BREAST CANCER CLASSIFICATION USING NEURAL NETWORK
APPROACH: MLP AND RBF


*Faculty of Arts and Sciences, Islamic University of Lebanon, Khalde Highway Lebanon, BP 30014.
Ali.raad@iul.edu.lb

**Faculty of Economic and Business, Lebanese University, Hadath, Lebanon.
Akalakech@ul.edu.lb

***Faculty of Engineering, Department of Biomedical, Islamic University of Lebanon, Khalde Highway Lebanon, BP 30014.
Mohammad.ayache@iul.edu.lb

ABSTRACT
The classification of breast cancer is a medical application that poses a great challenge for researchers
and scientists. The use of learning machine and artificial intelligence techniques has revolutionized the
process of diagnosis and prognosis of the breast cancer. The aim of our study is to propose an approach for
breast cancer distinguishing between different classes of breast cancer. This approach is based on the
Wisconsin Diagnostic and Prognostic Breast Cancer datasets for feature selection, and the classification of
different types of breast cancer using neural network approach, and especially the multi layer perceptron
MLP and the radial basis function RBF. The data set consists of nine features that represent the input layer to
the neural network. The neural network will classify the input features into two classes of cancer type (benign
and malignant). The proposed approach tested on the database, resulted in 97 % success rate of
classification using RBF neural network. Neural network approach and especially the RBF technique
seems an efficient method for classification in medical applications and especially for the breast cancer
classification.

Keywords: NEURAL NETWORK, MLP, RBF, BREAST CANCER, CLASSIFICATION.

1. INTRODUCTION
Breast cancer, is cancer that affects today more women
in the world. Thus, the fight against cancer is far from
completed. Medicine advances on all fronts to improve
the care of patients and defeat this disease of the
century. Because of this, it is essential that several
disciplines continue to make their contribution and
particularly data mining or artificial Intelligence.
To provide assistance to the medical, robust and reliable
diagnosis, neural networks can be a powerful tool for
distributed diagnosis [3].
In this paper, we tested the performance of the neural
networks based on the Wisconsin Breast Cancer
Database (WBCD). The problem of breast cancer
detection led researchers and experts in this field to
focus on other trends, such that new technologies other
human to address this social problem. The objective of
our work is to create a new approach that allows
whether a patient has a benign cancer or malignant
following several descriptors. To achieve this, we
propose a solution based on the concept of neural
networks [7] [8] [9].
Recently, the neural network has become a popular tool
in the classification of Cancer Dataset [1] [2] [4] [5].
This is particularly due to its ability to represent the
behavior of linear or nonlinear functions multidimensional and complex.
In this work, a parallel approach, which uses neural
network technique, is proposed to help in the diagnosis
of breast cancer. The neural network is trained with
breast cancer data by using feed forward neural network
model and backpropagation learning algorithm with
momentum and variable learning rate. The performance
of the network is evaluated.
The work of this paper will be presented in different
sections. In the second section, we introduce a general
overview about breast cancer anatomy and different
types of cancers to be classified. The representation of
data base used in our work will be in the section three
where we present the preprocessing tool of the database
before using it as inputs to the neural network.
Adaptation of parameters to the neural network and the
results obtained will be presented in the fourth section.
And finally, we present a conclusion that talk about the
performance of the method used in the purpose of breast
cancer classification.

2. BREAST CANCER: AN OVERVIEW
2.1 BREAST ANATOMY
The anatomy of the breast is quite complex. Figure 1
shows the most important structures of the breast. To
give an understanding of where and how different breast
tumors may develop, we will shortly describe the
structure of the breast. Each breast contains between 15
and 25 lobes that are connected to the nipple [D]
through converging ducts [A]. Each lobe is made up of
many smaller lobules [B]. Each lobule consists of 10 to
100 Terminal Duct Lobular Units (TDLU) where milk is produced. The most common area where breast cancer originates is in the TDLU [13].

![Figure 1: Important Structures of the Breast](image)

**2.2 BREAST CANCER TYPES**

We can distinguish three types of breast tumors: benign breast tumors, in situ cancers, and invasive cancers. The majority of breast tumors detected by mammography are benign. They are noncancerous growths and cannot spread outside of the breast to other organs. In some cases it is difficult to distinguish certain benign masses from malignant lesions with mammography.

If the malignant cells have not gone through the basal membrane but is completely contained in the lobule or the ducts, the cancer is called in situ or noninvasive. If the cancer has broken through the basal membrane and spread into the surrounding tissue, it is called invasive. Therefore, early detection of breast cancer is essential. In our study, we are focusing on the differentiation between benign and malignant tumors.

**3. DATABASE REPRESENTATION**

Breast cancer becomes one of the leading causes of death of women in the world. The mammography technique has been proved to be an effective tool for the detection of breast cancer in its earlier phase. Detection of Clusters is an important sign in the identification of micro calcification of mammograms. In our paper, a medical data based on breast cancer attributes was used for the purpose of classification between two types of cancers, benign and malignant.

**3.1 BREAST CANCER DATABASE**

The database used in our study is the Wisconsin breast cancer database. It has been done in the University of Wisconsin by Dr. William H. Wolberg [15]. The same database has been used by researchers for the purpose of classification and testing algorithms in the world of data mining. 699 patients form the total available database. 683 patients are used in our case of classification, due to missed values in 16 of all patients. 11 features or attributes represent the data for each patient, which are nine cytological characteristics of breast fine-needle aspirates and two other attributes contain the id number of each patient and the class label, that correspond to the type of breast cancer (benign or malignant). The values of cytological characteristics are in a range from one to ten, with one being the closest to benign and ten the most malignant. The table 1 shows us the complete information of the database.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clump thickness</td>
<td>1-16</td>
</tr>
<tr>
<td>Uniformity of cell size</td>
<td>1-16</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Class</td>
<td>2 for benign 1 for malignant</td>
</tr>
</tbody>
</table>

Table 1: Database Parameters Information.

To clarify more the database used in our study and the synonyms of the terms cited above. We can conclude that these terms are mostly used in any pathology report on fine needle aspirations to evaluate a lump or mass in a breast could be malignant or benign, in other terms cancerous or non-cancerous. These are terms used in a pathology report on fine needle aspirations to assess whether a lump in a breast could be malignant (cancerous) or benign (non-cancerous). In the case of cancer type, the cells have a variation in size and shape. For this reason, the uniformity of cell size or shape tends to a benign direction. Also, the other parameters like bare nuclei, bland chromatin and normal nucleoli are signs of benignity.

To understand more the database and to present the cancer type direction of each parameter, a statistical distribution for each term will be presented in the next step.

**3.2 STATISTICAL DATA REPRESENTATION**

To study the statistical distribution of each attribute of the database before entering all the attributes to the neural network for the classification way, a schematic representation for each of them shows us the frequency of each domain (1-10) in the attribute itself. Recall that 1 is closest to benign and 10 the most malignant.

![Graph 1: Statistical Distribution for the Clump Thickness Attribute](image)
3.3 DATA NORMALIZATION

The data normalization is considered as the most important preprocessing step using neural networks. To improve the performance of multilayer neural networks, it is better to normalize the data entry such that will be found in the interval of [0 1]. The original form of the data presented in our study was in the analog form with a range starting from 1 to 10. To transform the data into digital form, and use it as inputs of the neural network, scaling or normalization should be realized for each attribute. The nine numerical attributes, in the analog form, are scaled with a range of 0 and 1. There are many types of normalization that are found in the literature. The new values obtained after normalization, follow this equation:

\[
\text{New value (after normalization)} = \frac{\text{current value} - \text{min value}}{\text{max value} - \text{min value}}
\]

4. ARTIFICIAL NEURAL NETWORK

The application of neural network of the artificial intelligence world, in healthcare is relatively new [11] [12] [14]. The goal of this paper is to show how we can apply the data mining and especially the neural network to the medical databases, where this application can predict or classify the data with a given accuracy. To obtain a good prediction and especially a good classification, we should learn the algorithms with a good training set which rules or patterns are extracted to help the testing dataset classification [6][10].

4.1 ARCHITECTURE OF NEURAL NETWORK AND LEARNING PHASE

Multi layer feed forward neural network, such that multilayer perceptron MLP and radial basis function RBF, is the structure of network used in our study. Each network contains three layers, input, hidden and output layer. The features of the data base which correspond to the nine cytological parameters are the input of the networks. The number of hidden neurons should be determined experimentally. The output layer consists of two classes corresponding to the two types of breast cancer (0 for benign and 1 for malignant). The learning phase of the network constitutes of the utilization of back propagation algorithm. A supervised neural network based on back propagation algorithm was used in our study. Once a neural network architecture chosen, it is necessary to conduct a learning phase to determine the values of the weights for the output of the neural network to be as close as possible to the target. The flow chart below (figure 2) presents the learning phase of our neural network model.

4.2 ALGORITHM USED IN OUR STUDY

After the preparation of the data and normalization all the attributes to be ready for the input of the neural network, different steps have been done to test the performance of our model:

- Selecting the architecture of our network where it is constituted of an input layer, a hidden layer and an output layer. Nine neurons are in the input layer which corresponds to the nine features of each patient, and one neuron in the output layer (0 or 1). The number of hidden neurons is determined experimentally for each model.
- Initialization the neural network
- Network training using back propagation algorithm
- Network testing
- Presentation of the results obtained.

4.3 ADAPTATION OF PARAMETERS USING MLP

The neural network (Figure 3) is constituted of one input layer containing inputs corresponding to the nine attributes of the database, of one hidden layer with five neurons and of one output layer containing two outputs corresponding to our two classes, benign and malignant. Note that the optimal number of neurons of the hidden layer where it is equal to five was determined experimentally.

The activation functions used in the different layers of the neural network are the following:

- Input layer: no activation function (no calculation at the level of this layer)
- Hidden layer: simplified sigmoid function
  \[
  \phi(x) = \frac{1}{1 + e^{-x}}
  \]
- Output layer: simplified sigmoid function
  \[
  \phi_o(x) = \phi(x) \text{ or hyperbolic tangent}
  \]
  \[
  \phi_o(x) = \tanh\left(\sqrt{2}\right) = \frac{1 - e^{-x}}{1 + e^{-x}}
  \]

4.3 PERFORMANCE OF MLP NETWORK

Construction, learning and test are different phases used in classification problems modeled by neural networks. Three layers exist in a network of back propagation, including an input layer with nine parameters, a hidden layer containing five neurons and an output layer.
containing a single neuron. The value of the neuron of output layer indicates if the entry corresponds to a cancer case or not. The weights of the network connection are set randomly in the learning phase. The input parameters are normalized between 0 and 1.

Table 3 shows the experimental results of cancer dataset using MLP network.

<table>
<thead>
<tr>
<th>Training Samples</th>
<th>Test Samples</th>
<th>Classification Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Multi Layer Perceptron MLP</td>
</tr>
<tr>
<td>278</td>
<td>100</td>
<td>94%</td>
</tr>
<tr>
<td>180</td>
<td>198</td>
<td>90%</td>
</tr>
<tr>
<td>100</td>
<td>278</td>
<td>88%</td>
</tr>
</tbody>
</table>

Table 2: Experimental Results of Cancer Dataset Using MLP

Note that 88% of patients are classified correctly in the case where the training samples are more than the test samples. The percentage of error is 12%.

4.4 ADAPTATION OF PARAMETERS USING RBF

The figure 4 shows the architecture of the neural network RBF with Gaussian function as activation function.

![Figure 4: RBF Neural Network Model](image)

The neural network (Figure 4) is constituted of one input layer containing nine inputs corresponding to the nine attributes, of one hidden layer with five neurons and of one output layer containing two outputs corresponding to our two classes.

The activation functions used in the different layers of the neural network are the following:
- Input layer: no activation function (no calculation at the level of this layer)
- Hidden layer: Gaussian function
- Output layer: linear function

4.5 PERFORMANCE OF RBF NETWORK

The same phases discussed in the previous part are used to train and test the RBF network. Table 3 shows the experimental results of cancer dataset using RBF network.

Note that 97% of patients are classified correctly in the case where the training samples are more than the test samples. The percentage of error is 3%. A preliminary result shows the better classification of the RBF network than that of the MLP network.

<table>
<thead>
<tr>
<th>Training Samples</th>
<th>Test Samples</th>
<th>Classification Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Radial Basis Function RBF</td>
</tr>
<tr>
<td>278</td>
<td>100</td>
<td>99%</td>
</tr>
<tr>
<td>180</td>
<td>198</td>
<td>99%</td>
</tr>
<tr>
<td>100</td>
<td>278</td>
<td>97%</td>
</tr>
</tbody>
</table>

Table 3: Experimental Results of Cancer Dataset Using RBF

5. DISCUSSION

We have adopted approaches to build models to study neuronal breast cancer classification. We presented the results by modeling the attributes by neural networks. These results show the effectiveness of one technique over another and effectiveness of the extracted parameters. We can note that the RBF network, on the classification, performs better than the MLP network.

Finally we note that the model produced by the supervised technique RBF can be considered a successful model for the detection and classification of various breast cancers.

6. CONCLUSION

In this paper, a new approach is developed to study the breast cancer classification based on neural network technique.

The objective of this study is to create an effective tool for building neural models to help us making a proper classification of various classes of breast cancer.

The interest in neural networks is justified by their own properties: learning ability, generalization and reminiscence.

The multilayer structure of neural networks of type Feedforward Perceptron can be considered as a powerful tool in the field of nonparametric approximation. This is particularly due to its ability to represent any kind of multidimensional nonlinear functions. Choosing an appropriate architecture of the MLP and the type of neuron depends on the nature of the application domain. For this, a thorough study of the nature of this problem should occur prior to an architecture that can be found.

It has been shown theoretically that a multilayer neural network with one hidden layer is able to identify any linear or nonlinear multivariate with enough hidden neurons. Therefore, in general, we use a network with only one hidden layer. As we do not know the optimum number of hidden neurons, it is often necessary to run multiple learning by changing each time the number of hidden neurons to find the optimum number.

Since 1985, when presenting the back propagation gradient GBP, this algorithm has become the most widely used algorithm to stabilize a network Feedforward multilayer type. The merit of GBP is its simplicity.

The RBF network is an alternative to MLP, and the sigmoid activation function is replaced by a Gaussian function. This type of network has also shown
efficiency in the field of modeling, and in the domain of classification.
A detailed comparison between multilayer networks MLP and RBF showed that the model constructed from RBF neural network is much more efficient than other model based on the MLP.
In conclusion, the supervised RBF neural approach driven by the learning algorithm GBP works well, in terms of accuracy, efficiency and reliability.
Using this model, an automated classification of various types of breast cancer was performed by avoiding the question of the expert concerning the recognition of cancer required, improving the identification of breast cancer classification.
The study in this paper for the automatic classification of breast cancer based on RBF neural network is a new study that affects breast cancer development. Future studies will be realized using unsupervised learning technique for our problem of classification, as well as other techniques in data mining domains in order to improve the accuracy of classification.

ACKNOWLEDGEMENT
This project has been done at the Islamic University of Lebanon.

REFERENCES


