

# Determining the Efficient Structure of Feed-Forward Neural Network to Classify Breast Cancer Dataset

Ahmed Khalid

Najran University, Community College  
Department of Computer Science  
Najran, KSA

Noureldien A. Noureldien

University of Science and Technology  
Department of Computer Science  
Omdurman, Sudan

**Abstract**—Classification is one of the most frequently encountered problems in data mining. A classification problem occurs when an object needs to be assigned in predefined classes based on a number of observed attributes related to that object.

Neural networks have emerged as one of the tools that can handle the classification problem. Feed-forward Neural Networks (FNN's) have been widely applied in many different fields as a classification tool.

Designing an efficient FNN structure with optimum number of hidden layers and minimum number of layer's neurons, given a specific application or dataset, is an open research problem.

In this paper, experimental work is carried out to determine an efficient FNN structure, that is, a structure with the minimum number of hidden layer's neurons for classifying the Wisconsin Breast Cancer Dataset. We achieve this by measuring the classification performance using the Mean Square Error (MSE) and controlling the number of hidden layers, and the number of neurons in each layer.

The experimental results show that the number of hidden layers has a significant effect on the classification performance and the best classification performance average is attained when the number of layers is 5, and number of hidden layer's neurons are small, typically 1 or 2.

**Keywords**—Hidden Layers; Number of neurons; Feed Forward Neural Network; Breast Cancer; Classification Performance

## I. INTRODUCTION

Classification is one of the most frequently encountered problems in decision making tasks. A classification problem occurs when an object needs to be assigned in predefined classes based on a number of observed attributes related to that object. Many problems in business, science, industry, and medicine can be treated as classification problems.

Neural networks have emerged as one of the tools that can handle the classification problem. The advantage of neural networks is that, neural networks are data driven self-adaptive methods so that they can adjust themselves to the data without any explicit specification of functional form for the underlying model, and they can approximate any function with arbitrary accuracy.

Artificial neural networks consist of an input layer of nodes, one or more hidden layers and an output layer. Each

node in a layer has one corresponding node in the next layer, thus creating the stacking effect [1].

The hidden layer is a collection of neurons which provide an intermediate layer between the input layer and the output layer. Activation functions are typically applied to hidden layers.

Neural Networks are biologically inspired and mimic the human brain. A neural network consists of neurons which are interconnected with connecting links, where each link have a weight that multiplied by the signal transmitted in the network. The output of each neuron is determined by using an activation function such as sigmoid and step. Usually nonlinear activation functions are used. Neural networks are trained by experience. When an unknown input to the network is applied, it can generalize from past experiences and product a new result [2], [3].

Feed-forward neural networks (FNN) are one of the popular structures among artificial neural networks. These efficient networks are widely used to solve complex problems by modeling complex input-output relationships [4], [5].

Neural networks have been widely used for breast cancer diagnosis [6] [7] [8], and Feed-forward Neural Network (FNN) is commonly used for classification. Many researches evaluates the effect of the number of neurons in the hidden layer [9] [10][11] [12].

In this paper an experimental investigation was conducted to see the effect of the number of neurons and hidden layers of feed forward neural network on classification performance for the breast cancer dataset. The work of this paper will be presented in different sections. In the second section materials and methods are introduced. An experiment and results are presented in the third section. Section four gives discussion and conclusions.

## II. MATERIALS AND METHODS

The performance analysis of FFNN is to estimate the training and generalization errors. The result with the minimum estimated generalization error is used to determine an optimum for the application of neural network model [13].

The feed forward neural network is built using of Levenberg-Marquardt training algorithm which is widely used in classification literature [14,15,16].The network architecture used is composed of nine neurons for input layer and one neuron for the output layer. To achieve the paper objectives,

the number of hidden layers and the number of neurons per hidden layer are changed during the training and simulation of the network. The learning rate of 0.5 was used. The number of maximum allows epochs were 1000. The activation functions used in the different layers of the neural network is logsig. The performance of the classification is measured by the mean square error (MSE) which is calculated by equation 1.

$$MSE = \frac{1}{N} \sum_{k=1}^N e(k)^2 = \frac{1}{N} \sum_{k=1}^N (t(k) - a(k))^2$$

Where t(k) is the target output and a(k) is network output.

In this paper Wisconsin Breast Cancer Data (WBCD) is used, which have been analyzed by various researchers of medical diagnosis of breast cancer in the neural network literature [5], [16], [17], [18]. This data set contains 699 instances. The first attribute is the ID of an instance, and the next 9 attributes (Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli and Mitoses) represent different characteristics of an instance and the last attribute takes the values 2 and 4 (2 for benign, 4 for malignant). Each instance has one of 2 possible classes (benign or malignant).

In our experiments all 9 attributes are used. Each attribute has the domain 1-10. The data set was partitioned into two sets training and testing set. The testing set was not seen by neural network during the training phase. It is only used for testing the neural network after training.

In our classification experiments all 9 attributes are used, and 80% of the data is used for training, and the remaining 20% is used for testing.

### III. EXPERIMENTS AND RESULTS

Five experiments are carried out. In the first experiment the network is trained with one hidden layer; the number of neurons in this hidden layer is varies from 1 to 20. The network was trained 30 times for each structure.

Table (1) shows the ordered minimum mean of MSE for the 30 training trails for each number of neurons.

TABLE I. THE MINIMUM MEAN OF MSE IN THE 1ST EXPERIMENT

Number of neurons	MSE Mean of 30 trails
1	0.0805
9	0.0857
16	0.0858
17	0.0945
11	0.0958
3	0.0983
19	0.0985

From Table (1), the numbers of neurons that achieve best performance in terms of MSE mean are 1, 9, 16, 17, 11, 3 and 19.

In the second experiment, the network was trained with two hidden layers. In the first layer the numbers of neurons that achieve the best performance in the first experiment are used. For the second hidden layer the number of neurons is varied from 1 to 10. The network was trained 30 times for each possible pair of neurons. Table 2 shows the pair of number of neurons in the two layers that achieve the minimum MSE mean.

TABLE II. THE MINIMUM MEAN OF MSE IN THE 2ND EXPERIMENT

Mean of MSE	No. of Neurons in the 2nd hidden layer	No. of Neurons in first hidden layer
0.0894	1	17
0.0952	1	3
0.0943	1	11
0.0992	2	19
0.1020	2	16
0.1030	5	9
0.1435	2	1

Table (2), the pairs of numbers of neurons that achieve best performance in terms of MSE mean are (17, 1) (11, 1), (3, 1), (19, 2), (16, 2), (9, 5), (1, 2).

In the third experiment the network is trained with three hidden layers. The number of neurons in the first and second hidden layer is those pairs which give better performance in experiment two as listed above. The number of neurons in the third hidden layer is varied from 1 to 10. The network was trained 30 times for each possible triple set of neurons. Table (3) shows the number of neurons in the three layers that achieve the minimum MSE mean.

From Table (3), the three numbers of neuron numbers in the three layers that achieve best performance in terms of MSE mean are (1, 2, 1), (3, 1, 1), (9, 5, 2), (19, 2,4), (17, 1, 1), (16, 2, 5), and (11, 1, 1) sequentially.

In a similar manner, in experiments 4 and 5 the network is trained using four layers and five layers respectively. Tables (4) and (5) show the numbers of neurons in different layers that achieve the minimum MSE mean.

TABLE III. THE MINIMUM MEAN OF MSE IN THE 3RD EXPERIMENT

Mean of MSE	No. of Neurons in 3 <sup>rd</sup> Hidden Layer	Mean of MSE	No. of Neurons in 2 <sup>nd</sup> Hidden Layer	Mean of MSE	No. of Neurons in 1 <sup>st</sup> Hidden Layer
0.0810	1	0.1435	2	0.0805	1
0.1060	1	0.0951	1	0.0983	3
0.1115	2	0.103	5	0.0857	9
0.1188	4	0.0992	2	0.0985	19
0.1344	1	0.0894	1	0.0945	17
0.1366	5	0.102	2	0.0858	16
0.1835	1	0.0943	1	0.0958	11

TABLE IV. THE MINIMUM MEAN OF MSE IN THE 4TH EXPERIMENT

Mean of MSE	No. of Neurons in 4 <sup>th</sup> Hidden Layer r	Mean of MSE	No. of Neurons in 3 <sup>rd</sup> Hidden Layer	Mean of MSE	No. of Neurons in 2 <sup>nd</sup> Hidden Layer	Mean of MSE	No. of Neurons in 1 <sup>st</sup> Hidden Layer
0.0310	2	0.1188	4	0.0992	2	0.0985	19
0.0303	1	0.081	1	0.1435	2	0.0805	1
0.0468	1	0.106	1	0.0951	1	0.0983	3
0.0606	2	0.1366	5	0.102	2	0.0858	16
0.0660	2	0.1835	1	0.0943	1	0.0958	11
0.1183	4	0.1115	2	0.103	5	0.0857	9
0.1228	1	0.1344	1	0.0894	1	0.0945	17

TABLE V. THE MINIMUM MEAN OF MSE IN THE 5TH EXPERIMENT

Mean of MSE	No. of Neurons in 5 <sup>th</sup> Hidden Layer	Mean of MSE	No. of Neurons in 4 <sup>th</sup> Hidden Layer	Mean of MSE	No. of Neurons in 3 <sup>rd</sup> Hidden Layer	Mean of MSE	No. of Neurons in 2 <sup>nd</sup> Hidden Layer	Mean of MSE	No. of Neurons in 1 <sup>st</sup> Hidden Layer r
0.0300	2	0.0606	2	0.1366	5	0.1020	2	0.0858	16
0.0303	1	0.0303	1	0.081	1	0.1435	2	0.0805	1
0.0323	4	0.0468	1	0.106	1	0.0951	1	0.0983	3
0.0336	1	0.1183	4	0.1115	2	0.1030	5	0.0857	9
0.0821	2	0.0310	2	0.1188	4	0.0992	2	0.0985	19
0.1095	6	0.0660	2	0.1835	1	0.0943	1	0.0958	11
0.1231	4	0.1228	1	0.1344	1	0.0894	1	0.0945	17
<b>0.0623</b>		<b>0.0680</b>		<b>0.1245</b>		<b>0.1038</b>		<b>0.0913</b>	Average

#### IV. DISCUSSION AND CONCLUSIONS

From table (5) above, the last row shows the average performance of the best 7 number of neurons using different number of layers. This row shows that the increase of the number of hidden layers from 1 to 3 leads to a gradual lowering to the performance of classification, then the performance is increased when the number of layers is 4 and 5. Thus the number of hidden layers has a significant effect on the classification performance. And since the best performance and the best performance average is attained when the number

of layers is 5 this indicates that increasing in the number of hidden layers leads to better classification performance for Breast Cancer Data set.

All tables' show that the best performance is achieved when the number of neurons starting from layer two and up is small, typically 1 or 2.

So the final conclusion is that, to achieve better classification performance of Breast Cancer Dataset using a

FNN, with 5 hidden layers and the number of layer's neurons should be small, typically 1 or 2.

REFERENCES

- [1] Andina and Pham, "Computational Intelligence: For Engineering and Manufacturing", Springer, 2007, ISBN10 0-387-374507.
- [2] Martin T. Hagan, Howard B. Demuth, Mark H. Beale "Neural Network Design", ISBN: 0-9717321-0-8
- [3] Eric Kandel, James Schwartz, "Principles of Neural Science ", McGraw-Hill, fourth edition, 2000, ISBN 0-8385-7701-6
- [4] Karimi B, Menhaj M. B. and Saboori I, Multilayer feed forward neural networks for controlling decentralized large-scale non-affine nonlinear systems with guaranteed stability, International Journal of Innovative Computing, Information and Control, vol.6, no.11, pp.4825-4841, 2010.
- [5] Zare Nezhad B, and Aminian A. A multi-layer feed forward neural network model for accurate prediction of use gas sulfuric acid dew points in process industries, Applied Thermal Engineering, vol.30, no.6-7, pp.692-696, 2010
- [6] Tuba Kiyani, Tulay Yildirim, "Breast Cancer Diagnosis Using Statistical Neural Networks", Journal of Electrical and Electronics Engineering, 2004, vol. 4, Number 2, pp.1149-1153.
- [7] Sudhir D. Swarkar, Ashok Ghatol, Amol P. Pande, "Neural Network Aided Breast Cancer Detection and Diagnosis Using Support Vector Machine" Proceedings of the International conference on Neural Networks, Cavtat, Croatia, June 12-14, 2006, pp. 158-163.
- [8] Paulin F, A.Santhakumaran A, "Classification of Breast cancer by comparing Back Propagation training algorithms", International Journal on Computer Science and Engineering, ISSN: 0975-3397 Vol. 3 No. 1 Jan 2011 pp 327-332.
- [9] Bindu Garg, M.M. Sufian Beg, A.Q. Ansari, "Optimizing Number of Inputs to Classify Breast Cancer Using Artificial Neural Network" Journal of Computer Science & Systems Biology, JCSB, 2009, pp 247-254, ISSN: 0974-7230
- [10] M L Freedman and M L R Nierodzik, "Cancer and Age", 2007 by Elsevier Inc.
- [11] Jun Zhang MS, Haobo Ma Md MS, "An Implementation of Guildford Cytological Grading System to diagnose Breast Cancer Using Naïve Bayesian Classifier", MEDINFO 2004, M. Fieschi et al. (Eds), Amsterdam: IOS Press
- [12] ShuxiangXu, Ling Chen. "A Novel Approach for Determining the Optimal Number of Hidden Layer Neurons for FNN's and Its Application in Data Mining", 5th International Conference on Information Technology and Applications, ICITA2008 ISBN: 978-0-9803267-2-7, pp 683 - 686
- [13] Fujita O. Statistical estimation of the number of hidden units for feed neural networks, Neural networks, 1998, pp: 851-859.
- [14] Dario Baptista, Sandy Rodrigues, Fernando Morgado-Dias, "Performance comparison of ANN training algorithms for classification", Intelligent Signal Processing (WISP), 2013 IEEE.
- [15] Muhammad IbnIbrahimy, Md. Rezwanul Ahsan, Othman Omran Khalifa, "Design and Optimization of Levenberg-Marquardt based Neural Network Classifier for EMG Signals to Identify Hand Motions", Measurement Science Review, Volume 13, No. 3, 2013.
- [16] Reynaldi, A.; Lukas, S.; Margaretha, H., "Backpropagation and Levenberg-Marquardt Algorithm for raining Finite Element Neural Network", Computer Modeling and Simulation (EMS), Nov 2012 Sixth UK Sim, ISBN:978-1-4673-4977-2.
- [17] Rudy Setiono, Huan Liu, "Neural-Network Feature Selector" IEEE Transactions On Neural Networks, vol. 8, No. 3, May 1997, pg 664-662
- [18] Wlodzislaw Duch, Rafal Adamczak and Krzysztof Grabczewski, "A New methodology of Extraction, Optimization and Application of Crisp and Fuzzy Logic Rules" IEEE Transactions on Neural Networks, vol. 12, No. 2, March 2001, pp. 227-306.