International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-10, August 2019

Machine Learning Based Diagnosis and Prediction System for Congestive Heart Failure

Niharika Saxena, L.S.Maurya

Abstract: ***Recently, heart failure has become one of the major Causes of death. By 2030, if it is not controlled the toll will rise to twenty three million. Cardiologist can predict the disease with 70 % accuracy. Considering the limitation of cardiologist, a system can be provided to them to predict the disease with more accuracy. Machine Learning is frequently used in to days world to support healthcare industry. ML provides new opportunity to analyze the data with more accuracy. It bridges the gap between medical science and technology. Decision tree is one of the best classification techniques of machine learning which will analyze the data and predict the disease with accuracy. The main objective of my dissertation work is to predict the disease and analyze the result. So in this research work the DT technique is used for the prediction of disease and it gave result with more accuracy on comparison to previous work. Hence this study proved that DT algorithm gives the result with more accuracy in less time of execution. This research work is a growing range of efficient tools to assist healthcare industry and medical professionals for the betterment of patients.***

Keywords: Machine Learning; Healthcare Prediction system; Machine Learning Techniques; Classification.

I. INTRODUCTION

The technological age of today is the machine learning age. Previously in the age of organizing systems (1900s — 1940s) and the era of computer science (from the 1950s), computer systems could only conduct iterative, conditional and logical computations. Such processing systems, however, were created with the emergence of fresh techniques that could learn without implicit programming to execute particular tasks. This technology is called Machine Learning where systems take input from past experiences and make predictions on the basis of certain algorithms [1].

Heart Failure is one of the most basic human ailments on the planet and influences human life in all respects severely. For the prevention of heart failure, accurate and timely diagnosis of heart disease is essential. In many respects, the diagnosis of heart disease through traditional medical history was regarded unreliable. The methods of machine learning can quickly identify and classify people with heart disease [3].

In most cases, the diagnosis of heart disease relies on a complex mixture of clinical and pathological data, resulting in excessive medical costs impacting medical care. Thus, statistics and machine learning are two primary approaches to predicting the status of heart disease based on clinical information expression [2].

Manuscript published on 30 August 2019.
***Correspondence Author(s)**
Niharika Saxena, M.Tech, Dept. Of CSE SRMSCET Bareilly, India
Dr L.S.Maurya, Head of Department CSE & IT SRMSCET Bareilly, India

This research therefore demonstrated that the DT algorithm provides the outcome in less execution time with greater precision. This study work is a increasing range of effective tools to support patients ' improvement in the healthcare industry and medical practitioners.

II. LITERATURE REVIEW

1. In this article they researched multiple machine learning algorithms. Then these algorithms can be used for early forecast of various illnesses and epidemics in the field of healthcare. Multiple self-learning systems can be created using these algorithms to help physicians diagnose various illnesses faster and easier. Decisions can be made for the scope of effective outcomes related to patient healthcare.
2. This work gives the heart patients a better diagnosis than the prior one. We develop a prediction model for heart disease that can help medical professionals predict the status of heart disease based on patient clinical data. On the basis of these clinical data, they used J48 decision tree to classify heart disease against unpruned, pruned and pruned with a reduced approach to error pruning.
3. A hybrid smart machine-learning-based predictive scheme for the diagnosis of heart disease was suggested in this research study. On the Cleveland heart disease dataset, the scheme was evaluated. Seven well-known classifiers such as logistic regression, K-NN, ANN, SVM, NB, DT, and random forest were used to select significant characteristics using three feature selection algorithms Relief, mRMR, and LASSO.
4. In this paper, we have taken four administered machine learning calculations and thought about their competency as far as the exactness accomplished by them in anticipating the event of coronary illness. We look at what's more, differentiate the condition of workmanship frameworks utilized for Heart malady determination that utilization machine learning.
5. This paper meant to dissect the utilization of information mining in medicinal area and a portion of the calculations used to foresee maladies. It is seen that outcomes may fluctuate for various malady conclusion dependent on the devices and procedures utilized. Information mining gives great outcomes in ailment conclusion when suitable apparatuses and strategies connected.
6. This works include two undertaking in pipelined way to distinguish and removing the connection between the given MEDLINE dynamic. Firs errand include finding most reasonable model for expectation, the second undertaking is to discover great information portrayal. To accomplish this two

errand different prescient calculations and printed portrayal systems are considered.

7. In this Work author produced Systematic attempts to design a scheme that results in disease forecast such as diabetes. Three classification algorithms for machine learning are studied and evaluated on different measures during this work. Pima Indians Diabetes Database experiments are conducted. Using the Naive Bayes classification algorithm, experimental results determine the adequacy of the designed system with a achieved accuracy of 76.30 percent.
8. This paper surveys different Data Mining methods, for example, order, bunching, affiliation, relapse in wellbeing space. It additionally features applications, difficulties and future work of Data Mining in medicinal services.
9. Author discussed that Machine learning strategies have turned out to be progressively mainstream in directing prescient examination because of their remarkable execution in taking care of extensive scale datasets with uniform qualities and boisterous information.
10. In this paper Author concluded that human services is one of the quickest developing parts in the present economy ; more individuals require care and it is winding up increasingly costly. Machine learning can possibly support both patient and suppliers as far as better consideration and lower cost.

III. MACHINE LEARNING TECHNIQUE FOR PREDICTION

Machine learning is a utilization of Artificial Intelligence that paves the way for entirely new ability. It takes input and gives result without any need of modification by human. The core part of Machine Learning is to learn from past data experiences and make improvement as necessary for further references. The quantity and quality of data determines the effectiveness of the model. The cornerstone is to make machine learn automatically without any human interaction. Hence, many industries are getting benefit from it without any expense on human resource.

1. Logistic Regression: A regression of logistics is an algorithm of classification. For numeric classification issue, 0 is adverse class and 1 is favorable class to predict the value of predictive variable y when y [0, 1] is negative class. It also utilizes multi classification when y [0, 1, 2, 3] predicts the value of y [3].
2. Support Vector Machine: SVM is one of the common collections of classification-based supervised machine learning model. The objective of a support vector machine is to find the best highest margin separating hyper plane between the two classes, given a two-class learning data set [2].
3. Naïve Bayes: The NB is a learning algorithm supervised by classification. To determine the class of a fresh feature vector, it is based on conditional probability theorem. For a given class, the NB uses the training dataset to find out the conditional parameter probability value. The fresh vector class is calculated based on its probability of conditionality after calculating the probability conditional value of each vector. NB is used for classification of text-related problems [3].
4. Artificial Neural Network: The artificial neural network is a supervised algorithm for machine learning and is a mathematical equation that embeds messaging neurons. The ANN has three parts including features for inputs, outputs, and transfer. ANN is intended through neuron inclusion. This distinct neuronal mixture from different buildings is the same as multilayer perception [3].
5. Decision Tree: Decision tree algorithm is a supervised algorithm for learning the machine. It utilizes a decision tree to map objects to some predictive findings. It is based on the model of classification that categorizes the information input into an outcome [1]. A shape of a decision tree is a tree where each node is a node of leaf or a node of choice. A decision tree contained interconnected inner and external nodes. The inner nodes are the component of decision making that causes a decision to visit the next nodes and the child node. On the other side, the leaf node has no kid nodes and is linked to a label [3]. The algorithm's accuracy relies on the attributes chosen to train the machine.
6. K-NN: K-NN is an algorithm for supervised classification of learning. K-NN algorithm predicts a fresh input class label in the training sets; K-NN uses the relation of fresh input to its input samples. The samples in the training set are the same if the fresh entry. The output of classification K-NN is not nice. Let (x, y) be the results of practice and the role of learning h: X ⟶ Y, so that with observation x, h(x) can determine the value of y.

IV. METHODOLOGY

A. Datasets and Data Preprocessing: We took Cleveland heart dataset, which is accessible at UCI, beautifully prepared for this having 75 attributes with their domain values associated with them used for predict the disease. Following figure 1 shows the no of attributes and their description.

Sr. no	Attribute	Description
1	Age	Age in years
2	Sex	Male or Female
3	Cp	Chest pain type
4	Thestbps	Resting Blood Pressure
5	Chol	Serum cholesterol
6	Resecg	Resting electrographic result
7	Fbs	Fasting Blood Pressure
8	Thalach	Maximum heart rate achieved
9	Exang	Exercise induced agina
10	Oldpeak	ST depression induced by exercise relative to rest
11	Solpe	Slope of the peak exercise ST segment
12	Ca	No of major vessels colored by floursopy
13	Thal	Defect type

Fig 1 Data Set Description

B. Feature Extraction: Choosing features is crucial for the machine learning technique because irrelevant features sometimes affect the classification effectiveness of the classifier for machine learning. Highlight determination improves the characterization precision and decreases the model's implementation time.

C. *Classification of Data:* A decision-making tree is a supervised machine learning algorithm. A decision tree shape is a tree in which each node is a node of choice or a leaf node. Methods of decision-making tree are simple and simple to understand to make the decision.

The key decision-making algorithm called ID3, which utilizes a greedy top-down search without backtracking through the space of feasible branches. ID3 uses Entropy to create a DT and to collect information.

Using Information Gain, Gini Index and Gain Ratio Decision Trees, you can select the splitting attribute.

Information Gain: chooses the attribute to maximize entropy. It provides the measurement of how much data function a class can provide.

Entropy (S) = p+ (- log2 p+) + p- (-log2 p-)

V. PROPOSED APPROACH

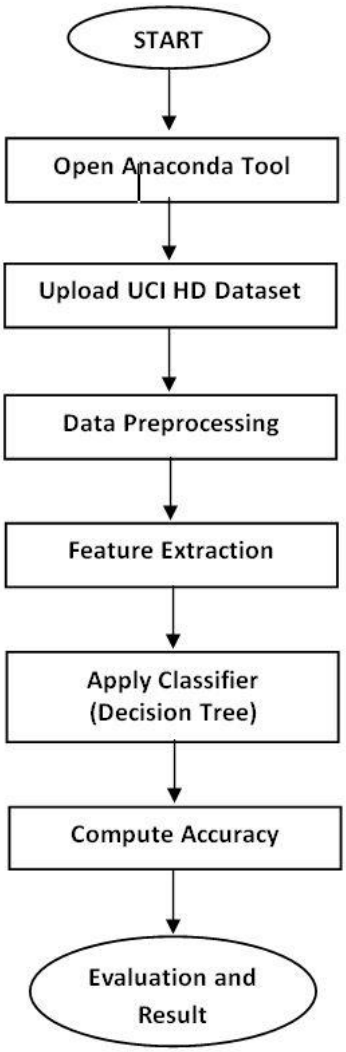

Fig 2 Proposed System

From the above figure 2 shows the proposed architecture used for our prediction system. Proposed approach consists of steps that are figured in the above diagram.

VI. RESULTS AND ANALYSIS

Here we show the utility of the forecast model to heart clinical information.

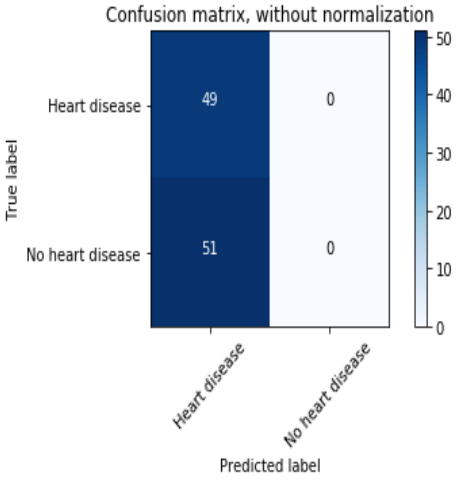

Figure 3

This figure 3 shows the matrix that predicts heart disease diagnosis. 49 percent of patients have heart disease and 51 percent do not have heart disease, according to this consequence.

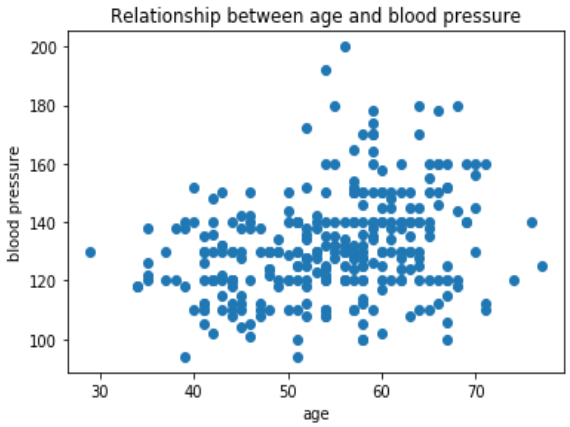

Figure 4

The connection between two characteristics is shown in the above figures 4, 5 and 6. It provides a relation between age and blood pressure in the first graph. Similarly, two other graphs demonstrate the connection between electrocardiography and age, respectively, and the maximum heart rate and age.

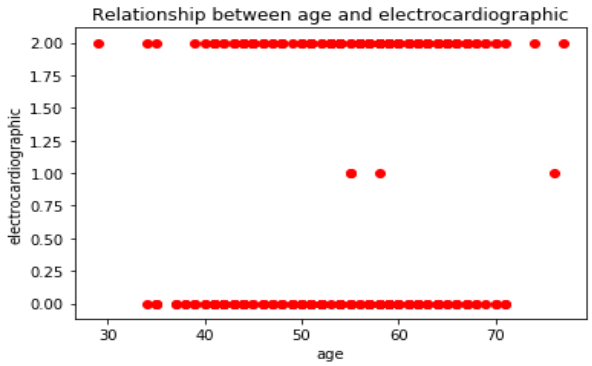

Figure 5

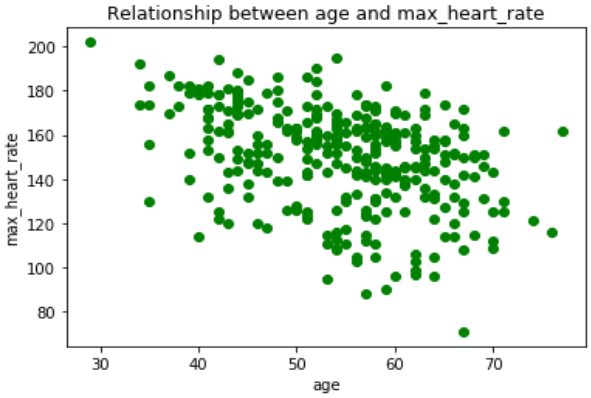

Figure 6

Result Analysis: From the consequence we mentioned above, it can be concluded that the decision tree algorithm calculated 90.32 percent precision in heart disease prediction.

```
The Accuracy is
90.32258064516128
Success Ratio
0.9032258064516128
The execution time is 0.1854689121246338 seconds ---
```

Figure 7

Success ratio is 0.9032 as shown in figure 7 and execution time is 0.1854 sec.

The assessment provides all information points in the following figure 8 CAP curve that are used to define how the decision tree works effectively. We're finally plotting the ideal model. A perfect is one that will detect all information points of class 1.0 in the same amount of attempts as the data point of class.

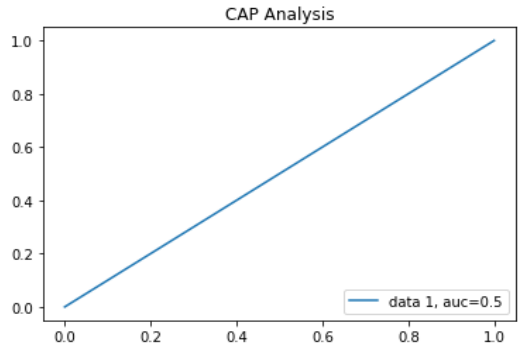

Figure 8

VII. CONCLUSION

The proposed methodology is partitioned into following fundamentals spaces and based on which following outcomes are required: Identifying the attribute-based CHF illness: here we determine whether or not the person is sick. Age, gender, cholesterol, blood sugar, etc. are the predictive features of heart disease. Using the DT algorithm, it is introduced. To determine the precision of the algorithm applied: in this chapter we discover the precision of the suggested method which is the choice tree. This outcome provides the highest precision result for the forecast system decision tree algorithm.

VIII. FUTURE WORK

To achieve more precision, the parameters can be decreased. We can also create a system for other illnesses such as diabetes; cancer; and so on. Other algorithms can also implement the scheme. We can also pave the way to get stakeholders ' attention, so we can make it more precise. Cloud computing, large data techniques, etc. can manage data storage.

REFERENCES

1. Akansha Dua and Arpit khanna, “A Study of Machine Learning Algorithms for Disease and Epidimic Prediction” in International Research Journal of Engineering and Technology (IRJET), 2018
2. Atul Kumar Pandey, Prabhat Pandey, K.L. Jaiswal and Ashish Kumar Sen, “A Heart Disease Prediction Model Using Decesion Tree” in IOSR Journal of Computer Engineering, 2013
3. Amin UI Haq, Jian Ping Li and Shah Najir, “A Hybrid Intelligent System framework for the Prediction of Machine Learning Algorithms” in Journal of Mobile Information System, 2018
4. Akanksha Gupta, Shubham Pathak and Prof. Kannadasan R ,”An Evaluation of Supervised Machine Learning Algorithms for Heart Disease Diagnosis ” in International Research Journal of Computer Science (IRJCS) (2017)
5. V.Kirubha and S.Manju Priya ,” Survey on Data Mining Algorithms in Disease Prediction”in International Journal of Computer Trends andTechnology (IJCTT) (2016
6. Pravin Shinde and Prof. samjay Jadhav ,”Health Analysis System Using Machine Learning” in International Journal of Computer Science and Information Technology(2014)
7. Deepti Sisodia and Dilip Singh Sisodia, “Prediction of Diabetes using Classification Algorithms” in International Conference on Computational Intelligence and Data Science, (2018)

8. Parvez Ahmad , Saqib Qamar and Syed Qasim Afser Rizvi , “Techniques of Data Mining In Healthcare : A Review” in International Journal of Computer Science (2015)
9. Ali Bou Nassif , Mohammad Azzeh , Shadi Banitaan and Daniel Neagu ,”Predictive Analytics Using Machine Learning” in Springer Neural Computing and Applications
10. Rohan Bhardwaj , Ankit R.Nambiar and Debojyoti Dutta ,”A Study of Machine Learning in Healthcare” in IEEE Conference (2017)