Journal of Theoretical and Applied Information Technology
20th September 2014. Vol. 67 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A SURVEY ON NEURAL NETWORK MODELS FOR HEART DISEASE PREDICTION

[1] SELVAKUMAR.P, [2] DR.RAJAGOPALAN.S.P
[1] Assistant Professor, ARM College of Engineering and Technology, Chennai Tamil Nadu, India
[2] Professor, GKM College of Engineering and Technology, Chennai Tamil Nadu, India
E-mail: [1] p.selvakumar1971@gmail.com, [2] sasirekarajj@gmail.com [1]

ABSTRACT

The healthcare organization (hospitals, medical centers) should provide quality services at affordable costs. Quality of service implies diagnosing patients accurately and suggesting treatments that are effective. To achieve a correct and cost effective treatment, computer-based information and/or decision support Systems can be developed to full-fill the task. The generated information systems typically consist of large amount of data. Health care organizations must have ability to analyze these data. The Health care system includes data such as resource management, patient centric and transformed data. Data mining techniques are used to explore, analyze and extract these data using complex algorithms in order to discover unknown patterns. Many data mining techniques have been used in the diagnosis of heart disease with good accuracy. Neural Networks have shown great potential to be applied in the development of prediction system for various type of heart disease. This paper investigates the benefits and overhead of various neural network models for heart disease prediction.

Keywords: *Artificial Neural Network, Data Mining, Heart Diseases, Knowledge Discovery, Heart Disease Prediction System*

1. INTRODUCTION

1.1 Data Mining

Data mining techniques are used for knowledge discovery in databases by extraction of interesting information such as non-trivial, hidden, previously unknown, potential useful and ultimately understandable knowledge or patterns from data's in large databases. Data mining provides different methodologies for decision-making, problem solving, analysis, planning, diagnosis, detection, integration, prevention, learning and innovation and forecasting

1.2 Knowledge Discovery

Knowledge Discovery from Databases (KDD) [27] is complex, iterative and interactive process. The process involves data preparation and selection, data cleaning, incorporating prior knowledge on data sets and interpreting accurate solutions from the results. For implementation purpose, several modules should be developed such as for data storage, processing data, data mining, evaluation and knowledge management. KDD application spreads in many areas such as fraud detection, marketing, manufacturing and telecommunication.

1.3 Application of Data Mining in Healthcare Systems

Knowledge Discovery can applied to clinical data which include the process of data collection, data integration, data analysis (preprocessing, missing values treatments, noise reduction, discretization, feature selection, classification, and model extraction).

Health care system must reduce the cost of clinical tests. It can be achieved by employing appropriate computer-based information [34] and/or decision support systems. Health care data includes patient centric data [34], resource management data and transformed data. Health care organizations must have ability to analyze these data.

In healthcare organization the quality of service at affordable cost is one of the major challenge. The effectiveness of quality of service can be guaranteed by diagnosing patients correctly and administering the correct treatments. The clinical decisions may lead to disastrous consequences if it is poorly diagnosed and it is not acceptable.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Currently diagnosing patients correctly and administering effective treatments have become quite a challenge. Poor clinical decisions may lead to patient death. The cost to treat a heart patient is quite high and not affordable by every patient. To achieve a correct and cost effective treatment computer-based information and/or decision support Systems can be developed to do the task. These generated information systems typically consist of large amount of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making in Healthcare System.

Data mining is used in healthcare for customer relationship management decisions, Physicians identify effective treatments and best practices, Patients receive better and more affordable healthcare services and finally Healthcare insurers detect fraud and abuse.

Data mining uses two strategies: supervised and unsupervised learning. In supervised learning, the training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object and a desired output value whereas in unsupervised learning no training set is used (e.g., k-means clustering is unsupervised) [12, 34].

1.4 Heart Disease Prediction

Researchers are using data mining techniques for the diagnosis of many diseases such as heart disease, diabetes, stroke and cancer. Many data mining techniques have been used in the diagnosis of heart disease with good accuracy.

The best example of Real Time Application is working on heart disease patients databases. The detection of a Heart Disease with several factors or symptoms is a multi-layered problem and if not detected correctly, might lead to false assumptions associated with erratic effects. Therefore the effective technique is to utilize the knowledge and experience of several specialists in assisting the diagnosis process [28], [46]. Researchers have been applying different data mining techniques such as naïve bayes, neural network, decision tree, bagging, kernel density, and support vector machine for prediction and diagnosis of heart diseases.

All the Heart Disease Prediction systems uses clinical dataset which consist of parameters and inputs from complex tests conducted in labs which is based on risk factors such as age, family history, diabetes, hypertension, high cholesterol, tobacco smoking, alcohol intake, obesity or physical inactivity, etc.

To reduce the diagnosis time and improve the diagnosis accuracy, Medical Diagnostic Decision Support Systems (MDDSS) has to be developed to deal with complicated diagnosis decision process[8]. The medical diagnosis is a complex and fuzzy cognitive process. Therefore Soft Computing methods, such as Neural Network can be applied for MDDSS.

1.5 Neural Networks

The Artificial Neural Network (ANN) is a technique that is commonly applied to solve data mining applications. Neural Network is a set of processing units when assembled in a closely interconnected network, produces rich structure exhibiting some features of the biological neural network. The structure of neural network provides an opportunity to the user to implement parallel concept at each layer level. The significant characteristic of ANN is fault tolerance. ANNs are well suited in situations where information is noisy and uncertain.

ANN can be divided into two types based on the training method: Supervised training and Unsupervised training [34]. Networks that are supervised require the actual desired output for each input where as unsupervised networks does not require the desired output for each input.

Neural network is an iterative learning process. In this process data cases are presented to the network one at a time, and the weights associated with the input values are adjusted each time [5]. During this learning phase, the network learns by adjusting the weights therefore it can predict the correct class label of input samples. Once a network has been structured for a particular application, it is ready to be trained. To start this process, the initial weights are randomly selected. Then the training or learning, begins.

There are many neural network techniques that have focused on heart disease prediction. The main research objective of this paper is to investigate recent Neural network models for various type of heart disease and prediction methodology. The following section discuss about benefits and overhead of each model.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. NEURAL NETWORKS FOR HEART DISEASE PREDICTION

A complex diagnosis system with a computational model based on a Multi Layer Perceptron (MLP) Neural Network with three layers is employed to develop a decision support system to diagnose all the known heart diseases, hypertension, coronary heart disease, rheumatic valvular heart disease, chronic corpulmonale, and congenital heart disease.

The different neural network techniques such as Multilayer Perceptron (MLP), Radial Basic Function (RBF), and Support Vector Machine (SVM) are tried with the aim of improving the performance of clinical decisions

For identification of patients with Congestive Heart Failure (CHF) from normal controls is investigated using spectral analysis and neural networks with new techniques..

Ischemic Heart Disease (IHD) should be detected at the earliest so that it will prevent severity and minimise mortality rate.Magneto Cardio Graphy (MCG) has been developed for the detection of heart malfunction. Physiologically defective heart emits abnormal patterns of magnetic field [11, 30, 45]. These patterns can be monitored by MCG and data interpretation is time consuming process and need highly trained domain experts. Machine learning approaches can be used to automate the interpretation of IHD pattern of MCG recordings. Two types of machine learning techniques, namely Back-propagation Neural Network (BNN) and Direct Kernel Self Organizing Map (DK-SOM) can be considered.

The wavelet neural network (W") [38] for classification is constructed for patients with coronary artery disease of different level. Additional layer in the whole neural network structure playing a role of linear or non-linear classifier realizing the Wavelet Transform (WT) is implemented for extraction of some features and delivering to the teaching process only that part of information, included in the input Heart Rate Variability (HRV) signal, which is most characteristic and significant for solution of the classification problem.

Wavelet Neural Networks (WNN) can be applied for disease classification which is useful to diagnose coronary artery disease in different levels. For the diagnosis of congenital heart diseases, Reategui et al [49] proposed a model by integrating Case-Based Reasoning (CBR) with a neural network. For discrimination of myocardial heart disease, fuzzy reasoning optimized by genetic algorithm can be utilized.

An Artificial Neural Network (ANN) to recognise 3 HRV patterns related to CHD risk via a Poincare plot encoding. HRV has been used as a non-invasive index of parasympathetic activity as well as a possible predictor of mortality and sudden cardiac death [1, 29].

In the following Section the various existing neural network model for heart disease prediction is discussed

2.1 Stochastic Back Propagation

The authors proposed [13] classification based on data mining techniques such as Rule based, Decision tree, Naïve Bayes and Artificial Neural Network effectively. One Dependency Augmented Naïve Bayes classifier (ODANB) and Naive Credal Classifier2 (NCC2) are used effectively for data pre-processing and decision making. This is an extension of naive Bayes to inaccurate probabilities that aims at delivering robust classifications when dealing with small or incomplete data sets.

The proposed [42] Stochastic back propagation algorithm is used for the construction of fuzzy based neural network. The three major task performed in their methodology is as follows, First, random values are assigned to initialize weights of connections. Second, for each unit net input value, output value and error rate is computed. Third, certainty measure (c) for each node is calculated to handle uncertainty. Finally the decision is made on the certainty measure.

The construction of network includes 3 layers namely an input layer, a hidden layer and an output layer. Sample trained neural network consisting of 9 input nodes, 3 hidden nodes and 1 output. When more than 75% of surface area of the lumen of an artery is occupied by a thrombus or blood clot then the expected result may be a prediction of cell death or heart disease according to medical guidelines.

The Bayesian Network type [39] is proposed and it is a kind of Bayesian Network (BN)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

retrieved by the algorithm which is called as Augmented Naïve BN, characterized mainly based on 1) All attributes have certain influence on the class. 2) The conditional dependency assumption is relaxed (certain attributes have been added a parent). Replacing missing values and Discretization are used as Pre-Processing Techniques[9]

WEKA tool is used for the experimental setup for large number of ARFF data sets, including the data sets from the UCI repository. The supervised discretization algorithm of Fayyad and Irani (1993) is used in the pre-processing step [22] which discretizes all the numerical features. The discretization intervals are calculated on the training set, and then applied on the unchanged test set.

The ODANB has been compared with other existing methods that improve the Naïve Bayes. The results of the comparison proved that the ODANB outperforms than other methods for the disease prediction not related to heart attack. The comparison criteria that have been introduced are accuracy of prediction. The dataset has three types namely Heart-c, Heart-h and Heart-statlog. For Heart-c ODANB is 80.46 and NB is 84.14, for Heart-h ODANB is 79.66 and NB is 84.05 and finally Heart-statlog ODANB is 80.00 and NB is 83.70. The results proved that for prediction of heart disease using Naïve Bayes observes better results.

2.2 Back Propagation Method

The authors [33] suggested supervised network for diagnosis of heart diseases and trained it using Back Propagation algorithm. The behavior of an ANN (Artificial Neural Network) depends on both the weights and the input-output function (transfer function) that is specified for the units. Any one of the following of three categories can be applied to this transfer function : Linear: Threshold: Sigmoid: All learning methods used for adaptive neural networks can be classified into two major categories**:**

Supervised learning, which incorporates an external expert trainer so that each output unit extracted is with response to input signals. In supervised learning error convergence, i.e. the minimization of error between the desired and computed unit values is a major concern. The main aim is to find set of weights, which minimizes the error. According to Scholtes (1991), One well-known method, which is common to many learning paradigms, is the Least Mean Square (LMS) convergence.

A set of experiments was performed on a sample database of 78 patients' records, 13 input variables (Age, Blood Pressure etc.) are used for training and testing of the Neural Network. The experiment is based on single input layer, hidden layer, output layer consist of 13 neurons respectively.

2.3 Multilayer Perceptron

The authors[20] developed a complex diagnosis system with a computational model based on a MultiLayer Perceptron (MLP) Neural Network with three layer. It is employed to develop a decision support system to diagnose all the known heart diseases like, hypertension, coronary heart disease, rheumatic valvular heart disease, chronic corpulmonale, and congenital heart disease. Back-Propagation (BP) algorithm [8] is used to train the system. The obtained network has a skeletal structure that reflects the regularity contained in the data, useful to improve the convergence and the network accuracy.

The MLP used in this system consists of three layers including an input layer, a hidden layer and an output layer [5, 17, 37]. The input layer possesses 38 nodes corresponding to the same number of input variables that are extracted from large number of patients. These input variables are divided into five categories (i)age, (ii) gender, (iii) clinical symptoms (14 in total), (iv) inducement and history *(5* factors), and (v) clinical examinations (17 in total).

The experimental results have shown that the adopted MLP neural network can achieve high accuracy level (63.6-82.9%) on the classification of heart diseases. It is an excellent decision support system that can be deployed in clinics. The system can be improved to have high accuracy by including careful variable expression and model parameter adjustment using more training records. The experimental result has indicates that the proposed MLP-based computational model has a strong capability to classify the five heart diseases with high accuracy, proving its usefulness in support of clinic decision process. The study shows that when the number reaches 20, the network performance has no further improvement.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The Authors [48] developed different neural network techniques such as MultiLayer Perceptron (MLP), Radial Basic Function (RBF), and Support Vector Machine (SVM) are tried with the aim of improving the performance of clinical decisions. The given data is transformed to the appropriate format for these neural network techniques. The data includes the physiology and operative scoring attributes and other relevant attributes useful in predicting patient risk.

WEKA software is used to develop the different neural classifiers In MLP, the selected learning rate is 0.3, the iteration is 500. The data set is divided into two. A test set is taken by using 50% of the overall pattern set or using a 10 fold cross validation partition. With the latter technique, the data set is divided into 10 partitions. One partition is used as a test set while the rest is for training; the procedure is repeated 10 times, so that each partition acts as a separate test set.

The mortality rate of cleaned data 14.34% (38 from 265 patterns with status= "dead"). The generation and analysis of a confusion matrix gives accuracy result. The predicted mortality rate for each neural network technique was lower than observed one. Percentage misclassification for each model is obtained by dividing the sum of the misclassification of "dead" or "alive" patient by the total number of patterns. The MLP model provides the best predications of patient risk with a MSE, and a misclassification (0.02, 3.7% with type 1, 0.01, 1.9% with type 2 respectively).

RBF has the worst classification compared to other models because almost medium and high risk patients are misplaced into lower levels of risk. For example, the 3 high risk patients are misplaced into low risk), and medium risk with 50% split of test set.

POSSUM and PPOSSUM are generic clinical tools that allow a metric factor to be used in assessing the severity of illness. The authors conducted experiments with POSSUM and PPOSSUM system with the ratio (O/E) for the range of predicted death rate of 20-30% is 0.26. and 2.25 for 265 patients.

The misclassification of each model is evaluated using confusion matrix. It is proved that using different models of neural network produces smaller misclassification errors. More interestingly, the models using the new outcome of risk (High, Medium, Low) had the smallest percentage of misclassification compared to the other risk predication models (i.e. mortality or morbidity).

2.4 Intelligent Heart Disease Prediction Systems

In IHDPS two methodologies are implemented, in the first method the authors [21] developed a prototype called Intelligent Heart Disease Prediction System (IHDPS) using three data mining modelling techniques, namely, Decision Trees, Naïve Bayes and Neural Network. IHDPS can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. With effective treatment, cost can be reduced. The results is displayed both in tabular and graphical forms for easy interpretation and visualization.

IHDPS can answer complex "what if" queries which is not supported in traditional decision support systems. Using medical factors such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease. It enables significant knowledge, e.g. patterns, relationships between medical factors related to heart disease. IHDPS is Web-based, user-friendly, scalable, reliable and expandable [7].

The system extracts hidden knowledge from a historical heart disease database. The DMX query language and functions are used to build and access these models. Five goals are defined based on business intelligence and exploration of data. The purpose of trained models is to evaluate the goals. The goals are evaluated against the trained models. The complex queries are answered by the three models each with its own potential with respect to accuracy, interpretation and access to detailed information.

The purpose of Lift Chart and Classification Matrix is to determine which model gives the highest percentage of correct predictions for diagnosing patients with a heart disease. IHDPS current version is based on the 15 attributes [Sex, Chest Pain Type, Fasting Blood Sugar, Age etc]. This list may need to be expanded to provide a more comprehensive diagnosis system. The limitation is that it only uses categorical data. For some diagnosis, the use of continuous data may be required. The other limitation is that it only uses three data mining techniques. Additional data mining techniques can be incorporated to provide better diagnosis.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The Another Intelligent Heart Disease Prediction System (IHDPS)[21] using three data mining modeling techniques, namely, Decision Trees, Naïve Bayes and Neural Network. IHDPS can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. It can answer complex queries for heart disease diagnosis and thus assist healthcare practitioners to make intelligent clinical decisions which cannot be performed by traditional decision support systems

In the second method IHDPS uses the CRISP-DM methodology to construct the mining models. This model consists of six major phases: business understanding, data understanding, data preparation, modelling, evaluation, and deployment [7].

Business understanding phase focuses on understanding the objectives and requirements from a business perspective, converting this knowledge into a data mining problem definition, and designing a preliminary plan to achieve the objectives. Data understanding phase uses the raw data and proceeds to understand the data, identify its quality, gain preliminary insights, and detect interesting subsets to form hypotheses for hidden information. Data preparation phase constructs the final dataset that will be fed into the modelling tools. This includes table, record, and attribute selection as well as data cleaning and transformation.

The modelling phase selects and applies various techniques. The parameters are calibrated to get optimal values. The evaluation phase evaluates the model to ensure that it achieves the business objectives. The deployment phase specifies the tasks that are needed to utilize the models [7].

Data Mining Extension (DMX), is a SQL-style query language for data mining and it is used for building and accessing the model contents. Tabular and graphical visualizations are incorporated to enhance analysis and interpretation of results.

For model creation, training, prediction and content access, DMX is used by this system. All parameters were set to the default settings except for parameters "Minimum Support = 1" for Decision Tree and "Minimum Dependency Probability = 0.005" for Naïve Bayes. The trained models were evaluated against the test datasets for accuracy and effectiveness before they were deployed in IHDPS. The models were validated using Lift Chart and Classification Matrix.

2.5 Probabilistic Neural Network (PNN) Model

The authors [39] proposed an efficient approach for the intelligent heart disease prediction based on Probabilistic Neural Network (PNN) technique. The dataset is clustered with the aid of k-means clustering algorithm [34]. The PNN with radial basis function is trained using the selected data sets.

With Existing approaches a new technique based on PNN for heart disease prediction is proposed. Probabilistic Neural Network which is a class of Radial Basis Function (RBF) network is useful for automatic pattern recognition, nonlinear mapping and estimation of probabilities of class membership and likelihood ratios.

The input layer consists of N nodes: one for each of the N input features of a feature vector. All the nodes in the pattern layer receive input from feature input layer. The hidden nodes are collected into groups: one group for each of the K classes. Each hidden node in the group for Class k related to a Gaussian function centered on its associated feature vector in the k^{th} class. All the Gaussians in a class group feed their functional values to the same output layer node for that class hence there are K output nodes.

The experimental setup shows that, the heart disease data is refined by removing duplicate records and supplying missing values. The number of parameters considered as input to the model is 13. The Proposed PNN along with existing three approaches is trained with different number of data cases. The resulting TP rate and FP rate on training of each set of data cases are recorded. It appears to be more effective as it has the highest number of correct prediction (94.6%) as compared to other technique. The Probabilistic Neural Network has better sensitivity (92.68%) and specificity (91.42%) respectively.

HDPS can be further enhanced and expanded. For example, the other medical attribute can be incorporated. Further in future one can also incorporate other data mining techniques, e.g., Time Series, Clustering and Association Rules. Continuous data can replace categorical data. Another area is to use Text Mining to mine the vast

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

amount of unstructured data available in healthcare databases. Another challenge is to integrate text mining and data mining.

2.6 Neural Networks Ensemble Model

The proposed system [36] uses neural networks ensemble model. It enables an increase in generalization performance by combining several individual neural networks to train task.

The architecture consists of the following components 1) Heart database component: It holds the features that are used to characterize patients and healthy persons. 2) Data partition component: It partitions the input data into train and validation data sets. Partitioning provides mutually exclusive data sets. Partitioning the input data reduces the computation time. 3) Variable selection component: It is used in reducing the number of inputs by setting the status of the input variables that are not related to the target as rejected. 4) Neural networks block component: It is used to classify the feature space. Three independent neural networks models were used to construct this component. Multi-layer feedforward neural network is the most widely used for prediction. The back propagation learning algorithm has been used in the feedforward, single hidden layer neural network. 5) Ensemble component: It is used to create new models by combining the posterior probabilities (for class targets) or the predicted values (for interval targets) from multiple predecessor models.

The proposed methodology is implemented with the SAS base software 9.1.3. SAS enterprise miner streamlines the entire data mining process from data access to model assessment. It supports all required tasks within a single, integrated solution by providing the flexibility for efficient Collaborations. SAS enterprise miner is designed for data miners, marketing analysts, database marketers, risk analysts, fraud investigators, business managers, engineers and scientists who play strategic roles in identifying and solving critical business or research issues. It can also be used in many machine intelligence applications.

The experimental results shows that, SAS enterprise miner 5.2 was used to construct a neural networks ensemble based methodology for diagnosing of the heart disease in a fully automatic method. Three independent neural networks models were used to construct the ensemble model. The experimental results gained 89.01% classification accuracy, 80.95% sensitivity and 95.91% specificity values for heart disease diagnosis.

2.7 Machine Learning and Neural Network Approaches

This technique [31] is used for identification of patients with Congestive Heart Failure (CHF) which is investigated using spectral analysis and neural networks. The identification system consists of two parts: feature extraction part and classification part.

The feature extraction part uses the method of approximate spectral density estimation of R-R-Intervals (RRI) data by implementing the soft decision sub-band decomposition technique. In the classification part, two different methods of machine learning approaches with neural networks are implemented and compared in their performances. Supervised neural network (back-propagation) and unsupervised neural network (Kohonen self organizing maps) are the basic approaches used in this system.

A data set of 17 CHF and 53 normal subjects is used as original training data set. The another set of 12 CHF and 12 normal subjects is used as original test data set. The classification features are the spectral density of 6 different regions covering the whole spectrum of the R-R-Intervals data obtained by 32-bands soft decision algorithm. A larger training data set, which is obtained by simulating 1000 CHF and 1000 normal subjects according to the spectral features obtained from the original training data and it is used to train the neural network. The neural network is used to test simulated data set of the same size of the training date set. The accuracy of the classification is found to be about 83.65% and 91.43% with supervised neural networks and unsupervised neural networks respectively.

Early detection of Ischemic Heart Disease IHD [43] may effectively prevent severity and reduce mortality rate.Magneto Cardio Graphy (MCG) has been developed for the detection of heart malfunction. MCG is capable of monitoring the abnormal patterns of magnetic field [11,30,45] as emitted by physiologically defective heart. Data interpretation is time consuming and requires highly trained professional. It can discriminate healthy and diseased cases at resting condition better than traditional ECG [16, 41].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A machine learning approach is used for automating the interpretation of IHD pattern of MCG recordings. Two types of machine learning techniques, namely Back-propagation Neural Network (BNN) and direct kernel Self-Organizing Map (DK-SOM), represents supervised and unsupervised learning techniques.

Analysis of myocardial infarction can be achieved by SOM, the BNN and robust Bayesian classifiers [4 ,6]. The usage of the DK-SOM along with application of the BNN as tools for concurrent identification of myocardial ischemia from MCG.

MCG data were acquired from 36 locations above the torso by making four sequential measurements in mutually adjacent positions. The cardiac magnetic field was recorded by nine sensors for 90s in each position using a sampling rate of 1000 Hz, resulting in 36 individual time series. WEKA 3.4.3 tool is used for implementation. Supervised learning neural network calculations were performed with the back-propagation. The data sets were initially normalized to a range of 0 and 1.

To achieve the highest prediction performance and reduction of computational time, the redundancy and multi-collinearity descriptors were discarded with UFS version 1.8. UFS is a software utilizing the variable reduction algorithm namely Unsupervised Forward Selection (UFS) [47].

The proposed machine learning models reveals higher sensitivity (86.2%) and specificity (73%) for detection of IHD. It is proved that the BNN provides less specificity and accuracy, but yields high sensitivity (90%). Therefore, BNN may be used initially to detect high risk individuals while false positive cases can subsequently be ruled out by the DK-SOM. The intelligent capabilities of both the BNN and the DK-SOM offer good support in detecting IHD via magneto cardiograms in an automatic manner.

2.8 Genetic and Fuzzy Expert System

Fuzzy set theory and fuzzy logic are highly suitable for developing knowledge based systems in healthcare for diagnosis of diseases .The proposed method [32] is an extended version of the model that combines the genetic algorithms for feature selection and fuzzy expert system for effective classification. In fuzzy logic process, initially fuzzification is performed by collecting the crisp set of input data and converting it to a fuzzy set using fuzzy linguistic variables, fuzzy linguistic conditions and membership functions. Inference is made based on a set of rules and finally, defuzzification step is performed. This system generates the fuzzy rules based on the support sets obtained.

The data classification technique based on various supervised machine learning algorithms, namely, Naive Bayes, Decision List and KNN. TANAGRA tool is used to classify the data and the data is evaluated using 10-fold cross validation. TANAGRA is a data mining tool for academic and research purposes. It proposes various data mining methods from exploratory data analysis, statistical learning, machine learning and databases.

The main objective is to reduce the number of attributes which were used for diagnosing heart disease. Earlier, 13 attributes were used for this prediction but in this system the number of attribute is reduced to six only using Genetic Algorithm [2] and Feature Subset Selection. In addition to the genetic algorithm, CFS Evaluator is also used.

Experiments are conducted in MATLAB using fuzzy tool. Mamdani model of fuzzy system is used. The fuzzy rules are generated based on expert's knowledge in this domain. The UCI machine learning repository dataset is used and only 6 attributes are found to be effective and necessary for heart disease prediction. In the proposed system, the input is the set of all the selected features and the output of the system is to achieve a value 0 or 1 that indicates the absence or presence of heart disease in patients.

Initially, missing values were identified in the dataset and they were replaced with appropriate values using Replace Missing Values filter using Weka tool. The analysis shows that Neural Network with 15 attributes has produced the highest accuracy i.e. 100%. The Decision Tree [19] has also performed well with 99.62% accuracy by using 15 attributes. The combination with Genetic Algorithm and 6 attributes, Decision Tree has shown 99.2% efficiency.

2.9 Genetic Algorithms

Intelligent Heart Disease Prediction System (IHDPS) [25] is built using data mining techniques like Decision Trees, Naïve Bayes and Neural Network. The system detects heart disease based on feed forward neural network architecture [40]

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and genetic algorithm. To train the neural network Hybridization has applied using Genetic algorithm and proved experimentally that proposed learning is more stable compare to back propagation.

The work is starting out with a randomly selected first generation. Every string in this generation is evaluated according to its quality, and a fitness value is assigned. Then, a new generation is produced by applying the reproduction operator. Pairs of strings of the new generation are selected and crossover is performed. Genes are mutated with a certain probability, before all solutions are evaluated again. This procedure is repeated until a maximum number of generations are reached. While doing this, the all time best solution is stored and returned.

Detail analysis has given with respect to genetic algorithm behavior and its relationship with learning performance of neural network. Affect of tournament selection has analyzed to get more detailed knowledge. With the proposed system hope that, design of diagnosis system for heart disease detection becomes easy, cost effective, reliable and efficient.

2.10Wavelet Neural Network Model

The authors [38, 44] developed wavelet neural networks (W") for classification purpose for patients with coronary artery disease of different level. Additional layer in the whole Neural Network (NN) structure playing a role of linear or non-linear classifier realizing the wavelet transform (WT) is implemented for extraction of some features and delivering to the training process. only that part of information, included in the input Heart Rate Variability (HRV) signal, which is most characteristic and significant for solution of the classification problem [14,15].

HRV signals classification system consists of following components.

2.10.1 Structure of the method

The HRV signal before the W" system input is delivered to preprocessing to obtain equally spaced signal samples. In WNN systems [35], the neural network plays a role of a classifier for features which are the outputs of the first layer realizing Wavelet Transform (WT). Depending upon the output layer of the WNN system, the decision about analyzed signal membership to particular class is made throughout the application of later described criteria.

2.10.2 Feature extraction

The feature extraction depends upon the calculation of the WT. In the WNN training process a simplest algorithm of error back propagation is used according to which each cycle of training process there is a need of partial derivatives calculation. This fact also explains the choice of the basic wavelet (i.e. a Morlet wavelet). Besides the weights and shifts in case of the first method for feature extraction the scale coefficient in WT also modify its value within the teaching process.

2.10.3 Neural network structures applied in classification

Single layer structures like Linear Activation Function, Tangsoidal Activation Function and Radial Basic Function are used.

The experimental setup shows that patients have been divided into four groups namely Control Group, 1 Vessel Group (1V), 2 Vessel Group (2V) and 3 Vessel Group (3V) according to the level CAD. Then, each group of patients has been divided further into two subgroups namely EF < 40% and EF >= 40% according to previously estimated ejection fraction (EF). The criterion was EF less or greater or equal to 40%. The HRV signals of 89 patients categorized as Control Group EF<40% is 1 No. and EF>=40% is 15 No. similarly for 1V 9and 15 nos., 2V 10 and 12 Nos. 3V 13 and 14Nos. respectively

The best results have been achieved in case of double layer NN structure with both tangsoidal and linear activation functions and when to the NN structure inputs, the feature vector, obtained through consecutive WT decompositions of HRV signals, have been delivered. Also the choice of basic wavelet function is important in case of both care about training time and classification quality. The algorithm of basic wavelet function selection will be a subject for further investigations.

2.11Pattern Recognition Methods

The authors developed [3] a system to train an Artificial Neural Network (ANN) to recognise HRV patterns related to CHD risk[26] via a Poincare plot encoding. HRV has been used as a non-invasive index of parasympathetic activity as well as a possible predictor of mortality and sudden cardiac death [1,29]. While conventional analysis can be used to measure the respiratory component of HRV, a new non-linear quantitative

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

method, the PoincarC plot, has been proposed for the evaluation of RSA.

The pattern recognition tasks are implemented by using ANNs by a supervised learning process which involves training and a test phase.

In the training phase a set of pre-determined classes exists and it is assumed that some mechanism can correctly label or classify each example. On receiving an example, some measurements are obtained from a preprocessing stage, known as features and this data is fed into a pattern recognition machine, known as the classifier. The training phase is terminated when the difference between the computed activation of the output units and the desired activation of the output units falls below a pre-specified value.

Thus because neurons in the ANN have learnt to respond to features as the network was trained with different examples, it develops the ability to generalize and can be tested to classify unknown examples. Standard screening tests were performed to identify selected modifiable and non-modifiable CHD risk factors. The CHD risk was calculated using a current scoring system which derives a CHD risk score from a combination of recognized CHD risk factors. Patients were categorized as low, medium and high risk groups ranging from 0 ("low risk" subject) to 32 ("high risk" subject).

The experimental setup shows the CHD risk classification (test phase) using the three RR interval files with 9 subjects were classified from the 3 pre-processed files. The best high and medium risk classifications (6 out of 9) were obtained from processing the records containing RR intervals acquired under controlled breathing frequency at 12 breaths.min and 6 breaths min added together. While the best low risk classification (6 out of 9) was obtained using the information recorded at 12 breaths.min.

The ANN together with Poincare plots can be successful in detecting individuals at varying risk of CHD using short-term RR interval measurements [23] under controlled breathing at 12 breaths.min and 6 breaths.min.

RSA is considered to be a non-invasive measure of parasympathetic activity [24] and is reduced in patients with CHD [1, 10]. The importance of controlled breathing frequency and tidal volumes for the measurement of RSA has been well documented [18].

3. RESEARCH CHALLENGES

The major research challenge in heart disease prediction using neural network model is accuracy in prediction. The study shows that when the input variable reaches 20, the network performance has no further improvement. It seems that using different models of neural network produces smaller misclassification errors.

Another limitation is that it only uses categorical data. For some diagnosis, the continuous data is required. Another limitation is that only three data mining techniques were used. Additional data mining techniques can be incorporated to provide better diagnosis.

As a future scope one can also incorporate other data mining techniques, e.g., Time Series, Clustering and Association Rules. Continuous data can also be used instead of categorical data. The other serious challenge is to use Text Mining to mine the vast amount of unstructured data available in healthcare databases. The integration of text mining and data mining can be done. The choice of basic wavelet function is important in case of both care about training time and classification quality. The algorithm of basic wavelet function selection will be a subject for further investigations. In the pre-processing of data information loss occurs when compression is performed on the information represented in the Poincare plots. The training patterns are not evenly distributed among the three classes. Different learning algorithms and ANN architectures can be implemented in order to improve the classification accuracy.

The following table compares the accuracy level of heart disease prediction using various neural network model:-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Performance Comparison Of Various Neural Network Models

Neural Network Models	Accuracy of prediction	Sensitivity	Specificity
Stochastic Back propagation	80.46 %	-	-
Multilayer Perceptron	63.6-82.9%	-	-
Probabilistic Neural Network (PNN) Model	94.6 %	92.68%	91.42%
Neural Networks Ensemble Model	89.01%	80.95%	95.91%
Wavelet Neural Network Model	-	78%	83%
Machine Learning and Neural Network Approaches	80.4%	86.2%	73%
Genetic and Fuzzy Expert System	99.2%	-	-

4. CONCLUSION AND FUTURE WORK

A number of works have investigated the performance and efficiency of the various neural network models for heart disease prediction. The major research challenge in existing neural network model is accuracy in prediction. The study shows that when the input variable reaches 20, the network performance has no further improvement. It seems that using different models of neural network produces smaller misclassification errors. Another limitation is that it only uses categorical data. For some diagnosis, the continuous data is required.

The limitation of this study is that the analysis deals only with neural network models for heart disease prediction system. The various other data mining techniques e.g Time Series, Clustering and association rules, reduced information loss, use of text mining can be considered for heart disease prediction system.

Our experience suggests that research challenges like more accuracy of prediction, less misclassification errors, use of continuous data instead of categorical data can be considered for developing prediction system. This survey has provided an analysis of the state-of-the-art in developing neural network models for heart disease prediction□□The description and comprehensive comparison of several main models has been provided, along with discussion of issues associated with their implementation. The goal of this survey has been to consolidate research in this area, to inform users about different models and to develop improved models that are based on previous experiences.

As a next step of this study we plan to develop a framework which will optimizes the above performance measures of neural network model for heart disease prediction system.

REFERENCES:

[1] K.Airaksinen, M.J. Ikaheimo, M.K. Linnaluoto, M. Neimela, J.T. Talclcunen. Control impaired vagal heart rate control in coronary heart disease. Brit. Heart J. 1988; 58592-597.

[2] M. Anbarasi, E. Anupriya, N.Ch.S.N.Iyengar, ―Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm; International Journal of Engineering Science and Technology, Vol. 2(10), 2010

[3] F. Azuaje', W. Dubitzky', Wu' X, P Lopes'73, N Black', K Adamson'.', JA White4 A Neural Network Approach to Coronary Heart Disease Risk Assessment Based on Short-Term Measurement of RR Intervals Computers in Cardiology 1997 Vol 24 0276-6547/97 1997 IEEE

[4] R. Bigi, D. Gregori, L. Cortigiani, A. Desideri, F.A. Chiarotto, G.M. Toffolo, Artificial neural networks and robust Bayesian classifiers for risk stratification following uncomplicated myocardial infarction, Int. J.Cardiol. 101 (2005) 481–487.

[5] C.M. Bishop, Neural nenvork for pollern recognition, Oxford University Press, Oxford, 1995.

[6] G. Bortolan, W. Pedrycz , An interactive framework for an analysis of ECG signals, Artif. Intell. Med. 24 (2002) 109–132.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[7] K. Charly, "Data Mining for the Enterprise", 31st Annual Hawaii Int. Conf. on System Sciences, IEEE Computer, 7, 295-304, 1998.

[8] W. David and W. Vivian, "Improving diagnostic accuracy using a hierarchical neural network lo model decision submks", hi. 3. Med. Inform., 57: 41-55.2000.

[9] P. Domingos, and M. Pazzani,On the optimality of the simple Bayesian classifier under zero one loss. Machine Learning, 1997.29(2-3): p. 103-30.

[10] D.L. Eckberg, M. Drabinsky, E. Braunvald Defective cardiac sympathetic control in patients with coronary artery disease. New England Journal Medicine 1971; 285: 877-883.

[11] R. Fenici, D. Brisinda, Bridging non-invasive and interventional electroanatomical imaging: role of magnetocardiography, J.Electrocardiol. 40 (2007) S47–S52.

[12] A. Figliola, E. Serrano, "Analysis of Physiological Time Series Using Wavelet Transform", IEEE EMB, May/June 1997, 889-898.

[13] Frawley and Piatetsky-Shapiro, 1996. Knowledge Discovery in Databases:An Overview. The AAAI/MIT Press, Menlo Park, C.A.

[14] Fu K.S, Syntactic Pattern Recognition and Applications", Prentice-Hall, Englewd Cliffs, NJ, 1982.

[15] K. Fukanaga, „Introduction to Statistical Pattem Recognition", Academic Press, NY, 1991.

[16] B.I. Hailer. S. Chaikovsky, Auth-Eisernitz, H. Schafer, P. Van Leeuwen, The value of magnetocardiography in patients with and without relevant stenosis of the coronary arteries using an unshielded system, Pacing Clin. Electrophysiol. 28 (2005) 8–16.

[17] D.J. Hand , Conrlruclion and assessment of classijicorion rules, Wiley, New York, 1997.

[18] J.A. Hirsch, B. Bishop. Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate. American Journal Physiology 198 I: 241:H62O-H629.

[19] T. J. Ho, "Data Mining and Data Warehousing", Prentice Hall, 2005

[20] Hongmei Yan', Jun Zheng, Yingtao Jiang, Chenglin Pengl, and Qinghui Li " Development of A Decision Support System For Heart Disease Diagnosis Using Multilayer Perceptron", 0-7803-7761-31031117.00 02003 IEEE.

[21] S.H. Ishtake, S.A. Sanap "Intelligent Heart Disease Prediction System Using Data Mining Techniques" International J. of Healthcare & Biomedical Research, Volume: 1, Issue: 3, April 2013,

[22] Juan Bernabé Moreno, One Dependence Augmented Naive Bayes, University of Granada, Department of Computer Science and Artificial Intelligence.

[23] P.W. Kamen, H. Krum, A.M. Tonkin Poincare plot of heart rate variability allo\!s quantitative display of parasympathetic nervous activity in humans. Clinical Science BME -3311 149-1 156.

[24] P.G. Katona, F. Jih. Respiratory sinus arrhythmia: noninvasive measure of parasympathetic cardiac control. Journal Applied Physiology 1975; 39:801-805.

[25] K.S. Kavitha, K.V. Ramakrishnan, Manoj Kumar Singh Modeling and design of evolutionary neural network for heart disease detection IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010 ISSN (Online): 1694-0814

[26] P.L. Lopes, R.H. Mitchel, J.A. White. The relationships between respiratory sinus arrhythmia and coronary heart disease risk factors in middle aged males. Automedica 1994;

[27] LUKASZ A KURGAN and PETR MUSILEK "A survey of Knowledge Discovery and Data Mining process models" The Knowledge Engineering Review Cambridge University Press Printed in the United Kingdom.

[28] K. Mehmed, "Data mining: Concepts, Models, Methods and Algorithms", New Jersey: John Wiley, 2003

[29] G.A. Myers, G.J. Martin, N.M. Magid, P.S. Barnett, J.W. Schaad, J.S. Weiss, M. Lesch, D.H. Singer Power spectral analysis of heart rate variability in sudden death: comparison to other methods. IEEE Transctions of Biomed. Engineering 1986.

[30] K. Nakai, H. Izumoto, K. Kawazoe, J. Tsuboi, Y. Fukuhiro, T. Oka, K. Yoshioka, M. Shozushima, M. Itoh, A. Suwabe, M. Yoshizawa,cThree-dimensional recovery time dispersion map by 64-channel magnetocardiography may demonstrate the location of a myocardial injury and heterogeneity of repolarization, Int. J. Cardiovasc. Imag. 22 (2006) 573–580.

[31] Nazar Elfadila and Abdulnasir Hossenb, Identification of patients with congestive heart failure using different neural networks

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

approaches Technology and Health Care 17 (2009) 305–321 305 DOI 10.3233/THC-2009-0542 IOS Press 0928-7329/09 2009

[32] Nidhi Bhatla Kiran Jyoti GNDEC, Ludhiana, India GNDEC, Ludhiana, India An Analysis of Heart Disease Prediction using Different Data Mining Techniques International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 8, October - 2012

[33] Niti Guru, Anil Dahiya and Navin Rajpal Decision Support System for Heart Disease Diagnosis using Neural Network Delhi Business Review _ Vol. 8, No. 1 (January - June 2007)

[34] M.K. Obenshain, "Application of Data Mining Techniques to Healthcare Data", Infection Control and Hospital Epidemiology, 25(8), 690–695, 2004

[35] Qmghua Bang, Albert Benveniste, "Wavelet Networks", IEEE Trans. On Neural Networks, Vol. 3, No. 6, Nov. 1992, p.

[36] Resul Das, Ibrahim Turkoglu, Abdulkadir Sengur Effective diagnosis of heart disease through neural networks ensembles Expert Systems with Applications 36 (2009) 7675–7680 0957-4174/$ - see front matter _ 2008 Elsevier Ltd.

[37] B.D. Ripley, Potfern recognition und neural network, Cambridge University Press, Cambridge, 1996.

[38] R. Schakoff, "Pattem recognition: Statistical, Structural and Neural Approaches", New York, 1992.

[39] Sellappan Palaniappan, Rafiah Awang, Intelligent Heart Disease Prediction System Using Data Mining Techniques, 978-1-4244-1968-5/08 ©2008 IEEE.

[40] M.K. Singh, 'password based a generalized robust security system design using neural network', IJCSI,vol4,No.9,pp.1-9,2009.

[41] F.E Smith, P. Langley, P.V. Leeuwen, B. Hailer, L. Trahms, U. Steinhoff, J.P. Bourke, A.Murray, Comparison of magnetocardiography and electrocardiography: a study of automatic measurement of dispersion of ventricular repolarization, Europace 8 (2006) 887–893.

[42] K. Srinivas, B. Kavihta Rani and Dr.A. Govrdhan "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks" K.Srinivas et al. / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010, 250-255

[43] Tanawut Tantimongcolwata, Thanakorn Naennab, Chartchalerm Isarankura-Na-Ayudhyaa, Mark J. Embrechtsc, Virapong Prachayasittikula, Identification of ischemic heart disease via machine learning analysis on magnetocardiograms Computers in Biology and Medicine 38 (2008) 817 – 825 0010-4825 Elsevier Ltd.

[44] E.J. Tkacz', P. Kostka "An Application of Wavelet Neural Network for Classification Patients with Coronary Artery Disease Based on HRV Analysis" Proceedings of the 22"d Annual EMBS International Conference, July 23-28, 2000, Chicago IL. 0-7803-6465-1/ IEEE

[45] K. Tolstrup, B.E. Madsen, J.A. Ruiz, S.D. Greenwood, J. Camacho, R.J. Siegel, H.C. Gertzen, J.W. Park, P.A. Smars, Non-invasive resting magnetocardiographic imaging for rapid detection of ischemia in subjects presenting with chest pain, Cardiology 106 (2006) 270–276.

[46] F. Weiguo, L. Wallace, S. Rich, Z. Zhongju, "Tapping the Power of Text Mining", Communication of the ACM. 49(9), 77-82, 2006.

[47] D.C. Whitley, M.G. Ford, Unsupervised Forward Selection: a method for eliminating redundant variables, J. Chem. Int. Comput. Sci. 40 (2000)1160–1168.

[48] W.N. Davis, Thuy Nguyen Thi Thu, "Predicting Cardiovascular Risks Using Possum, Ppossum And Neural Net Techniques"

[49] E. B. Reategui, J. A. Campbell, B. F. Leao. "Combining a neural network with ease-based reasoning in a diagnostic system". Artif lntel Medicine, 9; 5-27, 1997.