USING ARTIFICIAL NEURAL NETWORK IN DIAGNOSIS OF THYROID DISEASE: A CASE STUDY

Farhad Soleimanian Gharehchopogh¹, Maryam Molany² and Freshte Dabaghchi Mokri³

¹Computer Engineering Department, Urmia Branch, Islamic Azad University, Iran
Bonab.farhad@gmail.com
²,³Computer Engineering Department, Science and Research Branch, Islamic Azad University, West Azerbaijan, Iran
molanymaryam@gmail.com
freshte.dabaghchi@gmail.com

ABSTRACT

Nowadays, one of the main issues to create challenges in medicine sciences by developing technology is the disease diagnosis with high accuracy. In the recent decades, Artificial Neural Networks (ANNs) are considered as the best solutions to achieve this goal and involve in widespread researches to diagnose the diseases. In this paper, we consider a Multi-layer Perceptron (MLP) ANN using back propagation learning algorithm to classify Thyroid disease. It consists of an input layer with 5 neurons, a hidden layer with 6 neurons and an output layer with just 1 neuron. The suitable selection of activation function and the number of neurons in the hidden layer and also the number of layers are achieved using test and error method. Our simulation results indicate that the performed optimization in MLP ANNs can be reached the accuracy level to 98.6%.

KEYWORDS

Artificial Neural Network, Back Propagation Learning Algorithm, Thyroid disease, Classification, Multi-Layer Perceptron.

1. INTRODUCTION

Nowadays, by developing technology and information in medical sciences, the computer science professionals are capable of providing expert systems to diagnose different kinds of diseases with high accuracy. The medical professionals are made to use these systems due to the some possible occurred errors during diagnosis process [1, 2]. Disease diagnosis operation using expert systems are performed based on a sets of disease symptoms [2, 3]. These systems are based on artificial intelligence which helps the physician to minimize the costs and time and expert in effective diagnoses [4]. Among these, ANN is a family of artificial intelligence, the researchers could reach to the big success using them in diagnosing diseases such as Diabetes, Heart disease, Thyroid and so on [5,6,7]. ANNs due to advantages such as self-learning, associative memory, high parallelism strength and high speed and error tolerance against noises which might be in parameters and also their cheapness in reuse of available solutions is the best option to do this [8]. The ANN called connection-oriented networks which include a set of processors act as parallel, take the sets of input in a time and produce output based on processing algorithm [9].
Feed forward ANNs which trained by back propagation learning algorithm is a model of ANNs [11, 12, 13]. These networks are classified to 3 groups which include MLP, Radical Bases function (RBF) and Possible ANN (PANN) [10]. Back propagation learning algorithm which applies under-supervision learning and become trained like a human being. It means that the network is training based on a set of previous data input and output. It creates such a relationship between these inputs and outputs in which the error be minimized (the difference between output and anticipation results) [2]. To achieve this goal, it is used decreasing Gradient method [5, 6, and 14]. In this paper, we used the dataset of UCI machine learning by using MLP ANN to classify Thyroid disease. Due to learning capability, better performance in classification issues as efficient and decision making based on diagnoses features (like human intelligence), MLP ANNs are the best system to do this task [15,16].

In this paper, it is used MATLAB 2012.a software due to the flexibility, many prepared functions and proper performance to ANN simulation. This software also causes the accuracy of results and comparing the performance of the networks become increased and it is achieved the best performance to learning network in the shortest possible time. In the section 2, we begin to discuss and about several papers to diagnose disease using ANN. In section 3, we design the general process diagram. In (3.1) section, it is reviewed the selected architecture to study. In (3.2) section, dataset collecting and effective parameters in Thyroid disease diagnoses is discussed and described. The proper selection of Train and Test data in (3.3) section, normalizing input data in (3.4) section and training and learning of ANNs are discussed in (3.5). Section (4) allocates to review and assess simulation results of network and finally conclusion is provided in section (5).

2. PREVIOUS WORKS

By studying literature, it seems that most researchers have been used widely ANN techniques to diagnose Thyroid disease. Most used techniques are under-supervision techniques in this feature in which the researchers increase the speed and accuracy of network by applying network architecture, proper Initialized of weights and choosing proper activating function due to the type of their diagnoses. For example, Dey et al [6] has been used ANN techniques to diagnose Diabetes disease. The applied data in this paper are taken from Manipal Sikkim Institution of Medical Science Hospital which includes 530 patients. The output includes 2 classes of 0 and 1. They suggest two feed forward architectures in which the first one includes the number of neurons in three layers as (6-10-1) and the second involves two hidden layers and the number of neurons in (6-14-14-1) layers. They indicate that log sigmoid activating function in the hidden layer neuron is considerably increase the speed and performance of networks. So, if the number of the layers increases, the abstract error will be increased. The accuracy of performance in this network for the architecture with a single hidden layer is 92.5%. In section [2], Kabari et al provide a framework to diagnose skin disease using MLP ANN and can reach to an acceptable level of accuracy. In this paper, researchers have been used 4 neurons in input layer, 3 neurons in hidden layer and 2 neurons in output layer, respectively to reach the accuracy to 90%.

In [5], F.S.Gharechopogh et al. have been used ANN to diagnose heart disease. Their architecture includes MLP ANN with 60 nodes in input layer, 4 nodes in hidden layer and 2 nodes in output layer. It is back propagation learning algorithm for network learning. The value of assigned parameters for rate learning and momentum are considered 0.2 and 0.3 respectively. The achieved accuracy of performance for training and test set is 0.95% and 0.85%, respectively. In [11], Shukla et al could achieve to the acceptable results using ANN techniques to diagnose Thyroid disease. They use 3 architectures of back propagation learning algorithm, RBF network and Learning Vector Quantization Networks. The number of layers for back propagation learning algorithm is considered with a single hidden layer and the number of neuron in hidden layer is 46
and learning rate is equal to 0.8. RBF network has a single hidden layer, learning rate of 0.9 and the number of neuron in hidden layer is 100. Comparing the performance of these three architecture networks indicate that LVQ network has the best accuracy rate equal to 0.98%. Because LVQ like back propagation learning algorithm doesn’t fall into the local minimum trap as well as RBFN network which doesn’t require full cover of input space. But, among three applied ANNs in this paper, RBFN network has the least learning time. In [16], Isa et al have been used ANN to diagnose Thyroid disease. By selecting the proper activating function, they could improve the performance of ANN. The given data related to the data set of UCI site is used 215 data, MLP architecture, sigmoid activation function and a hidden layer with 7 neurons and reached to the accuracy rate of 97.6%. They also indicate that hyperbolic tangent function of MLP ANN is suitable to classify data to two classes and neural function for three classes.

Ozyilmaz et al in [17] begin to classify Thyroid disease using three architectures of MLP, RBF and CSFNN. The accuracy average of UCI data set based on MLP architecture is 88.3%; RBF with two hidden layer and full connection is 81.69% and CSFNN architecture (combination of MLP and RBF models) is equal to 85.915%, respectively. They proved that CSFNN architecture has the best classification rate with the least learning time for this data collection in compared with the others. RBF ANN didn’t act well in this feature and indicated that if MLP provides good results averagely, but it won’t guarantee the same performance for the others. It depends to the weight initialization. Researchers in [7] begin to compare the performance of several neural system architectures (LVQ, PANN, RBF and SVM) for dataset of thyroid disease. The used data set in this paper includes 5 classes (hypothyroidism, hyperthyroidism, normal, sub clinical hypothyroidism and sub clinical hyperthyroidism).

In RBF architecture, we reach to the accuracy average of 97.3%, for GRNN architecture to 97.0%, for PANN to 97.16% and for LVQ to 92.7%, respectively. They showed that PANN and RBF networks have better performance dataset than SVM, LVQ and GRNN networks. We discuss about Thyroid disease diagnoses using MLP ANN. We use 5 neurons of input layer, 6 neurons of hidden layer and one neuron of output layer. Comparing the obtained results with the above literature, we can increase the accuracy to 98.6% which is considerable.

3. PROPOSED METHOD

The work process in diagnosing the disease includes three basic stages. The first stage is the data collecting and classifying. The second include architecture selection and ANN learning and the third stage is to compare network performance and reaching to the best answer as indicated in Fig 1. The full descriptions of the flowchart are investigated in 3.1 and 3.5 sections.

3.1. MLP ANN

MLP ANNs architecture is shown in fig 2. The most popular architecture has three layer includes an input layer, a hidden layer and an output layer and all the connections are full in this architecture.
It means that the output of each neuron is linked and connected to the all neurons of the next layer. The first layer consists of a set of neurons which enters the input data to the network and
called input layer [18]. The second layer known as hidden layer can be consisted of one or a few layers which varied based on data under training process [16].

\[ Y = \text{purelin} \left( \sum_{i=1}^{6} \left( \text{logsig} \left( \sum_{l=1}^{3} W_{l,i} x_l + b_1 \right) \right) W_{2,i} + b_2 \right) \]

The final layer called output layer and indicates the output resulted in network. The calculation method of network general output is shown in Equation 1. The functional base of this network is neuron which is considered as a processing unit. A single neuron can't be used to solve problems due to its limited capabilities. But, as a set of neurons are connected according to a certain topology to provide a complicated system, the possibility to create a meaningful solution for network learning is also provided [31].

![Fig2- MLP ANN architecture](image)

### 3.2. Data Set Description

The Thyroid gland is one of the main glands of human body which looks like a butterfly and locates in front of windpipe. It absorbs the available iodine in food and produces T_3 and T_4 hormones (Triiodothyronine, tetraiodothyronine and thyroxine) [19]. It is the task of controlling and how metabolism, produce protein and body sensitivity to hormones. These hormones regulates metabolism base of the body and affects the growth and function of other systems in the body [1]. The output hormones of Thyroid gland are provided by Thyroid Stimulating Hormone.
International Journal on Computational Sciences & Applications (IJCSA) Vol.3, No.4, August 2013

(TSH) which pituitary produces. The Thyroid gland itself is regulated by Thyrotropin Releasing Hormone (TRH) which is produced by the hypothalamus. It can be noted to the most popular problems of Thyroid gland such as overactive thyroid gland known as hyperthyroidism and under activity gland as hypothyroidism [19]. The type of disease based on its parameters is classified in Table (1).

Table 1- classifying thyroid disease based on parameters

<table>
<thead>
<tr>
<th>Type of thyroid disease</th>
<th>T3</th>
<th>T4</th>
<th>TSH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypo</td>
<td>Increase</td>
<td>Increase</td>
<td>Reduction</td>
</tr>
<tr>
<td>Hyper</td>
<td>Reduction</td>
<td>Reduction</td>
<td>Increase</td>
</tr>
</tbody>
</table>

The applied data in this paper is related to UCI site which collected by James Cook University, Townsville of Australia, in 1992. In this site, thyroid disease classifies to 3 groups of hypothyroidism, hyperthyroidism and normal. The total numbers of laboratory samples are 215. It is also used 5 laboratory tests to distinct healthy and patient persons [20]. These parameters include as follow:

1. Class attribute (1 = normal, 2 = hyper, 3 = hypo)
2. T3-resin uptake test. (A percentage)
3. Total Serum thyroxin as measured by the isotopic displacement method.
4. Total serum triiodothyronine as measured by radioimmuno assay.
5. Basal TSH as measured by radioimmuno assay.
6. Maximal absolute difference of TSH value after injection of 200 micro grams of thyrotropin-releasing hormone as compared to the basal value.

All features are connected and there is no loosed one. In Table 2, the numbers of laboratory samples are determined based on the type of disease.

<table>
<thead>
<tr>
<th>Type disease</th>
<th>Number of samples</th>
<th>Type class of the UCI classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>150</td>
<td>1</td>
</tr>
<tr>
<td>Hyper</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>Hypo</td>
<td>35</td>
<td>3</td>
</tr>
<tr>
<td>Total:</td>
<td>215</td>
<td></td>
</tr>
</tbody>
</table>

Table 2- the numbers of laboratory samples based on the type of disease

3.3. Test and Learning Data Selection

The used data to enter the network are taken from data set of UCI machine learning which includes 215 samples and each one has 5 effective parameters to classify the type of disease. The given data is saved as 215*6 arrays. The first 5 columns relate to the input and the last one to the network desired output. To begin network training process, it is necessary to determine data set classification to 2 groups of test and training. The method of classification affects the performance of network. To reach a suitable classification and generally increasing network
performance, it is used trial and error method. Using different classifications, comparing network performance and choosing best classification and finally following training process is used by this classification. In Table 3, the obtained results of different classifications are presented.

<table>
<thead>
<tr>
<th>Train</th>
<th>Test</th>
<th>Performance for Tansig</th>
<th>Performance for Logsig</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>30</td>
<td>0.00179</td>
<td>9.19e-9</td>
</tr>
<tr>
<td>75</td>
<td>25</td>
<td>0.032</td>
<td>1.33e-5</td>
</tr>
<tr>
<td>80</td>
<td>20</td>
<td>0.0085</td>
<td>0.056</td>
</tr>
<tr>
<td>85</td>
<td>15</td>
<td>0.90e-10</td>
<td>5.94e-10</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>2.6e-7</td>
<td>0.046</td>
</tr>
<tr>
<td>95</td>
<td>5</td>
<td>5.26e-5</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

Due to the obtained results of Table 3, the best performance relates to the classification is 85% for Train data and 15% for Test data, respectively. The way of choosing data based on this classification is as random (using Randperm function).

3.4. Input Data Normalization

The goal of data pre-processing is to increase network performance. By doing so, it transforms the input to the form which becomes suitable to use in network [21]. The normalization process of input data to optimal training affects the network which considered as input pre-processing functions [22]. The researchers indicate that using different normalization methods in back propagation learning algorithm ANNs under training process will increase the training capabilities. Without it (normalization process), network learning will perform slowly [21]. The various types of normalization methods [21] are as follow:

1. Statistical or Z-Score Normalization
2. Min-Max Normalization
3. Median Normalization
4. Sigmoid Normalization
5. Statistical Normalization

Proper selection of normalization method increases the network performance [22]. Before ANN training, we normalize each data base input parameters. It is used min-Max normalization method in this paper and the calculation method is provided in equation 2. This method performs rescaling as the features or outputs are transmitted from a range of values to the new one. Often the ranges of values are mostly between 0 to 1 or -1 to 1 [21].

\[ x' = \left( x_{\text{min}} \right) + \left( x_{\text{max}} - x_{\text{min}} \right) \frac{(x_i - x_{\text{min}})}{(x_{\text{max}} - x_{\text{min}})} \]  \hspace{1cm} (2)

3.5. ANN training and learning

The techniques used to training an ANN are wide in which the most important is supervision or non-supervision learning techniques. Back propagation learning algorithm which uses under-supervision learning technique could learn well based on a set of training examples and then capable of diagnosing each pattern [32]. Basic MLP constructive unit is a simple model of artificial neuron. This unit adds the inputs collection with the value of threshold. And the sum depending on type activation function (generally sigmoid) passes [24]. The network begins to
train based on the random values of weights and biases [10, 25] and the training process will continue until the error is minimized [2, 10, 25]. Updating the values of all weights is performed based on error value as far as the difference between network input and output reaches to the desirable one. It gradually achieves from output layer, through hidden layer to input layer toward back [6]. Beside weights optimal adjustment, proper selection of activation function According to [16] Cause more optimal network performance. An activating function of MLP ANN has several major features such as continuous, differentiable, and monotonically non-decreasing [10, 26, and 16]. The applied activation function in this paper is considered for hidden layer neurons is Logsig and for each output layer neuron is Purelin.

4. EVALUATION AND DISCUSES

The main goal of this paper is to compare the performance of MLP ANN with the changes of activation function and the number of hidden nodes and to reach to the high accuracy. The back propagation learning algorithm performance is based on Gradient descent technique. It is considered to regulate weight connections among neurons to minimize system error between real output and target output [33]. Although, back propagation learning algorithm is the most popular algorithm to ANN training but sometimes can be inefficient. One of the main training drawback whit the algorithm is the slow convergence. It is proposed methods to improve convergence rate which includes proper selection of activating function in neurons and accurate determination of size parameter of learning rate [16].

Table 4- selecting the most appropriate numbers of hidden layer nodes

<table>
<thead>
<tr>
<th>Number Of Hidden Neurons</th>
<th>Performance for Tansig</th>
<th>Performance for Logsig</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.026</td>
<td>0.117</td>
</tr>
<tr>
<td>3</td>
<td>8.42 e-05</td>
<td>0.128</td>
</tr>
<tr>
<td>4</td>
<td>0.12</td>
<td>1.17 e-07</td>
</tr>
<tr>
<td>5</td>
<td>4.54 e-06</td>
<td>0.00011</td>
</tr>
<tr>
<td>6</td>
<td>4.19 e-08</td>
<td>1.45 e-10</td>
</tr>
<tr>
<td>7</td>
<td>0.0078</td>
<td>0.031</td>
</tr>
<tr>
<td>8</td>
<td>2.17 e-06</td>
<td>2.51 e-10</td>
</tr>
<tr>
<td>9</td>
<td>0.036</td>
<td>0.00014</td>
</tr>
<tr>
<td>10</td>
<td>1.99 e-05</td>
<td>0.0019</td>
</tr>
</tbody>
</table>

The numbers of input and output nodes are determined at the beginning of network learning process. The numbers of hidden nodes are achieved through trial and error method. As based on [28,29] and widespread experiments which indicate that single-layer ANN performance are better than networks that their hidden layer has more than a single layer and involves a level of complication. The reason for this difference is that second is more prone to falling into the weakness of local minimum. The obtained results from table 4 indicate that 6 neurons of ANN for hidden layer reach to the better performance. In this paper, we are used the architectures of (5, 6, 1) with a single hidden layer. The diagram of the given results indicates it obviously.
After determining the numbers of nodes, we begin to select appropriate activation function for network. Also it is achieved based on trial and error method and comparing the network performance. It is considered different activation functions for ANNs which resulted in better performance. In general, there are two linear and non-linear activation functions. The activation function in the hidden nodes of network are non-liner and usually all neurons of a layer have same activation function. The most popular applied activation function in hidden nodes is Sigmoid [34]. But according to the obtained performance in tables 3 and 4, it is indicated that Logsig activating function will reach us to the best answer. This function increases learning speed of back propagation learning algorithm or convergence in MLP ANNs [30]. In general, the obtained results of simulation are provided in table 5.

Table 5- The obtained results of network simulation

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>2.28 e-08</td>
</tr>
<tr>
<td>Gradient</td>
<td>5.63 e-05</td>
</tr>
<tr>
<td>Mu</td>
<td>1.00 e-11</td>
</tr>
<tr>
<td>Validation</td>
<td>6</td>
</tr>
<tr>
<td>Iterations</td>
<td>19</td>
</tr>
</tbody>
</table>

The performance is measured by mean square error (mse) which calculated according to Equation 3.

\[
\text{mse} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]  

In Fig 4, the performance of improved network during training is indicated. The performance is indicated for each of training, testing and validation set.
To stable validation of results (network performance), the correlation coefficient of R is calculated. Whatever the value of R becomes closer to 1 (along 45°), the performance will be good and trustful [6]. Figures 5 and 6 indicate the correlation diagram.

Fig 4- the performance of improved network during training

Correlation diagram for test and train data. For train and Test data the R value is equal to 0.99697 and 0.98297, respectively.

Fig 5- correlation diagram of train data
5. CONCLUSION

The importance of using ANNs to diagnose disease is to increase the accuracy of performance. The appropriate selection of ANN architecture affects the network performance effectively to reach the high accuracy. In this paper, we consider the type of appropriate activation function, correct selection of layer number and the network complexity so as to achieve the best result by comparing their performance to reach the best possible answer. By selecting a hidden layer and Logsig activation function for hidden layer and 6 neurons in the hidden layer, we can reach the classification accuracy for Thyroid disease to 98.6%. The proposed method in this paper can be a solution to increase the performance of ANN. So, it can be generalized to the other disease diagnoses systems of ANN.

REFERENCES


AUTHORS

Farhad Soleimanian Gharechopog is currently Ph.D. candidate in department of computer engineering at Hacettepe University, Ankara, Turkey. And he works an honour lecture in computer engineering department, Science and Research and Urmia branches. Islamic Azad University, West Azerbaijan, Iran. He is a member of editorial board and review board in many international journals and international Conferences. His interested research areas are in the Operating Systems, Software Cost Estimation, Data Mining and Machine Learning techniques and Natural Language Processing. For more information please visit www.soleimanian.net

Maryam Molany is a M.Sc. student in Computer Engineering Department, Science and Research Branch, Islamic Azad University, West Azerbaijan, Iran. Her interested research areas are Biomedical, Artificial Neural Network, Data Mining and Machine learning Techniques

Freshte Dabaghchi Mokri is a M.Sc. student in Computer Engineering Department, Science and Research Branch, Islamic Azad University, West Azerbaijan, Iran. Her interested research areas are Biomedical, Artificial Neural Network Meta Heuristic Algorithms, and Machine learning Techniques.