Decision Support System for Heart Disease based on Support Vector Machine and Artificial Neural Network

Mrudula Gudadhe, Kapil Wankhade
Department of Information Technology
Priyadarshini College of Engineering
Nagpur, India
mrudula_thakre@yahoo.co.in,kaps.wankhade@gmail.com

Snehlata Dongre, Lecturer
Department of Information Technology
Sinhgad Institute of Technology
Lonavala, Pune, India
dongre.sneha@gmail.com

***Abstract*—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. This paper presents a decision support system for heart disease classification based on support vector machine (SVM) and Artificial Neural Network (ANN). A multilayer perceptron neural network (MLPNN) with three layers is employed to develop a decision support system for the diagnosis of heart disease. The multilayer perceptron neural network is trained by back-propagation algorithm which is computationally efficient method. Results obtained show that a MLPNN with back-propagation can be successfully used for diagnosing heart disease than support vector machine.**

Keywords-back-propagation, clinical decision support system, heart disease, multilayer perceptron, support vector machine.

I. INTRODUCTION

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 and it is the leading cause of death in United States [17]. 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.

To reduce the diagnosis time and improve the diagnosis accuracy, it has become more of a demanding issue to develop reliable and powerful medical decision support systems (MDSS) to support the yet and still increasingly complicated diagnosis decision process. The medical diagnosis by nature is a complex and fuzzy cognitive process, hence soft computing methods, such as neural networks, have shown great potential to be applied in the development of MDSS of heart diseases. The number of medical decision support systems [1, 2, 4, 5, 6, 8, 9, 10, 11, 14, 16] are implemented using different approaches. A system for automatic diagnosis of heart diseases using neural network is described in [3]. 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%). Support Vector machines have also been utilized in decision support systems such as [13]. A genetic algorithm to select optimal feature subset for use with back propagation artificial neural networks has been described in [7]. The experiments however, have been performed taking the Cleveland database as a 2 class dataset. A genetic algorithm for feature selection as well as for optimization of SVM parameter has been proposed in [12]. Very recently, a real coded Genetic algorithm for critical feature analysis for heart disease diagnosis has been described in [15].

This paper presents a decision support system for heart disease classification using support vector machine and neural network. The dataset used is the Cleveland Heart Database taken from UCI learning data set repository which was donated by Detrano [18]. The dataset is being divided into five classes, 0 corresponding to absence of any disease and 1,2,3,4 corresponding to four different types of diseases. To the best of our knowledge, all the published works have used the dataset to differentiate between the absence (0) and presence (1, 2, 3 or 4) of a disease. In the present work, classifying the data into 5 classes using artificial neural network and using support vector machine classifying data into 2 classes.

The rest of the paper organized as, support vector machine described in section 2. Section 3 includes neural network theory and algorithm. Experiment and results explain in section 4 & follows conclusion in section 5.

II. SUPPORT VECTOR MACHINE

After SVM are invented by Vapnik in 1979, it is proposed by Vapnik in 1995 for classification and regression. Support Vector Machine (SVM), is a useful technique for data classification. Support Vector Machine, a promising new method for the classification of both linear and nonlinear data. A Support Vector Machine is an algorithm that works as follows. It uses a nonlinear mapping to transform the original training data into higher dimension. Within this new dimension, it searches for the linear optimal separating hyperplane that is a "decision boundary" separating the tuples of one class from another. With an appropriate nonlinear

978-1-4244-9034-/10/$26.00©2010 IEEE

mapping to a sufficient high dimension, data from two classes can always be separated by a hyperplane. The Support Vector Machine finds this hyperplane using support vectors ("essential" training tuples) and margins (defined by the support vectors).

Advantages of SVM are,

- Although the training time of even the fastest SVM's can be extremely slow they are highly accurate, owing to their ability to model complex nonlinear decision boundaries.
- They are much less prone to overfitting than other methods.
- The support vectors found also provide a compact description of the learned model.
- They have been applied to a number of areas, including handwritten digit recognition, object recognition, and speaker identification, and so many areas.
- SVM can be used for prediction as well as classification.

Support Vector Machine (SVM) is a class of supervised learning algorithms first introduced by Vapnik. SVM is a learning technique which trades off accuracy for generalization error. SVMs build a hyperplane which divides examples such that examples of one class are all on one side of the hyperplane, and examples of the other class are all on the other side.

Let the data set D be given as (X_1, y_1) , (X_2, y_2),……., (X_D, y_D), where X_i is the set of training tuples with associated class labels, y_i. Each y_i can take one of the two values, either +1 or -1 (i.e., $y_i \in \{+1,-1\}$). In an SVM, the idea is to find the hyperplane that maximizes the minimum distance from any training data point Fig 1. So we expect hyperplane with larger margin to be more accurate at classifying future data tuples than hyperplane with the smaller margin. This is why , the SVM searches for the hyperplane with the largest margin

A separating hyperplane can be written as,

$$W.X+b=0$$

where W is a weight vector and b is a bias.

Thus any point that lies above the separating hyperplane satisfies

$$W.X+b>0$$

Similarly, any point that lies above the separating hyperplane satisfies

$$W.X+b<0$$

The weights can be adjusted so that the hyperplanes defining the sides of margin can be written as

H_1: $W.X+b \geq 1$ for $y_i = +1$,

H_2: $W.X+b \geq 1$ for $y_i = -1$

That is, any tuple that falls on or above H_1 belongs to class +1, and any tuple that falls on or below H_2 belongs to class -1. Combining the two inequalities of equations we get

$$y_i(W.X+b) \geq 1, \text{ for all } i$$

The above problem can be solved by introducing the Lagrange multipliers ($\alpha_i \geq 0$ (i=1,2,..,m))

The patterns x_i which correspond to non-zero Lagrange coefficients are called support vectors. The resultant decision function has the following form

$$y(x) = \operatorname{sgn}\left[\sum_{i=1}^{m} \alpha_i y_i \langle x_i, x \rangle + b\right] \qquad (1)$$

Thus the optimal margin hyperplane is represented as a linear combination of training points. Consequently, the decision function for classifying points with respect to the hyperplane only involves dot products between points. The algorithm that finds a separating hyperplane in the feature space can be stated entirely in terms of vectors in the input space and dot products in the feature space.

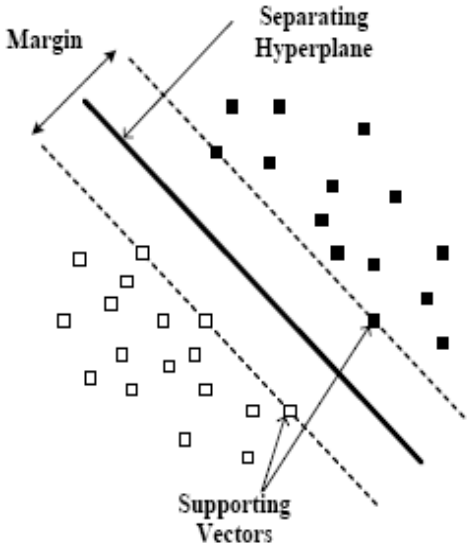

Figure 1. Maximum margin and optimal hyperplane

When the samples are not linearly separable, a kernel function is used to transforms the data to a higher dimensional space where it is linearly separable. The kernel function gives the dot product of the two examples in the higher dimensional space without actually transforming them into that space. This notion, dubbed the kernel trick, allows us to perform the transformation for purposes of classification to large dimensional spaces. In the nonlinear case, resultant decision function has the following form

$$y(x) = \operatorname{sgn}\left[\sum_{i=1}^{m} \alpha_i y_i K\langle x_i, x \rangle + b\right] \qquad (2)$$

where the kernel function $K(x_i, x) = \langle \varphi(x_i), \varphi(x) \rangle$ and $\varphi(x)$is the nonlinear map from original space to the high dimensional space. Following four basic kernels:

- linear: $K(x_i, x_j) = x_i^T x_j$.
- polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$, $\gamma > 0$.

- radial basis function (RBF): $K(x_i, x_j) = \exp(-\gamma|| x_i - x_j||^2)$, $\gamma > 0$.
- sigmoid: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$.
  Here γ, r, and d are kernel parameters.

The maximum margin allows the SVM to select among multiple candidate hyperplanes; however, for many data sets, the SVM may not be able to find any separating hyperplane at all, either because the kernel function is inappropriate for the training data or because the data contains mislabeled examples.

III. ARTIFICIAL NEURAL NETWORK

A. Multilayer Perceptron

The Multilayer perceptron was first introduced by M. Minsky and S. Papert in 1969. A neural network with one or more layers of nodes between the input and the output nodes is called *multilayer network*. The multilayer *network structure*, or *architecture*, or *topology*, consists of an input layer, one or more hidden layers, and one output layer. The input nodes pass values to the first hidden layer, its nodes to the second and so on till producing outputs shown in Fig. 2.

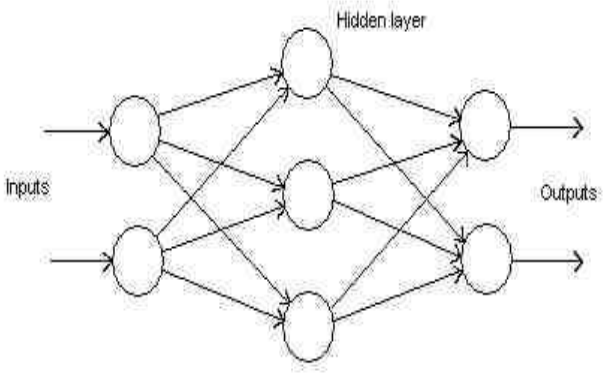

Figure 2. Multilayer Perceptron (MLP)

These are the hidden units that enable the multilayer network to learn complex tasks by extracting progressively more meaningful information from the input examples. The multilayer network MLP has a highly connected topology since every input is connected to all nodes in the first hidden layer, every unit in the hidden layers is connected to all nodes in the next layer, and so on. The input signals, initially these are the input examples, propagate through the neural network in a forward direction on a layer-by-layer basis, that is why they are often called feedforward multilayer networks.

Remember that the function inside units take as input the weighted sum, S, of the values coming from the units connected to it shown in Fig. 3. The function inside sigmoid units calculates the following value, given a real-valued input S:

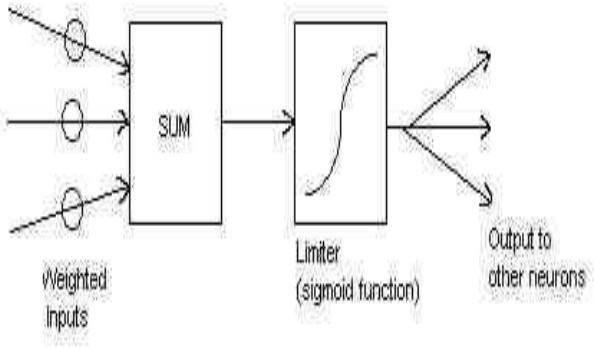

Figure 3. Function of one node of MLP

Sigmoid Function:

$$\sigma(S) = \frac{1}{1+e^{-s}} \qquad (3)$$

where e is the base of natural logarithms, e = 2.718...
Two kinds of signals pass through these networks:
- *function signals*: the input examples propagated through the hidden units and processed by their activation functions emerge as outputs
- *error signals*: the errors at the otuput nodes are propagated backward layer-by-layer through the network so that each node returns its error back to the nodes in the previous hidden layer.

B. Back-propagation

Since the real uniqueness or 'intelligence' of the network exists in the values of the weights between neurons, we need a method of adjusting the weights to solve a particular problem. For this type of network, the most common learning algorithm is called Back Propagation (BP). A BP network learns by example, that is, we must provide a learning set that consists of some input examples and the known-correct output for each case. So, we use these input-output examples to show the network what type of behavior is expected, and the BP algorithm allows the network to adapt.

The BP learning process works in small iterative steps: one of the example cases is applied to the network, and the network produces some output based on the current state of its synaptic weights (initially, the output will be random). This output is compared to the known-good output, and a mean-squared error signal is calculated. The error value is then propagated backwards through the network, and small changes are made to the weights in each layer. The weight changes are calculated to reduce the error signal for the case in question. The whole process is repeated for each of the example cases, then back to the first case again, and so on. The cycle is repeated until the overall error value drops below some pre-determined threshold. At this point we say that the network has learned the problem.

Properly trained backpropagation networks tend to give reasonable answers when presented with inputs that they have never seen. Typically, a new input leads to an output similar to the correct output for input vectors used in training that are similar to the new input being presented. This generalization property makes it possible to train a network on a representative set of input/target pairs and get good results without training the network on all possible input/output pairs.

IV. EXPERIMENTS AND RESULTS

The experiments are carried out on real heart data using support vector machine and artificial neural network

A. Experiment using Support Vector Machine

In order to test our approach we use the Cleveland Heart Database taken from UCI learning data set repository which was donated by Detrano [8-9]. The data set consists of 13 numeric attributes which include age, sex, chest pain type, resting blood pressure, cholesterol, fasting blood sugar, resting ECG, maximum heart rate, exercise induced angina, oldpeak, slope, number of vessels coloured and thal respectively. The classes include integers valued from 0 (no presence) to 4 (types of heart diseases). As a two class classification problem, 0 corresponding to absence of any disease and 1,2,3,4 corresponding to presence of heart disease Total number of patients instances is 303 from them 6 tuples having missing value so we ignore those tuples which are having missing values so we consider only 297 instances and 200 of them are used for training and rest are used for testing the SVM. Table 1 depicts the distribution of disease records for SVM. The experimental results are provided in Figure 4.
We have used the following procedure:

- Conduct simple scaling on the data
- Consider the RBF kernel $K(x; y) = \exp(-\gamma \| x_i - x_j \|^2)$, $\gamma > 0$
- Use cross-validation to find the best parameter C and γ
- Use the best parameter C=8 and γ=0.001953 to train the whole training sets
- Test

The 13 input parameters are

1. Age
2. Sex
3. Chest pain type
4. Resting blood pressure
5. Serum cholesterol
6. Fasting blood sugar
7. Resting electrocardiographic result
8. Maximum heart rate achieved
9. Exercise induced angina
10. Old peak
11. Slope
12. Number of vessels coloured
13. Thal

TABLE I. DISTRIBUTION OF HEART DISEASE RECORDS FOR SVM

Disease Diagnosis	No. of records	Training data	Test data
Presence of heart disease	140	90	50
Absence of heart disease	157	110	47
Total	297	200	97

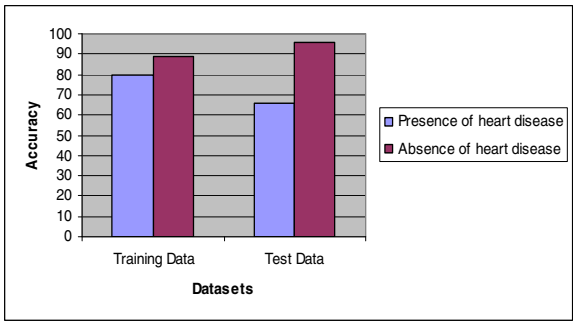

Figure 4. Comparison of accuracy rate obtained by training set and test set on Real Data

B. Experiment using Artificial Neural Network

In this subsection verify our algorithm on same real life heart data set from UCI machine learning repository [18]. The input layer possesses 13 nodes corresponding to the same number of input variables, extracted from a large number of patient cases.The 13 input variables are categorized into 2 groups and they are encoded accordingly (i) Variables with two independent attributes such as gender is encoded with binary values (0, 1); (ii)) The rest variables are encoded using continuous variables such as age, resting blood pressure are normalized on to the interval [0, 1].

The "The Disease Diagnosis" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Here the H0 denoting no presences of heart disease and H1, H2, H3, H4 are presenting the presence of heart disease. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values H1, H2, H3, H4) from absence (value H0). Table 2 depicts the distribution of disease records for ANN. The experimental results are provided in Figure 5.

TABLE II. DISTRIBUTION OF HEART DISEASE RECORDS FOR ANN

Disease diagnosis	No. of records	Training data	Test data
H0	164	122	42
H1	55	40	15
H2	36	20	16
H3	35	27	8
H4	13	11	2
Total	303	220	83

Figure 5. Comparison of accuracy rate obtained by training set and test set on Real Data

V. CONCLUSION

This paper has proposed a decision support system for heart disease classification based on Support Vector Machine and MLP neural network architecture. For training of MLP, Back propagation algorithm is used which is the famous learning algorithm for training of MLP. Support Vector Machine classifises 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 classifises the data into 5 categories of heart disease with 97.5% accuracy. This shows that both the methods gives high accurary to classify the data but Artificial Neural Network classifies the data more accurately as compared to Support Vector Machine. So decision support systems using SVM and Artificial Neural Network both can be deployable in any clinics to help the physician.

REFERENCES

[1] C. M. Bishop, Neural network for pattern recognition, Oxford University Press, Oxford, 1995.

[2] B. D. Ripley, Pattern recognition and neural network, Cambridge University Press, Cambridge, 1996.

[3] N. Kumaravel, K. S. Sridhar, N. Nithiyanandam, "Automatic diagnoses of heart diseases using neural network", Proccedings of the fifteenth Biomedical Engineering Conference, pp 319-322, 29-31 March,1996.

[4] D. J. Hand, Construction and assessment of classification rules, Wiley, New York, 1997.

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

[6] D. Y. Tsai and S. Watanabe, "Method for optimization of fuzzy reasoning by genetic algorithms and is application to discrimination of myocardial heart disease", IEEE Nuclear Science Symposium and Medico1 Imaging Conference, Pages 1756-1761, Nov. 1998.

[7] J. Yang, V. Honavar, "Feature Subset Selection using a Genetic Algorithm", IEEE Intelligent Systems, pp 44-49, March-April 1998.

[8] F. Azuaje, W. Dubitzky, P. Lopes, N. Black, K. Adamsom et al., "Predicting coronary disease risk based on short-term RR interval measurements: A neural network approach", Artificial Intelligence Medicine, IS: 275.197, 1999.

[9] E. J. Tkacz and P. Kostka, "An application of wavelet neural network for classification patients with coronary artery disease based on HRV analysis", Proceedings of Annual Internoliono1 Conference of the IEEE Engineering in Medicine and Biology. July 2000, pages 1391 - 1393.

[10] W. David and W. Vivian, "Improving diagnostic accuracy using a hierarchical neural network to model decision subtasks", Int. J. Med. Inform., 57: 41-55, 2000.

[11] Mercedes Fernandez and Carlos Hernandez-Espinosa, Weight Initialization Methods for Multilayer Feedforward, European Symposium on Artificial Neural Networks, Bruges (Belgium), 25-27, April 2001.

[12] C. L. Huang, C. J. Wang, "A GA-based feature selection and parameters optimization for support vector machines", Expert Systems with applications, Vol.-31, pp -231-240, 2006.

[13] E. Comak, A. Arslan, T. Ibrahim, " A decision support system based on support vector machines for diagnosis of the heart valve diseases", Computers in biology and Medicine, pp 21-27, 2007.

[14] H. Temurtas, N. Yumusak and F. Temurtas, A comparative study on diabetes disease diagnosis using neural networks, Elsevier Ltd, 2008.

[15] H. Yan, J. Zheng, Y. Jiang, C. Peng, C. Xiao, " Selecting critical clinical features for heart diseases diagnosis with a real-coded genetic algorithm", Applied Soft Computing, Vol.-8, pp 1105-1111, 2008.

[16] M. Rouhani and Mehdi motavalli haghighi, The Diagnosis of Hepatitis diseases by Support Vector Machines and Artificial Neural Networks, IEEE, 2009.

[17] Medline Plus: Heart Diseases. http://www.nlm.nih.gov/medlineplus/heartdiseases.html

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