IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

Heart Diseases Detection System

Ms Lolakshi[1], Gowthami K M[2], Rashmi S K[3], Sheekha[4], Vaishnavi A S[5]
Assistant Professor, Department of Information Science and Engineering[1]
Students Department of Information Science and Engineering[2,3,4,5]
Alva's Institute of Engineering and Technology, Mijar, Mangalore, Karnataka, India

Abstract: *Heart plays significant role in living organisms. Diagnosis and prediction of heart related diseases requires more precision, perfection and correctness because a little mistake can cause fatigue problem or death of the person. The prediction of heart disease is critically significant for diagnosis of diseases and treatment. The data mining techniques that can be applied in medicine, and in particular some machine learning techniques including the mechanisms that make them better suited for the analysis of medical databases. Heart disease is a significant problem in recent times; the main reason for this disease is the intake of alcohol, tobacco, and lack of physical exercise.*

Keywords: Heart Diseases Detection System.

I. INTRODUCTION

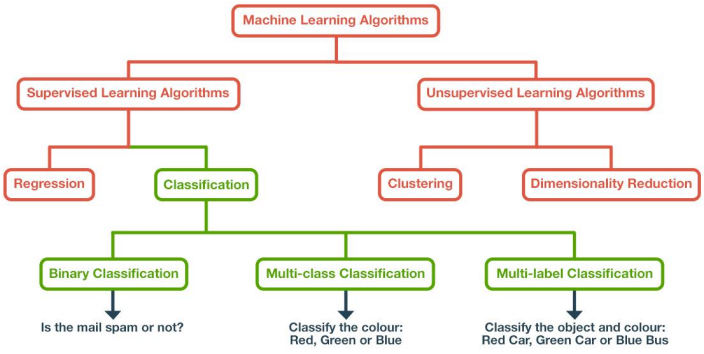
Heart is one of the most extensive and vital organ of human body so the care of heart is essential. Most of diseases are related to heart so the prediction about heart diseases is necessary and for this purpose comparative study needed in this field, today most of patient are died because their diseases are recognized at last stage due to lack of accuracy of instrument so there is need to know about the more efficient algorithms for diseases prediction.

Data mining technology provides a deep insight providing a user oriented approach to discover novel and hidden patterns in the data. This helps in evaluating the effectiveness of medical treatments. The data generated by healthcare transactions is enormous. This medical data containing patients' symptoms is analyzed to perform medical research [8]. With the development of information technology, extensive medical data is available. Medical data classification plays a significant role in various medical applications. Medical classification can be widely used in hospitals for the statistical analysis of diseases and therapies. It addresses the problems of diagnosis, analysis and teaching purposes in medicine.

To avoid these errors and to achieve better and faster results, we need an automated system. Over the past years, researchers find out that machine learning algorithms perform very well in analyzing medical data sets. These data sets will be directly given to machine learning algorithms, and machine learning algorithms will perform according to their nature, and those algorithms will give some outputs.

II. MACHINE LEARNING

Machine Learning is one of efficient technology which is based on two terms namely testing and training i.e. system take training directly from data and experience and based on this training test should be applied on different type of need as per the algorithm required. There are three type of machine learning algorithms:



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

III. PROPOSED METHOD

The proposed methodology integrates with supervised machine learning technique which is based on a hybrid approach for providing a better decision system using dual decision tree and genetic algorithm. Genetic algorithms are one of the best methods for search and optimization problems. A decision tree is a tree structure classifier that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. Pros of Decision Trees (DTs):

- DTs do not require any domain knowledge.
- DTs are easy to comprehend.

The learning and classification steps of a DT are simple and fast. Tree pruning is performed in order to remove anomalies in the training data due to noise or outliers. The pruned trees are smaller and less complex. Tree Pruning can be done through two approaches:

- Pre-pruning—The tree is pruned by halting its construction early.
- Post-pruning—This approach removes a sub-tree from a fully grown tree.

The cost complexity of a decision tree is measured by two parameters, the number of leaves in the tree and the error rate of the tree. Genetic algorithms (GA) were invented by John Holland in 1975. Genetic algorithms can be applied for search and optimization problems. GA uses genetics approach as its model for problem solving. Each solution in genetic algorithm is represented through chromosomes. Chromosomes are made up of genes, which are individual elements that represent the problem. The collection of all chromosomes is called the population [1, 27]. In general, there are three operators that can be applied in GA.

Selection: This operator is used in selecting individuals for reproduction with the help of fitness function. Fitness function in GA is the value of an objective function for its phenotype. The chromosome has to be first decoded, for calculating the fitness function.

Crossover: This is the process of taking two parent chromosomes and producing a child from them. This operator is applied to create better string.

Mutation: This operator is used to alter the new solutions in the search for better solution. Mutation prevents the GA to be trapped in a local minimum. The proposed system architecture (Fig. 1) consists of an ensemble classifier characterized by genetic algorithm with dual decision tree facilitates as follows, in the first stage multiple risk factors such as age, hypercholesterolemia, hypertension, diabetes, obesity, stress level, alcohol taken, etc., are taken as input. This input is preprocessed to fill up the missing values, remove noise and inconsistencies if any in the data and then is given to the hybrid scheme which consists of genetic algorithm and decision tree. Here, the features are initialized through decision tree and fitness is evaluated via genetic algorithm. The output from this hybrid scheme gives the optimized feature. This output is then given as the input to the decision tree classifier for obtaining the type of heart disease.

III. LITERATURE SURVEY

This Paper predicts heart disease for Male Patient using Classification Techniques. The detailed information about Coronary Heart diseases such as its Facts, Common Types, and Risk Factors has been explained in this paper. The Data Mining tool used is WEKA (Waikato Environment for Knowledge Analysis), a good Data Mining Tool for Bioinformatics Fields. The all three available Interface in WEKA is used here. Naive Bayes, Artificial Neural Networks and Decision Tree (J48) are Main Data Mining Techniques and through this techniques heart disease is predicted in this System [2]. Through this paper the information about Data Mining and heart diseases has been gathered. The detailed information about heart diseases, symptoms of heart attack and heart disease types are presented in this paper, the three main data mining techniques namely Decision Tree, Neural Networks and Naive Bayes Classifier are used. The main task of data Prediction is done using these three techniques.

The core concept of this paper is predicting heart disease using data mining Techniques. The main Methodology used for prediction is KNN Algorithms, Decision Trees like CART, C4.5, CHAID, J48, ID3 Algorithms, and Naive Bayes Techniques. This system uses 13 medical attributes as input and with that input, Data sets it to process the data mining techniques and shows the most accurate one.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

IV. METHODOLOGY

4.1 Data Source

The dataset used here for predicting heart disease is taken from UCI Machine learning repository. UCI is a collection of databases that are used for implement machine learning algorithms. The dataset used here is real dataset. The dataset consists of 300 instance of data with the appropriate 14 clinical parameters. The clinical parameter of dataset is about tests which are taken related to the heart disease as like blood pressure level, chest pain type, electrocardiographic result and etc.

4.2 Data Collection

First step for predication system is data collection and deciding about the training and testing dataset. In this project we have used 73% training dataset and 37% dataset used as testing dataset the system.

4.3 Attribute Selection

Attribute of dataset are property of dataset which are used for system and for heart many attributes are like heart bit rate of person, gender of the person, age of the person and many for predication system.

4.4 Preprocessing of Data

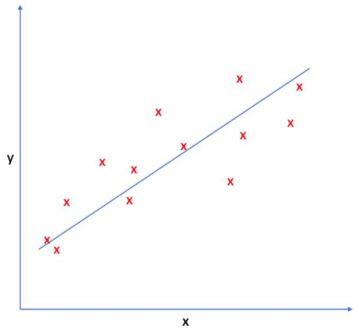
Preprocessing needed for achieving prestigious result from the machine learning algorithms. For example Random forest algorithm does not support null values dataset and for this we have to manage null values from original raw data. For our project we have to convert some categorized value by dummy value means in the form of "0"and "1" by using following code:

4.5 Data Balancing

Data balancing is essential for accurate result because by data balancing graph we can see that both the target classes are equal. Fig.3 represents the target classes where "0" represents with heart diseases patient and "1" represents no heart diseases patients.

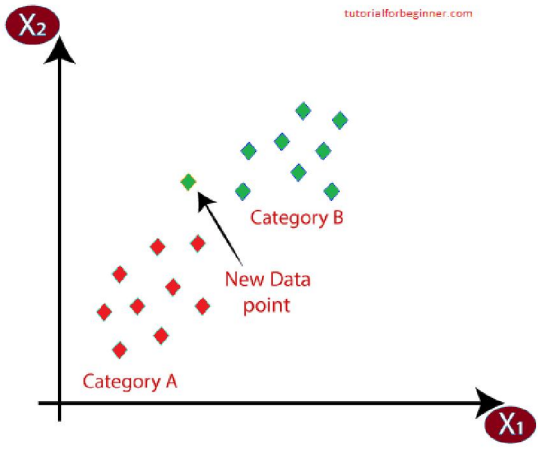
V. MACHINE LEARNING ALGORITHMS

Linear Regression: It is the supervised learning technique. It is based on the relationship between independent variable and dependent variable as seen in Fig.5 variable "x" and "y" are independent and dependent variable and relation between them is shown by equation of line which is linear in nature that why this approach is called linear regression.



K-nearest Neighbour: It work on the basis of distance between the location of data and on the basis of this distinct data are classified with each other. All the other group of data are called neighbor of each other and number of neighbor are decided by the user which play very crucial role in analysis of the dataset.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252



In the above Fig. k=3 shows that there are three neighbor that means three different type of data are there. Each cluster represented in two dimensional space whose coordinates are represented as

(Xi,Yi) where Xi is the x-axis, Y represent yaxis and i= 1,2,3,….n.

VI. CONCLUSION

Heart disease is a very critical issue in the present growing world. So, there is a need for an automated system to predict heart disease at earlier stages. So that it will be useful for the physician to diagnose the patients efficiently, and it will be useful to the people also because they can track their health issues by using this automated system. Some of the expert automated systems were summarized in this paper.

The Decision tree model has predicted the heart disease patient with an accuracy level of 91% and Naïve Bayes classifier has predicted heart disease patient with an accuracy level of 87%. Thus I conclude Decision Tree Classifier Naïve Bayes Classifier Accuracy Rate 91% 87% 85% 86% 87% 88% 89% 90% 91% 92% RATE IN PERCENTAGE Accuracy Rate of Classification Techniques this project by saying Decision tree Classification algorithm is best and better for handling medical data set. In the future, the designed system with the used machine learning classification algorithm can be used to predict or diagnose other diseases. The work can be extended or improved for the automation of heart disease analysis including some other machine learning algorithms.

REFERENCES

[1]. Likhitha KN, Nethravathi. R, Nithyashree. K, Rithika Kumari, Sridhar N and Venkateshvaran K, "Heart Disease Detection using Machine Learning Technique," *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2021, pp. 1738-1743, doi: 10.1109/ICESC51422.2021.9532705.

[2]. Archana Singh and Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms," *2020 International Conference on Electrical and Electronics Engineering (ICE3)*, 2020, pp. 452-457, doi: 10.1109/ICE348803.2020.9122958.

[3]. P. Motarwar, A. Duraphe, G. Suganya and M. Premalatha, "Cognitive Approach for Heart Disease Prediction using Machine Learning," *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 2020, pp. 1-5, doi: 10.1109/ic-ETITE47903.2020.242.

[4]. RahulKatarya and P. Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey," *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2020, pp. 302-305, doi: 10.1109/ICESC48915.2020.9155586.

[5]. Jian Ping Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," in *IEEE Access*, vol. 8, pp. 107562-107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[6]. Shah, D., Patel, S. & Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. *SN COMPUT.*

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

SCI. **1,** 345 (2020).

[7]. B. Keerthi Samhitha, M. R. Sarika Priya., C. Sanjana., S. C. Mana and J. Jose, "Improving the Accuracy in Prediction of Heart Disease using Machine Learning Algorithms," *2020 International Conference on Communication and Signal Processing (ICCSP)*, 2020, pp. 1326-1330, doi: 10.1109/ICCSP48568.2020.9182303.

[8]. ApurbRajdhan ,Avi Agarwal , Milan Sai , Dundigalla Ravi, Dr. Poonam Ghuli, 2020, Heart Disease Prediction using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 04 (April 2020),

[9]. Sameer S Yadav, Shivaji M Jadhav, S. Nagrale and N. Patil, "Application of Machine Learning for the Detection of Heart Disease," *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 2020, pp. 165-172, doi: 10.1109/ICIMIA48430.2020.9074954.

[10]. Mangesh Limbitote , Dnyaneshwari Mahajan , KedarDamkondwar , Pushkar Patil, 2020, A Survey on Prediction Techniques of Heart Disease using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 06 (June 2020).