International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 09 | Sept -2020 ISSN: 2582-3930

PREDICTION SYSTEM FOR HEART DISEASE USING NAIVE BAYES AND PARTICLE SWARM OPTIMIZATION

[1]R.M.Suruthi,[2]G.Naveenprabu,[3]R.Sandhiya Mathi,[4]K.Sashtika,[5]V.Sofiya

[1]Assistant Professor
[2,3,4,5]Student Pursuing Bachelor of Engineering
Department Of Computer Science and Engineering
Coimbatore Institute of Technology,
Coimbatore, India

.

Abstract

The term "heart disease" is often referred as "cardiovascular disease" is a condition that includes narrowing or blocking of blood vessels that leads to stroke ,chest pain and heart attack .Using the datasets available from the health care industry, any disease can be detected ,predicted and cured at the earliest .Many researchers have found that people died due to detection of heart disease at later stages.Therefore accurate prediction of heart disease is very important.In the proposed modulation the medical data such as blood pressure, cholesterol level, heart rate are taken as input and specific features are selected and modelled for prediction. Feature selection is done using PSO (particle swarm optimization) algorithm.The prediction model is built with Naive Bayes and prediction of heart disease can be done with the built model.

Keywords: *Heart disease, Naive based classifier, Feature selection,Particle swarm optimization,.*

I INTRODUCTION

Heart disease is a type of disease that involves the heart or blood vessels connecting to the heart. Heart disease which is also referred to as Cardio Vascular Disease includes Coronary Artery Diseases (CAD) such as angina and myocardial infarction (commonly known as heart attack). Heart disease is the leading cause of death mostly in men. However, women also have the same risk after menopause. Discomfort in chest and pain are the common symptoms. The risk of heart attack increases when a man is around 45 and woman is around 55. Also, there are some clear symptoms of heart attack that require immediate medical attention.

Some of major symptoms of heart disease are chest tightness, shortness of breath, nausea, indigestion, heart burn or stomach pain, sweating and fatigue, pressure in the upper back pain that spreads to the arm.

The other risk factors include age, gender, stress and unhealthy diet. Chance of having a heart disease increases when a person is getting older. Men have a greater risk of heart disease. However, women also have the same risk after menopause.But with the early detection of the disease it can be treated and cured. Since numerous factors are involved in heart disease, accurate prediction of this disease is challenging.

In the past few decades, medical data mining have played an important role to explore the hidden patterns which can be

used for clinical diagnosis of any disease dataset [3].

Classification is one of the data mining technique to classify the patient class as normal or heart disease but classification use all attributes either relevant or irrelevant features which may reduce the classification performance. Feature subset selection is one of the dimensionality reduction techniques use to improve the accuracy. Particle swarm optimization is a feature selection technique which removes the redundant features to improve the classifier performance.

So, in this paper based on the factors mentioned above we try to predict the risk of heart disease .Our proposed model identifies the relevant features and removes the irrelevant features, to predict the heart disease effectively.

II RELATED WORK

H.Benjamin Fredrick David and S.AntonyBeley[1] have used data set of 14 attributes.Using the classification algorithms like Random forest,Decision tree, Naive Bayes models are built and the accuracy of each model is tested.

Navdeep Singh and SonikaJindhal [2] used the most effective algorithms Navie Bayes and Genetic Algorithm for their performance analysis on heart disease prediction

M.Marimuthu,S.Deivarani , Gayathri.R[3] say that are seven key factors for heart disease such as smoking, physical inactivity, nutrition, obesity, cholesterol, diabetes and high blood pressure. Considering all theseparameters they have built a model and algorithms like naivbayes, SVM, Decision tree are used to predict the possible coronary heart disease patient in next 10 years .

Amandeep Kaur and Jyoti Arora [4] made analysis on different classification algorithms such as Decision Tree, Naïve Bayesian, Support Vector Machine (SVM), N classifier,Hybrid Approach, ArtificialNeural Network ANN.

III CLASSIFICATION

Classification consists of predicting outcomes based on several inputs given. In order to predict the outcome, the algorithm processes a set of attributes as a training set and the respective outcome is usually called as goal or prediction attribute. The algorithm aims in discovering relationships between the attributes that would make it possible to predict the result. Next the algorithm is given a data set which contains the same set of attributes, except for the prediction attribute (i.e. the outcome or the resultant of the prediction) . The algorithm analyses the input and predicts the existence of disease in any individual. Machine learning algorithms are used for classification.

IV CLASSIFICATION TECHNIQUE

There are primarily two types of classification

1.Supervised classification

2.Unsupervised classification

In Supervised classification, distribution-free process or statistical process is used in extracting class dependencies.Supervised learning is a Data mining task which infers a function from labeled training data.The training data contains a set of training examples. In supervised learning, each example is a pair which consists of an input object and a desired output value. A supervised learning algorithm aims in analysing the training data and provides an inferred function, which will be useful in mapping new examples.

V NAIVE BAYES CLASSIFIER

Naive Bayes classifiers are based on Bayes theorem with appropriate independence inference between the features. The Naive Bayes classifiers are a group of simple probabilistic classifiers. Naive Bayes classifiers are highly scalable by requiring a number of parameters linear for the number of features or predictors as variable in a learning problem. It is the simplest and the fastest probabilistic classifier especially for the training phase.

VI FEATURE SELECTION

Feature selection is a process of discarding the inappropriate and redundant features from dataset based on evaluation principle which is used in improvising accuracy. There are two approaches namely the individual evaluation and other one is the subset evaluation. The feature selection process is classified into three broad classes. One is filter and another one is wrapper and third one is embedded method based on how the feature selection is deployed by supervised learning algorithm.

VII PARTICLE SWARM OPTIMISATION (PSO)

PSO is an Evolutionary Computation technique proposed by Kennedy and Eberhart in 1995 . PSO was inspired by social behaviors such as bird flocking and fish schooling. In PSO ,population swarm contains “n” particles, and the position of each particle emphasize the potential solution in dimensional space D.
The particles change its condition based on three aspects:

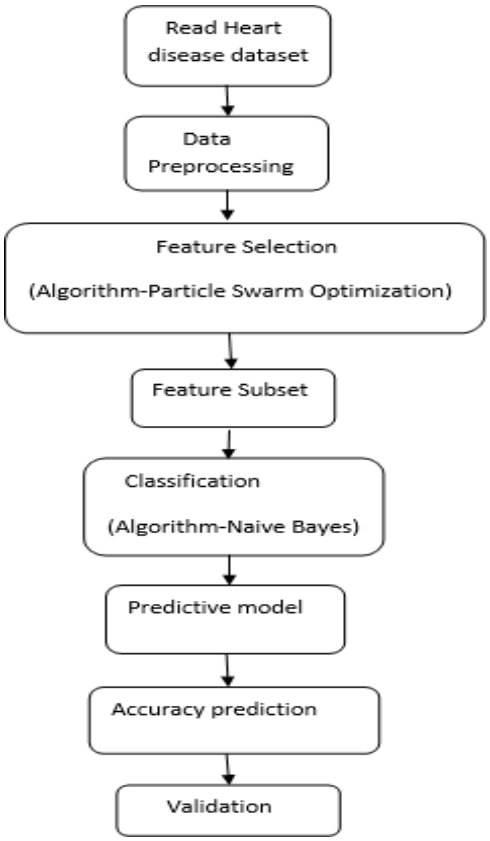
- To keep its inertia
- To change the condition according to its most optimist position
- To change the condition according to the swarm’s most optimist position.

In PSO, a population are encoded as particles in the search space dimensionality D. PSO starts with the random initialisation of a population of particles. Based on the best experience of one particle (pbest) and its neighbouring particles (gbest), PSO searches for the optimal solution by updating the velocity and the position of each particle.

VIII PROPOSED SYSTEM

In proposed methodology, performance of Naive Bayes classifier for prediction of Heart disease is improved using Particle Swarm Optimization (PSO) algorithm.
Algorithm:
Heart disease prediction using Naive Bayes classifier and Particle Swarm Optimization.
Input: Heart disease dataset.
Output: Patient dataset classified into heart disease presence and absence (normal).
Step 1: ead the heart disease dataset. Step 2: Preprocess the heart disease dataset.
Step 3: Apply Particle Swarm Optimization algorithm for feature selection.
Step 4: Remove the Features with low value of PSO.
Step 5: Apply Naive Bayes classifier on dataset with selected features.
Step 6: Evaluate the performance of the model interms of accuracy.

IX FLOW CHART



X EXPERIMENTAL RESULTS

In the proposed method, the accuracy of Naive Bayes classifier is improved by using the particle swarm optimization. PSO is used for feature selection. The dataset that is being usedcontains 12 features including 1 target class and 457 instances.

The features in the dataset are as follows:

i.Age: Age in years
ii.Sex: Sex in number
iii.cp: Chest pain type
iv.trestbps: Resting blood pressure
v.chol: Serum cholesterol in mg/dl
vi.fbs: Fasting blood sugar level
vii. restecg: Resting electrocardiographic results
viii. thalach: Maximum heart rate achieved
ix.exang: Exercise induced angina
x. oldpeak: ST depression induced by exercise
xi. num: Target class that denotes the identification of heart disease
xii. place: Address of the patient.

Heart disease presence is predicted based on the features selected by PSO algorithm. The selected features are age, sex,cp, trestbps, chol, fbs, restecg, exang, oldpeak, place.

The predictive model with Naïve Bayes accuracy is 77.17% and Naive Bayes+PSO is 79.34%. Improvement in accuracy in proposed model is 2.17%.

XI CONCLUSION

It is concluded that the proposed model is effective and efficient to improve the accuracy of the Naive Bayes classifier using the particle swarm optimization for feature subset selection which achieves similar or even better classification performance. The goal was successfully achieved by maximizing the classification performance and minimizing the number of features. From simulation results, it is analysed that this algorithm could automatically evolve a feature subset selection with a less number of features and increase classification performance than using all the features of a dataset. In the future, we want develop recommendation system for early prediction of heart disease diagnosis. Also the use of PSO for feature selection on datasets with a huge number of features can also be studied to note the various aspects of PSO in feature selection.

XII REFERENCES

[1] H. Benjamin Fredrick David and S. Antony Belcy ,Heart disease prediction using data mining techniques.

[2] Navdeep Singh, Sonika Jindal Heart Disease Prediction System using Hybrid Technique of Data Mining Algorithms.

[3]M. Marimuthu, S.Deivarani , Gayathri.RAnalysis of Heart Disease Prediction using Various Machine Learning Techniques.

[4] Amandeep Kaur and Jyoti Arora Heart disease prediction using data mining techniques.