Heart Disease Prediction Using Machine Learning Technique

[1] **Madhuri Rajkumar Howal**   [2] **Prof. Antosh. M. Dyade,**

[1] Computer Science & Engineering SVERI's College of engineering Pandharpur

[2] Computer Science & Engineering, SVERI's College of engineering Pandharpur

Abstract: - Diseases associated with Heart i.e. Cardiovascular Diseases (CVDs) are the primary motive for the wide variety of deaths within side the route of the maximum current couple of a long time and has advanced because the maximum perilous ailment, in India and within side the complete world. In this way, there may be a want for accurate, viable and dependable device to research such maladies in time for valid treatment. Machine Learning algorithms and processes were applied to numerous clinical datasets to numerous clinical datasets to research of widespread and complicated information. Numerous analysts, as of late, were the usage of numerous techniques to allow the fitness care enterprise and the specialist's within side the prognosis of coronary heart associated diseases. This paper demonstrates a survey of numerous fashions primarily based totally on such algorithms and strategies and analyzes their performance. Models rely on supervised mastering algorithms along with Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, Decision Trees (DT), Random Forest (RF) and ensemble fashions are located extraordinarily distinguished the various researchers.

Keywords— **Cardiovascular, datasets, supervised learning algorithms, Support Vector Machines, K-Nearest Neighbors, Naive Bayes, Decision Trees, Random Forest**

I. INTRODUCTION

Data Mining is a non-minor extraction of certain, beforehand obscure and potentially valuable information about data [1]. In short, it is a procedure of analyzing information from the substitute perspective of view and assembling the knowledge of it [2]. The discovered information can be used for various applications, for example healthcare industry. The Healthcare industry is"data rich", however lamentably not every one of the information is dug which is required for finding hidden patterns and effective decision making. Data Mining Techniques such as Propelled data mining techniques are

Utilized to find learning in the database and for medicinal research, especially in the Heart disease prediction. A major challenge facing the healthcare industry is the nature of the administration. Poor analysis can prompt appalling outcomes which are unacceptable. The datasets are overwhelming for human personalities to fathom, can be effectively investigated utilizing different machine learning techniques. Accordingly, these algorithms have become very useful, in recent times, to predict the presence or absence of heart related diseases accurately. Now a day's doctors are embracing many Scientific technologies. Our project's objective is to foresee the odds of heart disease based on the patient's dataset and the doctor's perspective in check-up using machine learning. By utilizing the patient's medical records, a new system is proposed to foresee the chances of heart attack Attributes such as Blood pressure (bp), age, thickness of the artery, etc. are sustained into the dataset and algorithm [3].

II. LITERATURE SURVEY

The motivation to do this problem comes from the World Health Organization estimated. As per the World Health Organization estimation till 2030, practically 23.6 million people will pass on due to Heart illness. Along these lines, to limit the threat, desire for coronary disease should be done. Investigation of coronary sickness has been regularly in perspective on signs, appearances and physical examination of a patient.

The most troublesome and complex assignment in medicinal services area is finding of right ailment or right illness. In late patterns investigation on these broad datasets has been able to be fundamental because of monetary weights on medicinal services commercial enterprises.(business endeavors).

Chen et al. Proposed the prediction system for heart disease, i.e. coronary illness using learning vector quantization algorithm. Another study probed on sample database of patients' records. The Neural Network is prepared, tested, trained with 13 input factors such as Age, Blood Pressure, Kumaravel et al. Have proposed automatic diagnosis framework for heart diseases using a neural network system with an accuracy of 63.6–82.9%. The heart is an important organ of human body part and it is similar to a pump, which circulates blood through the body [4] and If the circulation of blood in the body is inefficient, then vital organs like brain suffer and if heart quits working, demise happens within minutes. Life is totally reliant on the successful working of the heart. The term Heart disease alludes to disease of heart and blood vessel

framework inside 51it. Several factors have been demonstrated that increases the chances of Heart disease: Family history, Smoking, Poor diet, High blood pressure (Hyper Tension), High blood cholesterol (Caused by Obesity), Physical inactivity. True assurance of coronary sickness can't be conceivable by utilizing just human comprehension. There are stores of parameters that can impact the precise end like less exact outcomes, less experience, time subordinate execution, data up degree and substantially more.

III. PROPOSED SYSTEM ARCHITECTURE

MATH OR EQUATION

We can compute distance i.e. gap between two points using distance functions as d(x,y) where x, y are scenarios composed of features, such that x=x1,. . . ..,xN, y=y1,. . . . . . ,,yN. Two functions can be used:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

We use Manhattan Distance if we need to calculate the distance between two data points in a grid-like path. As mentioned above, we use the Minkowski distance formula to find Manhattan distance by setting p's value as 1.

$$d = \sum_{i=1}^{n} |x_i - y_i|$$

Where, n- number of variables, xi and yi are the variables of vectors x and y respectively, in the two-dimensional vector space. i.e. x = (x1,x2,x3,…) and y = (y1,y2,y3,…).Here we have used Euclidean distance measuring function to find distance between two points, absolute distance measuring is also called as Mahnattan distance

KNN Implementation

KNN classifier algorithm stands for k-nearest neighbor algorithm. This classifier is utilized to partition dataset parameters in different clusters based on the distance between points. In 1951, Hodges et al. introduced a nonparametric technique for pattern classification which is popularly known the Nearest Neighbor rule. K-Nearest Neighbor technique is a standout amongst the most rudimentary but very effective classification techniques. It makes no suspicions about the data and is generally be used for classification tasks when there is very less or no prior knowledge about the data distribution. This algorithm involves finding the k nearest data points in the training set to the data point for which a target value is unavailable and assigning the average value of the found data points to it

This is the classifier function for KNN .It uses following functions:

a) n neighbours: It hold the value of k which is an integer value which needs to be passed if required as user defined otherwise its value is 5 by default.

b) data.fit(): It is used to fit the model in which has two parameters as X, Y where X is training data with features and Y is training data with label. Predict().

c) Accuracy score: It is the function which is used to find accuracy of KNN algorithm. Accuracy means the proportion of the correctly predicted data points to all the anticipated data points.

Advantages of KNN are that cost of the learning procedure cost is zero, No presumptions about the attributes of the ideas to learn must be done and complex ideas can be learned by nearby guess utilizing basic strategies, it tends to be considered as baseline method.
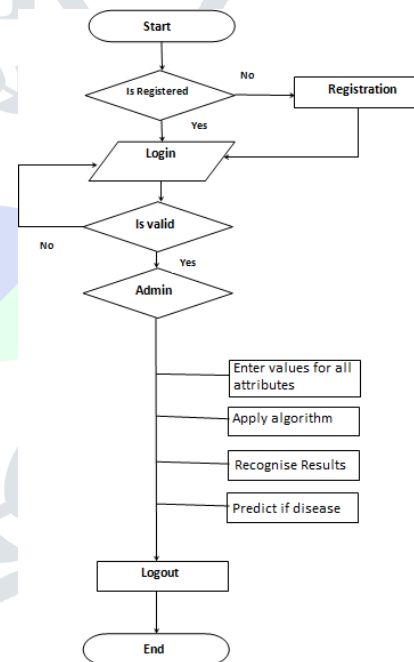

Fig. 1. Flowchart of our Proposed System

Fig1. Represents flowchart of our project which explains that from starting, after checking that person is registered or not then login page is provided after successful login. Then admin module starts which includes steps as entering value, applying algorithms, computing results and then predicting disease.

| Sr. No.	Attributes	Description	Values
1	Age	Age in years	continuous
2	Sex	Male or female	1=male 0=female
3	Cp	Chest Pain type	1=typical type 2=typical type angina 3=non-angina pain 4=asymptomatic
4	Thestbps	Resting blood pressure	Continuous value in mm hg
5	Chol	Serum Cholesterol	Continuous value in mm/dl
6	Restecg	Resting electrographic results	0=normal 1=having ST T wave abnormal 2=left ventricular hypertrophy
7	Fbs	Fasting blood sugar	1¿=120 mg/dl 0¡= 120 mg/dl
8	Thalach	Maximum heart rate achieved	Continuous value
9	Exang	Exercised induced angina	0=no 1=yes
10	Oldpeak	ST depression induced by exercise relative to rest	Continuous value
11	slope	Slope of the peak exercise ST segment	1=unsloping 2=flat 3=downsloping
12	Ca	Number of major vessels colored by floursopy	0-3 value
13	thal	Defect type	3=normal 6=fixed 7=reversible defect

TABLE I ATRRIBUTES FOR SYSTEM
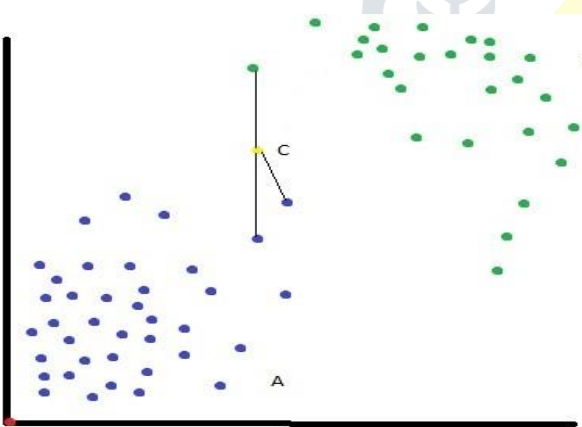
Fig. Architecture of System

Fig 3.C. represents graph of KNN algorithm in which two classes are differentiated in different clusters shown in different color. New point which needs to be predicted in which cluster it is C. Heart diseases when aggravated spiral way beyond control. When the early symptoms of heart diseases are ignored, the patient might end up with drastic consequences in a short span of time [7].

Data Sources: The datasets are gathered from the four following locations: 1. Cleveland Clinic Foundation (Cleveland. Data) [10] 2. Hungarian Institute of Cardiology, Budapest (Hungarian. Data) 3. V.A. Medical Centre, Long Beach, CA (long-beach-va .data) 4.University Hospital, Zurich, Switzerland (Switzerland. Data) 5. Statlog Heart Disease database [11]. All attributes are numeric-esteemed.

IV. RESULT & ANALYSIS

After evaluating attributes with different algorithms, i.e. we came to result that k-nearest neighbor (KNN) is easier to implement than Support vector machine (SVM). The classification of multiclass data is easier in this algorithm. While checking accuracy score, we observed that even SVM has slightly higher accuracy score than KNN then also KNN is useful due to multiclass data in taken datasets for sample. . However, it is always suggested to exercise daily and discard unhealthy habits at the earliest [6]. In future an expert system may be developed that can lead to the selection of proper treatment methods for a patient diagnosed with heart disease

CONCLUSION

Heart diseases are difficult to cure, and it takes away lots of lives every year. Sedentary lifestyle and excessive stress are major reason which worsened the situation. If the disease is detected early, then its curing becomes easy [5] .There are few treatment methods for a patient once diagnosed with a form of heart disease. Data mining can be of very good help in deciding the line of treatment from such suitable machine learning techniques, datasets and algorithms. Datasets dealing with the same medical issue as Coronary artery disease (CAD) may demonstrate distinctive outcomes while applying the similar machine learning technique. The classification results and accuracy score result the selected important features are based mainly on the efficiency of the medical diagnosis and analysis [6]

REFERENCES

[1] K. Sudhakar, Dr.M. Manimekalai "Study of Heart disease prediction using data mining" ISSN: 2277 128X

[2] M. Revathi "Review of Heart Disease Prediction using Data mining techniques"

[3] Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, August 2008

[4] Ankita Pimputkar, Jitendra S. Dhobi "A Survey on Heart Disease Prediction using Hybrid Technique in Data Mining" IJARIIE-ISSN(O)-2395- 4396

[5] Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee "Heart Diagnosis and prediction using machine learning and data mining : Review"

[6] Burak Kolukisa , Hilal Hacilar, Gokhan Goy, Mustatfa Kus, Burcu BakirGungor, Atilla Aral, Vehbi Cagrigungor "Evaluation of classification algorithms, Linear discriminate

Analysis, and a New hybrid Feature selection methodology for the diagnosis of Coronary Artery Disease"

[7] S. B. Patil and Y. S. Kumaraswamy, "Intelligent and effective heart attack prediction system using data mining and artificial neural network," European Journal of Scientific Research, 2009, p. 642-656.

[8] Quinlan, J.R, "Induction of decision trees". Journal of Machine Learning, 1986, pp. 81-106. 53

[9] Hanen Bouali and Jalel Akaichi et al. "Comparative study of Different classification techniques, heart Diseases uses Case." 2014 13th International Conference on Machine Learning and Applications

[10]Clevelanddatabase:http://archive.ics.uci.edu/ml/datasets/ Heart+Disease

[11] Statlog database e:http://archive.ics.uci.edu/ml/machine Learning databases/statlog/heart/

[12] Aljaaf AJ, Al-Jumeily D, Hussain AJ, Dawson T, Fergus P, Al-Jumaily M. Predicting the likelihood of heart failure with a multi-level risk assessment using decision tree. Third international conference on technological advances in electrical, Beirut, Lebano

[13] http://www.nhlbi.nih.gov/health/healthtopics/ topics/cad/

[14] https://ieeexplore.ieee.org/document/8093512/

[15] Heart Disease Prediction using machine learning techniques : A Survey: V,V Ramalingam, Ayantan Dandapath, M.Karthik Raja

[16] Prediction of Heart Disease using data mining and machine learning algorithms and techniques , Nikhil Kumar Mutlaya, K.V.s. Kaushik DOI: 10.13140/RG.2.2.18185.75362