International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 4th Feb 2014. Link :http://www.ijiemr.org/downloads.php?vol=Volume-03&issue=ISSUE-02

Title **A REVIEW ON ANN BASED HEART DISEASE PREDICTION ANALYSIS**

Volume 03, Issue 02, Pages: 14–21.

Paper Authors

K. SRINIVAS, B. KAVITHA RANI

Jyothishmathi Institute of Technology & Science Karimnagar, Telangana, India

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

A REVIEW ON ANN BASED HEART DISEASE PREDICTION ANALYSIS

[1]K. SRINIVAS,[2] B. KAVITHA RANI
[1,2]Associate Professor, Department of CSE, Jyothishmathi Institute of Technology & Science
Karimnagar, Telangana, India

ABSTRACT

Medical image analysis, symptom based disease prediction is the part where the most sought after brains are working. In this paper we aim to present a model on the prediction on diagnosis of cardio vascular disease with ECG analysis and symptom based detection.The model aims to be researched and advance in further to become robust and end to end reliable research tool. We will discuss about the classical methods and algorithms implemented on CVD prediction, gradual advancements, draw comparison of performance among the existing systems and propose an enhanced multi-module system performing better in terms of accuracy and feasibility. Incorporating the techniques of classification in these intelligent systems to achieve accurate diagnosis. Neural Networks has emerged as an important method of classification. Multi-layer Perceptron Neural Network with Back-propagation has been employed as the training algorithm in this work. This paper proposes a diagnostic system for predicting heart disease. For diagnosis of heart disease 14 significant attributes are used in proposed system as per the medical literature.

INTRODUCTION

Data mining is the field of storing, structuring and analyzing massive scale historical and authenticated data to find the frequent or very unexpected patterns and correlations among the part objects which seem to be unrelated and iteratively continue the paradigm style approach to get the knowledge derived from the huge amount of data on a scalable manner. As the disease diagnosis is known to be the most crucial part of clinical medicine, application of data driven methods and using the doctors‟ knowledge and experiences to design the machine learning algorithms to carry over the patients‟ and data makes it easier, time efficient, affordable and more accurate. So hereby accuracy of the prediction is the prime concern of ours while going through a predictive methodology and model.The most complicated and complex task in the field of medical sciences is the prediction of heart disease. Heart is considered to be the most vital organ of the human body [1]. There is an intense need in predicting the level and seriousness of heart disease that provide an accurate treatment to the patients. Heart disease can be referred to various conditions that lead to abnormal functioning of heart, which may involve blood vessels, arteries etc.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Effective diagnosis of heart disease results in an appropriate treatment to a patient. This requires a deep study of cardiovascular analysis of the patient that includes symptoms such as chest pain chest tightness, chest pressure, and discomfort in breathing, numbness etc. [2]. The cardiovascular diagnosis involves certain decisions to be taken based upon the health history and the clinical test results of a person. The process of decision making is a challenging task to the medical practitioners which has to be done accurately and efficiently where a mere negligence may lead to the life risk of a patient.

Traditional diagnosing approaches have no proper automated tools use for the purpose of heart disease diagnostic system. The commonly used data mining algorithms for predicting diseases are:

- Genetic algorithm
- K-means algorithm
- MAFIA algorithm

Several methods proposed the implementation of classification algorithms in diagnosis of heart disease and resulted with an accuracy of 88.33%. They used algorithms such as Naive Bayes algorithm, Decision list algorithm and KNN algorithm with the ECG attributes and clinical symptoms to detect the heart disease

2. LITERATURE SURVEY

There are thirty-five research papers that explore the computational methods to predict heart diseases. The summaries of them have been presented in a nutshell.

Shaikh Abdul Hannan et al. [5] used a Radial Basis Function(RBF) to predict the medical prescription for heart disease. About 300 patient's data were collected from the Sahara Hospital, Aurangabad. RBFNN (Radial Basis Function–Neural Network) can be described as a three-layer feed forward structure. The three layers are the input layer, hidden layer and output layer. The hidden layer consists of a number of RBF units (nh) and bias (bk). Each neuron on the hidden layer uses a radial basis function as a nonlinear transfer function to operate on the input data. The most often used RBF is usually a Gaussian function. Designing a RBFNN involves selecting centers, number of hidden layer units, width and weights. The various ways of selecting the centers are random subset selection, k-means clustering and others. The methodology was applied in MATLAB. Obtained results show that radial basis function can be successfully used (with an accuracy of 90 to 97%) for prescribing the medicines for heart disease.

AH Chen et al. [6] presented a heart disease prediction system that can aid doctors in predicting heart disease status based on the clinical data of patients. Thirteen important clinical features such as age, sex, chest pain type were selected. An artificial neural network algorithm was used for classifying heart disease based on these clinical features. Data was collected from machine learning repository of UCI. The artificial neural network model contained three layers i.e. the input layer, the hidden layer and the output layer having 13 neurons, 6 neurons and 2 neurons respectively. Learning Vector Quantization (LVQ) was used in this study. LVQ is a special case of an artificial neural network that applies a prototype-based supervised classification algorithm. C programming language was used as a tool to implement heart disease classification and prediction trained via artificial neural network. The system was developed in C and C# environment. The accuracy of the proposed method for prediction is near to 80%.

Mrudula Gudadhe et al. [7] presented a decision support system for heart disease classification. Support vector machine (SVM) and artificial neural network (ANN) were the two main methods used in this system. A multilayer perceptron neural network (MLPNN) with three layers was employed to develop a decision support system for the diagnosis of heart disease. This multilayer perceptron neural network was trained by back-propagation algorithm which is computationally an efficient method. Results showed that a MLPNN with back-propagation technique can be successfully used for diagnosing heart disease.

Manpreet Singh et al. [8] proposed a heart disease prediction system based on Structural Equation Modelling (SEM) and Fuzzy Cognitive Map (FCM). They used Canadian Community Health Survey (CCHS) 2012 dataset. Here, twenty significant attributes were used. SEM is used to generate the weight matrix for the FCM model which then predicts a possibility of cardiovascular diseases. A SEM model is defined with correlation between CCC 121(a variable which defines whether the respondent has heart disease) along with 20 attributes. To construct FCM a weight matrix representing the strength of the causal relationship between concepts must be constructed first. The SEM defined in the previous section is now used as the FCM though they have achieved the required ingredients (i.e. weight matrix, concepts and causality). 80% of the data set was used for training the SEM model and the remaining 20% for testing the FCM model. The accuracy obtained by using this model was 74%.

Carlos Ordonez [9] has studied association rule mining with the train and test concept on a dataset for heart disease prediction. Association rule mining has a disadvantage that it produces extremely large number of rules most of which are medically irrelevant. Also in general, association rules are mined on the entire data set without validation on an independent sample. In order to solve this, the author has devised an algorithm that uses search constraints to reduce the number of rules. The algorithm then searches for association rules on a training set and finally validates them on an independent test set. The medical significance of discovered rules is then evaluated with support, confidence and lift. Search constraints and test set validation significantly reduce the number of association rules and produce a set of rules with high predictive accuracy. These rules represent valuable medical knowledge.

Prajakta Ghadge et al. [10] have worked on an intelligent heart attack prediction. Sudhakar et al. [13] studied heart disease prediction using data mining. The data generated by the healthcare industry is huge and “information rich”. As such, it cannot be interpreted manually. Data mining can be effectively used to predict diseases from these datasets. In this paper, different data mining techniques are analyzed on heart disease database. Classification techniques such as Decision tree, Naïve Bayes and neural network are applied here. Associative classification is a new and efficient technique which integrates association rule mining and classification to a model for prediction and achieves maximum accuracy. In conclusion, this paper analyzes and compares how different classification algorithms work on a heart disease database. Shantakumar B. Patil et al. [14] obtained important patterns from heart disease database for heart attack prediction. Enormous amount of data collected by the healthcare industry is unfortunately not 'mined' properly to find concealed information that can predict heart attack. Here, the authors have proposed MAFIA algorithm (Maximal Frequent Item set Algorithm) to do so using Java. The data is preprocessed first, and then clustered using k-means algorithm into two clusters and the cluster significant to heart attack is obtained. Then frequent patterns are mined from the item set and significance weightages of the frequent data are calculated. Based on these weightages of the attributes (ex- age, blood pressure, cholesterol and many others), patterns significant to heart attack are chosen. This pattern can be further used to develop heart attack prediction systems. n system using big data. Heart attack needs to be diagnosed timely and effectively because of its high prevalence. The objective of this research article is to find a prototype intelligent heart attack prediction system that uses big data and data mining modeling techniques. This system can extract hidden knowledge (patterns and relationships) associated with heart disease from a given historical heart disease database. This approach uses Hadoop which is an open-source software framework written in Java for distributed processing and storage of huge datasets. Apache Mahout produced by Apache Software Foundation provides free implementation of distributed or scalable machine learning algorithms. Record set with 13 attributes (age, sex, serum cholesterol, fasting blood sugar etc.) was obtained from the Cleveland Heart Database which is available in the web. The patterns were extracted using three techniques i.e. neural network, Naïve Bayes and Decision tree. The future scope of this system aims at giving more sophisticated prediction models, risk calculation tools and feature extraction tools for other clinical risks.

3. HEART DISEASES

Heart disease is the leading cause of death in the U.S. At some point in your life, either you or one of your loved ones will be forced to make decisions about some aspect of heart disease. Knowing something about the anatomy and functioning of the heart, in particular how angina and heart attacks work, will enable you to make informed decisions about your health. Heart disease can strike suddenly and require you to make decisions quickly.

A. Heart Disease Facts

1) Heart disease is the leading cause of death for both men and women. More than half of the deaths due to heart disease in 2009 were in men.

2) About 610,000 Americans die from heart disease each year—that's 1 in every 4 deaths.

3) Coronary heart disease is the most common type of heart disease, killing more than 370,000 people annually.

4) In the United States, someone has a heart attack every 43 seconds. Each minute, someone in the United States dies from a heart disease-related event.

5) Heart disease is the leading cause of death for people of most racial/ethnic groups in the United States, including African Americans, Hispanics, and whites. For Asian Americans or Pacific Islanders and American Indians or Alaska Natives, heart disease is second only to cancer.

6) Coronary heart disease alone costs the United States $108.9 billion each year. This total includes the cost of health care services, medications, and lost productivity.

B. Risk Factors

High blood pressure, high LDL cholesterol, and smoking are key heart disease risk factors for heart disease. About half of Americans(49%) have at least one of these three risk factors. Several other medical conditions and lifestyle choices can also put people at a higher risk for heart disease, including:

1) Diabetes
2) Overweight and obesity
3) Poor diet
4) Physical inactivity
5) Excessive alcohol use

This paper analyzes the heart disease predictions using classification algorithms. These hidden patterns can be used for health diagnosis in medicinal data. Data mining technology afford an effective approach to latest and indefinite patterns in the data. The information which is identified can be used by the healthcare administrators to get better services. Data mining classification techniques like J48, Decision Trees, and Naive Bayes are used to analyze the dataset based on disease attribute.

4. DESIGN OF MODEL

The project so far discussed about 2 modules:

1. ECG report analysis and prediction of atrial fibrillation with convolutional neural network.
2. Predicting risk of heart disease by multiclass artificial neural network.

Module-1: ECG analysis:

Electrocardiography (ECG):

Electrocardiography depicts the electrical activity of a large mass of atrial and ventricular cells and to summarize, ECG actually records two types of events:

(1) Depolarization- the spread of electrical stimuly across the muscles of heart and (2) Repolarization: The event of return of stimulated muscles to the state of rest. ECG is expressed in a continuous graph with peaks and depressions and we will discuss about a structural unit of an ECG report below.

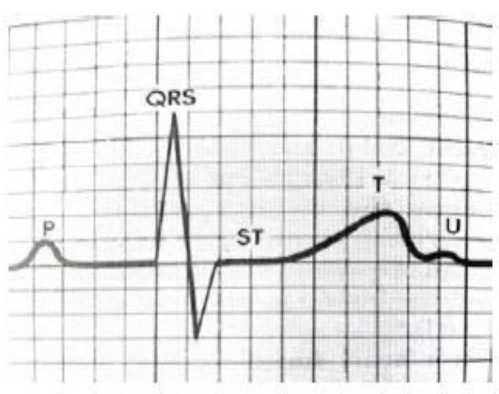

fig:1- model ECG wave segments [3]

ECG report:

Above is a model wave form structure of an ECG report. The graph comprises of 3 complex or specified waves which signifies 3 discrete events.

Table-2:

Complex	Clinical significance
P-wave	Atrial depolarization
QRS complex	Ventricular depolarization
ST segment, T wave, U wave	Ventricular repolarization

The data:

The dataset used in our project is MIT-BIH arrhythmia dataset which is downloadable from Physionet official site [9]. The data contains two channel ambulatory ECG recording of 48 patients. The detailed description of the dataset can be accessed through the source mentioned. Input requirement of this model is in mat file format. One should refer to the research paper describing the data [10].

Analytical model:

The model is a 1-D convolutional network, with a kernel size 16, and 64 filters and a default stride of 1. The neural network is built in python with Tensorflow framework. It takes real-time or consolidated ECG result fetched by any smart fitness device or ECG machines as input. We use the MIT-BIH chest arrythmia database provided by Physionet.org for training and testing the model. We split and merge the whole dataset into 2 parts one for training purpose and the other for testing the model. The operation is done randomly by a "merge-dataset" program written in python. We run the train dataset first to train the model on the given files consisting of the ECG results and detection remarks and then try the model on test.mat file to check the accuracy, prediction result, false positive and false negative score, the recall, precision and hence calculate the F-1 score. After repeated run and 24 fold cross validation our model has reached an F1 score of 86 which is quite better when compared to the existing models. The model detects possible existence of atrial fibrillation in patients" heart, provided his ECG result as input. We have built 2D convolutional network with the input ECG and transformed into a 2D array. The 2D convolutional network comprises of 2 layers, one with 32 and the other with 64 feature maps with 1 x 5 shape. Each layer has a max pooling index of 2. In our developed CNNs we have used ReLU as actiavtion function of neurons with categorical cross entropy loss. To fight the problem of overfitting we have applied batch normalization and with a dropout rate of 0.5. The 2D convolutional neural network, if applied on the transformed dataset, promises an F1 score of more than 0.92 as experimented in our project.

Module-2: Predicting risk of heart disease:

After being done with the proximal analysis of patients" ECG results, we move on to the next part of our project which is the prediction of risk of cardio vascular diseases for which we use a deep learning approach on a multiclass neural network which we train upon a typical dataset which consists of 14 health attributes.

The data:

The dataset is derived from UCI data repository. It has a consolidated research result of 303 patients across 3 countries and their respective four eminent hospitals. Each instance has 76 attributes out of which 14 are the principle ones and others are derived as subset of the former. Out of 14 attributes, 2 are holistic for prediction (i.e.: Age, Sex),

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

11 are numerical graded attributes and 1 is a predicted attribute, derived from 10 class attributes and signifies the angiographic disease status by the extent of diameter narrowing of vessels (less than or greater than 50%). For the detailed description of the dataset, one should refer to the UCI heart-disease dataset [11].

Predictive model:

We have built a multiclass deep neural network written in crystal. Instead of using matrices, it uses object oriented method in building a neural network. Our multilayer model has 10 input layer nodes, 10 hidden layer nodes, and 2 output layers following the convention:

number of hidden layer nodes = 2/3* number of input layer nodes + number of output layer nodes.

We provide the vivid description about the specifications of our model below.

Sigmoid activation function:

As an activation function, attached to the output layer of a neural network generally determines the type of the output the layer ejects, we have carefully chosen a sigmoid of S-shaped activation function for the implementation of our model because sigmoid activation functions are the best ones used for predictive output of multiclass neural networks. Our desired output is a possibility index of risk probability of heart disease which can vary from 0 as the lowest to 1 as highest considering all possible intermediate values as depicted in fig-2. Sigmoid functions are also differentiable and monotonic which adds an advantage for processing of our main function.

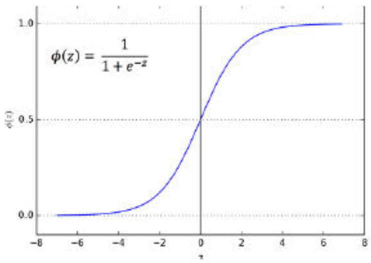

fig:2- Sigmoid activation function

Mean Square Error cost function:

The MSE cost function measures the mean square average of the outputs of a predictive model those vary from the correct output. The cost increases when the performance of the output is worse on the training dataset.

Architecture of the neural network:

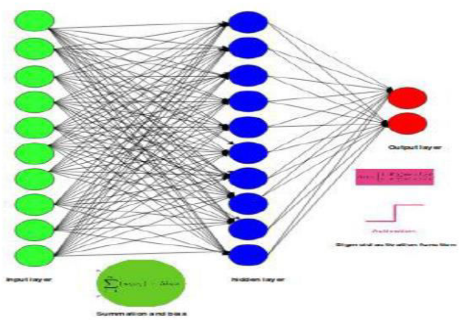

fig-3: Architecture-neural network

CONCLUSION

The proposed system of heart disease prediction with appropriate diagnosis has been framed up using Multilayer Perceptron Neural Network. For effective prediction, back propagation algorithm was applied to train the data and compare the parameters iteratively. Throughout the paper, we have evaluated and discussed about both existing and proposed methods, statistical models and systems for analyzing ECG signals and predicting risk of heart disease evaluated on attributes. Further research will undergo on adding more analytical model of other diagnostic tests for detection of other types of heart related disease.

REFERENCES

[1]Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications (0975 – 888) Volume 47–No.10, June 2012.

[2]K.Srinivas, Dr.G.Raghavendra Rao, Dr. A.Govardhan,"Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques", The 5th International Conference on Computer Science & Education Hefei, China. August 24–27, 2010.

[3]R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," Expert Syst. Appl., vol.36, no. 4, pp. 7675–7680, 2009.

[4]Yanwei Xing, Jie Wang and Zhihong Zhao, "Combination data mining methods with new medical data to predicting outcome of Coronary Heart Disease", 2007 International Conference on Convergence Information Technology.

[5]UCI Machine Learning Repository [http://archive.ics.uci.edu/ml/datasets/Heart+Disease

[6]Jiawei Han, Micheline Kamber & Jian Pei-Data Mining: Concepts and Techniques; 3rd ed; 2011.

[7]K.Srinivas, B.Kavihta Rani and A.Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", *(IJCSE) International Journal on Computer Science and Engineering,*, ISSN: 2347-2693, Vol. 02, No. 02, (2010), pp.250-255.(UGC) (Citations -269)