ISSN (Online): 2456-0448
International Journal of Innovative Research in Management, Engineering and Technology
Vol. 6, Issue 3, March 2021

Heart Disease Prediction Using Hybrid Machine LearinG

[1] Manikandan B,[2] Nishanthinidevi D,[3] Epsy Millaniya C,[4] Thenmadhuram S,[5] Arun Kumar V

[1] Assistant Professor, Department of information technology, Hindusthan Institute of Technology, Coimbatore, India

[2][3][4][5] Students, Department of Information Technology, Hindusthan Institute of Technology, Coimbatore,India

Abstract: ***Heart disease is the most significant cause of death worldwide. Heart disease prediction can be done by this system. This system finds out the chances of heart disease occuring in humans by the highest percentage. This system is classified using data mining technique which is most often used. This data mining technique is used to reduce the vast amount of data in Health care Field. This accuracy of prediction can be done with the help of R Programming. This system provides better accuracy and time consuming for instant prediction of heart disease.***

Keywords— PSO,KNN.

INTRODUCTION

Heart is an crucial organ of all living creatures, which pumps the blood from the heart to all other body parts through the blood vessels of circulatory system. It is very problematic to find heart disease risk factors such as diabetes, high cholesterol, high blood pressure, abnormal pulse rate and other factors. All human creatures life depends on the working of heart. If the function of the heart is not in good condition, it may lead to affect the other organs of the body such as kidney, liver, brain, etc...

The clinical symptoms of the Heart Disease complicate the prognosis, as it is influenced by many factors like functional and pathologic appearance. This could subsequently delay the prognosis of the disease. Hence, there is a need for the invention of newer concepts to improve the prediction accuracy with short span. Disease prognosis through numerous factors or symptoms is a multi-layerd problem, even that could lead to a false assumption. Therfore, an attempt is made to bridge the knowledge and the experience of the experts and to build a system that fairly supports the diagnosing process.

SYSTEM IMPLEMENTATION

EXISTING SYSTEM

Various methods have been used for knowledge abstraction by using known methods of data mining for prediction of heart disease. In this work, numerous readings have been carried out to produce a prediction model using not only distinct techniques but also by relating two or more techniques. A diagnosis system for the Alzheimer's disease where the objective of feature selection was completed through the use of PCA with combination to linear discriminant analysis and Fisher discriminant ratio.The final classification was performed by using artificial neural network and support vector machine (SVM).

Integrated local fisher discriminant analysis (FDA) and SVM for hepatitis disease diagnosis. Initially dimension of feature set was reduced through FDA and classification was achieved through SVM. Authors tested their technique over UCI dataset. To improve the performance of solving the text categorization problem used feature selection and extraction. They argued that due to increase in the number of documents and in the presence of irrelevant noisy features, the automated categorization is getting hard.

`PROPOSED SYSTEM

Prediction of Heart disease is necessary for the machine learning process because sometimes irrelevant features affect the classification performance of the machine learning classifier. In proposed system using K-Means clustering and K-NN algorithms are used to predict the heart disease. K-means clustering is one of the simplest and popular unsupervised machine learning algorithms.

Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known, or labelled, outcomes.To process the learning data, the k-means algorithm in data mining starts with a first group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative (repetitive) calculations to optimize the positions of the centroids. k-nearest neighbour is one of the most basic yet essential classification algorithms in machine learning.

It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection. It is widely disposable in real-life scenarios since it is non-parametric, meaning, it does not make any underlying assumptions about the distribution of data. K-nearest neighbour (KNN) algorithm uses 'feature similarity' to predict the values of new data points which further means that the new data point will be assigned a value based on how closely it matches the points in the training set.

MODULES

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. The multi-class variable is used to check the presence or absence of heart disease. In the instance of the patient having heart disease, the value is set to 1, else the value is set to 0 indicating the absence of heart disease in the patient. The pre-processing of data is carried out by converting medical records into diagnosis values. The results of data pre-processing for 297 patient records indicate that 137 records show the value of 1 establishing the presence of heart disease while the remaining 160 reflected the value of 0 indicating the absence of heart disease.

FEATURE SELECTION AND REDUCTION:

From among the 13 attributes of the data set, two attributes pertaining to age and sex are used to identify the personal information of the patient. The remaining 11 attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of heart disease. As previously mentioned in this experiment, several (ML) techniques are used namely, NB, GLM, LR, DL, DT, RF, GBT and KNN and K-means. The experiment was repeated with all the ML techniques using all 13 attributes.

Advantages:

Relatively simple to implement.
Scales to large data sets.
Guarantees convergence.
Can warm-start the positions of centroids.
Easily adapts to new examples.
Generalizes to clusters of different shapes and sizes, such as elliptical clusters

CLASSIFICATION MODELIN

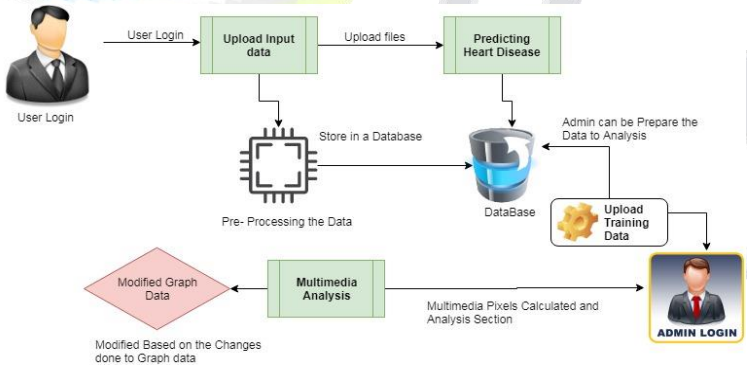
The clustering of datasets is done on the basis of the variables and criteria of Decision Tree (DT) features. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error. The performance is further optimized by choosing the DT cluster with a high rate of error and extraction of its corresponding classifier features. The performance of the classifier is evaluated for error optimization on this

data set. In the instance of the patient having heart disease, the value is set to 1, else the value is set to 0 indicating the absence of heart disease in the patient. The pre-processing of data is carried out by converting medical records into diagnosis values. The results of data pre-processing for 297 patient records indicate that 137 records show the value of 1 establishing the presence of heart disease while the remaining 160 reflected the value of 0 indicating the absence of heart disease.

PERFORMANCE MEASURES:

Several standard performance metrics such as accuracy, precision and error in classification have been considered for the computation of performance efficacy of this model. Accuracy in the current context would mean the percentage of instances correctly predicting from among all the available instances. Precision is defined as the percentage of corrective prediction.

SYSTEM DESIGN ARCHITECTURE:



IMPLEMENTATION RESULT

In the prediction of heart disease we used UCI Dataset .Its Consist of 303 Patient Health Data. We analysis with the help of R Studio . The detection method is used our proposed algorithm K-Means and KNN.

CONCLUSION

In this paper we studied how data mining techniques brings with set of techniques to find out hidden patterns for making decision in healthcare organizations. We focused on classification methods of data mining used in data discovery. Different classification techniques of data mining have merits and demerits for data classification and knowledge extraction

LITERATURE REVIEW

1) A.sheik abdullah, r.r.rajalaxmi – data classification using random forest classifier
It produces a highly accurate classifier, but it is much harder and time consuming
.
2) Azhar hussein alkheoush, mariam zormodi moghodam, inas ai mansoori – particle swarm optimization (pso) algorithm
It is the most powerful evolutionary algorithm to generate rules for heart disease, but it is difficult to definite initial design parameters.

3) Nabeel ai-milli – neural network and back propogation algorithm - over other function approximation methods on any domain that use the neural networks, but it is not accurate.

4) Durairaj m, revathi v – back propogation mlp (multilayer perception) of artificial neural network. This is considered to be the best algorithm for prediction, but mlp is sensitive to feature scaling and requires a number of hyper parameters such as number of hidden neurons, layers and iterations.

5) Nagaraj m. lutimath, chethan c, basavaraj s pol – machine learning (decision tree, neural networks, naive bayes classification, genetic algorithms, regression, and support vector machines)

REFERENCES

1)Amin Ul Haq ,Jian Ping Li ,Muhammad Hammad Memon ,Shah Nazir ,and Ruinan Sun,"A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms". Hindawi,Mobile Information Systems,Volume 2018.

2)Avinash Galande, Pavan Kumar T," Heart Disease Prediction Using Effective Machine Learning Techniques".International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June. 2019.

3)Azhar Hussein Alkeshuosh and Mariam Zomorodi Moghadam, "Using PSO Algorithm for Producing Best Rules in Diagnosis of Heart Disease". International Conference on Computer and Applications, 2017.

WEB REFERRENCE
https://www.analyticsvidhya.com/blog/2020/09/web-application/
https://www.mattioli1885journals.com/index.php/progressinnutrition/article/view/983
https://digitalcommons.unl.edu/libphilprac/4032/