

A MACHINE LEARNING APPROACH FOR PREDICTION OF HEART DISEASES

Gudipudi Akhilesh Teja ¹ P.V. Sridevi ² Department of Electronics and communication Engineering, Andhra University College of Engineering, Andhra University

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, there are numerous death cases related to heart and their counting is increasing exponentially day by day. To deal with the problem there is essential need of prediction system for awareness about diseases. Machine learning is the branch of Artificial Intelligence (AI), it provides prestigious support in predicting any kind of event which take training from natural events. In this paper, the accuracy of machine learning algorithms for predicting heart disease was calculated, for this algorithms K-nearest neighbor, Random Forest, Logistic Regression and Support vector machine(SVM) by using UCI repository dataset for training and testing. For implementation of Python programming Anaconda (jupyter) notebook is best tool which have many type of library, header file, that make the work more accurate and precise.

KEYWORDS: K –Nearest Neighbor, Logistic Regression, Random Forest, and Support Vector Machine Algorithms.

1. 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.

Machine Learning is one of the efficient technologies for the testing, which is based on training and testing. It is the branch of Artificial Intelligence (AI) which is one of broad area of learning where machines emulating human abilities, machine learning is a specific branch of AI. On the other hand machine learning systems are trained to learn how to process and make use of data hence the combination of both technologies is also called as Machine Intelligence.

As the definition of machine learning, it learns from the natural phenomenon, natural

things so in this paper the biological parameter as testing data such as cholesterol, blood pressure, sex, age, etc. are used and on the basis of these and comparison is done in the terms of accuracy of algorithms such as in this paper four algorithms are used which are decision tree, linear regression, k-neighbor and SVM.

In this paper, the accuracy of four different machine learning approaches are calculated and on the basis of calculation it was concluded that which one is best among them. This paper consist of introduction about the machine learning, classification, related work of researchers, methodology, algorithms used and the summary of this paper with slight view about future scope of this paper.

EXISTING SYSTEM

Very few systems use the available clinical data for prediction purposes and even if they do, they are restricted by the large number of association rules that apply. Diagnosis of the condition solely depends upon the Doctor's intuition and patient's records. The decision support system and will prove to be an aid for the physicians with the diagnosis. The algorithm, Fuzzy c means uses clustering and makes use of clusters and data points to predict the relativity of an attribute .Each data point is associated with multiple

clusters depending upon the membership degrees.

DISADVANTAGES

1. Detection is not possible at an earlier stage
2. Practical use of various collected data is time consuming

2. PROPOSED SYSTEM

To deal with the problem there is essential need of prediction system for awareness about diseases. Machine learning is the branch of Artificial Intelligence (AI), it provides prestigious support in predicting any kind of event which take training from natural events. The accuracy of machine learning algorithms are calculated for predicting heart disease, for this algorithms are k-nearest neighbor, decision tree, random forest classifier, logistic regression and support vector machine(SVM) by using UCI repository dataset for training and testing.

2.1 RANDOM FOREST: A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems.

2.2 KNN: KNN algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

2.3 LOGISTIC REGRESSION: Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

Logistic regression predicts the output of a categorical dependent variable.

2.4 SVM: A support vector machine (SVM) is a type of deep learning algorithm that performs supervised learning for classification or regression of data groups. In AI and machine learning, supervised learning systems provide both input and desired output data, which are labeled for classification.

2.5 DECISION TREE: A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.

ADVANTAGES

1. High performance and accuracy rate
2. Machine Learning Algorithms is very flexible and is widely used in various domains with high rates of success

3. LITERATURE SURVEY

1. Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

AUTHORS: Senthil kumar mohan, chandrasegar thirumalai and Gautam Srivastava

Heart disease is one of the most significant causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. The ML techniques are being used in recent developments in different areas of the Internet of Things (IoT). Various studies give only a glimpse into predicting heart disease with ML techniques. In this paper, a novel method that aims at finding

significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques. This produced an enhanced performance level with an accuracy level of 88.7% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM).

2) Prediction of Heart Disease using Machine Learning Algorithms: A Survey

AUTHORS: Himanshu Sharma and M A Rizvi

Health care field has a vast amount of data, for processing those data certain techniques are used. Data mining is one of the techniques often used. Heart disease is the Leading cause of death worldwide. This System predicts the arising possibilities of Heart Disease. The outcomes of this system provide the chances of occurring heart disease in terms of percentage. The datasets used are classified in terms of medical parameters. This system evaluates those parameters using data mining classification technique. The datasets are processed in python programming using two main Machine Learning Algorithms namely Decision Tree Algorithm and Naive Bayes Algorithm which shows the best algorithm among these two in terms of accuracy level of heart disease.

3) Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools

AUTHORS: M. Nikhil Kumar, K. V. S. Koushik, K. Deepak

Objectives: The objective of our work is to analyze various data mining tools and techniques in health care domain that can be employed in prediction of heart disease system and their efficient diagnosis. **Methods/Statistical Analysis:** A heart disease prediction model, which implements data mining technique can help the medical

practitioners in detecting the heart disease status based on the patient's clinical data. Data mining classification techniques for good decision making in the field of health care addressed are namely Decision trees, Naive Bayes, Neural Networks and Support Vector Machines. Hybridizing or combining any of these algorithms helps to make decisions quicker and more precise. **Findings:** Data mining is a powerful new technology for the extraction of hidden predictive and actionable information from large databases that can be used to gain deep and novel insights. Using advanced data mining techniques to excavate valuable information, has been considered as an activist approach to improve the quality and accuracy of healthcare service while lowering the healthcare cost and diagnosis time. Using this technique presence of heart disease can be predicted accurately. Using more input attributes such as controllable and uncontrollable risk factors, more accurate results could be achieved. **Applications/Improvements:** This method can be further expanded. It can use many of input attributes. Other data mining techniques are also be used for predication such as Clustering, Time series, Association rules. The unstructured data available in healthcare industry database can also be mined using text mining.

4) Application of Machine Learning in Diseases Prediction

AUTHORS: Pahulpreet Singh Kohli and Shriya Arora

The application of machine learning in the field of medical diagnosis is increasing gradually. This can be contributed primarily to the improvement in the classification and recognition systems used in disease diagnosis which is able to provide data that aids medical experts in early detection of fatal diseases and therefore, increase the survival rate of patients significantly. This paper has different classification algorithms, each with its own advantage on three separate databases of disease (Heart, Breast cancer, Diabetes) available in UCI repository for disease prediction. The feature selection for each dataset was accomplished by backward modeling using the p-value test. The results of the study strengthen the idea of the application of machine learning in early detection of diseases.

5)An Extensive Review on Swarm Robotics
AUTHORS:S. Kumra, R. Saxena, and S. Mehta.,

Swarm robotics is a new approach to the coordination of multi-robot systems which consist of large numbers of relatively simple robots which takes its inspiration from social insects. The most remarkable characteristic of swarm robots are the ability to work cooperatively to achieve a common goal. In this paper, classification of existing researches, problems and algorithms aroused in the study of swarm robotics are presented. The existing studies are classified into major areas and relevant sub-categories in the major areas.

4. MODULES:

- User
- Data Collection
- Attribute Selection
- Preprocessing of data.

User:

Users add the data to the database and view the data to the view data and predict the heart disease using ml.

Data Collection:

First step for predication system is data collection and deciding about the training and testing dataset. In this paper 73% training dataset and 27% dataset used as testing dataset the system.

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 more predication system.

Preprocessing of data:

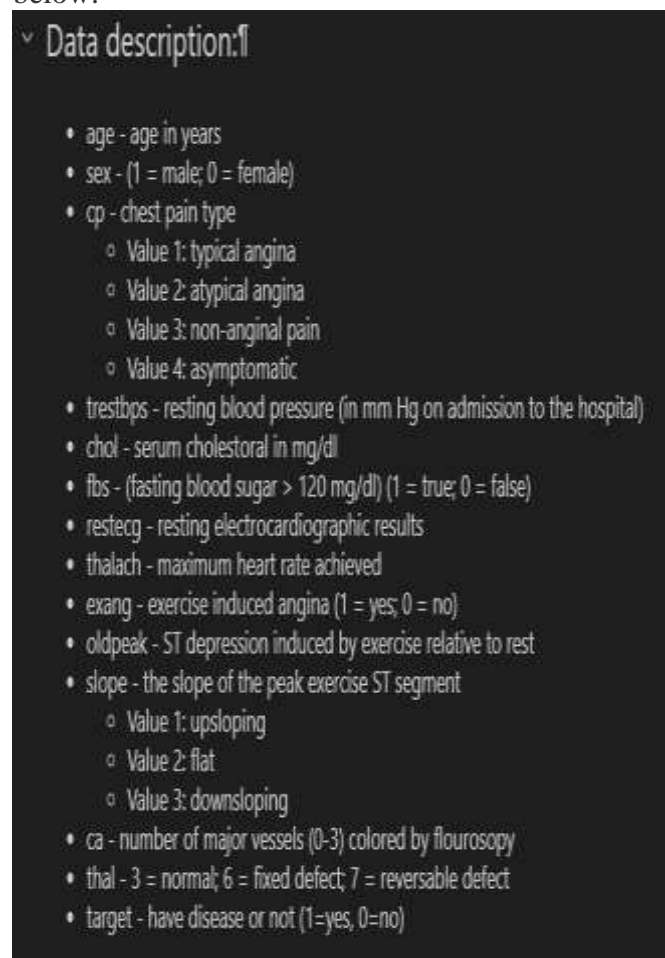
Preprocessing needed for achieving prestigious result from the machine learning algorithms. For example Random forest algorithm does not support null values dataset and have to manage null values from original raw data. For this paper some categorized values are converted by dummy value means in the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

Admin:

Admin will give authority to Users. In order to activate the users admin can Predict Heart Disease.

5. DATA SET

This paper uses the heart disease dataset that is obtained from UCI Machine Learning Repository. UCI Machine Learning Repository is one of the largest available datasets, having over 417 various datasets. The Cleveland dataset from UCI Machine Learning Repository is one of the datasets on heart disease, which is widely used by researchers to date. This paper will also use this dataset of which contains 303 rows. The dataset contains 76 features in which 14 attributes including class label are used. The 14 features together with their descriptions and data types are shown below.



6. RESULTS: This paper aim is to know whether the patient has heart disease or not. The records in the dataset are divided into the training set and testing tests. After preprocessing the data, the data classification namely k-nearest neighbor, random forest classifier, logistic regression and support vector machine were applied. Then 4 models were trained and tested with maximum scores as follows:

Logistic Regression 92.31%

KNN Model 87.91%

SVM 91.21%

RANDOM FOREST 89.01%

CONCLUSION

Heart is one of the essential and vital organs of human body and prediction about heart diseases is also important concern for the human beings so that the accuracy for algorithm is one of parameter for analysis of performance of algorithms. Accuracy of the algorithms in machine learning depends upon the dataset that used for training and testing purpose. When performed, the analysis of algorithms on the basis of dataset and on the basis of confusion matrix Logistic Regression is best one.

FUTURE ENHANCEMENT

For the Future Scope more machine learning approach will be used for best analysis of the heart diseases and for earlier prediction of diseases so that the rate of the death cases can be minimized by the awareness about the diseases.

REFERENCES

[1] Santhana Krishnan J and Geetha S, "Prediction of Heart Disease using Machine Learning Algorithms" ICICT, 2019.
[2] Aditi Gavhane, Gouthami Kokkula, Isha Panday, Prof. Kailash Devadkar, "Prediction of Heart Disease using Machine Learning", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology(ICECA), 2018.

[3] Senthil kumar mohan, chandrasegar thirumalai and Gautam Srivastva, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" IEEE Access 2019.
[4] Himanshu Sharma and M A Rizvi, "Prediction of Heart Disease using Machine Learning Algorithms: A Survey" International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 8 , IJRITCC August 2017.
[5] M. Nikhil Kumar, K. V. S. Koushik, K. Deepak, "Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools" International Journal of Scientific Research in Computer Science, Engineering and Information Technology ,IJSRCSEIT 2019.
[6] Amandeep Kaur and Jyoti Arora, "Heart Diseases Prediction using Data Mining Techniques: A survey" International Journal of Advanced Research in Computer Science , IJARCS 2015-2019.
[7] Pahulpreet Singh Kohli and Shriya Arora, "Application of Machine Learning in Diseases Prediction", 4th International Conference on Computing Communication And Automation(ICCCA), 2018.
[8] M. Akhil, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," Procedia Technol., vol. 10, pp. 85–94, 2013.
[9] S. Kumra, R. Saxena, and S. Mehta, "An Extensive Review on Swarm Robotics," pp. 140–145, 2009.
[10] Hazra, A., Mandal, S., Gupta, A. and Mukherjee, " A Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review" Advances in Computational Sciences and Technology , 2017.
[11] Patel, J., Upadhyay, P. and Patel, "Heart Disease Prediction Using Machine learning and Data Mining Technique" Journals of Computer Science & Electronics , 2016.
[12] Chavan Patil, A.B. and Sonawane, P. "To Predict Heart Disease Risk and Medications Using Data Mining Techniques with an IoT Based Monitoring System for Post-Operative Heart Disease Patients" International Journal on Emerging Trends in Technology, 2017.
[13] V. Kirubha and S. M. Priya, "Survey on Data

Mining Algorithms in Disease Prediction,”
vol. 38, no. 3, pp. 124–128, 2016.