International Journal of Biomedical Data Mining

Zribi and Boujelbene, Biomedical Data Mining 2016, 5:1
DOI: 10.4172/2090-4924.1000118

Review Open Access

The Neural Networks with an Incremental Learning Algorithm Approach for Mass Classification in Breast Cancer

Zribi M* **and Boujelbene Y**

Faculty of Economic Science and Management, Sfax University, Tunisia

***Corresponding author:** Manel Zribi, Faculty of Economic Science and Management, Sfax University, Tunisia, Tel: +216 74 242 951; E-mail: manel0306@yahoo.fr

Rec date: Feb 22, 2016; **Acc date:** Mar 29, 2016; **Pub date:** Mar 31, 2016

Abstract

As breast cancer can be very aggressive, only early detection can prevent mortality. The proposed system is to eliminate the unnecessary waiting time as well as reducing human and technical errors in diagnosing breast cancer. The correct diagnosis of breast cancer is one of the major problems in the medical field. From the literature it has been found that different pattern recognition techniques can help them to improve in this domain. This paper uses the neural networks with an incremental learning algorithm as a tool to classify a mass in the breast (benign and malignant) using selection of the most relevant risk factors and decision making of the breast cancer diagnosis To test the proposed algorithm we used the Wisconsin Breast Cancer Database (WBCD). ANN with an incremental learning algorithm performance is tested using classification accuracy, sensitivity and specificity analysis, and confusion matrix. The obtained classification accuracy of 99.95%, a very promising result compared with previous algorithms already applied and recent classification techniques applied to the same database.

Keywords: Artificial neural networks (ANN); Breast cancer, Incremental learning algorithm; classification risk factors

Introduction

Breast cancer is a leading fatality cancer for woman. According to epidemiological data, breast cancer accounts for 20-25% of female malignant tumor, with is expected to increase. These facts have driven us to select this deadly cancer as our domain. Breast cancer has four early signs; micro-calcification, mass, architectural distortion and breast asymmetries. However, only data regarding mass will be used as a pilot project to test our system later on. Masses of 2 cm in diameter are palpable with regular breast self-examination while mammogram images can capture it from 5 mm in diameter. However, these images were to be determined by an expert radiologist who is familiar with breast cancer. Generally, there are 2 types of breast cancer which are in situ and invasive. In situ starts in the milk duct and does not spread to other organs even if it grows. Invasive breast cancer on the contrary, is very aggressive and spreads to other nearby organs and destroys them as well. It is very important to detect the cancerous cell before it spreads to other organs, thus the survival rate for patient will increase to more than 97%. However, time taken from taking mammogram images to biopsy dangerous result varies between 2weeks to a month in average.

As breast cancer can be very aggressive, only early detection can prevent mortality. The proposed system is to eliminate the unnecessary waiting time as well as reducing human and technical errors in diagnosing breast cancer. The etiologies of breast cancer remain unclear and no single dominant cause has emerged [1,2]. Prevention is still a mystery and the only way to help patients survive is by early detection. A major class of problems in medical science involves the diagnosis of disease, based upon various tests performed upon the patient. For this reason the use of classifier systems in medical diagnosis is gradually increasing. There is no doubt that evaluation of data taken from patients and decisions of experts are the most important factors in diagnosis. But, the different artificial intelligence techniques for classification also help experts a great deal. There has been research on medical diagnosis of breast cancer with WBCD using Artificial Neural Networks (ANNs) in literature, and most has reported high classification accuracy.

- Guo and Nandi [3], proposed a multilayer perceptron as a classifier with retro propagation of error algorithm for breast cancer diagnosis with WDBC database (Wisconsin Diagnosis Breast Cancer) they got a rate of classification of 96.21%.
- Erdem [4], used an artificial neural network with a back-propagation algorithm of the gradient. The sample of 699 WBCD patients (Wisconsin Breast Cancer Dataset), and is divided into two sub-sample 490 for learning and 209 to test the network which is composedof 9 inputs, 4hidden neurons and 2 outputs (malignant and benign) provides a good percentage of 99.28.
- Marcano [5], should on a sample of 410 patients WBCD compound for learning and 273 for the test sample, the network consists of an input terminal and 9 input layer neurons, 4 hidden neurones associated with a sigmoidal activation function in the output layer (malignant and benign) and a learning algorithm AMMLP (Artificial Metaplasticity algorithm with multi-layer perceptron) provides a good percentage grading 99.63.
- Murat [6], uses the same data base as the ones used in WBCD. The network comprises 9 inputs, one hidden layer with 11 hidden neurons and an output layer with a linear function. The learning algorithm is one of the gradient back-propagation for 95.2% of well classified patients.

In this study, the ANN with incremental algorithm is proposed for classifying the breast cancer lesions as benign or malignant.

Brief review of artificial neural networks

ANNs are biologically inspired and mimic the human brain. The neurons are interconnected with connection links which have weights which are 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. NN's are trained by experience. When an unknown input to the network is applied, it can generalize from past experiences and product a new result [7-10]. The multilayer Perceptron (MLP) is the most used network. It is a type of feedforward network composed of successive layers and very efficient for the classification problems. The idea consists in classify the neurons by interconnected layers. The first, which is called the input layer, is composed of a number of neurons with the task of receiving information from the outside. This information is processed and then transmitted to the neurons of the inter-layers which will in turn treat it and then send the results to a final layer called the output layer (Figure 1).

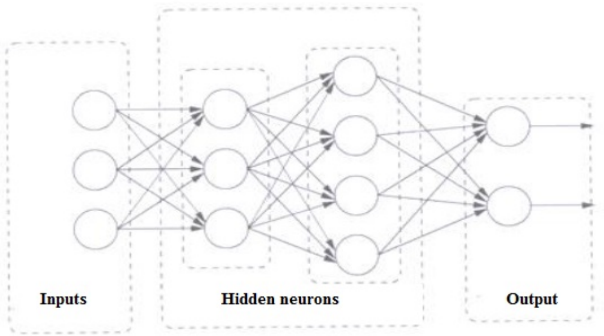

Figure 1: Multilayer perceptron.

Given the lack of theoretical rules to identify the optimal architecture of a neural network for a particular situation, several authors proposed empirical rules based on a number of tests obtained by varying the number and size of the intermediate layers. Finally, it should be noted that some attempts to automatically identify the best architecture have been proposed. For the researcher, the main advantage of ANNs is that there is no need to specify the functional relation between variables. Since they are connectionist-learning machines, the knowledge is directly imbedded in a set of weights through the linking arcs among the processing nodes. In order to train a neural network properly, one needs a large set of representative 'good quality' examples. In the case of clustering problems, the researcher should be cautious when drawing conclusions from neural networks trained with only one hidden neuron, as observed in most previous studies by Cheng [11]. Beyond the network's ability to recognize and properly classify the individuals presented at the input through a battery of variables, we tried to better make use of the results so as to give other elements for a more accurate analysis. Furthermore, the undertaken learning network enables us to study the contribution of each input variable in the explanation of the studied phenomenon.

Problem statement

For this research, data are taken from a single source that is from University of Wisconsin Hospitals. The dataset is actually a record regarding mass in breast from mammogram images, and the values are entered by an expert radiologist based on the mammogram images. Preliminary examination on the chosen dataset is compulsory before the cleaning process take place. The dataset consist of 699 data with 65.5% classified as benign and 34.5% as malignant. As medical data are best not to be tampered, all data with; missing value are eliminated. There are nine criteria recorded for masses in this dataset. They are clump thickness (clump), uniformity of cell size (ucsiz),uniformity of cell shape (ucsha), marginal adhesion(marga), single epithelial cell size (secsi), bare nuclei(baren), bland chromatin (bland), normal nucleoli (normn)and mitoses (mitos) with one class attribute that is either benign or malignant.

Network architecture: The problem of reducing the model was initiated for whom complicity model depends on the number of elements, connections and model calculations. In our research process about the network structure that best fits our application, we use an incremental algorithm-based technique [12] performed in 4 steps:

The first step consists in training a minimum network made up of a single neuron on its hidden layer. As soon as the learning process stabilizes, in other words, the improvement of the error compared to the previous step is inferior to a given threshold, the network stands still and a new neuron is added to the hidden layer (Figure 2). The learning process starts again but only the weights of the last neuron are fixed. Then, the process passes onto adding neurons to the hidden layer whenever the network improvements are not significant. The process stops once the global network error reaches the desired threshold or when adding a new neuron does not improve the error obtained in the previous iteration.

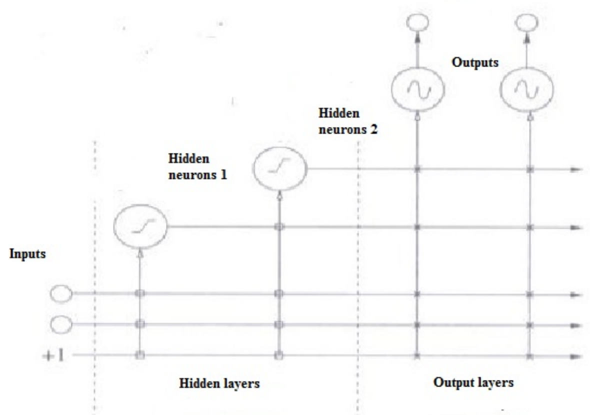

Figure 2: Incremental algorithm.

This work is intended to identify the most effective learning algorithm in our study. This type of algorithm will be specified for each type of variable.

The overall sample was divided into three sub samples:

- The first sub-sample, which includes 70% of the starting sample, is a learning sample.
- The second sub-sample, which is called a test sample and represents 15 % of the overall sample, is used to examine the learning sample results.
- The third sub-sample is a validation sample which represents 15% of the overall sample and is used to validate the results.

The evaluation of the network performance is carried out on the basis of:

- The minimum squared error (MSE)
- Good clearance rate in the 3 sub-samples

Generally, a network may have different activation functions for different nodes in the same or different layers. However, almost all the networks use the same activation functions particularly for the nodes of the same layer. While most researchers use logistic activation functions for the hidden nodes shows that the sigmoid activation functions are more effective for the classification problems. Even if the work of [11,13-15] showed that one hidden layer nothing is known a priori about the used number of the hidden neurons. Similarly, there is no consensus on which the activation function should be used for the output nodes several researchers use linear output nodes. We use the toolbox of MATLAB software; we get after several tests the optimum architecture (Figure 3). The optimal network consists of four hidden neurons with a good ranking of 99.95% (Figure 4).

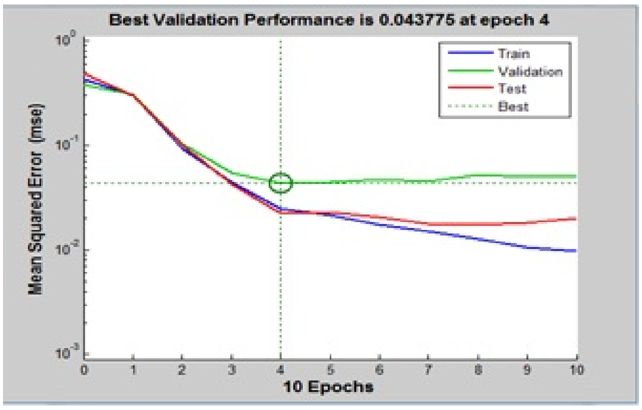

Figure 3: Performance with MSE.

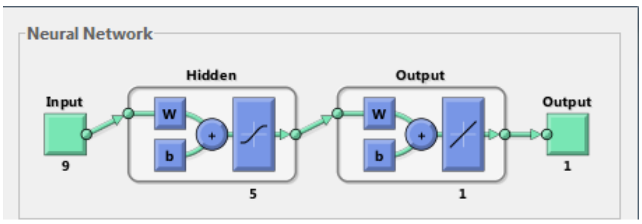

Figure 4: Optimum neural network.

Our algorithm enabled us to achieve a better percentage compared to what existed in previous work with the same database (WBCD), mainly that of Marcano [5]. Our research will be improved and deepened by a sensitivity analysis to rank the variables according to their degree of influence on the type of the tumor mass.

Sensitivity analysis

Determining the relevant variables is an essential step in identifying the models. The number of variables must be sufficient so that the selected model can adequately explain the studied phenomenon. A first solution consists in studying all the possible variable combinations. It is obvious that this procedure becomes extremely heavy as soon as the number of the variables that come into operation becomes high. An optimal approach requires a tool with which we can assess the importance of each variable which helps compare different subsets, define a selection strategy for the exploration of the space of the variables combinations, and finally set a criterion for the procedure termination. The variable selection technique adopted here consists, once the best architecture is determined, in removing au input variable each time then proceeding to learn the network and evaluate the new performance. The weaker this performance is, compared to that of the starting network, the more relevant the variable in the chosen model is. Therefore, we can classify the system input variables according to their relevance for the explanation of the phenomenon in question. In fact, the more important the variable is, the higher the performance difference is, after the elimination of the variable and vice versa. This technique helps us not only rank the variables, but also find the best combination among the variables presented at the system input, which reduces the complexity of the model through a non-significant loss in terms of performance. Table 1 summarizes the results of this procedure.

	Removes criteria	Pi with Total sample (Pi)	P^*-Pi= (1 : 9)	
Criteria	none	P^*=0,0193	0,0000	class
clump	x1	0,0218	-0,0025	7
ucsiz	x2	0,0188	0,0005	3
ucsha	x3	0,0229	-0,0036	5
marga	x4	0,0214	-0,0021	8
secsi	x5	0,0162	0,0031	1
baren	x6	0,0252	-0,0059	4
bland	x7	0,0227	-0,0034	6
normn	x8	0,0213	-0,0020	9
mitos	x9	0,0166	0,0027	2

Table 1: Class of criteria.

Based on these results, three variables can be a first variable group having great importance in the problem of having a breast cancer, it is:

Single epithelial cell size: since the epithelial cells are not naturally present in the bone marrow and are not detected in healthy individuals, bone marrow may thus be considered an indicator of metastatic disease in patients with breast cancer at the primary stage.

Mitosis: is a process of a controlled cell division to reproduce daughter cells genetically identical to the parent one. Malignant cells are characterized by uncontrolled and intense cell division compared to a normal cell population.

Uniformity of cell size: cancer cells are characterized by anisocytosis, namely, inequality in terms of size compared to healthy cells.

The second group includes less significant:

Bare nuclei: in the normal state, the nucleoli are inside the nucleus. When they are combined with the cytoplasm, this means that the cell is abnormal and it is likely to become cancerous.

Uniformity of Cell Shape: the cancerous cells are marked by irregular contours and incisures.

Bland chromatin: it is a protein caused by an excessive estrogen reception. H2az is a protein which induces the expression of the estrogen receptor gene. Over production of this protein is a sign of breast cancer cells since these are hormone-dependent.

Clump thickness: Thickness of the membrane: the thickness of the plasma membrane of a cancerous cell is greater than that of a normal cell.

And finally the variables that have not submitted explanatory power important:

Marginal adhesion shape: this is an over-expression of the integrin beta3 protein at the surface of the cancerous cell.

Normal nucleoli: the DNA is naturally protected by a nuclear membrane. If failure is observed in this membrane, this means that there is tumor growth.

Conclusion

Our project is in line with the branch of the decision-making support and health economics. The desire to acquire new knowledge about the biology of breast cancer requires intelligent approaches that can be adaptable to high-dimensional data and uncertainties. Breast cancer development is often confusing. Many studies based on large series of patients identified the significant prognostic parameters for the overall survival time. These parameters will be highlighted in our study due to the importance of the early diagnosis in improving the cancer prognosis. Knowing the prognostic factors is of a great importance because it helps measure the risk, adapt the given treatment, and especially improve the knowledge about the natural history of the disease. Based on past research and real applications to date, the many methods and approaches are still not enough to correctly identify the type of tumors. With the wide, proven and effective applications of the ANN in solving multicriteria decision making models both by academicians and practitioners, it is hoped that the incremental algorithm approach can help in deciding the mass classes, given the many available criteria.

Indeed, identifying breast cancer patients having a risky pattern of reactions makes it possible to give these patients specific care. Undoubtedly, this could significantly improve their quality of life. For this reason, we had better conduct our research in this context that. The sensibility analysis enabled us to classify the breast cancer factors depending on their degree of importance and influence on the type of disease (Single Epithelial Cell Size: Mitosis and Uniformity of Cell Siz), which may help clinician experts in medical decision making. Our research work can be developed in medical imagery to determine the change caused by breast cancer, and compared with the work of Zhang et al. [16] Developed an accurate computer-aided diagnosis (CAD) system of MR brain images, is essential for medical interpretation and analysis. In this study, we propose a novel automatic CAD system to Distinguish abnormal brains from normal brains; and In Wang et al. [17] study we proposed a novel computer-aided diagnosis system for Detecting abnormal breasts combine in the ACP and K-means approach.

References

1. Evalgelos D, George K (2000) Fast detection of masses in computer-aided mammography. IEEE Signal Processing Magazine 17: 54–64.
2. Rodrigues PS, Chang R, Suri JK (2006) Non-extensive entropy for cad systems of breast cancer images. Brazilian Symposium, Brazil.
3. Hong G, Nandi AK (2006) Breast cancer diagnosis using genetic programming generated feature. Elsevier Pattern recognition 39: 980-987.
4. Alkım E, Gürbüz E, Kılıç E (2012) A fast and adaptive automated disease diagnosis method with an innovative neural network mode. Neural Netw 33: 88-96.
5. Marcano-Cedeño A, Quintanilla-Domínguez J, Andina D (2011) WBCD breast cancer database classification applying artificial metaplasticity neural network. Expert Systems with Applications 38: 9573-9579.
6. Karabatak M, Ince MC (2009) An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications 36: 3465-3469.
7. Andina K, Jevtic A, Marcano-Cedeño A, Barrón-Adame MJ (2007) Error weighting in artificial neural networks learning interpreted as a metaplasticity model. Lecture notes in computer science 45: 244-252.
8. Haykin S (1999) Neural networks, a comprehensive foundation. (2nd edition), Pearson, Canada.
9. Hagan MT, Demuth HB,Beale MH (1996) Neural network design. (2nd edition), PWS Pub, Boston, USA.
10. Kandel ER, Schwartz JH, Steven SA, Jessell TM (2000) A principles of neural science. (5th edition), McGraw-Hill, USA.
11. Cheng SNC, Chan HP, Helvie MA, Goodsitt MM, Adler DD, et al. (1994) Classification of mass and non-mass regions on mammograms using artificial neural networks. Imaging SeiTechnol.
12. Dunkin N, Taylor JS, Koiran P (1996) A new incremental learning technique. Neural Nets WIRN VIETRI. Springer, NewYork.
13. Cybenko G (1989) Approximation by superpositions of a sigmoidal function; Math control systems signals, 2: 303-314.
14. Funahashi KL (1989) On the approximate realisation of continous mapping by neural networks. Neural networks 2: 183-192.
15. Verma B (2008) Novel network architecture and learning algorithm for the classification of massabnormalities in digitized mammograms. J artmed 42: 67-79.
16. Zhang Y, Dong Z, Wang S, Ji G, Yang J (2015) Preclinical Diagnosis of Magnetic Resonance (MR) Brain Images via Discrete Wavelet Packet Transform with Tsallis Entropy and Generalized Eigenvalue Proximal Support Vector Machine (GEPSVM). Entropy 17: 1795-1813.
17. Wang SH, Zhang YD, Liu G, Yang J (2016) Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform.aime 8:1-11.