Comparing The Performance Of MLP With One Hidden Layer And MLP With Two Hidden Layers On Mammography Mass Dataset

Venu Azad

Department of Computer Science, Govt. girls P.G. College
Sec 14, Gurgaon, Haryana, India

Abstract: Nowadays soft computing techniques such as fuzzy logic, artificial neural network and neuro-fuzzy networks are widely used for the diagnosis of various diseases at different levels. In this paper, a multilayer perceptron neural network classifier is introduced to classify the mammography mass data set into two classes benign and malignant on the basis of mammography mass data set attributes. The performance of the MLP neural network in two different configurations is measured. In first configuration one hidden layer is used and in second two hidden layers are used. A four–fold cross validation method is used for the assessment of generalization of the system. The result shows that the proposed MLP with two hidden layer achieve the accuracy of 89% approx., proving its usefulness in classification of mammography masses.

Keywords: Mammography, Multilayer perceptron network, Cascade learning, four - fold cross validation.

1. INTRODUCTION

Breast cancer is the most common type of cancer that found among the females, so the detection of this disease at an early stage is very necessary. There are various ways of detecting the breast cancer which includes physical examination by a physician, microscopy or biopsy (FNA biopsy, core needle biopsy etc.), mammogram digital images. Here we use a mammography mass data set to classify the breast cancer into benign and malignant classes. A mammogram is a kind of X-ray from which a physician reads different attributes like shape, mass density, mass margin etc., these attributes here are used to classify the mammography mass data set. In this system we use mammography data set for classification of breast cancer.

A lot of research has been done by researcher to diagnose breast cancer. In [1] statistical neural network like RBF(radial basis function), GRNN(general regression neural network) PNN(Probabilistic neural network) and MLP(multilayer perceptron) are used to classify WCBD(Wisconsin breast cancer data) data set the overall accuracy of these systems > 96%. In [2] a self-organizing map neural network is used to examine breast sonography tumor data set with accuracy 85.6 %. In [3] Digital mammogram is used to extract the features and an auto associator neural network is used to classify these extracted features and achieve the accuracy of 94%. In [4] Biopsy images are used to examine the breast cancer, these images are first preprocessed using image processing techniques such as adaptive thresholding based segmentation and watershed segmentation method then a classification algorithm based on feed –forward neural network is used to classify each cancer object into four object types. In [5] SVM along with its variant like L1- SVM, L2-SVM and μ-SVM as well as combinations of SOM-RBF is also used to improve the classification accuracy of malignant tumor in WBCD data set. In [6] wavelet analysis and hybrid network i.e. fuzzy-neural network is used to classify mammography image taken from MIAS data set. In [7] a soft cluster neural network is proposed for classification of digital mammograms, the concept of soft cluster is introduced as a pattern may fall in more than one group. The digital mammograms are first preprocessed to extract the features than soft cluster neural networks are used to classify. In[8] a single layer multilayer perceptron is used for the classification of mammographic masses and in [9] the performance of single layer MLP is compared with the FNN(fuzzy neural network) on mammography mass dataset. In [10] a constructive algorithm that creates compact neural network architecture was developed to classify the early breast cancer in patients.

In the proposed system we utilize mammography mass data set taken from the UCI learning repository for classification of masses and implemented an MLP neural network because Multilayer perceptron is one of the most popular neural networks due to its simple and clear architecture; it uses a simple back propagation algorithm to train the network. In this paper, we implemented this simple MLP network to support the classification of mammography masses. For classification of masses an MLP network is developed which consist of three layers i)one input layer ii) one hidden layer , and iii) one output layer the input layer of the system consists of 5 variables each representing a mammography mass attribute. In this paper, MLP with one hidden layer and MLP with two hidden layers are used to classify the masses and their performance is compared. The number of nodes in the hidden layer is determined using a cascade learning algorithm [11] .The output layer consists of one node which gives an output as 0 or 1 according to the pattern.
presented to it. We have applied a cross validation method to assess the generalization of the system.

In this paper we first study the architecture of MLP network, in section 3 we discuss the architecture of the proposed system. In section 4, the MLP neural network with two hidden layer is introduced. In section 5 the performance evaluation and experimental result analysis are summarized. Finally in section 6 conclusions about the proposed system is summarized.

2. THE MULTILAYER PERCEPTRON ARCHITECTURE

In this section we study basic concept about the architecture of MLP network. An MLP is a neural network and a neural network can be defined as an artificial neural network consists of a large number of interconnected processing elements known as neurons that act as a microprocessor.

![Architecture of multilayer perceptron network](image)

**Figure 1.** Architecture of multilayer perceptron network

NN are in consideration due to its self-adaptation, robustness, and performs the nonlinear mapping between the input feature and the desired output [12].

A multilayer perceptron is a mathematical model for classification of non-linear data into different classes. It is the most popular and frequently used neural network architecture [13] - [15].

The MLP is feed-forward network architecture consists of two layers with one or more than one hidden layers; the layer is named as the input layer, hidden layer, the output layer.

The hidden layer and output layer is the processing layer unlike the input layer. The input is presented to input layer, the weighted sum of input and the presence bias is calculated and it serves as the input to the hidden layer neurons transformation function, at hidden layer a transformation function is used to map the weighted sum input to intermediate output. This intermediate output act as input to next hidden layer or to the output layer, again a transformation function is used in output layer to calculate the final output.

Each node in MLP is a processing element which performs following function i) Compute the weighted sum of the input along with the present bias ii) process this weighted sum of input using an activation function to compute the output generated by that neuron

\[ V_j = \sum_{i=1}^{n} W_{ji} X_i + \theta_j \]  \hspace{1cm} (1)

\[ Y_j = f(V) \]  \hspace{1cm} (2)

Where \( V_j \) is the weighted sum of inputs \( x_1, x_2, x_3, \ldots, x_p \) and bias \( \theta_j \) for jth neuron , \( W_{ji} \) is the connection weight between input \( x_i \) and neuron \( j \), and \( f(\cdot) \) is the activation function of the jth neuron, and \( Y_j \) is the output of the jth neuron

3. PROPOSED SYSTEM ARCHITECTURE

In this section we describe the details of the system developed, the MLP used here for classification consist of three layers including input layer, a hidden layer and an output layer

3.1 Input layer

In this system data set under study is a mammography mass data set which is taken from the UCI learning repository. The data set consist of six attributes i) BI-RADS assessment ii) Age iii) Mass shape iv) Mass margin v) Mass density and vi) The severity (i.e class benign or malignant). The data set contains total 961 instances in which 516 instances belong to a benign class whereas 445 instances belong to malignant class. The data set contains the missing values such as BI-RADS assessment attributes contain 2 missing values, age contain 5, Shape contain 31, Margin contain 48, and density contain 76 missing values.

Before feeding the input into the neural network we have to process these missing values this can be done by either removing the records with missing values or filling the missing values. There are different methods for filling the missing values, we use a substitution mean method [16] in which the missing value is replaced by the mean or average of the other observed values of the attributes.

In the input layer of the system there are five nodes each corresponds to an attribute of mammography mass data.

3.2 Hidden layer

In the proposed system there is only one hidden layer with 16 nodes. The number of nodes in the hidden layer plays a significant role in network’s ability to classify the input.

So, we have to carefully choose the number of hidden layer neuron for the system here we used a cascade learning algorithm [11] to find the number of nodes(neuron) in the hidden layer. In the cascade learning algorithm [11] consider two parameter accuracy and convergence speed which they want to optimize but we
concentrated only on the accuracy of the network to classify the data. The algorithm is as follow:

**Step 1:** Initialize the number of neurons in hidden layer with small value hidden layer neuron $i=5$ for current network

**Step 2:** for $i=1:20$
- Create a multilayer perceptron network with $i$ number of neurons
- Train the network with current configuration (given in table 1)
- Test the network with test data set and compute the average accuracy.
- Increment the hidden layer neuron by 1.

**Step 3:** End

In figure 2 we can see the result of accuracy rate with respect to the number of hidden layer neuron.

Here we can see that network achieve a high accuracy rate when the number of hidden layer neuron is 16 and the highest accuracy is 87.91% . As a result we choose 16 neurons as hidden layer neuron.

### 3.3 Output layer

The output layer consists of one node whose output is used to determine whether the input pattern presented to network belong to benign class or malignant class. Here output 0 represents class benign and output 1 represent class a malignant.

### 3.4 Summary about the system

The classifier designed for the classification of a mammography mass data set is a MLP neural network with 5 input variables, 16 hidden layer nodes and one output node . The training parameter used in the system is given in table 1.

![Figure 2](image)

**Figure 2** Accuracy rate vs. number of hidden layer neuron

![Table 1](image)

**Table 1** the training parameters

<table>
<thead>
<tr>
<th>The training parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of epochs</td>
<td>5000</td>
</tr>
<tr>
<td>Preset learning rate</td>
<td>0.07</td>
</tr>
<tr>
<td>Error precision target</td>
<td>0.6</td>
</tr>
<tr>
<td>Transfer function</td>
<td>Logsig</td>
</tr>
</tbody>
</table>

The other training parameter values are the default values such as performance function is mse (mean square error) transfer function used in hidden layer and output layer is logsig, with training function TRIANLM.

### 4. MLP WITH TWO HIDDEN LAYERS

Generally, a single hidden layer is sufficient to simulate the problem using MLP neural network. Two hidden layer are used where we have data with discontinuity such as saw tooth wave form [17]. Here we have data with discontinuous values so we decided to introduce two hidden layers in the MLP architecture. The number of neurons in each hidden layer is determined by using an iterative algorithm. The algorithm is as follows:

**Step 1:** Initialize the number of nodes in hidden layer $l_1=i$ and number of nodes in hidden layer $l_2=j$

**Step 2:** For hidden layer $l_2=j$ repeat the following step
- Train with hidden layer $l_1$ neuron $i$;
- Measure the accuracy of the network
- Increment $i$ as $i+1$;
- Again Train the network with hidden layer neuron $=i$
- Measure the accuracy
- Increment the layer 2 neuron by 1 i.e $j=j+1$;

**Step 3:** End

Where $i=5$ and $j=5$

![Figure 3](image)

**Figure 3** Architecture of MLP with two hidden layer

We can see in figure 4, that network achieve the highest accuracy of 83.33 % with neuron in hidden layer $l_1=6$ and in hidden layer $l_2=5$. The other network parameters are given in table 1.

If we train the network with more number of neuron in hidden layer 1 and hidden layer 2 the time to train the network will increases and we can see from the graph that as the number of hidden layer neuron increases the performance start degrading so we stop at hidden layer neuron 12 and hidden layer 2 neuron at 11.

### 5. PERFORMANCE EVALUATION AND RESULT ANALYSIS
In the present system we have used the measure of specificity, sensitivity and accuracy for the performance evaluation of the system constructed [18] as follow.

\[ \text{Specificity} = \frac{TN}{TN + FP} \]  
\[ \text{Sensitivity} = \frac{TP}{TP + FN} \]  
\[ \text{Accuracy} = \frac{\text{Total no. of correct classification}}{\text{Total no. of instances}} \]

The above mentioned performance metrics are calculated as follows

1. Specificity: it is the percentage of healthy people classified correctly.

2. Sensitivity: It is the percentage of the abnormal patient (patient who is malignant i.e. suffering from the breast cancer) classified correctly.

3. Accuracy: It is the percentage of the correct classification.

The confusion matrix is constructed by using the average values of four experiments carried out on different set of data. The accuracy rate is calculated for each set of experiment and average accuracy is calculated which is 87.91%. The accuracy for each set of experiment is noted down in table 2.

Table 2. Confusion matrix for two configuration of MLP network a) MLP with one hidden layer b) MLP with two hidden layer

<table>
<thead>
<tr>
<th>Class</th>
<th>Malignant</th>
<th>Benign</th>
<th>Row sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malignant</td>
<td>94</td>
<td>12</td>
<td>106</td>
</tr>
<tr>
<td>Benign</td>
<td>17</td>
<td>117</td>
<td>134</td>
</tr>
<tr>
<td>Column sum</td>
<td>111</td>
<td>129</td>
<td>240</td>
</tr>
</tbody>
</table>

a. MLP with one hidden layer

<table>
<thead>
<tr>
<th>Class</th>
<th>Malignant</th>
<th>Benign</th>
<th>Row sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malignant</td>
<td>95</td>
<td>12</td>
<td>107</td>
</tr>
<tr>
<td>Benign</td>
<td>15</td>
<td>118</td>
<td>133</td>
</tr>
<tr>
<td>Column sum</td>
<td>110</td>
<td>130</td>
<td>240</td>
</tr>
</tbody>
</table>

b. MLP with two hidden layer

<table>
<thead>
<tr>
<th>Class</th>
<th>Malignant</th>
<th>Benign</th>
<th>Row sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malignant</td>
<td>95</td>
<td>12</td>
<td>107</td>
</tr>
<tr>
<td>Benign</td>
<td>15</td>
<td>118</td>
<td>133</td>
</tr>
<tr>
<td>Column sum</td>
<td>110</td>
<td>130</td>
<td>240</td>
</tr>
</tbody>
</table>

Table 3 Accuracy measure for four experiments

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP(1 hidden layer)</td>
<td>84.68</td>
<td>90.69</td>
</tr>
<tr>
<td>MLP (2 hidden layer)</td>
<td>86.36</td>
<td>90.76</td>
</tr>
</tbody>
</table>

The sensitivity, specificity and accuracy are calculated which are summarized in table 3 and graphically represented in figure 5.

6. CONCLUSION AND FUTURE SCOPE

In this paper, a classifier is developed using an MLP neural network for the classification of mammographic masses in breast cancer. In this study mammography mass data set is obtained from the UCI learning repository with 961 cases out of which 516 cases belong to benign class and 445 cases belong to malignant classes.
MLP with one hidden layer with 16 neurons is trained using back propagation algorithm and a second configuration in which we uses two hidden layer neuron with number of neurons 6, 5 respectively. The result obtained from experiment shows that the network achieves high accuracy of 87.91% in first configuration and 88.75% in second configuration .The results shows that by increasing the number of hidden layers in the architecture improves the performance of the system but the complexity of the system also increases.

References

AUTHOR
Ms Venu Azad is currently working as an Extension Lecturer (Computer Science Department) in Govt. girls college Gurgaon Sector 14, Gurgaon. She has completed her B.tech in computer science from GCEW College Gurgaon and M.tech in computer science from ITM University Gurgaon. She has 3 Research Papers in different Conference including IEEE Explorer.