Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021

Application of Different Machine Learning Techniques for Predicting Heart Dieses

Tejaswini Zope
Department of Computer Engineering
Pimpri Chinchwad College of Engineering, Akurdi
Pune, India

Dr. K. Rajeswari
Department of Computer Engineering
Pimpri Chinchwad College of Engineering, Akurdi
Pune, India

Prof. Sushma Vispute
Department of Computer Engineering
Pimpri Chinchwad College of Engineering, Akurdi
Pune, India

***Abstract*— In this work, heart disease is regarded as one of the main causes in the world today. Doctors cannot easily predict it because it is a difficult task that requires experience and more predictive knowledge. There is a lot of knowledge available in the healthcare system on the web. However, there is a lack of effective analysis tools to capture the patterns and relationships hidden in the data. The automatic diagnosis system will improve medical efficiency and will reduce costs. The aims to predict the occurrence of disease-supporting data collected from medical research, especially in heart disease. The goal is to apply data processing technology to extract hidden patterns on the data set. These patterns are known for heart disease, and to predict whether patients have heart disease. The existence of these patterns is scored according to the scale.**

Keywords: SVM, Navie bayes, Decision tree,Random Forest, Logistic regression

I. INTRODUCTION

Due to heart problems, the highest death rate in India and abroad is especially

. According to the World Health Organization (WHO), heart-related diseases claim 17.7 million lives each year, accounting for 31% of all deaths worldwide. Therefore, this is usually an important time to check this mortality rate by correctly identifying the disease at the initial stage. We may use data processing techniques to gain insight into the data set. Health administrators often use the discovered knowledge to improve service levels. Clinicians can also use the knowledge discovered to reduce the number of adverse drug reactions and propose equivalent alternatives to cheaper treatments. Predicting the future behavior of patients in a given history is one of the important applications of knowledge extraction techniques that can be used for health management. A huge challenge facing healthcare organizations (hospitals, medical centers) is to provide quality services at an affordable cost. Quality services include correct diagnosis of patients and implementation of effective treatment. Bad clinical decisions can have disastrous consequences and are therefore unacceptable. Hospitals must also minimize the value of clinical trials. They can use appropriate computer-based information to achieve these results Considering the impact of cardiovascular disease on the world's population

, the machine learning model for early detection becomes very useful. Constantly trying to use a differenttechnological advancements are made to affect this rising gigantic problem. Different bioengineering techniques are developed in recent years to cope-up with the ever-growing

Health problems. Continuing research in this area has shown the benefits of increasing the reduction rate.

II. LITERATURE REVIEW

The emergence of artificial intelligence in the field of health sciences has encouraged many studies aimed at reducing mortality through the application of different data processing techniques.

Shantakumar B.atil et al. proposed effective attack prediction by extracting important patterns from data sets. The Kmeans clustering algorithm is used. Use the MAFIA algorithm to calculate the weight of each article. The calculated weight is supported, and the mode with a value greater than the edge is considered for prediction. Jyoti Soni et al. proposed the use of 15-attribute data sets and data processing techniques such as RNA, statistics, grouping rules, and association rules to predict heart conditions. R. Chitra et al. proposed to reduce the size of information to improve accuracy by applying genetic algorithm in the computer-aided system

for diagnosis and prediction. This paper considers a neural network with reduced feature preprocessing and normalized data for the classification of heart conditions. Hlaudi Daniel Masethe et al. proposed experiments on various algorithms such as j48, SIMPLE CART and reptree. The prediction rate is compared, so this article proposes the best method.

G. Purusothaman et al. Proposed a wide range of data processing classification techniques, such as ANN, symbolic logic, neural network, decision tree, genetic algorithm for data mining, and nearest neighbor method. [5]. The document proposes the application of hybrid data processing methods. Cheryl Ann Alexander and others discussed the importance of large-scale data analysis in the prediction, prevention, and treatment of chronic diseases.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021

[6]. The paper presents the thoughts of the Internet of Things and cloud computing technology in the medical field.

III. RELATED WORK

Unlike diseases in which the entire cardiovascular system has problems, heart disease only affects the heart. According to the Centers for Disease Control and Prevention (CDC), heart disease is the leading cause of death in the United States. Heart disease accounts for about a quarter of deaths in the United States and affects all genders and all races and ethnic groups.

A. Types

There are several different types of heart disease, and they affect the heart in different ways.

1. Coronary artery disease

Coronary artery disease, also called coronary heart disease, is the most common type of heart disease. It occurs when the arteries that supply blood to the heart are blocked by plaque. This causes them to harden and narrow. Dental plaque contains cholesterol and other substances. As a result, the blood supply is reduced, and the heart receives less oxygen and nutrients. Over time, the heart muscle will weaken, and there is a risk of heart failure and arrhythmia.

2. Congenital heart defects

People with congenital heart defects are born with heart problems. There are many types of congenital heart defects, including:

• Heart valve abnormalities-the valves may not open properly or blood may leak.

• Diaphragm defect: There is a hole in the wall between the lower or upper chamber of the heart.

• Atresia: One of the heart valves is missing.

Congenital heart disease may involve serious structural problems, such as missing ventricles and problems with the main arteries leading from the heart. Many congenital heart diseases do not cause any obvious symptoms and only become obvious during routine medical examinations.

3. Arrhythmia

Arrhythmia is an irregular heartbeat. It occurs when the electrical pulses that coordinate the heartbeat do not work properly. Therefore, the heart may beat too fast, too slow, or irregularly.

OBJECTIVE

Ease of use: The main goal of the project is to develop a simple and user-friendly platform, because here the medical details of the patient must be provided and the extracted characteristics supported, and then the algorithm will detect the type of intestinal diseases and spots. . Since the algorithm here does the job, the number of well-trained models is small and they will definitely make mistakes in predicting intestinal diseases and their types. So, in short, the accuracy is improved, which also saves time and allows doctors and patients to be easier to predict if they are susceptible to some form of heart disease, otherwise it would be difficult for us to test without the involvement of a doctor.

No human intervention is required-to detect

types of heart disease, medical details such as age and cholesterol must be provided. Here, the algorithm will provide the result of extracting feature support, so because there is no human intervention, the possibility of error is very small It also saves a lot of time for patients or doctors, who can perform more treatments or other procedures more quickly. This is usually to prevent results from being obtained faster. This, in turn, will accelerate the heart disease treatment prevention / prevention process by 4,444 tons, while saving critical time for physicians and patients to continue treatment and take preventative measures to reduce the impact of heart disease.

Not only does it detect the types of intestinal diseases, but also proposes preventive measures: In this project, our goal is not only to find and predict the type of heart disease, but to point out the preventive measures that must be taken to reduce the impact of intestinal diseases. disease. Getting advice on taking preventive measures will help doctors and patients make progress easily with additional treatment steps.

Efficient use of available annotated data samples - Successful training of machine learning algorithms requires thousands of annotated training samples, which is widely recognized. Therefore, we use powerful networks and training strategies that rely on data preprocessing to use the available annotation samples more effectively. Since medical data is not available for during large batches (more than or up to thousand samples, according to machine learning standards), we use data pre-processing to form the use of available data more effectively . Data preprocessing is essential for data processing techniques that involve converting data to a clear format.

Real world medical data is often incomplete, inconsistent, and / or lacking in certain behaviors or tendencies, and can contain many errors. Data preprocessing can be an effective method of solving these problems. Data pre-processing prepares the data for further processing.

DATASET

Heart disease prediction can also be a web-based machine learning application, trained on the ICU dataset. The user enters his specific medical data to ask him to make a prediction of heart disease. The algorithm will calculate the probability of heart disease. The results will be displayed on the website itself. Therefore, the value and time required to predict disease are minimized. The knowledge format plays a vital role in this application.

Data Collection and Preprocessing

The dataset used is the Heart Disease dataset, which is a combination of 4 different databases, but only the UCI Cleveland dataset is used. The database contains a total of 76 attributes, but all published experiments refer to the use of Only a subset of 14 functions. Therefore, we used the processed sites in the UCI Cleveland data set for analysis. A complete description of the 14 attributes used in the proposed work is mentioned in Table 1 below.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021

Sr.no	Attribute	Attribute description
1	Id	patient identification number
2	Ccf	social security number (I replaced this with a dummy value of 0)
3	Age	age in years
4	Sex	sex (1 = male; 0 = female)
5	painloc	chest pain location (1 = substernal; 0 = otherwise)
6	painexer	(1 = provoked by exertion; 0 = otherwise)
7	relrest	(1 = relieved after rest; 0 = otherwise)
8	pncaden	(sum of 5, 6, and 7)
9	cp:	chest pain type -- Value 1: typical angina -- Value 2: atypical angina -- Value 3: non-anginal pain -- Value 4: asymptomatic
10	trestbps	resting blood pressure (in mm Hg on admission to the hospital)
11	chol	serum cholestoral in mg/dl
12	smoke	I believe this is 1 = yes; 0 = no (is or is not a smoker)
13	cigs	(cigarettes per day)
14	years	(number of years as a smoker)

METHODOLOGY

Logistic Regression

Logistic regression is a classification algorithm that is used primarily for binary classification problems. In logistic regression, the logistic regression algorithm does not fit a straight line or a hyperplane, but uses a logistic function to compress the output of the linear equation between 0 and 1. There are 13 independent variables, which makes the logistic regression is good for classification.

Naive Bayes

Bayes 'naive algorithm is based on Bayes' rule. The independence between the attributes of the data set is the main assumption and the most important assumption for the classification. When the independence assumption is established, it can be easily and quickly predicted and maintained at its best. Bayes' theorem calculates the posterior probability of event (A), given some prior probability of event B denoted by P (A / B)

Classification

The attributes mentioned in Table 1 are provided as inputs to different ML algorithms, such as random forest, decision tree, logistic regression, and naive Bayes classification techniques. The input data set is divided into 80% of the training data set and the remaining 20% of the test data set. The training data set is the data set used to train the model. The test dataset is used to verify the performance of the trained model. For each algorithm, the performance is calculated and analyzed based on different indicators used, such as accuracy, precision, recall, and Fmeasure score, as described below.

Random Forest

Random forest algorithm is used for classification and regression. Create a tree for the data and make predictions based on it. The random forest algorithm can be used for large data sets, and the same results can be produced even if the registry value is missing in the large data set. The generated decision tree samples can be saved for other data. There are two stages in a random forest. First, create a random forest, and then use the random forest classifier created in the first stage to make predictions.

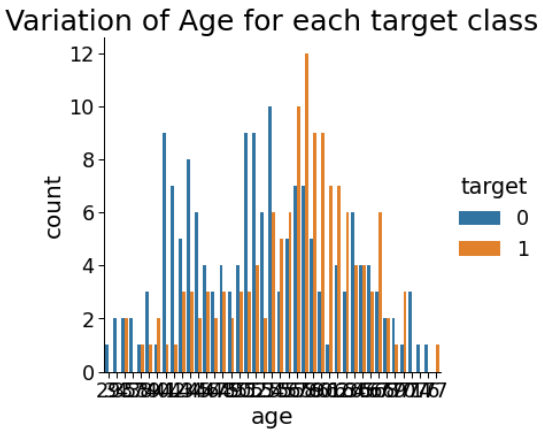
Decision Tree

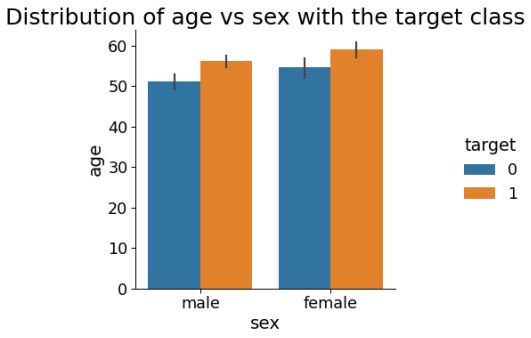
The decision tree algorithm takes the form of a flowchart, where internal nodes represent the attributes of the data set, and external branches are the results. Decision trees were chosen because they are fast, reliable, easy to interpret, and require very little data preparation. In a decision tree, the class label prediction is derived from the root of the tree. The value of the root attribute is compared with the record attribute. From the comparison result,

$$P(A|B) = (P(B|A)P(A)) / P(B) \quad (1)$$

RESULT AND ANALYSIS

Algorithm	Accuracy
Svm	0.92
Naive Bayes	0.86
Logistic Regression	0.86
Decision Tree	1.0
Random forest	0.99





CONCLUSION

With the increasing number of deaths caused by heart disease, the development of an effective and accurate system for predicting heart disease has become a mandatory requirement. Find the most effective ML algorithm to detect heart disease. This compares the accuracy scores of decision trees, logistic regression, random forest, and naive Bayes algorithms for predicting heart disease using the UCI machine learning repository dataset. The results of this study show that random forest and decision tree algorithms are the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021

most effective algorithms, with accuracy scores of 0.99 and 1.0 for predicting heart disease.

REFERENCES

[1] Cognitive Approach for Heart Disease Prediction using Machine Learning a Pranav Motarwar, a Ankita Duraphe, b G Suganya* c M Premalatha a UG student, Vellore Institute of Technology, Chennai, Tamilnadu, India b Associate Professor, Vellore Institute of Technology, Chennai, Tamilnadu, Indi

[2] T.Nagamani, S.Logeswari, B.Gomathy, Heart Disease Prediction using Data Mining with Mapreduce Algorithm, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.

[3] Prediction of Heart Disease Using Machine Learning Aditi Gavhane Department of Information Technology Sardar Patel Institute of Technology Mumbai, India

[4] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, Design And Implementation Heart Disease Prediction Using Naives Bayesian, International Conference on Trends in Electronics and Information(ICOEI 2019).

[5] Theresa Princy R,J. Thomas,Human heart Disease Prediction System using Data Mining Techniques, International Conference on Circuit Power and Computing Technologies,Bangalore,2016.

[6] Nagaraj M Lutimath,Chethan C,Basavaraj S Pol.,Prediction Of Heart Disease using Machine Learning, International journal Of Recent Technology and Engineering,8,(2S10), pp 474-477, 2019.

[7] UCI, Heart Disease Data Set.[Online]. Available (Accessed on May 1 2020): https://www.kaggle.com/ronitf/heart-disease-uci.

[8] Sayali Ambekar, Rashmi Phalnikar,Disease Risk Prediction by Using Convolutional Neural Network,2018 Fourth International Conference on Computing Communication Control and Automation.

[9] C. B. Rjeily, G. Badr, E. Hassani, A. H., and E. Andres, Medical Data Mining for Heart Diseases and the Future of Sequential Mining in Medical Field, in Machine Learning Paradigms, 2019, pp. 7199.

[10] Jafar Alzubi, Anand Nayyar, Akshi Kumar. "Machine Learning from Theory to Algorithms: An Overview", Journal of Physics: Conference Series, 2018

[11] Fajr Ibrahem Alarsan., and Mamoon Younes Analysis and classification of heart diseases using heartbeat features and machine learning algorithms,Journal Of Big Data,2019;6:81.