A DIAGNOSIS OF HEART DISEASE USING MACHINE LEARNING

[1]P. Prasanna Kumar, [2]N. Bhargav reddy, [3]V. Dorababu, [4]N. Jeswanth,

Pursuing B. Tech [ECE] in NBKR Institute of Science and Technology - Vidyanagar,
Kota Mandal Tirupati District AP.

[5]Dr. Abraham Mathew
Associate Professor
department of "Electronics and Communication Engineering",
NBKR Institute of Science & Technology (NBKRIST), Vidyanagar, Nellore, AP.

Abstract: **Day by day the cases of heart diseases are increasing at a rapid rate and it's very Important and concerning to predict any such diseases beforehand. This diagnosis is a difficult task i.e. it should be performed precisely and efficiently. The research paper mainly focuses on which patient is more likely to have a heart disease based on various medical attributes. We prepared a heart disease prediction system to predict whether the patient is likely to be diagnosed with a heart disease or not using the medical history of the patient. We used different algorithms of machine learning such as logistic regression and SVM to predict and classify the patient with heart disease. A quite Helpful approach was used to regulate how the model can be used to improve the accuracy of prediction of Heart Attack in any individual. The strength of the proposed model was quiet satisfying and was able to predict evidence of having a heart disease in a particular individual by using SVM, KNN, Decision Tree Classifier, Random Forest Classifier, Naïve Bayes and Logistic Regression which is used to predict the model with above Machine Learning Algorithms.**

Among the above the SVM showed a good accuracy in comparison to the previously used classifiers. So, a quiet significant amount of pressure has been lifted off by using the given model in finding the probability of the classifier to correctly and accurately identify the heart disease. The given heart disease prediction system enhances medical care and reduces the cost. This project gives us significant knowledge that can help us predict the patients with heart disease it is implemented on the. pynb format.

Keywords: **Prediction, heart disease, symptoms, machine learning**

1.INTRODUCTION

In this project I am going to analyze dataset about Heart Attack and use 6 ML algorithms to make a prediction. I am also going to choose the best algorithm in terms of accuracy and cross validation score. Moreover, i will do deeper analysis about these algorithms.

World Health Organization has estimated 12 million deaths occur worldwide, every year due to Heart diseases. Half the deaths in the United States and other developed countries are due to cardio vascular diseases. The early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes in high risk patients and in turn reduce the complications. This research intends to pinpoint the most relevant/risk factors of heart disease as well as predict the overall risk using SVM. There is no dearth of records regarding medical symptoms of patients suffering heart strokes. However, the potential they have- to help us foretell similar possibilities in seemingly healthy adults are going unnoticed. For instance: As per the Indian Heart Association, 50% of heart strokes occur under 50 years of age and 25% of all heart strokes occur under 40 years of age in Indians. Urban population is thrice as vulnerable to heart attacks as rural population.

We thus propose to collect relevant data pertaining all elements related to our field of study, train the data as per the proposed algorithm of machine learning and predict how strong is there a possibility for a patient to contract a heart disease. For the purpose of patients entering data, we suggest to make use of the easily available sensors in watches and cell phones to measure the simple factors. Our main aim behind developing the system is to make it user friendly so that regular monitoring of the patient is made possible. Thus, it is of utmost importance that the factors required in the input are most accurate and easily available.

The algorithm we have used is explained in greater detail. Next, we put forth the results of our experiment on sample dataset using the proposed algorithm. Also, the statistics using different methods are explained. Finally, the paper is described briefly with the conclusion at the end.

2.LITERATURE SURVEY

[1] Senthil Kumar Mohan. discussed as follows

Senthil Kumar Mohan proposed Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques in which strategy that objective is to finding critical includes by applying Machine Learning bringing about improving the exactness in the expectation of cardiovascular malady. The expectation model is created with various blends of highlights and a few known arrangement strategies. We produce an improved exhibition level with a precision level of 88.7% through the prediction model for heart disease with hybrid random forest with a linear model (HRFLM) they likewise

[2] Sonam Nikhar discussed as follows

Sonam Nikhar has built up the paper titled as Prediction of Heart Disease Using Machine Learning Algorithms by This exploration plans to give a point by point portrayal of Naive Bayes and decision tree classifier that are applied in our examination especially in the prediction of Heart Disease. Some analysis has been led to think about the execution of prescient data mining strategy on the equivalent dataset, and the result uncovers that Decision Tree beats over Bayesian classification system.

[3] Abhay Kishore discussed as follows

Abhay Kishore developed Heart Attack Prediction Using Deep Learning in which This paper proposes a heart attack prediction system using Deep learning procedures, explicitly Recurrent Neural System to predict the probable prospects of heart related infections of the patient. Recurrent Neural Network is a very ground-breaking characterization calculation that utilizes Deep Learning approach in Artificial Neural Network. The paper talks about in detail the significant modules of the framework alongside the related hypothesis. The proposed model deep learning and data mining to give the precise outcomes least blunders. This paper gives a bearing and point of reference for the advancement of another type of heart attack prediction platform.

[4] Avinash Golande discussed as follows

Avinash Golande proposed Heart Disease Prediction Using Effective Machine Learning Techniques in which Specialists utilize a few data mining strategies that are available to support the authorities or doctors distinguish the heart disease. Usually utilized methodology utilized are decision tree, k- closest and Naive Bayes. Other unique characterization-based strategies utilized are packing calculation, Part thickness, consecutive negligible streamlining and neural systems, straight Kernel self- arranging guide and SVM (Bolster Vector Machine). The following area obviously gives subtleties of systems that were utilized in the examination.

[5] Lakshmana Rao discussed as follows

Lakshmana Rao Machine Learning Techniques for Heart Disease Prediction in which the contributing elements for heart disease are more. So, it is difficult to distinguish heart disease. Different systems in data mining and neural systems have been utilized to discover the seriousness of heart disease among people. Not doing early identification, may impact the heart or cause sudden passing. The perspective of therapeutic science furthermore, data burrowing is used for finding various sorts of metabolic machine learning a procedure that causes the framework to gain from past information tests, models without being expressly customized. Machine learning makes rationale dependent on chronicled information.

3.REASEARCH WORK

To initiate with the work, we have started collecting data in each and every aspect towards the goal of the system. In the first place, the research was in the direction of the main causes or the factors which have strong influence on the heart health. Some factors are unmodifiable like age, sex and family background but there are some parameters like blood pressure, heart rate etc. which can be kept in control by following certain measures. Many doctors suggest healthy diet and regular exercise to keep the heart healthy. Following are the parameters which are considered for the study in designing the system which have major risk percentage with respect to CAD.

1. Age
2. Sex
3. Resting Blood Pressure
4. Heart Rate
5. Diabetes
6. Hyper cholesterol

The next step was to collect dataset. For this we have used Cleveland dataset from UCI library. The dataset contains as many as 76 parameters describing the complete health status of heart. These parameters are obtained by expensive clinical tests like ECG, CT scan etc. Out of these, the traditional heart disease prediction system uses 13 major parameters. Since these parameters require expensive lab tests to find ECG, chest pain type, ST depression etc. To avoid these and to make system less complex we selected above mentioned parameters which can be easily measured using different sensors available in the market. The following research work briefly explains the latest sensors in the market used to measure different parameters.

A. AliveKor:

It comes as a touchpad connected to your cell phone (through wireless network) or as a wristband. The touchpad simulates an ECG of the patient on his cell phone via Bluetooth. Thus all the necessary parameters like pulse rate, blood pressure are easily available. The wristband on the other hand, uses finger touch to display the pulse function on the dial. It can also notify an atrial fibrillation.

B. My Heart in this system, a number of on-body sensors are used to collect physiological data that are sent wirelessly to a PDA. The information is analyzed and health recommendations are given to the user based on the analysis. [1]

C. Health Gear is an application to track most common indexes like lab tests and physical parameters. Fields include: - [Physical] Indexes like Height, Weight, BMI - Blood Pressure, Hemoglobin, WBC, RBC, Platelets - [Lipids]: Cholesterol, HDL, LDL, VLDL, triglycerides, - [Sugar]: Fasting Glucose, after meals, HbA1C.

D. Fitbit This sensor is used to keep the track of health which has features of measuring pulse rate, BP, calories burned. After this study, we have concluded with using Fitbit to collect the data which is easily available and less expensive and Health Gear for all the other parameters.

4. PROPOSED MECHANISM:

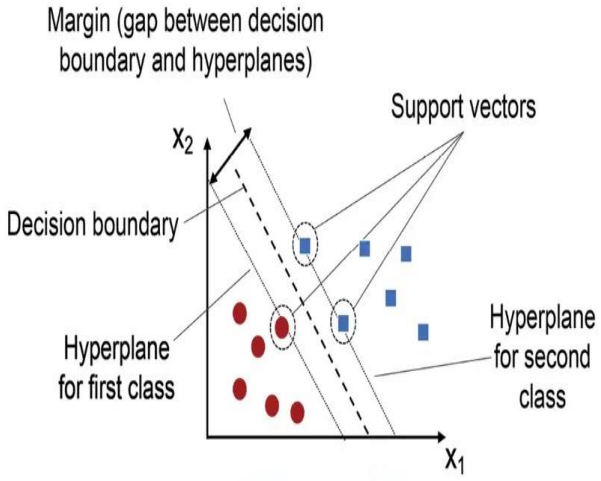
The support vector machine has proven to be the most popular and evolving branch of machine learning in recent studies. Now in the proposed system we used the Support Vector Machine (SVM) to train and test the data set.

5. SUPPORT VECTOR MACHINE:

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 goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence algorithm is termed a Support Vector Machine.

Consider the below diagram in which two different categories are



Linear SVM:

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 linearly separable data, and a classifier is used called Linear SVM classifier.

Non-linear SVM:

Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed non-linear data, and the classifier used is called a Non-linear SVM classifier.

To produce a non-linear support vector machine we make use of what is called a **kernel function**. Depending on the kernel used, the kernel function transforms our data to a feature space where the data becomes more likely to be linearly separable. This is known as the 'kernel trick'.

The system is developed using python code using PyCharm IDE. With the help of python library sci-kit learn, the system is implemented. The sample parameter values taken for two different categories in (SVM) is as shown below figure.

```
from sklearn.svm import SVC

#finding the best parameters
grid =
{"C":np.arange(1,10,1),'gamma':[0.00001,0.00005,0.0001,0.0005,0.001,0.005,0.01,0.05,0.1,0.5,1,5]}
svm0 = SVC(random_state=42)
svm_cv = GridSearchCV(svm0,grid,cv=10)
svm_cv.fit(x,y)
#print("Best parameters of SVC:",svm_cv.best_params_)

svm = SVC(C=svm_cv.best_params_["C"],gamma=svm_cv.best_params_["gamma"],random_state=42)
svm.fit(train_x,train_y)
print("SVC Accuracy :",svm.score(test_x,test_y))
```

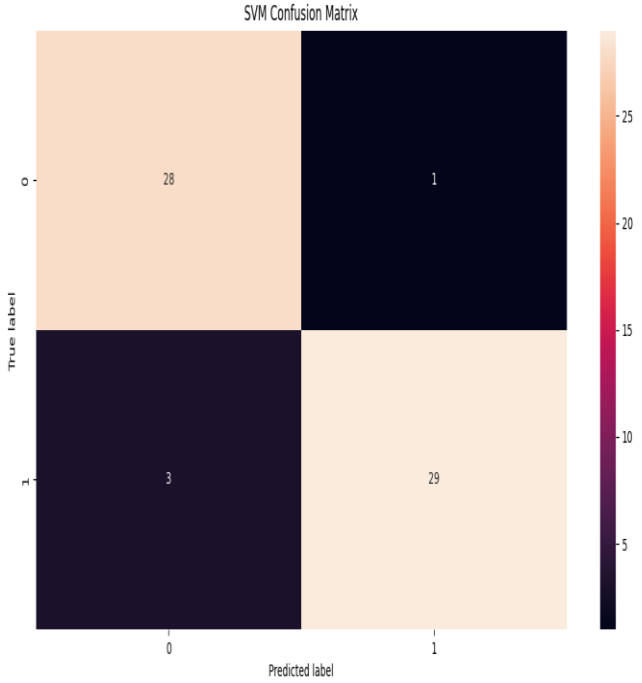
For further details about the parameters visit the sci-kit learn library. Different python libraries such as SciPy, NumPy and panda are used to support the algorithm.

6. RESULTS:

The output of the system will give a prediction result if the person has a heart disease, in terms of Yes or No. The system gives an idea about the heart status leading to CAD beforehand. If the person is prone to have heart disease, then the result obtained will be Yes and vice versa. In case of a positive output, he needs to consult a cardiologist for further diagnosis. The statistics of the results obtained during the testing of the dataset is shown in the following table.

SVC Classification Report:

	Precision	recall	f1-score	support
0	0.90	0.97	0.93	29
1	0.97	0.91	0.94	32
accuracy			0.93	61
macro avg	0.93	0.94	0.93	61
weighted avg	0.94	0.93	0.93	61



7. CONCLUSION

The Heart Disease Prediction System using Machine learning algorithm, viz. SVM provides its users with a prediction result that gives the state of a user leading to CAD. Due to the recent advancements in technology, the machine learning algorithms are evolved a lot and hence we use Support Vector Machine in the proposed system because of its efficiency and accuracy. Also, the algorithm gives the nearby reliable output based on the input provided by the users. If the number of people using the system increases, then the awareness about their current heart status will be known and the rate of people dying due to heart diseases will reduce eventually.

I used 6 different algorithms in this notebook. SVC are better than the others in terms of accuracy score and cross validation score. Finally, i got % 93 accuracy from SVC

8. REFERENCES

[1] Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. Online: 25 March 2017 DOI: 10.1007/s10462-01

[2] Prerana T H M, Shiva Prakash 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

[3] 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

[4] F BrainBoudi, 'Risk Factors for Coronary Artery Disease',2016. [Online]Available:https://emedicine.medscape.com/article /164163-overview.

[5] National Health Council, 'Heart Health Screenings',2017. [Online]Available:https://www.heart.org/HEARTORG/ Conditions/Heart-HealthScreenings_UCM_428687_Article.jsp#. WnsOAeeYPIV

[6] ScikitLearn,'SVM', [Online]
Available: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html
[7] Prediction System for heart disease using Naïve Bayes *Shahab Adam Pattekari and Asma Praveen Department of Computer Science and Engineering Khaja Banda Nawaz College of Engineering
[8] Comak E, Arslan A (2012) A biomedical decision support system using LS-SVM classifier with an efficient and new parameter regularization procedure for diagnosis of heart valve diseases. J Med Syst 36:549–556
[9] Ahmed Fawzi Otoom, Emad E. Abdallah, Yousef Kilani, Ahmed Kef aye and Mohammad Ashour (2015) Effective Diagnosis and Monitoring of Heart Disease ISSN: 1738-9984 IJSEIA
[10] N. Oliver and F. F. Mangas, "Health Gear: a real-time wearable system for monitoring and analyzing physiological signals", IEEE International Workshop on Wearable and Implantable Body Sensor Networks, Cambridge, USA, (2006).
[11] r. Amit Mittal, 'Increasing heart attacks in young Indians',2017, Available:https://timesofindia.indiatimes.com/life - style/health-fitness/every-heart-counts/increasing-heart-attacks-inyoung-indians/articleshow/56295257.cms
[12]AliveKor,[Online]Available:https://www.alivecor.com/how-it-works.