Prediction of Coronary Heart Diseases using Computational Algorithms

Prof Anu Thomas,SIES(NERU)COLLEGE OF ARTS,SCIENCE AND COMMERCE

Abstract

Cardio Vascular Diseases (CVD) is caused due to the disorder in the heart and blood vessels which resulted in coronary heart diseases, cardiac arrest, stroke etc. For diagnosis of Coronary Heart Diseases(CHD), expensive as well as time consuming test like electrocardiogram ,stress testing, echocardiography and coronary angiography is usually been done .Researchers are therefore forced to find an alternative method through different methodologies such as machine learning algorithms which can predict and detect the coronary heart diseases at the early stages. This research is aimed at prediction and detection of coronary heart diseases by hybridizing computational intelligence algorithm. The suggested research will be carried out through three major levels mainly preprocessing, feature selection, classification . Preprocessing is the initial level, where the data set is pre-processed by using numeric to nominal and replacing missing value techniques as well as the conversion of unstructured data into structured one. The next stage is feature selection which is used for improving the prediction performance, giving quicker and budgeted predictions . Dimensionality reduction also plays an important role in the field of unstructured data as it contains multiple attributes which inevitably will be a part of feature selection. And the last stage is hybridising the computational algorithm which will be used for the classification .The proposed system will be compared and analysed with the existing system based on the accuracy and adaptability.

Introduction

Coronary heart diseases continues to be the greatest problem of death for the last few decades. In recent years, computer technology and machine learning techniques are used to build software to help doctors in making decisions on heart diseases at primary stages. Heart diseases prediction model can

help medical professional in predicting heart diseases status, grounded on the clinical data of the patients. For this, data mining with intelligence algorithm is used to handle the problem of prediction in medical data sets consisting of multiple inputs.

According to the new survey[1], heart diseases are the major reason for deaths. The existence of life is depending on the smooth working of human heart, used for supplying the blood to the entire human body. Any heart diseases will create problems to the other parts of the body mainly brain, kidney, lungs etc. There exist several factors that can raise the risk of heart diseases, mainly varying blood pressure, smoking ,alcohol consumption, obesity, lack of exercise. World Health Organization(WHO) has assessed that 17.5 million death occurred universally in the year 2012 and they are predicting that by 2030,it may rise up to 23.6 million.

In medical field, researchers mainly opted for Back Propagation Neural Network algorithm(BPNN) to represent a problem which is not structured because of its possibility to link composite nonlinear relationship amongst input and output attributes. Back propagation algorithm is also known as native search algorithm.it is used to produce various sets of weights and biases for each execution in training phase. Because of this, each execution will give different prediction output and convergence speed.

To overpower this shortcomings, Gray Wolf Optimizer(GWO) is utilised to calculate the desirable opening weights and biases of BNN algorithm[15].GWO algorithm is a biologically motivated optimization algorithm. Multi Objective Gray Wolf Optimiser(MOGWO) is a variant, in which two modules remained united to the GWO algorithm to permit it to accomplish multi objective optimization. The principal module will acts as a library for storage and recovering the greatest non controlled results so far in optimization and the subsequent module will be a leader selection mechanism. In this research ANN is hybridised with MOGWO. To develop a perfect ANN model and to delimit the shortcomings of back propagation algorithm.

Literature Review

Emina Alickovic &Abdulhamit Subasi(2016) [1], proposed a system for diagnosing the heart arrhythmia by using Discrete Wavelet Transform(DWT) and Random Forests(RF) classifier. Heart

arrhythmias or heart rhythm problem occur when the heart beats too fast, too slow or irregularly. In this research work, Random Forests (RF) classifiers is recommended for ECG's signal arrangement in analysis of heart arrhythmia. Discrete Wavelet Transform (DWT) is utilized to divide ECG sign into various consecutive frequency groups. Precise ECG signal ordering is the major requisite for recognition of all arrhythmia forms. Performance of the planned structure is calculated using two diverse datasets, specifically MIT-BIH datasets and St Petersburg Institute of Cardiologcal Technics 12-lead Arrhythmia datasets. For MIT-BIH database, RF classifier has generated a total accuracy of 99.33 percentage against 98.44 percentage and 98.67 percentage for the C4.5 and CART classifier. For St Petersburg Institute of Cardiological Technics 12-lead Arrhyhmia database,RF classifier has generated a total accuracy of 99.95 percentage against 99.80 percentage for both C4.5 and CART classifier.

Reddy et al.,(2016) [2] ,proposed a data mining technique using decision tree in order to find out the accuracy of the prediction of heart diseases. They have used classification techniques . According to the author, it is possible to predict heart diseases in diabetic patients with a reasonable accuracy by using decision tree and by using these classifiers, prediction can be done well in advance. Here they have implemented two steps in this technique for building and applying the tree to the data sets. They have used many famous tree algorithms like CART, ID3, C4.5, CHAID, and J48.In conclusion, the authors had explained that decision tree algorithm is the most remarkable classification model, but they fail to prove it through a data set. Neither they could predict the accuracy of this algorithm using any technique. They have mentioned that simple k-means algorithm is used. The author had also mentioned the future enhancement by using stacking techniques to increase the accuracy of decision tree and reduce the number of leaf nodes.

Ilayaraja M, Meyyappan T (2015) [3], proposed an effective mining methods to forecast the risk of heart problems through frequency item sets. In this research work ,the authors have developed a system to forecast the risk level of the patients having heart problems through frequency item sets.

The database of several heart disease patients are collected for this research work. Frequency item sets are created based on the selected symptoms and minimum support value. The mined frequency item sets help the medical consultant to make investigative decisions and decide the risk level of patients at an initial stage. The have proposed a system which can be applicable to any medical database to predict the risk factors with risk level for any patients based on selected factor.

Archana and Sandeep (2015) [4], have developed a prediction algorithm using missing value imputation. The research suggested a combinational model built on K means clustering with different perceptron. The result of the planned procedure is explored on 3 medicinal data sets specifically Wisconsin Brest Cancer , Hepatitis and Prima Indians Diabetics from UCI repository . The output of the research is strong and the projected model has given a good output when number of misplaced values are huge in the data base.

Mrutyunjaya Panda , Ajith Abraham(2015)[5] , proposed an innovative techniques to discover the finest pertinent feature subclass using fuzzy rough set based attribute selection using biologically motivated algorithm quest such as ant colony and particle swarn optimization. The authors have proposed a K nearest neighbour(KNN) based classifier (FRNN) with hybrid fuzzy rough set to classify the designs in the condensed datasets, received from the fuzzy rough bio inspired algorithm search. The authors have proposed a hybrid method which can be consequently validated using real time database acquired from California university, machine learning repository Irvine. Authors have developed a simulation system. Simulation results provided a good classification accuracy. They have also conducted the parametric and non-parametric statistical test of significance to witness uniformity of the classifiers.

M.Akhil jabbar , Priti Chandra B.L Deekshatulu (2013)[6] , proposed a unique algorithm which merges KNN with generic algorithm for remarkable classification. Generic algorithm is hybridised to perform a universal exploration in large and multimodal data sets and to give a best solutions The proposed algorithm is divided into two parts, first part deals with calculating variables using

genetic search and the second part handles with developing classifier and calculating accuracy of the classifier. The performance of the proposed system was checked with six medical databases and one non-medical database. Out of this 7 datasets,6 data sets are taken from UCI repository and one is taken from hospitals in Andhra Pradesh. The main gap of this research is that the proposed method was not able to account for the misplaced or redundant attributes present in the datasets.

Resul Das,Ibrahim Turkoglu,Abdulkadir sengur(2009)[7], proposed an effective way of diagnosing the heart diseases through neural network ensembles. In this research work, the authors have introduced a method in which they have used SAS base software 9.13 for detecting heart diseases. This collaboratively constructed technology makes a novel model by merging the subsequent possibilities or the foreseen values from different ancestor models. They also performed many investigation with the projected tool and have obtained 89.01 percentage classification accuracy from the investigation made on the data taken from the heart diseases database,Cleveland. They have obtained 80.95 percentage and 95.91 percentage sensitivity and specificity values in heart disease diagnosis.

Markos G. Tsipouras et al,(2008) [8] ,proposed the analysis of coronary artery disease using fuzzy modelling and data mining. The proposed system is developed from an original explained database, by a 4 step methodology. Firstly, they did induction of a decision tree from data. Secondly, they removed a set of procedures from the decision tree, and formulated a crisp model. Thirdly, they changed the rules of crisp set into a fuzzy model and finally they optimised the attributes of the fuzzy model. The authors have taken tenfold cross validation and the average specificity and sensitivity received is 62 percentage and 54 percentage individually, using the fixed rules taken from decision tree, while the average sensitivity and specificity increase to 80 percentage and 65 percentage individually, when optimization and fuzzification stages were used.

Sellappan Palaniappan ,Rafiah Awang(2008)[9] , projected an heart diseases forecast system by mining methods. The researchers have developed a model of intelligent heart diseases prediction system using data mining methods, specifically, naïve bayes,neural network, and decision.

Outcome of this model showed an exclusive power in understanding the objectives of the mining goals. This prediction system can answer 'what if ' queries which any decision support systems can not

Gang Kou et.al (2007)[10], proposed a privacy preserving data mining techniques using data separation based methods. This research paper uses a technique for data separation for preserving privacy in classification of medicinal data. The author have taken two methods to protect privacy, firstly to vertically partition the medicinal data and mine these divided data at different locations and the next method is to horizontally partition data across different locations. They have implemented these two techniques using UCI KDD archive.

Harleen Kaur and Siri Krishan Wasan(2006)[11] , proposed an experimental study on the data mining applications in healthcare. Author have used classification based data mining methods such as decision tree, artificial neural networks and rule based artificial neural networks to huge amount of data. The aim of this classifications is to allocate a class to find earlier hidden records. The primary idea is to discover a classification model for class attributes,where a test set is used to find the accurateness of the system. The data set is divided into training and test datasets. Classification process consists of training datasets which is analysed by classification algorithm.Test data sets are used in classification rules to calculate the accuracy.

Vili Podgorelec et al.(2005)[12],proposed an algorithm using machine learning to knowledge discovery and data mining based on the classification rules induction. This is a method for automatic rules induction called Automatic Rule Extractor(AREX) using automatic programming and decision tree is introduced. The projected method is applied to cardiovascular data sets containing different clusters of variables which will probably disclose the presence of certain specific cardiovascular difficulties in young patients. After doing the literature review, it is clear that prediction and detection

of heart diseases is a need for the medical domain .From this source, this research can be extended by hybridizing the computational intelligence algorithm using clinical data set.

Need for the study

- Machine Learning algorithms capabilities are much higher than explored, the usage of machine learning on medical domain will be of great help.
- In line with the literature review attempted , it is very clear and concise that prediction of heart diseases is the need of the hour. Development of a system that will support existing medical growth is a mandate.
- Any human being gets throttled when it comes to medical diagnosis, this research will have a social impact and will be a boon for the healthcare.

Major Concepts

1.Artificial Neural Networks

Artificial Neural Network(ANN) is a method used to solve data mining applications. Neural networks is an interconnected network, which provides rich structure giving some features of the biological neural networks. Neural Networks gives an advantage to the user by implementing parallel concepts at every layer .Neural networks has 3 layers: input , hidden and output layers. Amount of hidden layers and the amount of neurons in every layer depends on the complexity of system. ANN are trained based on two methods. Supervised training and Unsupervised training. For supervised training, external is available which gives the neural network input data and desired output. In unsupervised training, the input data and computation function is given and the output is calculated.

2.Gray Wolf Optimiser

Gray Wolf Optimiser(GWO) is mainly implemented for discovering the initial optimal weights.GWO represents the process of hunting among gray wolves in the jungle.They live in pack and there are

mainly two gray wolves which manages the other wolves in the pack.GWO is also called as a universal search methods while back propagation is local search one.

3.Multi Objective Gray wolf optimiser

Multi Objective Gray Wolf Optimiser (MOGWO) is used, in which mainly two new modules are united with the GWO algorithm to accomplish multi objective optimization. The first components is a leader selection mechanism and the second component is an archive for storing and retrieving the best non dominated obtained solution so far during optimization .

Statement of the problem

Design and develop an efficient method to predict and detect heart diseases from clinical data by using a hybridized computational intelligence algorithm.

Objective of the study

Primary Objective

- To explore the prediction and detection of the heart diseases with the help of Computational intelligence

Secondary Objective

- To select the best feature subset that contributes most to accuracy and efficiency.
- To develop an user friendly tool to convert an ECG signal into a human understandable format.

Design of Study

In heart diseases dataset, initially the data is preprocessed followed by feature selection and this predefined data sets is used to train the ANN with an initial weight, and the trained ANN is tested with rest of the data set to get an accuracy based results. Weight is updated and performed with ANN and MOGWO and then hybridized ANN with MOGWO. And in the last stage, performance in terms of accuracy is compared.

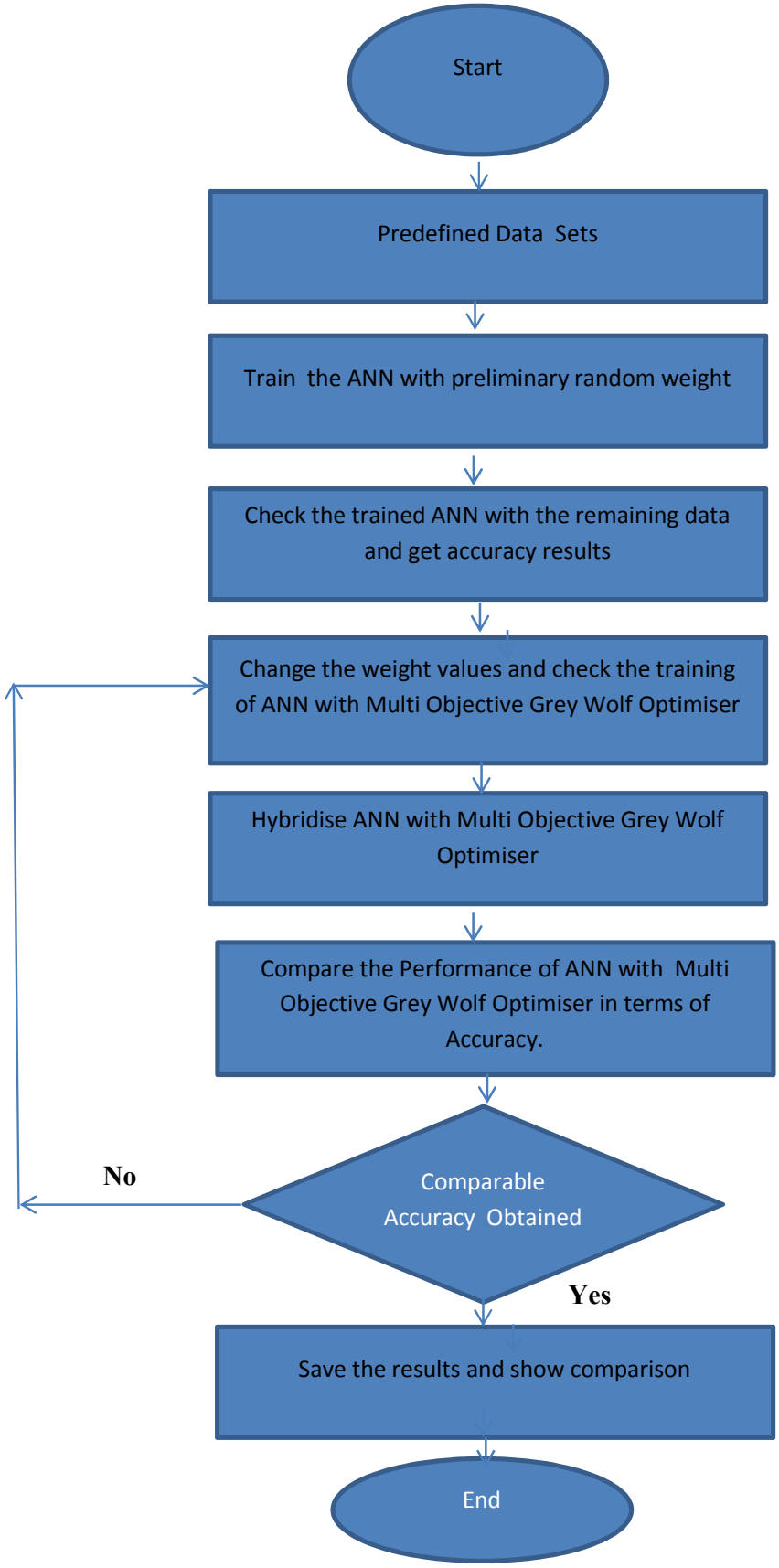

Fig 1:Draft of research approach

References

[1]"Cardiovascular diseases (CVDs)", *World Health Organization*, 2016. [Online]. Available: http://www.who.int/cardiovascular_diseases/en/. [Accessed: 07- Nov- 2016].

[2] E. Alickovic and A. Subasi, "Medical Decision Support System for Diagnosis of Heart Arrhythmia using DWT and Random Forests Classifier", *J Med Syst*, vol. 40, no. 4, 2016.

[3] R. R.V.K, R. K.P and S. C.H, "Prediction of Heart diseases using decision tree approach", vol. 6, no. 5, pp. 530-532, 2016**.**

[4] Ilayaraja M. and Meyyappan T., "Efficient Data Mining Method to Predict the Risk of Heart Diseases Through Frequent Itemsets", *Procedia Computer Science*, vol. 70, pp. 586-592, 2015.

[5]A. Purwar and S. Singh, "Hybrid prediction model with missing value imputation for medical data", *Expert Systems with Applications*, vol. 42, no. 13, pp. 5621-5631, 2015.

[6]A. Purwar and S. Singh, "Hybrid prediction model with missing value imputation for medical data", *Expert Systems with Applications*, vol. 42, no. 13, pp. 5621-5631, 2015.

[7]M. jabbar, B. Deekshatulu and P. Chandra, "Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm", *Procedia Technology*, vol. 10, pp. 85-94, 2013.

[8]D. R, T. I and S. A, "Effective diagnosis of heart disease through neural networks ensembles", *Expert Systems with Applications,Elsevier*, vol. 36, pp. 7675-7680, 2009.

[9]T. M. G, E. T. P, F. D. I and M. L.K, "Automated Diagnosis of Coronary Artery Disease Based on Data Mining and Fuzzy Modeling", vol. 12, no. 4, pp. 447-458, 2008.

[10]P. S, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", vol. 8, no. 8, pp. 343-350, 2008.

[11]K. G, S. Y and C. Z, "Privacy-Preserving Data Mining of Medical Data using Data Separation – Based Techniques", vol. 6, pp. 429-434, 2007.

[12] W. S.K, B. V and K. H, "The Impact of Data Mining Techniques on Medical", vol. 5, pp. 119-126, 2006.

[13]P. V, K. P, S. M.M, H. M and R. I, "Knowledge Discovery with Classification Rules in a Cardiovascular Dataset", p. 80, 2005.

[14]H. Turbieh, "A hybrid ANN-GWO algorithm for prediction of heart diseases", *American Journal of operations Research*, vol. 6, pp. 136-140, 2016.