing Data Mining Techniques Used For Heart Disease", International Research Journal of Engineering and Technology, Volume: 04 Issue: 06
- [13] Abhishek Taneja - "Heart Disease Prediction System Using Data Mining Techniques", oriental journal of computer science & technology, Volume: 06