

Survey on Decision Support System For Heart Disease

Jayshri S.Sonawane ¹, Dharmaraj R. Patil ² and Vishal S. Thakare ³

¹ Department of Computer Engg., North Maharashtra University, RCPIT
Shirpur, Maharashtra, India

² Department of Computer Engg., North Maharashtra University, RCPIT
Shirpur, Maharashtra, India

³ Department of Computer Engg., North Maharashtra University, RCPIT
Shirpur, Maharashtra, India

Corresponding Author Email: jayshri.sonawane@gmail.com

Abstract

The diagnosis of heart disease in most cases depends on a complex combination of clinical and pathological data. Because of this complexity, there exists a significant amount of interest among clinical professionals and researchers regarding the efficient and accurate prediction of heart disease. In case of heart disease time is very crucial to get correct diagnosis in early stage. Patient having chest pain complaint may undergo unnecessary treatment or admitted in the hospital. In most of the developing countries specialists are not widely available for the diagnosis. Hence, automated system can help to medical community to assist doctor for the accurate diagnosis well in advance. So the decision support systems play an important role in the diagnosis of heart disease. However, accurate diagnosis at an early stage followed by proper subsequent treatment can result in significant life saving.

Keywords: *Decision Support System (DSS) .*

1. Introduction

1.1 Heart disease

The heart is the organ that pumps blood, with its life giving oxygen and nutrients, to all tissues of the body. If the pumping action of the heart becomes inefficient, vital organs like the brain and kidneys suffer and if the heart stops working altogether, death occurs within minutes. Life itself is completely dependent on the efficient operation of the heart.

The term heart disease applies to a number of illnesses that affect the circulatory system, which consists of heart and blood vessels. It is intended to deal only with the condition commonly called "Heart Attack" and the factors, which lead to such condition. Cardiomyopathy and Cardiovascular disease are some categories of heart diseases. The term cardiovascular disease includes a wide range of conditions that affect the heart and the blood vessels and the manner in which blood is pumped and circulated through the body. Cardiovascular disease (CVD) results in severe illness, disability, and death. Narrowing of the coronary arteries results in the reduction of blood and oxygen supply to the heart and leads to the Coronary heart disease (CHD). Myocardial infarctions, generally known as a heart attacks, and angina pectoris, or chest pain are

encompassed in the CHD. A sudden blockage of a coronary artery, generally due to a blood clot results in a heart attack. Chest pains arise when the blood received by the heart muscles is inadequate. High blood pressure, coronary artery disease, valvular heart disease, stroke, or rheumatic fever/rheumatic heart disease are the various forms of cardiovascular disease[1].

1.2 Introduction To Heart Disease Diagnosis

Medical diagnosis is considered an art regardless of all standardization efforts made, which is greatly due to the fact that medical diagnosis necessitates an expertise in coping with uncertainty simply not found in today's computing machinery. The researchers are encouraged by the advancement in computer technology and machine learning techniques to develop software to assist doctors in making decision without necessitating the direct consultation with the specialists.

The medical diagnosis process can be interpreted as a decision making process, during which the physician induces the diagnosis of a new and unknown case from an available set of clinical data and from his/her clinical experience. This process can be computerized in order to present medical diagnostic procedures in a rational, objective, accurate and fast way.

The diagnosis of heart disease in most cases depends on a complex combination of clinical and pathological data. Because of this complexity, there exists a significant amount of interest among clinical professionals and researchers regarding the efficient and accurate prediction of heart disease . According to the statistic data from WHO, one third population worldwide died from heart disease; heart disease is found to be the leading cause of death in developing countries by 2010. It shows one third American adult have one or more types of heart diseases based on American Heart Association report. Computational biology is often applied in the process of translating biological knowledge into clinical practice, as well as in the understanding of biological phenomena from the clinical data. The discovery of biomarkers in heart disease is one of the key contributions using computational biology. This process involves the development of a predictive model and the integration of different types of data and knowledge for diagnostic purposes. Furthermore, this process requires the design and combination of different methodologies from statistical analysis and data mining [2].

In modern times, the number of people suffering from heart disease is on a rise. A large number of people die every year due to heart disease all over the world. However, accurate diagnosis at an early stage followed by proper subsequent treatment can result in significant life saving. Unfortunately, accurate diagnosis of heart diseases has never been an easy task. As a matter of fact, many factors can complicate the diagnosis of heart diseases, often causing the delay of a correct diagnosis decision. For instance, the clinic symptoms, the functional and the pathologic manifestations of heart diseases are associated with many human organs other than the heart, and very often heart diseases may exhibit various syndromes. At the same time, different types of heart diseases may have similar symptoms. Hence, there is a pressing need to develop medical diagnostic decision support systems which can aid medical practitioners in the diagnostic process[3].

2. Literature Survey

The numbers of medical decision support systems are implemented using different approaches. George et al. have proposed decision support system to define and detect agitation transition. In this system support vector machines is used for detection. This system is for Dementia patients. This system presents a decision confidence measure and two new SVM architectures, which were applied to agitation detection and agitation transition detection. An accuracy of 91.4% was achieved, in comparison with 90.9% for the traditional SVM [4].

Haitham and Alan have proposed automated recognition of obstructive sleep apnea syndrome using SVM classifier. In this study, they evaluated features from the magnitude and phase of the thoracic and abdominal respiratory effort signals for OSA detection. This is based on the physiological fact that during normal breathing the abdominal and thoracic efforts happen simultaneously. The goal of this study is to evaluate classification of whole night normal and apneic epochs using extracted features from the phase and magnitude of the respiratory efforts signals, compared and combined with some other features from HRV and oxygen saturation signals [5] [6].

Support Vector machines have also been utilized in decision support systems such as [7]. An intelligent system based support vector machine along with a radial basis function network is presented for the diagnosis. The support vector machine with sequential minimal optimization algorithm is applied to India based patients' data set. Then, the Radial Basis Function(RBF) network structure trained by Orthogonal Least Square (OLS) algorithm is applied to same data set for predictions [8].

Tsai and Watanabe proposed a genetic algorithm(GA) - based method and implemented for determining the set of fuzzy membership functions that can provide an optimal classification of myocardial heart disease from ultrasonic images. In this method an average classification rate of 96% is achieved [9]. In another approach genetic algorithm is used to determine the attributes which contribute more towards the diagnosis of heart ailments which indirectly reduces the number of tests which are needed to be taken by a patient[10]. Yang and Honavar have proposed a feature subset algorithm using genetic algorithm. A genetic algorithm to select optimal feature subset for use with back propagation artificial neural networks has been described[11].A genetic algorithm for feature selection as well as for optimization of Support Vector Machine(SVM) parameter has been proposed by Haung. The proposed method performs feature selection and parameters setting in an evolutionary way [12]. Very recently, a real coded Genetic algorithm for critical feature analysis for heart disease diagnosis has been described [13].

Diagnosis Of Heart Disease Using Data mining Algorithm proposed by Rajkumar and Sophia. In this approach the initial diagnosis of a heart attack is made by a combination of clinical symptoms and characteristic electrocardiogram (ECG) changes. The accuracy with this technique is 52.33% [14]. Palaniappan and Awang proposed Intelligent Heart Disease Prediction System Using Data Mining Techniques. This research has developed a prototype Intelligent Heart Disease Prediction System(IHDPS) using data mining techniques, namely, Decision Trees, Nave Bayes and Neural Network. Results show that each technique has its unique strength in realizing the objectives of the defined mining goals[15].

The features of the artificial neural network (ANN), high accuracy and learning rate, make it worth trying as an algorithm to the prediction of heart disease [16]. A system for automatic

diagnosis of heart diseases using neural network is described by Kumaravel et al. The system uses features extracted from the ECG data of the patients. The system is used for classifying 5 major heart diseases using 38 input variables with an appreciable accuracy level (63.6 - 82.9%) [17].

Decision support system for heart disease based on support vector machine and artificial neural network is proposed by Gudadhe, Wankhede and Dongre. This paper proposed a decision support system for heart disease classification based on Support Vector Machine and MLP neural network architecture. Support Vector Machine classifies the heart disease data into two classes which shows presence of heart disease or absence of heart disease with 80.41% accuracy. Artificial Neural Network classifies the data into 5 categories of heart disease with 97.5% accuracy. This shows that both the methods gives high accuracy to classify the data but Artificial Neural Network classifies the data more accurately as compared to Support Vector Machine [3].

HDPS: Heart disease prediction system proposed by Chen, Huang and Hong. This heart disease prediction system can assist medical professionals in evaluating a patients heart disease based on the clinical data of the patient. The approach include three steps. Firstly select 13 important clinical features, i.e., age, sex, chest pain type, trestbps, cholesterol, fasting blood sugar, resting ecg, max heart rate, exercise induced angina, old peak, slope, number of vessels colored, and thal. Secondly, develop an artificial neural network algorithm for classifying heart disease based on these clinical features. The accuracy of prediction is near 80%[2]

3. Methodology

In the proposed decision support system for heart disease we will use two approaches. In first approach multilayer perceptron architecture of artificial neural network with back propagation algorithm is used and in the second approach Learning vector quantization algorithm is used[2][3].

3.1 Multilayer Perceptron

The single layer perceptron is not able to solve non- linearly separable problems. For that purpose one or more layers are added in single layer perceptron i.e. multi-layer perceptron. Multilayer perceptron network is a feed-forward neural network as shown in Fig. 1. They are widely used for pattern classification, recognition, prediction and approximation.

Above network have an **input layer** (on the left) with three neurons, one **hidden layer** (in the middle) with three neurons and an **output layer** (on the right) with three neurons. There is one neuron in the input layer for each predictor variable. In the case of categorical variables, $N-1$ neurons are used to represent the N categories of the variable.

Input Layer — a vector of predictor variable values $(x_1...x_p)$ is presented to the input layer. The input layer (or processing before the input layer) standardizes these values so that the range of each variable is -1 to 1. The input layer distributes the values to each of the neurons in the hidden layer. In addition to the predictor variables, there is a constant input of 1.0, called the *bias* that is fed to each of the hidden layers; the bias is multiplied by a weight and added to the sum going into the neuron.

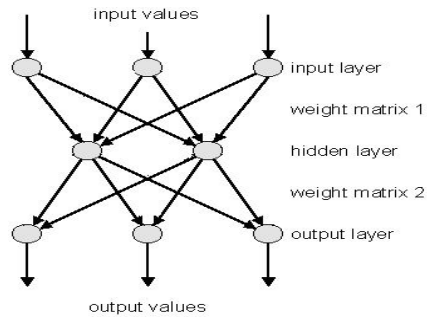


Fig. 1. Multi-layer Perceptron Network [20].

Hidden Layer — arriving at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight (w_{ji}), and the resulting weighted values are added together producing a combined value u_j . The weighted sum (u_j) is fed into a transfer function, σ , which outputs a value h_j . The outputs from the hidden layer are distributed to the output layer.

Output Layer — Arriving at a neuron in the output layer, the value from each hidden layer neuron is multiplied by a weight (w_{kj}), and the resulting weighted values are added together producing a combined value v_j . The weighted sum (v_j) is fed into a transfer function, σ , which outputs a value y_k . The y values are the outputs of the network.

3.2 Back-Propagation Network

The back-propagation algorithm can be employed effectively to train neural networks; it is widely recognized for applications to layered feed-forward networks, or multi-layer perceptrons. The back-propagation algorithm is capable of adjusting the network weights and biasing values to reduce the square sum of the difference between the given output (X) and an output values computed by the net (X') with the aid of gradient decent method as follows:

The back-propagation algorithm consists of four steps:

1. Compute how fast the error changes as the activity of an output unit is changed. This error derivative (EA) is the difference between the actual and the desired activity.

$$EA_j = \frac{\partial E}{\partial y_j} = y_j - d_j \quad \text{---(1)}$$

2. Compute how fast the error changes as the total input received by an output unit is changed. This quantity (EI) is the answer from step 1 multiplied by the rate at which the output of a unit changes as its total input is changed.

3. Compute how fast the error changes as a weight on the connection into an output unit is changed. This quantity (EW) is the answer from step 2 multiplied by the activity level of the unit from which the connection emanates.

$$EI_j = \frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j} \times \frac{\partial y_j}{\partial x_j} = EA_j y_j (1 - y_j) \quad \text{---(2)}$$

$$EW_{ij} = \frac{\partial E}{\partial Wij} = \frac{\partial E}{\partial xj} \times \frac{\partial xj}{\partial Wij} = EI_j y_i \quad \text{---(3)}$$

4. Compute how fast the error changes as the activity of a unit in the previous layer is changed. This crucial step allows back propagation to be applied to multilayer networks. When the activity of a unit in the previous layer changes, it affects the activities of all the output units to which it is connected. So to compute the overall effect on the error, we add together all these separate effects on output units. But each effect is simple to calculate. It is the answer in step 2 multiplied by the weight on the connection to that output unit.

$$EA_i = \frac{\partial E}{\partial yi} = \sum_j \frac{\partial E}{\partial xj} \times \frac{\partial xj}{\partial yi} = \sum_j EI_j Wij \quad \text{---(4)}$$

By using steps 2 and 4, we can convert the EAs of one layer of units into EAs for the previous layer. This procedure can be repeated to get the EAs for as many previous layers as desired. Once we know the EA of a unit, we can use steps 2 and 3 to compute the EWs on its incoming connections.

3.3 Learning Vector Quantization Algorithm

LVQ can be understood as a special case of an artificial neural network. It applies a winner-take-all Hebbian learning-based approach. The network has three layers, an input layer, a Kohonen classification layer, and a competitive output layer. The network is given by prototypes $W=(w(i),...,w(n))$. It changes the weights of the network in order to classify the data correctly. Learning Vector Quantization (LVQ) is a supervised version of vector quantization, similar to Selforganising Maps (SOM). As supervised method, LVQ uses known target output classifications for each input pattern of the form. It directly defines class boundaries based on prototypes, a nearest-neighbor rule and a winner-takes-it-all paradigm.

In terms of neural networks a LVQ is a feedforward net with one hidden layer of neurons, fully connected with the input layer. A CV can be seen as a hidden neuron ('Kohonen neuron') or a weight vector of the weights between all input neurons and the regarded Kohonen neuron respectively

LVQ uses that same internal architecture as SOM: a set of n -dimensional input vectors are mapped onto a two-dimensional lattice, and each node on the lattice has an n -dimensional reference vector associated with it. The learning algorithm for LVQ, i.e., the method of updating the reference vectors, is different from that for SOM. Because LVQ is a supervised method, during the learning phase the input data are tagged with their correct class. The input vectors are compared to the reference vectors and the closest match is found using the formula:

$$|w_{i*} - x| \leq |w_i - x| \quad \text{---} \quad \text{--- (5)}$$

where x is a input vector, w_i are the reference vectors, and w_{i*} is the winning reference vector. The reference vectors are then updated using the following rules:

$$\Delta w_{i*} = \alpha(t)(x - w_{i*}) \quad \text{--- (6)} \quad \text{if } x \text{ is in same class of } w_{i*}$$

$$\Delta w_{i*} = -\alpha(t)(x - w_{i*}) \quad \text{--- (7) if } x \text{ is in different class from } w_{i*}$$

$$\Delta w_{i*} = 0 \quad \text{--- (8) if } i \text{ is not the index of the winning reference vector}$$

The learning rate $0 < \alpha(t) < 1$ should generally be made to decrease monotonically with time, yielding larger changes for early iterations and more fine tuning as convergence is approached. There are several versions of the LVQ algorithm for which the learning rules differ in some details. When the learning phase is over, the reference vectors can be frozen, and any further inputs to the system will be placed into one of the existing classes, but the classes will not change.

LVQ offers an alternative which adapts few prototypes based on a given set of training data. Basic LVQ is given by a set of prototypes w_r together with class information. An input x is mapped to the class of the winner, i.e. the prototype w_r with smallest distance $d(w_r, x)$. Standard LVQ learning moves the respective winner into the direction of the considered pattern x or into the opposite direction, depending on the fact whether the classification is correct. Thus, LVQ shares the aspects of SOM which are relevant for a more general metric: besides a choice of the metric, the representation and adaptation of the prototypes is to be defined. LVQ itself does not possess a cost function in the continuous case, thus adaptations of original LVQ to more general metrics are often based on heuristics as proposed. Usually, the metric d is substituted by a problem specific version, but adaptation of the prototypes takes place as in standard LVQ using Hebbian learning[20].

4. Conclusion

The research shows that the decision support systems are useful tool for physician to detect the heart disease. The neural networks can be effectively used in the decision support system. The main characteristics of neural networks are that they have the ability to learn complex nonlinear input-output relationships, use sequential training procedures, and adapt themselves to the data. So that it is choice of most of the researchers in decision support systems.

References

- [1] "Medline plus: Heart diseases," <http://www.nlm.nih.gov/medlineplus/heartdiseases.html>.
- [2] A.H. Chen, S.Y. Huang, P.S. Hong, C.H. Cheng and E.J. Lin, "Hdps: Heart disease prediction system", IEEE Conference on Computing in Cardiology, pp. 557-560, December 2011.
- [3] M. Gudadhe, K. Wankhade and S. Dongre, "Decision support system for heart disease based on support vector machine and artificial neural network", IEEE International conference on computer and communication technology, pp. 741-745, November 2010.
- [4] G. E. Sakr, I. H. Elhaji and H. A. Huijer, "Support vector machines to define and detect agitation transition," IEEE Transactions On Affective Computing, vol. 1, pp. 98-108, December 2010.
- [5] M. Haitham, A. Angari and A. V. Sahakian, "Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier," IEEE Transactions On Information Technology In Biomedicine, vol. 16, pp. 463-468, May 2012.
- [6] F. Azuaje, W. Dubitzky, P. Lopes, N. Black and K. Adamsom, "Predicting coronary disease risk based on shortterm rr interval measurements: A neural network approach," Artificial Intelligence in Medicine, vol. 15, pp. 275-297, March 1999.

- [7] E. Comak, A. Arslan and T. Ibrahim, "A decision support system based on support vector machines for diagnosis of the heart valve diseases," *Computers in biology and Medicine*, vol. 37, pp. 21-27, January 2007.
- [8] S. Ghumbre, C. Patil, and A. Ghatol, "Heart disease diagnosis using support vector machine," *International Conference on Computer Science and Information Technology (ICCSIT)*, pp. 84-88, December 2011.
- [9] D. Y. Tsai and S. Watanabe, "Method for optimization of fuzzy reasoning by genetic algorithms and its application to discrimination of myocardial heart disease," *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 2239-2246, December 1966.
- [10] E. A. M. Anbarasi and N. Iyengar, "Enhanced prediction of heart disease with feature subset selection using genetic algorithm," *International Journal of Engineering Science and Technology*, vol. 2, pp. 5370-5376, November 2010.
- [11] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," *IEEE Intelligent Systems*, pp. 44-49, March 1998.
- [12] C. L. Huang and C. J. Wang, "A ga-based feature selection and parameters optimization for support vector machines," *Expert Systems with applications*, vol. 31, pp. 231-240, October 2006.
- [13] J. Z. H. Yan and C. Xiao, "Selecting critical clinical features for heart diseases diagnosis with a real-coded genetic algorithm," *Applied Soft Computing*, vol. 8, pp. 1105-1111, March 2008.
- [14] A. Rajkumar and G. S. Reena, "Diagnosis of heart disease using datamining algorithm," *Global Journal of Computer Science and Technology*, vol. 10, pp. 38-43, December 2010.
- [15] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," *International Journal of Computer Science and Network Security*, pp. 343-350, January 2008.
- [16] W. G. Baxt, "Application of artificial neural networks to clinical medicine," *Lancet*, vol. 346, pp. 1135-1138, October 1995.
- [17] N. Kumaravel, K. S. Sridhar and N. Nithiyanandam, "Automatic diagnoses of heart diseases using neural network," In *Proceedings of the fifteenth Biomedical Engineering Conference*, pp. 319-322, March 1996.
- [18] P. P. Sunila and N. Godara, "Decision support system for cardiovascular heart disease diagnosis using improved multilayer perceptron," *International Journal of Computer Application*, vol. 45, pp. 12-20, May 2012.
- [19] Q. K. AL-Shayea, "Artificial neural network in medical diagnosis," *International Journal of Computer Science*, vol. 8, pp. 150-154, March 2011.
- [20] "Neural network paradigms," <http://iopscience.iop.org/0067-0049/37020.text.html>.
- [21] UCI, "Uci machine learning repository: Heart disease data set," <http://archive.ics.uci.edu/ml/datasets/Heart+Disease.html>.