

Diagnosis of heart disease using a committee machine neural network

Kanita Karaduzovic-Hadziabdic^a, Raşit Köker^b

^aInternational University of Sarajevo, Faculty of Engineering and Natural Sciences,
Sarajevo, Bosnia and Herzegovina
kanita@ius.edu.ba

^bSakarya University, Faculty of Technology, Electrical and Electronics Engineering
Department 54187, Esentepe – Sakarya Turkey
rkoker@sakarya.edu.ba

Abstract

Heart disease is one of the leading causes of deaths worldwide. Several methods have been developed by researchers to support medical diagnosis of heart disease, including artificial intelligence methods. In the past, committee machines have been shown to achieve higher classification accuracy than a single classifier. This study uses a committee of classifiers consisting of a combination of feed-forward multi layer perceptron (MLP) and radial basis functions neural networks (RBF) to diagnose heart disease. The output of the committee has been obtained based on majority voting. Several MLP training algorithms have been analyzed from the viewpoint of learning performance based on the network topology to find the network with the best prediction results. Cleveland heart disease dataset has been used throughout the experiments. The results show that the committee machine approach gives significantly better results than a single neural network. The classification accuracy obtained by the proposed method achieves a high accuracy rate of 95,4545%. This result is better than the results achieved by other methods that use Cleveland dataset reported in literature to this date.

Keywords: Machine learning, neural networks, committee machines, cardiovascular diseases

1. Introduction

Heart disease (cardiovascular disease) is a disease involving heart or blood vessels. Heart disease remains one of the leading causes of deaths worldwide. The total number of deaths from heart disease in 1990 was 14.4 million. This number rose

to 17.5 million in 2005. By 2015, the World Health Organization estimates that this number will rise to 20 million, which is equivalent to 30% of worldwide deaths [1]. The main risk factors of heart disease include diabetes, family history of heart disease, stress, poor diet and nutrition, too much alcohol, smoking, hypertension, high LDL cholesterol and low HDL cholesterol [2]. Early and correct diagnosis of heart disease can allow proper treatment and as a result reduce the mortality rate. Among machine learning methods, neural networks are one of the most common methods used in medical diagnostics [3].

Many machine learning classification methods have been employed on a Cleveland heart disease dataset. Detrano, who created the Cleveland heart disease dataset, achieved a classification accuracy of 77% by using logistic regression derived discriminant function [4]. Chen et al. [5] used feed-forward multilayer perceptron (MLP) neural network for the classification of heart disease. Their achieved accuracy is near 80%. Cheung [6] used C4.5, Naive Bayes, BNND and BNNF algorithms and obtained classification accuracy of 81.11%, 81.48%, 81.11% and 80.96%, respectively. A. V. Senthil Kumar [7] applied the combination of artificial neural networks and fuzzy inference system to diagnose heart disease. He achieved an accuracy of 91.83%. Bhuvaneswari [8] applied the combination of genetic algorithm and neural networks for the classification of heart disease. He first applied genetic algorithm to obtain the weights that are used by the feed-forward neural network in heart disease classification. Classification accuracy obtained was 94.17%. An approach to use an ensemble of neural networks for heart disease diagnosis was performed by Das et al. [9]. They create an ensemble based method by combining the posterior probabilities or the predicted values from multiple predecessor multilayer feed-forward neural networks. They report an achieved accuracy of 89.01%. Can [10] applied principal component analysis and a system of parallel MLP neural networks to diagnose heart disease. He achieved a success rate of 88.5%. In this study, we use committee machines of different classifiers as an aid to medical heart disease diagnostics.

One of the keys to successfully design a committee is to create individual neural networks which perform better than random guessing and at the same time are not correlated regarding the generated error on the training data [11]. We propose a machine learning intelligent system composed of a combination of MLP and radial basis function (RBF) networks to distinguish the presence or absence of heart disease. We show that an appropriate combination of different classifiers can increase the performance of a single classifier. Committee machines have already been shown to outperform a single classifier in medical diagnostics as well as in other domains [12,13,14]. Experiments carried out in this research show that the final results of the committee based on majority voting increases the overall classification accuracy. The obtained classification accuracy is 95.4545%. Cleveland dataset from the University of California, Irvine (UCI) Machine Learning Repository [15] is used throughout the experiments.

2. Dataset description

The dataset was obtained from the University of California Irvine (UCI) Machine Learning Repository [19]. The repository contains four different heart disease databases out of which the Cleveland dataset is the most complete and is used in this study. It contains 303 cases. Six of the cases have incomplete data which has been discarded and 297 examples were used during the experiments. Out of these, 137 cases have heart disease and the remaining 160 do not. From the 76 available attributes all the published papers refer to the following 14 attributes including the class type (i.e. whether a patient has a heart disease or not):

Feature	Description
Age	Age in years
Sex	Instance gender (1= male; 0 = female)
Cp	Chest pain type (1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic)
Trestbps	Resting blood pressure in mm Hg
Chol	Serum cholesterol in mg/dl
Fbs	Fasting blood sugar > 120 mg/dl (1: true, 0: false)
Restecg	Resting ECG results (0: normal, 1: ST-T wave abnormality, 2: LV hypertrophy)
Thalach	Maximum heart rate achieved
Exang	Exercise induced angina (1: yes, 0: no)
Oldpeak	ST depression induced by exercise relative to rest
Slope	Slope of the peak exercise ST segment (1: up-sloping, 2: flat, 3: down-sloping)
Ca	Number of major vessels colored by fluoroscopy (values 0-3)
Thal	Defect type. Values: 3: normal, 6: fixed defect, 7 reversible defect
Num	Diagnosis of heart disease (0 for healthy, 1 for heart disease)

Table 1: Summary of Cleveland heart disease dataset

3. Artificial neural networks

Artificial neural networks constitute of interconnected artificial neurons that use a computational model to process information. They have the ability to learn and adapt. During the learning (training) process, the adjustment of weights between the connected neurons is performed. Learning rules can be divided into three categories: supervised learning, unsupervised learning and reinforcement learning. In supervised learning, the synaptic weights are adjusted to minimize the differences between the desired response and actual response of the network. The training of the network is repeated until there are no significant changes in the synaptic weights, indicating the successful learning phase. In unsupervised learning, desired results are not available and the weights are updated in response to the inputs only. In reinforcement learning, neural network takes inputs and the network interacts with its environment to measure its performance. The knowledge acquired during

the training phase is stored in the synaptic weights.

3.1. Multilayer perceptron networks

One of the most common used architecture of neural networks in pattern classification task is feed-forward multilayer perceptron neural network. It consists of an input layer, one or more hidden layers and an output layer. The input and output layers consist of input and output neurons respectively, whose numbers are set in task specification. However, the number of hidden layers and the number of neurons in each hidden layer that will lead to successful classification results is not trivial and is usually task dependant. Rumelhart et. al. in 1986 [16] proposed training of MLP neural networks based on an error-correction back propagation algorithm, which consists of a feed-forward pass and back propagation. During the feed-forward pass supervised learning is used where the network is trained using the data for which inputs as well as the desired outputs are known. Every node in the hidden layer of the network as well as the nodes in the output layers calculate the activation values. The differences between actual output and the desired output are used to calculate the error. During the back propagation, the error is propagated from the output layer to the input layer. This error at the output neuron is used to adjust the weights of the interconnecting neurons. Usually, algorithms based on the gradient descent training algorithm (also known as steepest descent), which is a first order optimization algorithm, are used to update the weights. After a certain error or the number of training epochs is reached, the training phase is completed and final weights are obtained. The weights that were computed during training contain the knowledge that are used during the testing phase. Figure 1 illustrates an example of MLP neural network used in this work. We use 14 attributes described in Table 1 as inputs to the neural network.

3.2. MLP training algorithms

Several training algorithms have been developed for training of MLP neural networks, majority which are based on the gradient descent algorithm. In this work we analyze the performance of 8 training algorithms. To choose the best one for our system, we created a neural network composed of 14 input nodes (attributes presented in Table 1), one hidden layer, and one neuron in the output layer (i.e. indicating whether a patient has a heart disease or not). The number of hidden neurons is important for accuracy and successful classification. For each training algorithm the optimal number of hidden nodes was found by experiment and used during training. The initial weights were chosen randomly for each neural network. We than tested this neural network architecture with 8 different training algorithms and noted the results. Table 2 displays training algorithms analyzed with the obtained classification accuracy.

Levenberg-Marquardt (LM) [17,18] back-propagation algorithm achieved the highest accuracy, 93.1818%, containing 9 neurons in the hidden layer. LM algorithm is a numerical minimization algorithm that was independently developed by

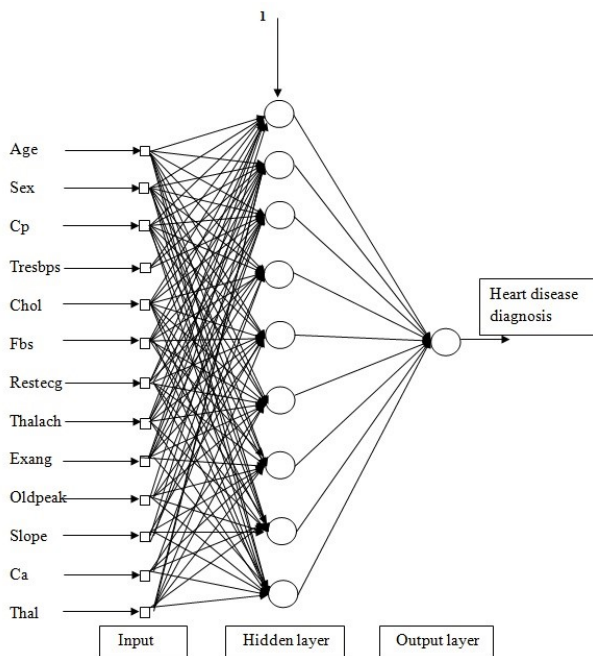


Figure 1: Architecture of MLP neural network

Kenneth Levenberg and Donald Marquardt. It is based on the gradient descent algorithm, which is a stable but very slow algorithm and Gauss-Newton algorithm which greatly improves the convergence speed of the gradient descent algorithm. The resultant LM algorithm is fast and has stable convergence. We use this algorithm in our study.

Training Algorithm	Accuracy %
Levenberg-marquardt back-propagation	93.1818%
Gradient descent	86.3636%
Gradient descent with momentum	81.8%
Gradient descent with momentum and adaptive rule back-propagation	86.36%
Resilient back propagation	79.55%
Scaled conjugate gradient back-propagation	77.27%
BFGS quasi-Newton back-propagation	81.82%
One step secant method	86.36%

Table 2: Test performance of different neural network training algorithms

3.3. Radial basis function networks

Radial basis function networks are feed-forward neural networks that have three layers: an input layer, a hidden layer containing a non-linear RBF activation function and an output linear layer. Only the weights between the hidden layer and the output layer are determined. The weights between the input and the hidden layer are fixed. The hidden layer consists of n neurons. RBF networks use radial basis functions as activation functions. Most commonly used radial basis function is Gaussian function:

$$G(x, c_i) = e^{-\|x - c_i\|/2\sigma_i^2}, \quad i = 1, 2, \dots, k$$

where x is the input vector, c_i is the center and σ_i is the width of the i^{th} neuron in the hidden layer. Since the Gaussian function responds only to a region of the input where the Gaussian is centered, it is important that the suitable centers are found. The weights between the input layer the hidden layer are the centers c_i . The distance between the input vector x and c_i is the Euclidean distance. RBF neural networks are fully connected. The output layer contains m neurons which is the number of possible outputs. The result of each output neuron is the linear weighted sum of the outputs of the middle layer:

$$z_j = \sum_{i=1}^k w_{ij} G(x, c_i), \quad j = 1, 2, \dots, m$$

where w_{ij} is the weight between the i^{th} hidden layer neuron and the j^{th} output layer neuron and m is number of neurons in the output layer. Since linear combiners often have fewer weights, RBF networks train fast and require fewer training samples [19]. Figure 2 illustrates the architecture of the RBF neural network:

3.4. Neural network committee machine design

In order to perform the classification, committee machine approach is taken. Committee machines have been shown to achieve increased classification accuracy as the result of the committee is much better than the result of the individual machine performing alone [20]. Even if an individual machine does not perform well, the overall decision of the committee may still result in a successful diagnosis based on the final decision. The outputs of individual machines of a committee may be combined in several ways. In this paper we use majority voting to combine the outputs of a committee. In majority voting, the majority number of neural networks, *the experts*, that have made a decision on the particular output is considered as the final decision. Committee machines designed in this research are composed of three trained experts, two MLP neural networks and one RBF neural network. Due to the scarcity of the data (only 297 samples), each individual neural network was trained on the same data set, input $x(n)$. The final decision made is based on the majority voting producing the overall output $y(n)$. The architecture of our system is depicted in Figure 3.

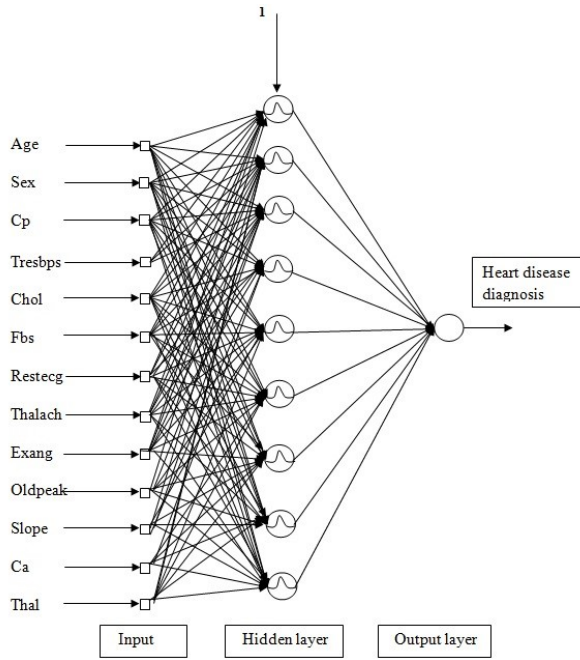


Figure 2: Architecture of the RBF neural network

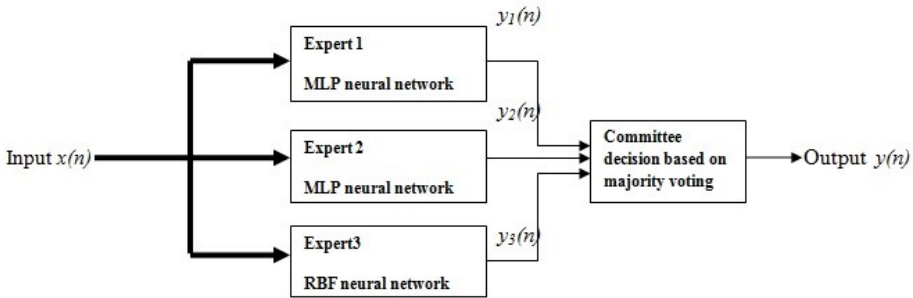


Figure 3: Block diagram of a neural network committee machine

As activation function, for the two MLPs, nonlinear anti-symmetric sigmoid hyperbolic tangent activation function was used. It is the most widely used activation function. It is defined by:

$$\varphi(v) = \tanh(v)$$

Anti-symmetric sigmoid tangent hyperbolic activation function is smooth differentiable function, which is important for the successful operation of the back-propagation stage of the network training algorithm.

As Figure 4 shows, the anti-symmetric sigmoid tangent hyperbolic activation

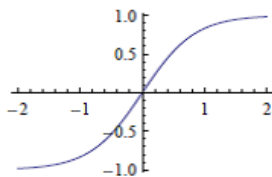


Figure 4: Anti-symmetric sigmoid hyperbolic tangent activation function

function accepts an input in any range and produces the output in a range $[-1; 1]$. The feature vector components containing the input data were scaled to an interval $[-1; 1]$. For RBF neural network a Gaussian radial basis function is used as the activation function. 70% of the data has been used for training of the neural network, 15% for the validation and 15% for the testing. Implemented code was done with the help of Matlab's Neural Network Toolbox.

4. Results

In our experiments we used the combination of MLP neural networks and RBF neural networks to perform heart disease classification. Table 3 shows the experimental results of each individual network and the results of the committee based on majority voting.

Expert	Accuracy %
Expert1 (MLP)	86.3636
Expert 2 (MLP)	93.1818
Expert 3 (RBF)	88.6364
Committee based on majority voting	95.4545

Table 3: Correct heart disease classification results

Increasing the number of MLP or RBF neural networks did not increase the performance results. The experimental results show that the proposed classification method to the best of our knowledge performs better than the other heart disease classification methods. From the results it is evident that the accuracy achieved by the committee is much better than the results of a single classifier. The committee was effective because each classification algorithm correctly recognized some heart disease risk factors with respect to the other ones. Tests were also performed by using more than three experts, as well as making all three experts to be of type MLP or RBF but this did not result in better performance of the system.

5. Conclusion

In this paper committee machines based on majority voting constituting of two MLP and one RBF neural networks have been applied to predict whether a patient has cardiovascular disease or not. We showed that the combination of these classifiers in an ensemble increases the accuracy and the robustness of the system as the committee was able to outperform a single classifier. After analyzing eight MLP training algorithms, it was found that Levenberg-Marquardt algorithm performs the best and was used for the training of MLP neural networks. The final result obtained using majority voting achieved high accuracy of 95.4545%. This result is higher than the results obtained by other methods that have also used Cleveland dataset recorded in the literature to this date.

References

- [1] Promoting Cardiovascular Health in the Developing World, Institute of Medicine (US) Committee on Preventing the Global Epidemic of Cardiovascular Disease: Meeting the Challenges in Developing Countries; Edited by Valentin Fuster and Bridget B Kelly, National Academies Press (US); 2010
- [2] <http://science.jrank.org/pages/3259/Heart-Diseases.html> (last accessed: June 25, 2013)
- [3] J. Piecha, The neural network selection for a medical diagnostic system using an artificial data set., *Journal of Computing and Information Technology*, 9(2), p.p. 123-132, 2001
- [4] <http://vk.cs.umn.edu/bsitutorial/uci/numeric/cholesterol.arff> (last accessed: June 25, 2013)
- [5] A.H. Chen and S.Y. Huang and P.S. Hong and C.H. Cheng and E.J. Lin, HDPS: Heart Disease Prediction System, *Computing in Cardiology* 2011
- [6] N. Cheung, Machine learning techniques for medical analysis. School of Information Technology and Electrical Engineering B. SC. Thesis University of Queensland, 2001
- [7] A.V. Senthil Kumar, Diagnosis of Heart Disease using Fuzzy Resolution Mechanism, *Journal of Artificial Intelligence* 5 (1): pp. 47-55, 2012
- [8] A. N. G. Bhuvanewari, Cardiovascular Disease Prediction System using Genetic Algorithm and Neural Network, *Computing, Communication and Applications (IC-CCA)*, 2012 International Conference on, vol., no., pp. 1-5, 22-24 Feb 2012
- [9] R. Das and I. Trukoglu and A. Sengur, Effective diagnosis of heart disease through neural networks ensembles, *Expert Systems with Applications*, Elsevier Lt, 2008
- [10] M. Can Committee Machine Networks to Diagnose Cardiovascular Diseases, *Southeast Europe Journal of Soft Computing*, pp. 76-93, 2013
- [11] T. G. Dietterich. Machine-learning research: Four current directions. *The Artificial Intelligence Magazine*, 18(4) pp. 97-136, 1998.
- [12] F. Astrom and R. Koker, A parallel neural network approach to prediction of Parkinson's Disease, *Expert systems with applications*, 38(10), pp.12470-12474, 2011.

- [13] A. Sboner and C. Eccher and E. Blanzieri and P. Bauer and M. Cristofolini and G. ZUiani and S. Forti A multiple classifier system for early melanoma diagnosis, *Artificial Intelligence in Medicine*, pp. 29-44, 2003
- [14] M. Can and K. K. Hadziabdic and N. M. Demir Teaching Neural Networks to Classify the Authors of Texts, *International Congress on Knowledge, Economy and Management*, 2011
- [15] <http://archive.ics.uci.edu/ml/datasets/Heart+Disease>, UCI Repository of Machine Learning Data bases, Cleveland Heart Disease Dataset
- [16] D. E. Rumelhart and G.E. Hinton and R. J. Williams, Learning internal representations by error propagation. In Rumelhart, D. E., McClelland, J. L., and the PDP Research Group, editors, *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 1: Foundations*, pp. 318-362. The MIT Press, Cambridge, MA, 1986
- [17] K. Levenberg, A method for the solution of certain problems in least squares, *Quarterly of Applied Mathematics*, pp. 164-168, 1944
- [18] D. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, *SIAM Journal on Applied Mathematics*, 11(2), pp. 431-441, 1963
- [19] S. Haykin, *Neural network: A comprehensive foundation*, Upper Saddle River, NJ: Prentice Hall, 1999
- [20] M. Skurichina and R. P. W. Duin, Bagging, boosting, and the random subspace method for linear classifiers, *Pattern Anal. Appl.* 5(2), pp. 121-135, 2002