International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072

Human Heart Disease Prediction Using Ensemble Learning And Particle Swarm Optimization

Latha.B[1], Nischitha.U[2], S.Srividhya[3]

[1] B.E Student, Dept. of Information Science and Engineering, BNM Institute of Technology, Bengaluru-560070(Karnataka).
[2]B.E Student, Dept. of Information Science and Engineering, BNM Institute of Technology, Bengaluru-560070(Karnataka). [3]Assistant Professor, Dept. of Information Science and Engineering, BNM Institute of Technology, Bengaluru-560070(Karnataka)

Abstract - *The technologies like analytics, artificial intelligence, machine learning have various benefits in the field of healthcare to diagnose various diseases. Now a days, heart disease has become more. Blood pressure, cholesterol and pulse rate are different for individual patients. In order to predict heart disease, machine learning algorithm such as particle swarm optimization and ensemble method (Ada Boost , Gradient descent , Random Forest , Decision Tree , Gaussian Naive Bayes) is used. Particle swarm optimization is used to extract important features from the datasets whereas ensemble is used to select the algorithm which gives maximum accuracy. Ada Boost and Particle swarm optimization has achieved the highest accuracy.*

***Key Words*: Heart Disease, Ensemble Learning, Particle Swarm Optimization, Data Pre-processing**

1. INTRODUCTION

Blockage of blood in the coronary arteries leads to heart disease. The supply of the blood to the heart muscle is reduced because the arteries are blocked with fatty deposits which affects the heart. Chest pain, discomfort in shoulder and back, fatigue, shortness of breath, nausea, heart burn are the symptoms of heart disease. If the Blood pressure, cholesterol and pulse rate exceeds the threshold value there is a chances of having heart disease. Smoking, drinking also causes heart disease.80% of death in the world is because of heart disease. World Health Organization has estimated that by 2030,23.6 million people will die due to Heart disease[1]. In the rural areas, though the doctors are available, patients are unable to reach the doctors in time. So data mining techniques are used to predict the heart disease.Data mining techniques such has classification, clustering, regression can be used to predict the heart disease. Particle Swarm optimization is an optimization technique used as feature selection so that heart disease can be predicted [2]. This work uses Ensemble learning techniques and Particle Swarm Optimization to predict the heart disease. Data preprocessing is done in the first step, data preprocessing means filling the missing values. Then Ensemble classifier is applied to select the classifier which gives maximum accuracy. After that Particle Swarm Optimization is applied to extract the important features that affect the heart disease.

2. RELATED WORK

RajwantKaur, SukhpreetKauret.al[3] uses SVM classifier to predict the heart disease based on the risk factors .This system saves both money and time to under medication. The dataset are taken from American Heart Association to predict the heart disease and which showed the accuracy of 95%.M.A.Jabbar, Dr.B.L. Deekshatulu et.al[4] uses Naive bayes algorithm to predict the heart disease. Naive bayes classifier classifies the presence and absence of the heart disease depending on a feature set. The datasets are taken from UCI machine learning repository to apply this method which gave the accuracy of 86.29%.M.A.Jabbar,Dr.B.L.Deekshatulu et.al[5] uses Lazy associative method to predict the heart disease. Lazy associative classification method induces class association rules specific to test instance. This approach uses dataset that are collected from UCI machine learning repository and one real life data set that is collected from Andhra Pradesh heart association and achieved an accuracy of 90%. V.Krishnaiah, M.Srinivas et.al[6] uses Fuzzy K-NN classifier method to predict the heart disease. Inorder to remove the uncertainity in the dataset fuzzy KNN classifier is used.The dataset has a 200 records with a collection of 13 attributes. Accuracy of this approach is 93%. Kathleen H. Miao1 , Julia H. Miao1 and George J. et.al[7] uses ensemble learning to predict the heart disease. The ensemble learning that is used is an adaptive boosting algorithm. The developed models were utilized to classify the presence and absence of heart disease. The datasets are collected from various corporate hospitals in Hyderabad. The proposed approach achieved an accuracy of 91.66%. M.A.Jabbar, Dr.B.L.Deekshatulu et.al[8] uses Associative classification to predict the heart disease. Associative classification is possible to build more accurate classifier if it is focused on limited set of association rules where the consequent of the rule is restricted to class. The data sets are collected from UCI machine repository. The proposed approach have an accuracy of 86.80%.

3. THEROTICAL BACKGROUND

This section discusses basic concept of Particle Swarm Optimization and Ensemble Learning.
Particle Swarm Optimization:

Particle Swarm Optimization is an optimization algorithms. Particle Swarm Optimization contains populations which are encoded as particles. In search space, the particles are randomly initialized. For best solution, the particles in search space need to repeatedly update their positions based on its own position and based on neighbouring particles. In PSO, important factor that affects the heart is selected.

Ensemble Learning:

Ensemble Classifier is used to select one model out of multiple models which gives the maximum accuracy.An ensemble model means a collection of an individual model.In this work five algorithms are used which is listed as below

- Adaboost
- Gradient descent
- Random Forest
- Decision Tree
- Gaussian Naïve Bayes

AdaBoost:

Adaptive Boosting is a type of ensemble classifier method which converts weak learner into strong learner. It works by selecting a base classifier such as decision tree. It can be improved by taking misclassified attributes from the dataset in an iterative method. Equal weights are assigned to the attributes and choose a base classifier.

Gradient Descent:

Gradient descent is the popular optimization strategy which is used in machine learning and deep learning. It is used for training the model.It is easy to understand and implement the algorithm. It is based on a convex function, that weaks the parameters iteratively to minimize a given function to its local minimum.

Random Forest:

Random Forest is a type of ensemble classifier which is used in medical applications for better prediction accuracy. Random Forest algorithm constructs N number of decision trees by selecting random attributes as their input. In Random Forest, the bias remains the same for all if the number of trees increases also. As the number of trees is increased in the forest the variance, of the model decreases. The prediction can be decided by taking the average voting or majority voting.

Decision Tree:

Decision tree is the important tool for classification and prediction. A Decision tree is a tree like structure, which has nodes. Nodes represents attributes and each branch represents an outcome of the datasets. Inorder to construct the decision tree classifier domain knowledge or parameter setting is not required. Decision trees can be applied on high dimensional data.

Gaussian Naive Bayes:

A Gaussian Naive Bayes algorithm is a type of Naive Bayes algorithm. It is applied on the features which have a continuous values. For all the features gaussian distribution is calculated. The Naive Bayes classifier is based on the simplifying assumptions that the attribute values are conditionally independent given the target value.

4. PROPOSED APPROACH

The main motive of this approach is to apply ensemble classifier for classification and apply particle swarm optimization as feature subset selection for prediction of heart disease. The dataset is containing 16 attributes initially which will be reduced to 5 attributes by applying PSO. Dataset are taken from kaggle.

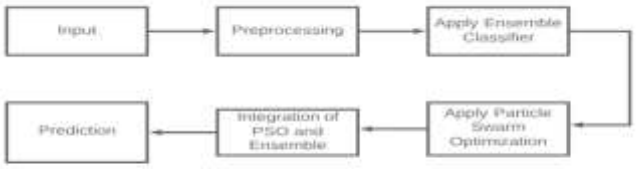

Figure 1 System Architecture

Algorithm

Step 1: Loading the dataset
Step 2: Apply data preprocessing Technique on heart disease data set
Step 3: Apply ensemble classifiers to select the algorithm which gives maximum accuracy.
Step 4: Apply Particle Swarm Optimization is used to select the important features from the datasets.
Step 5: Predicting the heart disease.

5. RESULT

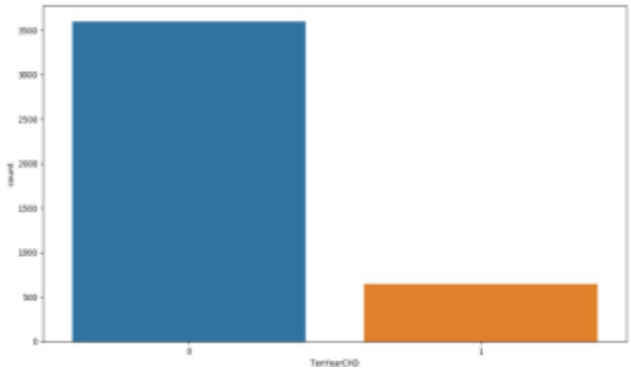

Figure 2 Results of Data preprocessing

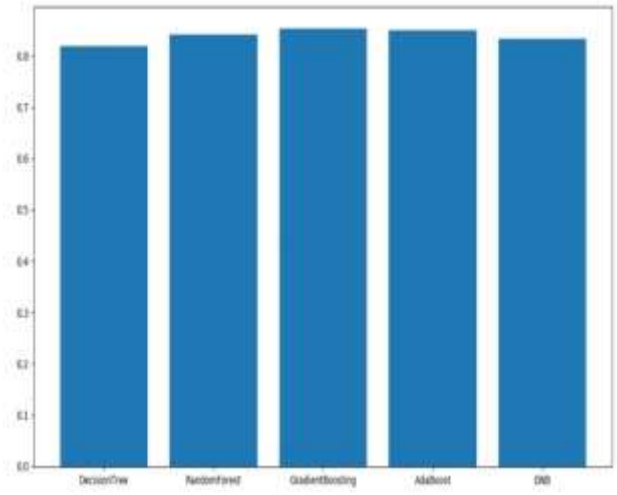

Figure 3 Accuracy obtained by various ensemble methods

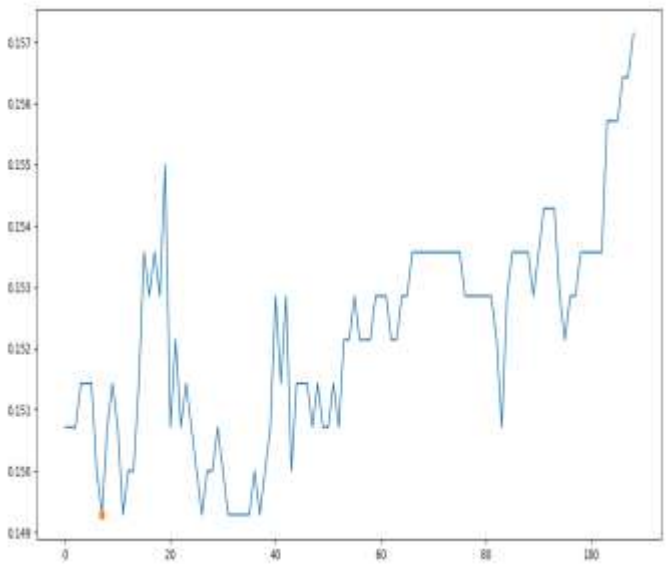

Figure 4 Implementation of Particle Swarm Optimization.

Figure 5 User Interface

The accuracy of the proposed approach 84.88% using AdaBoost and Particle Swarm Optimization. The error rate of the proposed approach is 4%.

6. CONCLUSION

In the proposed approach PSO is used to select the main features and remove other features from the training datasets. An ensemble classifier such has AdaBoost is applied on the pre-processed data to know the accuracy of the prediction. The learning accuracy of this system is significantly improved by using AdaBoost and Particle Swarm Optimization. This model will help doctors to accurately predict and early diagnosis of heart diseases using a subset of features.

REFERENCES

[1] Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, A data mining approach for prediction of heart disease using neural networks, international journal of computer engineering and technology, 2012.

[2] M.A.Jabbar,"Prediction of heart disease using k-nearest neighbor and particle swarm optimization", Biomedical Research 2017 ISSN 0970-938X, pp. 4154-4158.

[3] RajwantKaur,SukhpreetKaur,"Prediction of Heart Disease Based on Risk Factors Using Genetic SVM Classifier" IJARCSSE Volume 5 Issue 12 December 2015,ISSN:2277 128X,pp.205-208.

[4] M.A.Jabbar, Dr.B.L.Deekshatulu,"Computational Intelligence Technique for early Diagnosis of Heart Disease",IEEE International Conference on Engineering and Technology 2015.

[5] V.Krishnaiah, M.Srinivas,"Diagnosis of Heart Disease Patients Using Fuzzy Classification Technique",IEEE International Conference on Computer and Communications Technologies (ICCCT), 2014.

[6] M.A.Jabbar, Dr.B.L.Deekshatulu,"Heart Disease Prediction using Lazy Associative Classification", IEEE 2013, pp.40-46.

[7] Kathleen H.Miao1, Julia H. Miao1and George J. Miao," Diagnosing Coronary Heart Disease Using Ensemble Machine Learning" IEEE, Vol. 7, No. 10, 2016,pp.30-39.

[8] M.A.Jabbar, Dr.B.L.Deekshatulu,"Alternating decision trees for early diagnosis of heart disease", Proceedings of International Conference on Circuits, Communication, Control and Computing IEEE 2014.