International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

HEART ATTACK PREDICTION USING MACHINE LEARNING

T M Geethanjali [1], Arpitha R Hegde [2], Maheshwari [3], Surabhi B N [4]

[1] *Assistant Professor, Department of Information Science and Engineering, P E S College of Engineering, Mandya.*
[2] *Student, Department of Information Science and Engineering, P E S College of Engineering, Mandya.*
[3] *Student, Department of Information Science and Engineering, P E S College of Engineering, Mandya.*
[4] *Student, Department of Information Science and Engineering, P E S College of Engineering, Mandya.*

Abstract- According to the WHO report, heart disease is one of the main causes of death. Every year 17.9 million deaths are due to heart disease. As the population grows, diagnosing and initiating therapy at an early stage becomes more challenging. As a result, the goal of this project is to develop a machine learning model for predicting heart attack based on the relevant characteristics. For this analysis, we employed a Kaggle and Data World dataset which contains 14 distinct features linked to Heart Attack. The model is created using machine learning methods such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Random Forest. The results demonstrate that, when compared to other machine learning techniques, SVM predicts heart attack with high accuracy rate in less time.

Key Words: heart attack, random forest, SVM, KNN, dataset.

1. INTRODUCTION

Heart attack is a stoppage of blood stream to the heart muscle. Heart attack happens when blood coagulation blocks blood stream to the heart. The most complicated and complex task inside the field of clinical science is to anticipate heart attack. Heart is the most essential organ of the human body. There is an extreme need in anticipating the extent and earnestness of heart condition that give an exact treatment to the patients. This necessity a profound investigation of heart examination of the patient that incorporates indications like chest pain, chest tightness, chest pressure factor and inconvenience in breathing, numbness and so on .The heart analysis includes certain choices to be taken dependent on the wellbeing history and the clinical preliminary consequences of an individual. The clinical professionals must do their work precisely and effectively where a simple carelessness may cause hazard to a patient. For valid and exact diagnosis of heart attack there should be a keen computerized framework that should help clinical experts in settling on choices dependent on the current side effects and clinical history of a patient. Heart Attack is quite possibly the most dangerous diseases which can wreck one anytime of time with no greeting.

2. RELATED WORK

In [1] they worked on how to detect and prognosis heart disease in humans in the early stages. This helps the doctors to administer proper treatment for patients. This reduces the mortality rate caused due to heart attack. They have used the machine learning algorithm such as Random Forest, Naïve Bayesian Classification and Logistic Regression to predict the heart attack. Dataset were drawn out from the Cleveland heart disease in the UCI machine learning repository, cardiovascular disease dataset, and Framingham Heart study dataset and Kaggle respectively. Dataset consist of 17 features and 1990 observations. They have worked on identifying the key patterns and features by mixing data mining techniques with big data. They worked on finding accuracy for the above mentioned algorithms. The accuracy of Random Forest, Naive Bayesian and Logistic Regression are respectively as 92.44%, 61.96% and 59.7% .

In [2] the aim is to work on how to reduce the efforts and time put in by the doctor of see the patients. They predict the risk of the patient to heart attack using the binary classifier. They have developed the graphical user interface to enter the details

of patients and about their present health status. The interface developed is user friendly. The person with little or zero technical knowledge can use the interface with ease. The dataset was obtained from the University of California and Irvine's machine learning repository. The dataset comprises of 14 attributes. They made used of Naïve Bayes algorithm for building the model to. They have worked on finding the accuracy to build the model. The accuracy was up to 81.25%.

In [3] they have worked on finding the effective way in forecasting a patient's heart disease. The HDPS system established in this study is an anecdote method for classifying cardiac disease that can be employed. The classification and prediction was trained by learning Vector Quantization Algorithm which is one of Artificial Neural Network learning approach. The HDPS system is computer-aided system developed using C and C# environment. They have obtained the dataset from the UCI repository dataset consists of 13 clinical features. For classification, they employed an Artificial Neural Network method. Finally, the heart disease prediction model is developed by achieving the accuracy of the around 80%.

In [4] the aim is to work on when a person has chest discomfort or similar symptoms, it is possible to forecast the likelihood of a heart attack. This system has 25 attributes. They have used the machine learning algorithms such as decision tree and random forest. Beside random forest is implemented to upgrade the accuracy of the classification result of heart attack prediction. They have developed an android application of entering the patient heart details and many more. The developed application is user friendly. The system is organized in three modules: Data pre-processing, Classification and Optimization. After uploading the information of the patient on the android application it will help us to predict whether the person go through heart attack or not. The app works by taking the input from the user, calculations are done on the back end of the app and the gives the results as he/she has chest pain which leads to heart attack.

In [5] they have worked on providing a novel method which intends to improve the accuracy of cardiovascular disease prediction by employing machine learning approaches to uncover key information. The dataset is taken from the UCI Cleveland dataset, with 13 clinical features. They have built model using R-Studio platform. The prediction model has accuracy of 88.7% through the usage of the Hybrid Random Forest with Linear Model (HRFLM). They have anticipated that females have much less hazard for coronary heart sickness as compared to males.

3. PROPOSED WORK

The main intention is to detect heart attack of a person at a prior stage. This system may be useful for the doctors to treat the patients. The system may control the deaths that are caused due to heart attack. The project is developed using machine learning algorithms such as KNN, SVM and Random Forest. Along with that a web interface is developed, which take the input of heart condition data from the patient and predicts the results by comparing the input data with the dataset present in the database and display the result like whether the person is normal or may undergo heart attack in near future.

3.1Attributes description

The dataset used in our project are extracted from Kaggle [6] and Data World repository [7]. Kaggle dataset consists of 14 features and 303 rows. Data World dataset consists of 14 features and 753 rows. Along with that we have merged both the Kaggle and Data World dataset and obtain 14 features and 1056 rows. We have analysed the model for different datasets. The 14 features that are used in model building are shown in Fig 2.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
2	63	0	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	0	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	0	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	0	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	0	1	120	263	0	1	173	0	0	2	0	3	1
10	52	0	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	0	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	0	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	0	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	0	3	110	211	0	0	144	1	1.8	1	0	2	1

Fig1. Sample Dataset

Above Fig 1 shows the dataset that is stored in Comma-Separated Value format. This CSV is comfortable to analyze the features in the dataset.

Dataset Attributes	Description
age	Age in years
sex	0-female;1-male
Cp	Chest pain type 0(Typical angina);1(Atypical angina) 2(Non-anginal pain);3(Asymptomatic)
Trestbps	Resting blood pressure
Chol	Serum cholesterol in mg/dl
Fbs	Fasting Blood sugar >=120 mg/dl 0-False;1-True
Restecg	Rest electrocardiographic results 0:Nothing;1:ST-T Wave 2:Possible left ventricular hypertrophy
Thalach	Maximum heart rate achieved
Exang	Exercise induced angina 0-No;1-Yes
Oldpeak	ST depression induced by exercise
Slope	Peak exercise St segment 0:Upsloping;1:Flatsloping 2:Downsloping
ca	Major vessels coloured by fluoroscopy
Thal	Thallium stress result 3-normal;6-fixed defect;7-reversable defect
Target	0:No;1:Yes

Fig 2: Dataset Description

3.2 Data Preprocessing

This method covert the raw data into an understandable format. The format that we have used in our project is Comma-seperated values (CSV). The raw data may have the missing values and duplicate values. In order check whether dataset has missing and duplicate values we need to pre-process. If data contains missing value we try to fill the missing value using mean method. If the data contains duplicate values we remove it from the dataset and select only the unique data.

3.3 Machine Learning Algorithms Used

In this section we give information about the algorithms that were used to build our heart attack prediction model. The algorithms used are K-Nearest Neighbor, Support Vector Machine and Random Forest.

K-nearest neighbors (KNN) algorithm is a type of supervised Machine Learning algorithm which can be used for both classification as well as regression predictive problems. KNN algorithm uses feature similarity to predict the values of new data points which further means that the new data point will be assigned a value based on how closely it matches the points in the training set. In our project when the user enters the data regarding the heart condition, the entered data gets compared with trained dataset and which has the closet match to the input data, the value of that will be assigned to input data.

Support Vector Machine

The Support Vector Machine (SVM) is a supervised machine learning technique that is most commonly used to solve classification problems. Each data item is plotted as a point in n-dimensional space (where n is the number of features you have), with the value of each feature being the value of a certain coordinate in the SVM algorithm Then we classify the data by locating the hyper-plane that best distinguishes the two classes and predicts the heart attack.

Random Forest Classifier

A random forest is a supervised machine learning algorithm. A random forest consists of multiple random decision trees. This is helpful in analyzing the data and to explain the relationship between the dependent variable like Yes/No or True/False or 0/1. It is made used in our project to predict whether the person undergo heart attack or not.

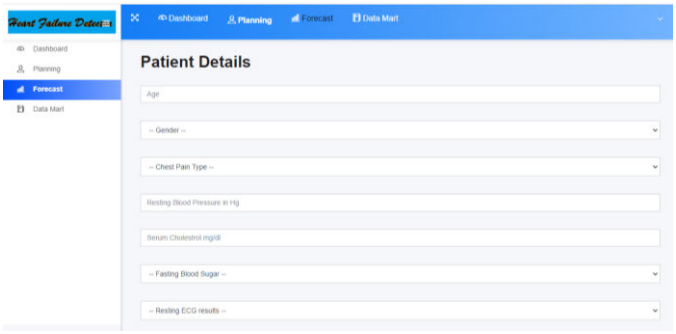

Fig 3: Forecast Window

Above Fig 3 is the forecast window, where the user enters the inputs such as age, gender, chest pain type, blood pressure level, cholesterol level, fasting blood sugar and many more. After entering the inputs user needs to save the data and after saving the data the output gets displayed. The output may be either normal or may undergo heart attack in near future.

3.4 Result

Algorithm	Accuracy
K-NN	92.6
SVM	94.4
Random Forest	93.7

Table 1 Kaggle dataset accuracy

Algorithm	Accuracy
K-NN	91.7
SVM	93.8
Random Forest	92.7

Table 2 Data World dataset Accuracy

Algorithm	Accuracy
K-NN	92.1
SVM	94.1
Random forest	93.2

Table 3 Merged dataset Accuracy

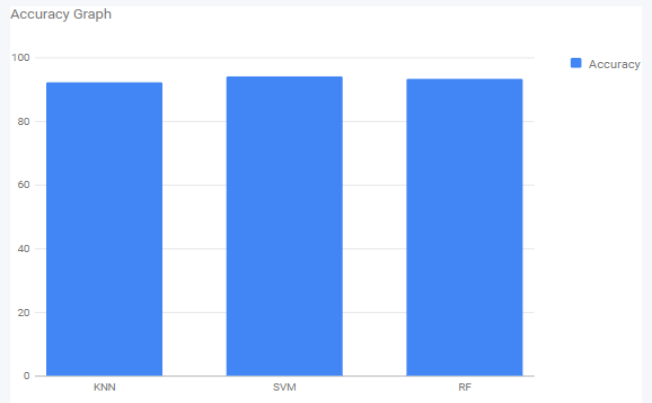

Fig4. Accuracy Graph

Above fig 4 displays the accuracy graph. The accuracy gets plot in a bar graph form.

4. CONCLUSION

The purpose of the project is to decide various ML techniques that are useful in effective heart disease predication. We have considered 14 essential attributes and applied three Machine Learning classification techniques KNN, SVM and random forest are algorithm. The main criterion is to get efficient and accurate predication with mention features and testing it. The data is pre- processed and then applied in the model. By analyzing the accuracy of all three algorithms we come to the conclusion that SVM algorithm displays the best result to predict the heart attack. In future we can use more complex machine learning algorithms and deep learning approach to build a model to get better result with higher accuracy in early stage of predication.

ACKNOWLEDGEMENT

We wish to express our deep sense of gratitude to our project guide T M GEETHANJALI Assistant Professor, Department of Information Science & Engineering (IS&E), PES College of Engineering for her guidance and useful suggestions throughout our project work and we are thankful for her time, effort and also for her continued encouragement in preparing this research paper. We also grateful to M L ANITHA (Professor, HOD, IS&E) for her insightful comments and reflection on previous version of this report. We want to show our gratitude to our family for their unwavering moral support. We want to express our gratitude to every one of our friends for their support and encouragement during the writing of our research paper.

REFERENCES

1. Thankgod Obasi, M. Omair Shafiq, "Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attack and Diseases", 2019 IEEE International Conference on Big Data(Big Data)

2. Sushmita Manikandan, "Heart Attack Prediction System", International Conference on Energy, Communication, Data Analysis and Soft Computing (ICECD-2017)

3. SY Huang, AH Chen, CH Cheng, PS Hong and EJ Lin, HDPS: Heart Attack Prediction System in March2012, published in IEEE paper journal with reference ID 12591935.

4.Procheta Nag, Saikat Mondal, Foysal Ahmed, Arun More, M.Raihan, A Simple Acute Myocardial Infarction (Heart Attack) Prediction System using clinical data and data mining techniques published in: 2017 20th International Conference of Computer and Information Technology (ICCIT) 22-24 December,2017.

5. Senthil Kumar Mohan, Chandrasegar Thirumalai and Gautam Srivastva. "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" published on (June 19, 2019 IEEE).

6. Kaggle Dataset Link
https://www.kaggle.com/namanmanchanda/heart-attack-eda-prediction-90-accuracy

7 Data World Dataset Link

https://data.world/informatics-edu/heart-disease-prediction