ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijsem.org
Vol 14, Issuse.4 Nov 2023

HEART STROKE DISEASE PREDICTION USING MACHINELEARNING

1.SHIVA BHAVANI K, 2. A. HARIPRIYA, 3. C. VAISHNAVI, 4. D. HARSHINI,5. K. NAGA HARATHI

ABSTRACT: Heart Attack is a term that assigns a large number of medical conditions related to heart. The key to Heart (Cardiovascular) diseases to evaluate large scores of data sets, compare information that can be used to predict, Prevent, Manage such as Heart attacks. The main objective of this research is to develop an Intelligent System using machine learning technique, namely, Naive Bayes, KNN, Random forest Decision tree. It is implemented as web based application in this user answers the predefined questions. Data analytics is used to incorporate world for its valuable use to controlling, contravasting and Manage a large data sets. It can be applied with a much success to predict, prevent, Managing a Cardiovascular Diseases. To solve this we aims to implement the Data Analytics based on SVM and Genetic Algorithm to diagnosis of heart diseases. This result reveal, which Algorithm is best, optimized Prediction Models. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions, which traditional decision support systems cannot. By providing effective treatments, it also helps to reduce treatment costs.

KEYWORDS: SVM, KNN, Cardiovascular disease etc.

1. INTRODUCTION

Heart is a vital organ of the humanoid body. It pumps blood to every part of our anatomy. If it miscarries to function correctly, then the brain and various other organs will stop functioning, and within few minutes, the person will die. Change in lifestyle, work related stress and wrong food habits add to the increase in rate of several heart related illnesses. Heart diseases have occurred as one of the most prominent cause of death all around the world. According to World Health Organization, heart associated diseases are responsible for the taking 17.7 million lives every year, 31% of all global deaths. In India too, heart related diseases have become the top cause of death. Heart diseases have killed 1.7 million Indians in 2016, according to the 2016 Global Burden of Disease Report, released on September 15, 2017. Heart related diseases increase the outlay on health care and reduce the efficiency of an

[1]ASSISTANT PROFESSOR, [2,3,4&5]UG SCHOLAR
DEPARTMENT OF IOT, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD

ISSN2454-9940
www.ijsem.org
Vol 14, Issuse.4 Nov 2023

individual. Estimates made by the World Health Organization (WHO), suggest that India have lost up to $237 billion, from 2005- 2015, due to heart related or cardiovascular diseases. Thus, reasonable and accurate prediction of heart related diseases is very important. Medical organizations, all around the world, collect data on various health related issues. These data can be oppressed using various machine-learning techniques to gain useful understandings.

But the data collected is very massive and, many a times, this data can be very noisy. These datasets, which are too devastating for human minds to comprehend, can be easily explored using various machine-learning techniques. Thus, these algorithms have become very useful, in recent times, to predict the presence or absence of heart related ailments accurately.

LITERATURE REVIEW

1)**Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques AUTHORS:** Senthil kumar mohan, chandrasegar thirumalai and Gautam Srivastva

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. We have also seen ML techniques 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, we propose 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. We produce 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 SurveyAUTHORS:** 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 Algorithm 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 analyse 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,

ISSN2454-9940
www.ijsem.org
Vol 14, Issuse.4 Nov 2023

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 PredictioAUTHORS:**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. In this paper, we apply 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 RoboticsAUTHORS:**S. Kumra, R. Saxena, and S. Mehta.,

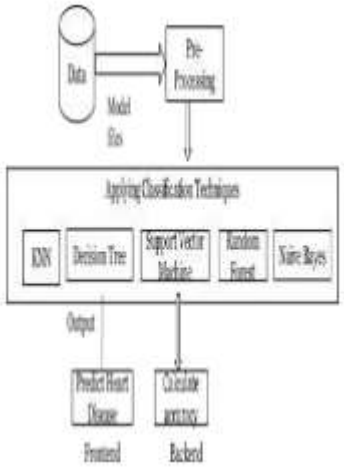
Swarm robotics is a new approach to the coordination of multi-robot systems which consist of large numbers of relatively simple robots which takes its inspiration from social insects. The most remarkable characteristic of swarm robots are the ability to work cooperatively to achieve a common goal. In 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.

EXISTING SYSTEM: The World Health Organization (WHO) has estimated that 12million deaths occur worldwide, every year due to the Heart diseases .About 25% deaths in the age group of 25-69 year occur because of heart diseases. In urban areas, 32.8%. Deaths occur because of heart ailments, while this percentage in rural areas is 22.9.Over 80% of deaths in world are because of Heart disease. WHO estimated by 2030, almost 23.6 million. People will die due to Heart disease. The diagnosis of diseases is a significant and tedious task in medicine. Treatment of the said disease is quite high and not affordable by most of the patients particularly in India.

PROPOSED SYSTEM In this system, we are implementing effective heart attack prediction system using Machine-learning algorithm. We can give the input as in CSV file or manual entry to the system. After taking input, the algorithms apply on that input to algorithms. After accessing data set the operation is performed and effective heart attack level is produced. The proposed system will add some more parameters significant to heart attack with their weight, age and the priority levels are by consulting expertise doctors and the medical experts. The heart attack

ISSN2454-9940
www.ijsem.org
Vol 14, Issuse.4 Nov 2023

prediction system designed to help the identify different risk levels of heart attack like normal, low or high and also giving the prescription details with related to the predicted result.



MAIN FLOW

1. Upload Training Data: The process of rule generation advances in two stages. During the first stage, the SVM model is built using training data during each fold; this model is utilized for predicting the class labels the rules are evaluated on the remaining 10% of test data for determining the accuracy, precision, recall and F-measure. In addition, rule set size and mean rule length are also calculated for each fold of cross-validation.

2. Data Pre- Processing: Heart disease data is pre-processed after collection of various records. The dataset contains a total of 303 patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 297 patient records are used in pre-processing. The multiclass variable and binary classification are introduced for the attributes of the given Dataset.

Predicting Heart Disease: The training set is different from test set. In this study, we used this method to verity the universal applicability of the methods. In k-fold cross validation method, the whole dataset is used to train and test the classifier to Heart Stoke.

IMPLEMENTATION

MODULES:

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

Users:

User 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 project we have used 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 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

Admin:

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

ALGORITHMS

Logistic regression Classifiers

ISSN2454-9940
www.ijsem.org
Vol 14, Issuse.4 Nov 2023

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name logistic regression is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name multinomial logistic regression is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does.This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

NAÏVE BAYES

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature . Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias). While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique.

Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2, Knime

ISSN2454-9940
www.ijsem.org
Vol 14, Issuse.4 Nov 2023

2.1.1, Orange 2.0b and RapidMiner 4.6.0). We try above all to understand the obtained results.

SVM In classification tasks a discriminant machine learning technique aims at finding, based on an independent and identically distributed (iid) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point x and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space. SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to genetic algorithms (GAs) or perceptrons, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

K-Nearest Neighbors (KNN)

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not "learn" until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

CONCLUSION This paper discusses the various machine learning algorithms such as KNN, support vector machine, Naïve Bayes, decision tree and k- nearest neighbor, which were applied to the data set. It utilizes the data such as blood pressure, cholesterol, diabetes and then tries to predict the possible coronary heart disease patient in next 10 years. Family history of heart disease can also be a reason for developing a heart disease as mentioned earlier. So, this data of the patient can also be included for further increasing the accuracy of the model. This work will be useful in identifying the possible patients who may suffer from heart disease in the next 10 years. This may help in taking preventive measures and hence try to avoid the possibility of heart disease for the patient. So the doctors can closely analyze when a patient is predicted as positive for heart disease, then

ISSN2454-9940
www.ijsem.org
Vol 14, Issuse.4 Nov 2023

the medical data for the patient. An example would be - suppose the patient has diabetes that may be the cause for heart disease in future and then the patient can be given treatment to have diabetes in control, which in turn may prevent the heart disease.

REFERENCES

[1].https://www.researchgate.net/publication/319393368_Heart_Disease_Diagnosis_and_Prediction_Using_Machine_Learning and_Data_Mining_Techniques_A_Review J

[2].Brownlee, J. (2016). Naive Bayes for Machine Learning. Retrieved March 4, 2019, from https://machinelearningmastery.com/naive-bayes-for-machine-learning

[3].Science, C., & Faculty, G. M. (2009). Heart Disease Prediction Using Machine learning and Data Mining Technique. Ijcsc 0973-7391, 7, 1–9

[4].https://dzone.com/articles/a-tutorial-on-using-the-big-data-stack-and-machine
[5].https://pythonhow.com/html-templates-in-flask

[6].Intelligent Heart Disease Prediction System Using Data Mining Techniques-Sellappan Palaniappan, Rafiah Awang 978-1- 4244-1968-5/08/ ©2008 IEEE.

[7].Intelligent Heart Disease Prediction System Using Data Mining Techniques-Sellappan Palaniappan, Rafiah Awang 978-1- 4244-1968-5/08/ ©2008 IEEE

[8].Blake,C.L.,Mertz,C.J.:"UCIMachineLearningDatabases",http://mlearn.ics.uci.edu/databases/heartdisease/, 2004

[9].Chapman, P., Clinton, J., Kerber, R. Khabeza, T., Reinartz, T., Shearer, C., Wirth, R.: "CRISP-DM 1.0: Step by step data mining guide", SPSS, 1-78, 2000