

## INTEGRATING ARTIFICIAL NEURAL NETWORKS FOR DEVELOPING TELEMEDICINE SOLUTION

Mihaela GHEORGHE<sup>1</sup>

Faculty of Economic Cybernetics, Statistics and Informatics,  
Bucharest University of Economic Studies,  
[Mihaela.Gheorghe@ie.ase.ro](mailto:Mihaela.Gheorghe@ie.ase.ro)

**Abstract:** Artificial intelligence is assuming an increasing important role in the telemedicine field, especially neural networks with their ability to achieve meaning from large sets of data characterized by lacking exactness and accuracy. These can be used for assisting physicians or other clinical staff in the process of taking decisions under uncertainty. Thus, machine learning methods which are specific to this technology are offering an approach for prediction based on pattern classification. This paper aims to present the importance of neural networks in detecting trends and extracting patterns which can be used within telemedicine domains, particularly for taking medical diagnosis decisions.

**Key words:** telemedicine, neural networks, machine learning, medical diagnosis, artificial intelligence

**JEL Classification Codes:** C6, C8, I1

### 1. INTRODUCTION

One of the most important tasks of the medical providers is to diagnose different diseases. For this, in order to establish a valid diagnosis, physicians need to choose between different type of illness based on a specific number of observations and knowledge. Neural networks can be used within any situation that involves a relationship between different variables and medicine is one of them. Artificial neural networks (ANNs) have the main advantage of being capable of solving problems that are too complex for a sequential algorithmic solution. Currently, ANN's topologies are widely used in a various areas of medicine subsets, including image analysis, modelling kinetics of drug release [1], monitoring health indicators (blood pressure, respiration rate or level of glucose). Their topologies can also be used within the medical diagnosis process. They provide an important instrument, complementary to statistical methods and can be used on large sets of data, especially when these are being characterized by noise, they are incomplete, missing different values for attributes or a high degree of independence exists between factors.

Therefore, ANN technology can be used to easily correlate different factors, to find hidden correlation among different inputs or non-linear system in which the relationships are unknown or very complex and also, to make prediction and classification based on them [2].

### 2. ARTIFICIAL NEURAL NETWORKS: THEORETICAL FRAMEWORK

Artificial neural networks were developed as a result of the process of simulating biological nervous system [3]. They are represented as a set of nodes also named neurons along with the connections between them. Each of these is being associated with weights which are

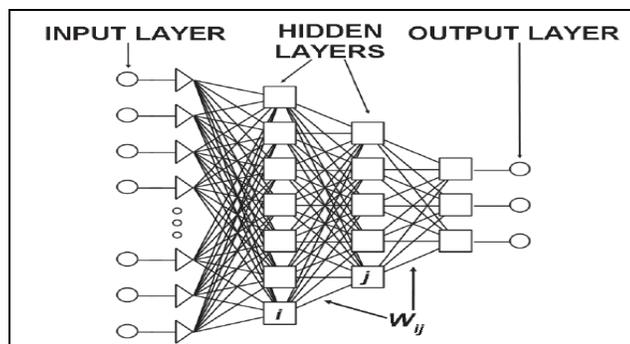
---

<sup>1</sup> PhD Student

representing the strength of neuron connections. In what concerns their topology, most of the artificial neural networks have three types of layers in their structure and within them the data is being propagated and transformed based on different algorithms. These layers are represented as follows:

- first layer represents the interface with the environment and contains the input neurons
- hidden layer is the one responsible with most of the computations that are done
- output layer represents the level at which the results are stored

The number of neurons in each layer depends on the complexity of the datasets. The general architecture of a neural network with three layers is illustrated in figure 1.



**Figure 1: An artificial neural network [4]**

Each neuron from a layer is connected to each neuron from the second layer through a weighted relationship shortened as  $w_{ij}$  which represents the strength of the connection between neuron  $i$  from the first layer and neuron  $j$  from the next one.

In what concerns the computations, these are handled within the hidden layer where the data is being mathematically processed by following some steps including [4]:

- calculating the network inputs based on weights according to (1):

$$net_j = \sum_{i=1}^m x_i * w_{ij} + \theta_j \quad (1)$$

where:  $x_i$  represents the incoming data from neuron  $i$ ,  $w_{ij}$  is the weight between neurons  $i, j$  and  $\theta_j$  is the bias or threshold value, for each  $i \in \overline{1, m}$  and  $j \in \overline{1, n}$ .

- transforming the network inputs through a transformation function
- transferring the outputs to the next layer which can be achieved based on different functions; in this study, the sigmoid function is used and is calculated with (2):

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

### 3. CASE STUDY: ANN MODEL FOR BREAST CANCER DIAGNOSIS

Artificial neural networks can be used within the medical field in order to assist physicians to make accurate diagnostic for breast cancer. In this chapter, two algorithms are described and used on a breast cancer dataset with the purpose of analysing their results and exemplify the ANN approach and these are represented by Multilayer Perceptron and Naïve Bayes neural networks.

### 3.1. DATASET DESCRIPTION

Breast cancer dataset [5] was obtained from the University Medical Centre, Institute of Oncology Ljubljana and contains a number of 286 instances. There are 9 attributes and a class, all described [6] in table 1.

**Table 1. Breast cancer database attributes**

Attribute name	Values	Description
Menopause	lt40, ge40, premeno.	Pre-menopausal status is a factor that leads to the assumption that a recurrence is not likely to occur.
Tumor-size	0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59.	Represents the diameter of the existing tumor.
Node involvement	0-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20, 21-23, 24-26, 27-29, 30-32, 33-35, 36-39.	Axillary lymph nodes represent common site of early metastasis. The more nodes are involved, the more likely recurrence is.
Degree of malignancy	1, 2, 3.	The tumor's histological grade affects recurrence. If 1 it is less likely to occur, if it's about 2 or 3, tumors consist in abnormal cells and recurrence is more likely.
Breast	left, right.	A slightly higher recurrence is more likely on the left side.
Breast-quadrant	left-up, left-low, right-up, right-low, central.	Breast cancer often occurs in the upper outer quadrant increasing the chances of recurrence.
Irradiant	yes, no.	Radiotherapy reduces the risks of recurrence.
Age	10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99	The strongest risk factor for breast cancer is the age.
Node-capsular invasion	yes, no.	If the tumor remains contained by the lymph nodes capsule, recurrence is more likely.
Class variable	no-recurrence-events, recurrence-events	Reappearance of cancer after 5 years.

Regarding the class distribution, there are two main categories differentiated by the possibility of breast cancer reoccurrence after a period of 5 years.

### 3.2. ALGORITHMS DESCRIPTION

#### 3.2.1. NAÏVE BAYES

Naïve Bayes is a probabilistic neural artificial classifier based on the theorem of Bayes illustrated by (3) and which considers the fact that all attributes independently contribute to the probability of a certain decision [7].

$$P(C|X_1, X_2, \dots, X_n) = \frac{P(C) \cdot P(X_1, X_2, \dots, X_n|C)}{P(X_1, X_2, \dots, X_n)} \quad (3)$$

where: P represents the probability function, C is the class variable and  $X_1, X_2, \dots, X_n$  are the attributes.

This classifier is based on the process of training data with the conditional probability that each variable  $F_k$  is given by the class labelled as C. Thus, the classifier used to train the described

datasets has the assumption that all variables are conditionally independent and the class distribution is calculated by (4):

$$P(C = c | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(C = c) * \prod_{f_k} P(X_n = x_k | C = c) \quad (4)$$

Naive Bayes classifier algorithm consists of the following [7]:

- calculating  $P(C = c)$  based on the maximum posterior probability as shown in (5)

$$P(C = c) = \max_{c_i \in C} (P(c_i | X)) \quad (5)$$

- estimating  $P(c_i | x)$  for each given class based on (6)

$$P(c_i | x) = \frac{(P(x | c_i) * P(c_i))}{P(x)} \quad (6)$$

- estimating  $P(x | c_i)$  based on (7)

$$P(x | c_i) = \prod_{j=1}^n P(x_j | c_i) \quad (7)$$

- estimating the probability for each  $P(c_i)$  and  $P(x_j | c_i)$  based on the Gaussian distribution calculated by (8)

$$P(x_j | c_i) = \frac{1}{\sqrt{2\pi}\sigma_{ji}} * e^{-\frac{(x_j - \mu_j)^2}{2\sigma_{ji}^2}} \quad (8)$$

where:  $\sigma$  is the standard deviation factor and  $\mu$  the mean value.

### 3.2.2. MULTILAYER PERCEPTRON BACK PROPAGATION

The multilayer Perceptron (back propagation) network is formed by processing elements modelled as neurons, which have the ability of computing the weighted sum of the input data based on a bias and transform this value through the activation function. The main steps of the back propagation algorithm are described as follows [8]:

- initialization of the connection weights  $w_{ij}$  with random values
- repeat the previous step until convergence by computing the updates using (9), modifying the old values based on the updates using (10) and computing error using (11) in order to check either if the error is less than a fix established limit value or if the gradient  $\frac{\partial E(t)}{\partial w}$  is smaller than a specific limit.

$$\Delta w(t) = -\eta * \frac{\partial E(t)}{\partial w} \quad (9)$$

$$w(t+1) = w(t) + \Delta w(t) \quad (10)$$

$$E = \frac{1}{2} * \sum_{j=1}^n (d_j - y_j) \quad (11)$$

where:  $t$  is the iteration/epoch number,  $w$  represents the connected weight,  $\eta$  is the learning rate of the algorithm and  $E$  is the mean square error between the current output  $y_j$  and the desired one  $d_j$ .

In order to achieve a better accuracy of the multilayer algorithm, a momentum term [9] is added to the back propagation equations. This is used for computing the weighted values and is calculated based on (12).

$$\Delta w(t) = -\eta * \frac{\partial E(t)}{\partial w} + \alpha \Delta w(t-1) \quad (12)$$

where  $\alpha$  represents the momentum term and can take values between (0,1).

### 3.3. EXPERIMENT RESULTS AND PERFORMANCE ANALYSIS

Before training the dataset based on the previously described algorithms, a normalization process [10] is being done as a pre-processing step in order to enhance learning performance and to reduce errors. In this study, the normalized values of each attribute are being calculated based on (13).

$$value = \frac{current\ value - min\ value}{max\ value - min\ value} \quad (13)$$

The results obtained by applying the multilayer perceptron algorithm (MLP) both simple and with a momentum term and also by applying Naïve Bayes classifier (NB) are represented within a confusion matrix [11] used to store the relationship between outcomes and predicted classification. A generalized representation of a confusion matrix is illustrated in table 2.

**Table 2. Confusion matrix**

Actual values	Predicted	
	Negative	Positive
Negative	a	b
Positive	c	d

The entries within the above tables have the following meaning:

- a represents the correct number of negative instances identified as negative
- b represents the incorrect number of negative instances identified as positive
- c represents the correct number of positive instances identified as positive
- d represents the incorrect number of positive instances identified as negative

The experiments results are represented in table 3 for the network trained with basic MLP, in table 4 for the network trained with MLP adjusted with a momentum term and in table 5 the ones obtained by training with NB.

**Table 3. Basic MLP results**

Actual values	Predicted	
	Negative	Positive
Negative	80	4
Positive	5	197

**Table 4. MLP with momentum**

Actual values	Predicted	
	Negative	Positive
Negative	79	1
Positive	6	201

**Table 5. NB results**

Actual values	Predicted	
	Negative	Positive
Negative	41	27
Positive	44	174

In order to evaluate the performance [11] of these machine learning techniques, the following indicators are being calculated: accuracy based on (14) and Mathews Correlation Coefficient (MCC) using (15).

$$Accuracy = \frac{a+d}{a+b+c+d} \quad (14)$$

$$MCC = \frac{a+d-b*c}{\sqrt{(a+c)*(a+b)*(a+c)*(a+b)}} \quad (15)$$

In table 6 the calculated values for accuracy (ACC) and MCC indicators are presented for each of the proposed algorithms described in 3.2.

**Table 6. ACC and MCC results**

Algorithm	Indicators	
	ACC	MCC
Basic MLP	96,8531%	92,44%
MLP with momentum	97,5524%	94,13%
Naïve Bayes	75,1748%	36,56%

The high percentage numbers obtained for MCC and ACC indicators regarding the Multilayer Perceptron algorithm with a momentum term prove that artificial neural networks are representing an important approach for developing informatics systems that can be used to optimize medical workflows and efficiency.

#### 4. CONCLUSIONS AND FUTURE RESEARCH

Within this paper, an artificial neural network approach has been developed and described in order to study the breast cancer classification based on several machine learning techniques. Based on the information reported in chapter 3, these can be considered a powerful instrument and an approach which presents a strong capability to accurately recognize breast cancer diseases. Future studies will be realized using other learning strategies and algorithms for classification issues in order to improve the performance and results. Also, this will be applied to different medical datasets in order to extend the constructed neural models.

#### ACKNOWLEDGMENT

This work was financially supported through the project "Routes of academic excellence in doctoral and post-doctoral research - READ" co-financed through the European Social Fund, by Sectoral Operational Programme Human Resources Development 2007-2013, contract no POSDRU/159/1.5/S/137926."

#### REFERENCES

1. <http://www.mdpi.com/1999-4923/4/4/531/pdf>, *Artificial Neural Networks in Evaluation and Optimization of Modified Release Solid Dosage Forms*, 2012.
2. D. Kriesel, *A brief introduction to Neural Networks*, 2011, Available: <http://www.dkriesel.com/media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf>
3. <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002063>, *Learning, Memory, and the Role of Neural Network Architecture*, 2011
4. [http://www.zsf.jcu.cz/jab\\_old/11\\_2/havel.pdf](http://www.zsf.jcu.cz/jab_old/11_2/havel.pdf), *Artificial neural networks in medical diagnosis*, 2013
5. <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer>
6. Longo L., Hederman P., *Argumentation Theory for Decision Support in Health-Care: A Comparison with Machine Learning*, Brain and Health Informatics: International Conference, BHI 2013, Maebashi, Japan, pp. 168-180, October 29-31, 2013. Proceedings
7. <http://researchpublications.org/IJCSA/NCAICN-13/189.pdf>, 2013
8. *The Backpropagation Algorithm*, <http://page.mi.fu-berlin.de/rojas/neural/chapter/K7.pdf>
9. *Improved Back propagation learning in neural networks with windowed momentum*, <http://synapse.cs.byu.edu/papers/IstookIJNS.pdf>
10. *Statistical Normalization and Back Propagation for Classification*, [http://www.researchgate.net/profile/Santhakumaran\\_A/publication/260024206\\_Statistical\\_Normalization\\_and\\_Back\\_Propagation\\_for\\_Clasification/links/00b4952f1e2762d491000000.pdf](http://www.researchgate.net/profile/Santhakumaran_A/publication/260024206_Statistical_Normalization_and_Back_Propagation_for_Clasification/links/00b4952f1e2762d491000000.pdf)
11. *An Overview of General Performance Metrics of Binary Classifier Systems*, [http://sebastianraschka.com/PDFs/articles/performance\\_metrics.pdf](http://sebastianraschka.com/PDFs/articles/performance_metrics.pdf), 2014