International Journal of Humanities and Social Science Invention (IJHSSI)
ISSN (Online): 2319 – 7722, ISSN (Print): 2319 – 7714
www.ijhssi.org ||Volume 10 Issue 9 Ser. II || September 2021 || PP 08-13

Heart Disease Prediction Using Intelligent Machine Learning System

[1] SHIVARAJ KUMAR T H [2] CHANDANA M G
[1] Assistant Professor, Department of Computer Science,
Bengaluru North University, Kolar
[2] PG Student, Department of Computer Science,
Bengaluru North University, Kolar

Abstract- *Heart disorder is one of the most complicated and life threatening disease, when we talk about detection of diseases related to heart it has to be done very efficiently as it is the basis on which entire treatment process will be decided .This project proposes the systematic process for predicting and diagnosing the cardio vascular disease and few more diseases of heart. In this project work, our concentration is on predicting and diagnosing the heart disorder using SVM, Decision tree and K-means algorithm, which provides various significant attributes in the medical literature .Those attributes are been used in the algorithm along with an dependent variable which can have various values 0 or 1(i,e. test positive or negative).The number of peoples taking up the test will be divided into a required ratio(eg:80:20)out of which 80 will be given for training model and 20 will be given for testing model.The database which will be collected from the medical organization will be compared with the training components and here the patients will be categorized .The classified output will be given to the testing model which will compare this result with the result generated by the dependant variable when these values matches the heart disorders will be predicted with good accuracy.*

Keywords *– Heart disorder, SVM, K-means, Decision tree.*

Date of Submission: 01-09-2021 Date of Acceptance: 14-09-2021

I. INTRODUCTION

Heart disease is a major life threatening disease that can cause either death or a serious long term disability. However, there is a lack of effective tools to get a hidden relationships and trends in a e-health data. Medical diagnosis is a complicated task and plays a vital role in saving human lives so it needs to be executed accurately and efficiently. An appropriate and accurate computer based automated decision support system is required to reduce the cost for achieving clinical tests. This provides an insight into machine learning techniques used in diagnosing various diseases. Various data mining classifiers have been discussed which has emerged in recent years for efficient and effective disease diagnosis.

The Heart is one of the most important organs in the human body. It is the center of the circulatory system. The heart functions as a pump that propels blood to different parts of the human body through a network of blood vessels, supplying a constant supply of oxygen as well as other vital nutritional components. If the heart ever stops functioning and ceases to pump blood, the body will shut down and within very less time a person will expire. The usage of information technology in health care industry is increasing day by day to aid doctors in decision making activities. It helps doctors and physicians in disease management, medications and discovery of patterns and relationships among diagnosis data. Current approaches to predict cardiovascular risk fail to identify many people who would benefit from preventive treatment, while others receive unnecessary intervention. Machine-learning offers opportunity to improve accuracy by exploiting complex interactions between risk factors. We assessed whether machine-learning can improve cardiovascular risk prediction.

Our study shows that artificial intelligence could significantly help in the fight against it by improving the number of patients accurately identified as being at high risk and allowing for early intervention by doctors to prevent serious events like cardiac arrest and stroke. Based on their results, it is clear that artificial intelligence and machine learning techniques have a key role in fine-tuning risk management strategies for individual patients.

Data Mining techniques available namely Classification techniques involving Naive Bayes (NB), Decision tree (DT), Neural network (NN), Genetic algorithm (GA), Artificial intelligence (AI) and Clustering algorithms like K-NN, and Support vector machine (SVM). Cardiovascular sickness is one of the most important human-threatening and life qualities reducing disease. Heart failure is the first cause of admission by healthcare professionals in their clinical practice. Building an effective disease management strategy requires analysis of large amount of data, early detection of the disease, assessment of the severity and early prediction

of adverse events. This will inhibit the progression of the disease, will improve the quality of life of the patients and will reduce the associated medical costs. Diagnosis is basically based on patient's Electro cardiogram (ECG), Echocardiography (ECHO) tests results and doctor's experience" Applying machine learning, Deep learning methods in heart disease research is a best approach for disease diagnosis, prediction, management and other related clinical administration aspects. Disease prediction and decision making plays a significant role in medical diagnosis. Hence, in the framework of this study, efforts were made to review the current literature on machine learning and data mining approaches in heart disease research. Efficient data capturing from health records can be achieved using deep learning techniques.

II. PROPOSED METHOD

Heart disease prediction using Machine learning algorithms is detecting whether the patient is having heart disorder or not by SVM and Decision tree and to suggest treatment level using K-means algorithm.

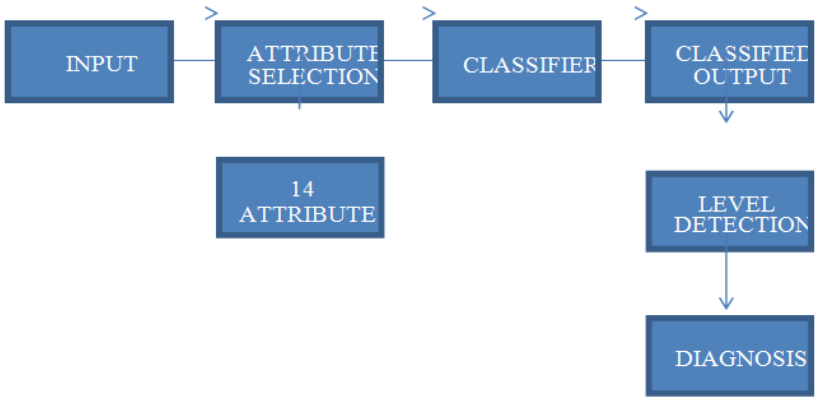

Figure 1: *Block diagram*

Input

The database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).

The names and social security numbers of the patients were recently removed from the database, replaced with dummy values. One file has been "processed", that one containing the Cleveland database. All four unprocessed files also exist in this four unprocessed files also exist in this directory. To see Test Costs (donated by Peter Turney), please see the folder "Costs".

ATTRIBUTE SELECTION

The database contains 76 attributes ,but all published experiments refer to use a subset of 14 attributes. The 14 selected heart disease attributes are age, sex, cp(chest pain), Trestbps, Chol, fbs, exang maximum heart rate, thalach, old peak ST, slope, ca, thal, target, human lives so it needs to be executed accurately and efficiently, supplying a constant supply of oxygen as well as other vital nutritional components. The selected 14 attributes are listed below .

Name	Type	Description
Age	Continuous	Age in years
Sex	Discrete	0=female 1=male
Cp	Discrete	Chest pain type: 1=Typical angina 2=atypical angina 3=non-anginal pain 4=asymptom
Trestbps	Continuous	Resting blood pressure (in mm Hg)
Chol	Continuous	Serum cholesterol in mg/dl
Fbs	Discrete	Fasting blood sugar>120 mg/dl: 1=true and 0=False

Exang continuous heart rate achieved	Discrete	Exercise induced angina: 1 = Yes and 0 = No
Thalach	Continuous	Maximum heart rate Achieved
Old peak ST	Continuous	Depression induced by exercise relative to rest
Slope	Discrete	The slope of the peak exercise segment :1 = up sloping 2 = flat 3 = down sloping
Ca	Continuous	Number of major vessels colored by fluoroscopy that ranged between 0 and 3.
Thal	Discrete	3 = normal 6 = fixed defect 7= reversible defect
Name	Type	Description
Target	Discrete	Diagnosis classes: 0 = No Presence 1=Least likely to have heart disease 2= >1 3= >2 4=More likely have heart disease

Classifier

Research on data mining has led to the formulation of several data mining algorithms. These algorithms can be directly used on a dataset for creating some models or to draw vital conclusions and inferences from that dataset. Some popular data mining algorithms are Decision tree, Naïve Bayes, k-means, artificial neural network etc.

Decision Tree

A Decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible

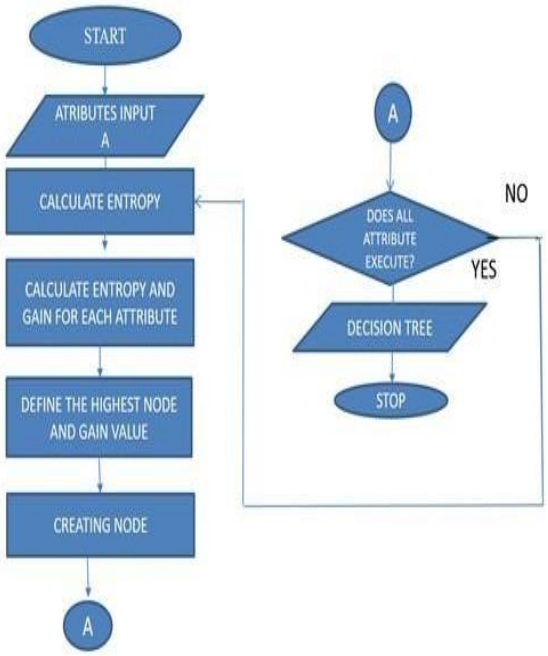

Figure 2: Block Diagram of Decision Tree

consequences including chance event outcomes and utility. It is one of the ways to display an algorithm. Decision trees are commonly used in operations research, specifically in decision analysis to help and identify a strategy that will most likely reach the goal. It is also a popular tool in machine learning. A Decision tree can easily be transformed to a set of rules by mapping from the root node to the leaf nodes one by one.

Finally by following these rules, appropriate conclusions can be reached.

- Formula to calculate entropy of the decision tree is;

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

Where; pi-probability of the data points, i-data points

- Since we have only two vales for c either positive or negative so,

E(s)=[-p+ log2 p+]+ [-p- log2 p-]

where p+**:**Data set with positive result p-:Data set with negative result

- Formula to calculate gain is;

$$I_G = 1 - \sum_{j=1}^{c} p_j^2$$

Where: Ig-is information gain

Pj-Probability of data point with highest entropy

Support vector Machine

It is a supervised learning method which classifies data into two classes over a hyper plane. Support vector machine performs a similar task like C4.5 except that it doesn‟t use Decision trees at all. Support vector machine attempts to maximize the margin (distance between the hyper plane and the two closest data points from each respective class) to decrease any chance of misclassification. plane that maximize the majority where peoples taking up the test are predicted to have the disorder or not. This output solely depends on the classifier if the algorithm gives test positive than the person is said to have heart disordered else the person does not have heart disorder.

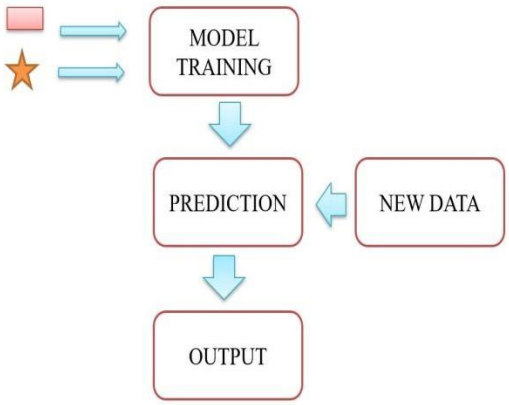

Figure 3: Block Diagram of SVM

LEVEL DETECTION

Once the heart disorder is predicted the next step is to detect the level or stage of the disorder which is done with the help of k-means algorithm.

K-means creates k groups from a set of given objects so that the members of a group are more similar. Other than specifying the number of clusters, k-means also “learns” the clusters on its own without any information about which cluster a particular observation should belong to. Thats why k-means can be called as semi-supervised learning method. K- means is specially effective over large datasets. K Nearest Neighbor

algorithm is used for classification as well as for regression. Here k is the data set items considered for the Classification.

K-means algorithm k-mean create k group from a set of given objects so that the members of a group are more similar. Other than specifying the number of clusters k-mean also learns the clusters on its own without any information about which cluster particular.

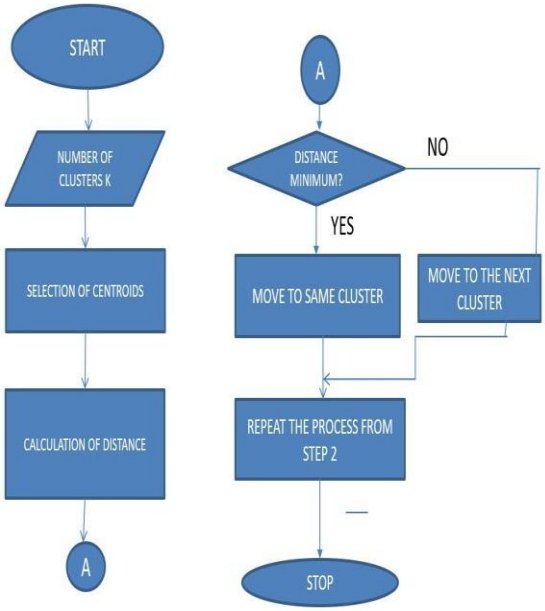

Figure 4: Block Diagram of KNN

- The distance in k-means algorithm is calculated using euclidean distance formula;

$$d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.$$

- Formula to measure the new centre point is;

$$\boldsymbol{v}_i = (1/c_i) \sum_{j=1}^{c_i} \boldsymbol{x}_i$$

Where; Ci-No. of data points in ith cluster
Xi-Minimum dtance from the centre point

DIAGNOSIS

Diagnosis is the output or the result stage which tells all about the patients detected with heart disorder with the level of there disease. An appropriate and accurate computer based automated decision support system is required to reduce cost for achieving clinical tests. This paper provides an insight into machine learning techniques used in diagnosing the heart disease accurately with reduced cost of clinical test.

FLOWCHART

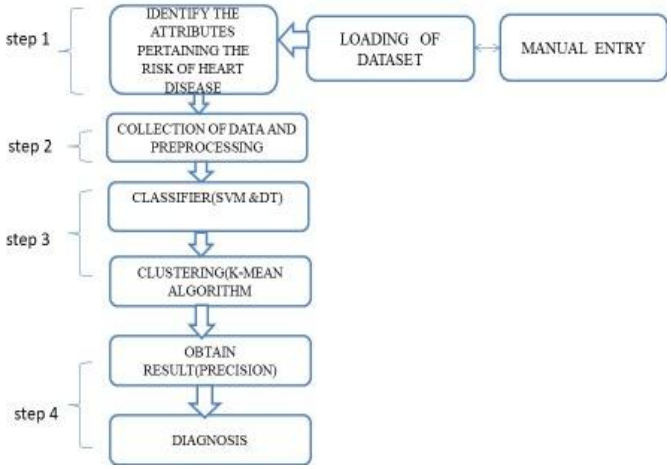

Figure 5: Flowchart of System

- The flowchart is divided into four parts: 1.Understaning domain data

2.Data preparation 3.Preprocessing modules 4.Evaluation and diagnosis

Here, understanding domain data is a step which selects the path to OS directory which is to ensure that the project requirements are met.

III. CONCLUSION

Heart Disease is one of the most complicated and life threatening disease, the detection of disease related to heart has to be done very efficiently. This project proposes the systematic process of predicting and diagnosing cardio vascular disease based on 14 attributes that relates the heart disease. The main objective of the project was to detect the heart disease

with better accuracy using machine learning algorithm. Decision tree algorithm and SVM algorithm where are used to predict the disorder in which for comparative analysis data was taken from an open source medical website. Later in order to find the stage of disorder k-means algorithm is used, which would group the patients in their respective category. This project uses Machine Learning Algorithm like Support Vector Machine (SVM) and Decision Tree to classify whether person is having disease are not, further it uses K – Means clustering algorithm to group the patients into level of disease, in this project we have considered numbers of levels to be there.

REFERENCES

[1]. Intelligent heart disease prediction system using machine learning by Tanvi Sharma , Sahil Verma , Kavita : vol 4 , 2017
[2]. Neural network based intelligent system for predicting heart disease by K.Subadra , Vikas : vol 8 , 2019
[3]. Heart disease prediction using data mining techniques by H.Bejamin , Anatony Belcy : vol 9 , 2018
[4]. Prediction of heart disease by clustering and classification by Reetu Singh , E.Rajesh : vol 7 , 2019
[5]. Heart disease prediction using data mining techniques by Amandeep Kaur , Jyoti Arora : vol 9 , 2018
[6]. Predictive data mining for medical diagnosis ; by Jyoti soni , ujma Ansari Dinesh : vol 11 , 2019
[7]. K. Subhadhra Vikas B presented "Neural Network Based Intelligent system for predicting Heart Disease". ". IEEE Computer Society Conference on computer vision and pattern recognition, vol. 2, pp. 446-451, 2018
[8]. H. Benjamin Fredrick David and S. Antony Belcy presented " Heart Disease Prediction using Data Mining Techniques". A Review, International Journal of Advanced Studies in Computer Science and Engineering IJASCSE, Vol 3. Issue 8, 2019.
[9]. Reetu Singh1, E. Rajesh presented" Prediction of Heart Disease by clustering and classification techniques".
[10]. Amandeep kaur, Jyothi Arora presented, "Prediction of Heart Disease using Data Mining Techniques". , ISSN 0973-4562 Vol. 10 No.30 (2015).
[11]. Himanshu sharma, M A Rizvi presented "Heart Disease Prediction using Machine Learning and Deep Learning Techniques".
[12]. Jaymin patel, prof. Tejal Upadhyay, Dr.Samir patel presented." Heart Disease Diagnosing and Prediction using Machine Learning and Deep Learning Techniques".
[13]. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmitha Mukherjee presented. "Heart Disease with Risk Prediction Using Machine Learning Algorithms". VOL. 50, NO. 10, OCTOBER 2012.
[14]. S. Kavitha, K, R. Baskaran, S. Sathyavathi presented."Predective Data Mining for Medical Diagnosis". VOL.1, NO.3.JULY- 2017
[15]. Jyoti Soni, Ujma Ansari, Dipesh Sharma and Sunita Soni presented." Heart Disease Prediction using K-means Algorithm Using Machine Learning". published in International Journal of Applied Engineering Research, ISSN 0973- 4562 Vol. 10 No.30 (2015)

SHIVARAJ KUMAR T H , et. al. "Heart Disease Prediction Using Intelligent Machine Learning System." *International Journal of Humanities and Social Science Invention (IJHSSI),* vol. 10(09), 2021, pp 08-13. Journal DOI- 10.35629/7722