

Detection of Breast Lesions in Medical Digital Imaging Using Neural Networks

Gustavo Ferrero, Paola Britos and Ramón García-Martínez

Software & Knowledge Engineering Center, Graduate School, Buenos Aires Institute of
Technology

Intelligent Systems Laboratory, School of Engineering, University of Buenos Aires.

rgm@itba.edu.ar

Abstract. The purpose of this article is to present an experimental application for the detection of possible breast lesions by means of neural networks in medical digital imaging. This application broadens the scope of research into the creation of different types of topologies with the aim of improving existing networks and creating new architectures which allow for improved detection.

1. Introduction

Breast cancer has been determined to be the second leading cause of cancer death in women, and the most common type of cancer in women; there are no official statistics in the Argentine Republic, but it is estimated that 22 in 100,000 women are affected by this illness, similarly to what is observed in other Western countries [Mols *et al*, 2005]. The mammography is the best method of diagnosis by images that exists at the present time to detect minimum mammary injuries, fundamentally small carcinomas that are shown by micro calcifications or tumors smaller than 1cm. of diameter that are not palpated during medical examination. [Antonie *et al*, 2001]. Currently, joint efforts are being made in order to be able to detect tissue anomalies in a timely fashion, given that there are no methods for breast cancer prevention. Early detection has proved an essential weapon in cancer detection, since it helps to prolong patients' lives. Physicians providing test results must have diagnostic training based on mammography, and must issue a certain number of reports annually. Double reading of reports increases sensitivity for detection of minimal lesions by about 7%, though at a high cost. The physician shall then interpret these reports and determine according to his/her best judgment the steps to be taken for the proper diagnosis and treatment of the patient. for this reason, physicists, engineers, and physicians are in search of new tools to fight cancer, which would also allow physicians to obtain a second opinion [Gokhale *et al*, 2003, Simoff *et al*, 2002]. The American College of Radiology having approved the use of new digital mammographs, digital photos have begun to be stored in databases together with the patient's information, for later processing via different methods [Selman, 2000]. Different methods have been used to classify and/or detect anomalies in medical images, such as wavelets, fractal

Please use the following format when citing this chapter:

Ferrero, G., Britos, P., García-Martínez, R., 2006, in IFIP International Federation for Information Processing, Volume 218, Professional Practice in Artificial Intelligence, eds. J. Debenham, (Boston: Springer), pp. 1–10.

theory, statistical methods and most of them used features extracted using image-processing techniques. In addition, some other methods were presented in the literature based on fuzzy set theory, Markov models and neural networks. Most of the computer-aided methods proved to be powerful tools that could assist medical staff in hospitals and lead to better results in diagnosing a patient [Antonie *et al*, 2001].

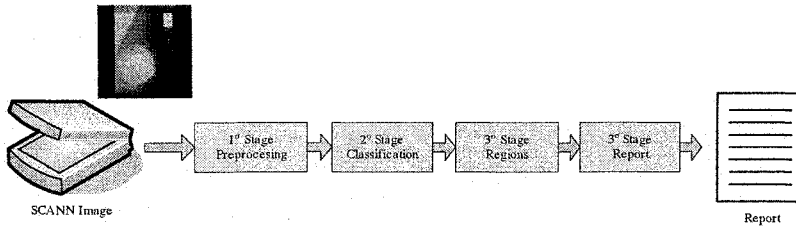


Fig. 1. Processing Steps

Different studies on using data mining in the processing of medical images have rendered very good results using neural networks for classification and grouping. In recent years different computerized systems have been developed to support diagnostic work of radiologists in mammography. The goal of these systems is to focus the radiologist's attention on suspicious areas. They work in three steps: i. analogic mammograms are digitized; ii. images are segmented and preprocessed; iii. Regions of Interests (ROI) are found and classified by neural networks [Lauria *et al*, 2003].

2. Proposed Method

Radiologists do not diagnose cancer versus benign nodules; they detect suspicious regions and send them for additional work up [Baydush *et al*, 2001]. Bearing in mind the way medical imaging specialists work, the system works as follows: i. capturing medical image, ii. storing image on the data base, iii. starting up processing, iv. generating report, v. validating report. The first and last steps generate information which provides a work environment where system users are given the ability to create new network topologies in order to validate the results obtained.

2.1. Mammography Processing

Figure 1 shows the stages that have been adopted for the image processing of a mammography. Stage 1: image pre-processing; this stage begins by acquiring the image, which is delivered to the following stage containing only the region of interest. Stage 2: image classification to determine whether or not it contains malignant lesions that require in-depth examination by specialists. Stage 3: if the classifier determines

that the image shows malignant lesions, suspicious areas are scanned for. Stage 4: mammography processing report generated.

2.2 Mammography Pre-processing

The first stage contains a set of steps which as a group serve the purpose of eliminating all information which is irrelevant for the classification. Step 1. Using median filter. Step 2. Cropping margins. Step 3. Eliminating isolated regions. Step 4. Equalizing. Order filters are based on a specific treatment of image statistics called order statistics. These filters operate in the neighborhood of a certain pixel, known as window, and they replace the value of the central pixel. Order statistics is a technique that organizes all pixels in a window in sequential order, on the basis of their grey level [Liew *et al*, 2005]. The M mean in a set of values is such that half of the values in the set are smaller than M and half of the values are greater than M. In order to filter the mean in the area around the neighborhood, we ranked the intensities in the neighborhood, we determined the mean, and assigned the latter to the intensity of the pixel. The main purpose of mean filtering is to cause the points with very different intensities to become similar to their neighbors, thus eliminating any isolated intensity peaks that appear in the area of the filter mask. The median filter is a nonlinear filter, used in order to eliminate the high-frequency filter without eliminating the significant characteristics of the image. A 3x3 mask is used, which is centered on each image pixel, replacing each central pixel by the mean of the nine pixels covering the mask. The window size allows the characteristics of the image to be preserved while at the same time eliminating high frequencies [Díaz, 2004]. Next the automatic cropping is performed. The purpose of this step is to focus the process exclusively on the relevant breast region, which reduces the possibility for erroneous classification by areas which are not of interest. Image segmentation is an important step in several image applications. A host of techniques and algorithms usually fall into this general category as a starting point for edge detection, region labelling, and transformations. All these techniques, region labelling, and analyses, are relatively simple algorithms that have been used for many years to isolate, measure, and identify potential regions [Jankowski and Kuska, 2004]. A stack method is used for region labelling, as it is one of the fastest and simplest to implement. After labelling, those areas which are not of interest to the study are eliminated. It is known that the surface covered by the breast is over 80%; therefore, isolated areas with surfaces smaller than 1% do not belong to the breast and are eliminated through the creation of masks obtained from neighboring pixels. Lastly, a uniform equalization is performed, which will essentially help enhance image contrast.

$$F(g) = [g_{\max} - g_{\min}] P_p(g) + g_{\min}$$

Where g_{\max} , and g_{\min} correspond to the maximum and minimum intensity values in the range of grey values of the image. Figure 2 shows the results of image pre-processing.

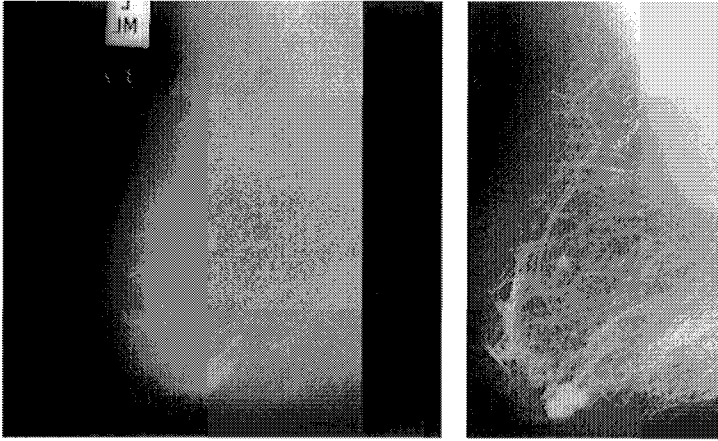


Fig. 2. Automatic Pre-processing

2.3 Classification

Neural networks are models which attempt to emulate the behavior of the brain. As such, they perform simplification, identifying the relevant elements in the system. An adequate selection of their features coupled with a convenient structure constitutes the conventional procedure utilized to build networks which are capable of performing a given task [Hertz *et al*, 1991]. Artificial neural networks offer an attractive paradigm for the design and analysis of adaptive, intelligent systems for a wide range of applications in artificial intelligence [Fiszelew *et al*, 2003]. Artificial neural networks are based on a rather simple model of a neuron. Most neurons have three parts, a dendrite which collects inputs from other neurons (or from an external stimulus); a soma which performs an important nonlinear processing step; finally an axon, a cable-like wire along which the output signal is transmitted to other neurons is called synapse [W. Gestner, NA]. Neurons are grouped in layers; these interconnected layers form a neural network, thus each neural network is composed of N number of layers (Figure 3). Depending on how these components (layers) are connected, different architectures may be created (feed forward NN, recurrent NN, etc.). The topology or architecture of a neural network refers to the type, organization, and arrangement of neurons in the network, forming layers or clusters. The topology of a multilayered neural network depends on the number of variables in the input layer, the number of hidden neuron layers, the number of neurons per each hidden layer, and the number of output variables in the last layer. All these factors are important when determining network configuration [Zurada, 1995].

Thus, neural network structures can be defined as collections of parallel processors interconnected in the form of an oriented lattice, arranged in such a way that the network structure is appropriate for the problem under consideration. The connections

between neurons in a neural network have an associated weight, which is what allows the network to acquire knowledge. The most commonly used learning écheme for the MLP is the back-propagation algorithm. The weight updating for the hidden layers adopts the mechanism of back-propagated corrective signal from the output layer.

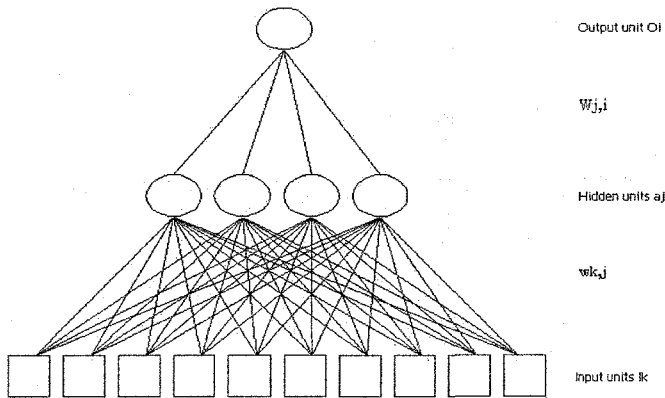


Fig. 3. Neural Network

It has been shown that the MLP, given flexible network/neuron dimensions, offers an asymptotic approximation capability. It was demonstrated that two layers (one hidden only) perceptrons should be adequate as universal approximators of any nonlinear functions [Kung *et al*, 1998]. A multilayer perceptron is structured as follows:

- Function signal: the signal that propagates from the input to output.
- Error signal: generated by output neurons and it backpropagates as an adjustment to the synaptical connections towards the input in order to adjust the output obtained to the expected output as faithfully as possible.

Thus all output neurons and those in the hidden layer are enabled to perform two types of calculations according to the signal they receive: If it is a function signal, it will be a forward calculation (forward pass); if it is an error signal, it will be a backward calculation (backward pass). The rule of propagation for neurons in the hidden layer is the weighted sum of the outputs with synaptic weights w_{ji} , then, a sigmoid transference function is applied to that weighted sum and is limited in the response. Basically, the backpropagation algorithm is based on error minimization by means of a traditional optimization method called gradient descent. That is, the key point consists in calculating the proper weights of the layers from the errors in the output units; the secret lies in evaluating the consequences of an error and dividing the value among the weights of the contributing network connections. Neural network learning can be specified as a function approximation problem where the goal is to learn an unknown function $f: \mathbb{R}^N \rightarrow \mathbb{R}$ (or a good approximation of it) from a set of input-output pairs $S = \{(x^N, y) \mid x^N \in \mathbb{R}^N, y \in \mathbb{R}\}$ [Parekh, *et al*, 2000]. Pattern

classification is a special case of function approximation where the function’s output y is restricted to one of M ($M \geq 2$) discrete values (or classes). A neural network for solving classification problems typically has N input neurons and M output neurons. The k th output neuron ($1 \leq K \leq M$) is trained to output one (while all the other output neurons are trained to output zero) for patterns belonging to the k th class. A single output neuron suffices in the case of problems that involve two category classifications. The multilayer perceptron facilitates the classification of nonlinear problems; the more hidden layers in a neural network, the simpler it will be to isolate the problem (Figure 4).

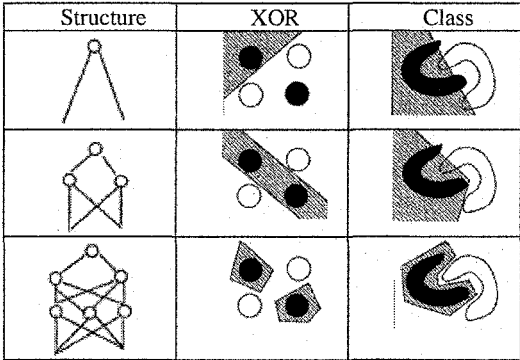


Fig. 4. Geometric Interpretation of the Role of Hidden Layers

2.4. Image Features

A radiological image is formed by tissue absorption when exposed to Roentgen radiation. Depending on the amount of radiation absorbed, an object (tissue) may be radiopaque (RO), radiolucid (RL), or radiotransparent (RT).

- When a low amount of X-ray radiation is absorbed by the object, virtually all the rays reach the film; the color appears dark and it is a radiotransparent body (air cavities).
- When a moderate amount of radiation is absorbed by the object, the color appears grey and it is a radiolucid body (noncalcified organic tissue).
- When a large amount of X-ray radiation has been absorbed or it has been completely absorbed, the color appears light or white and it is a radiopaque body (inorganic tissue, calcified tissue).

The range of colors is related to and depends on the extent of X-ray absorption by the tissue. The more radiation that is absorbed by the tissue, the less radiation will reach the film, signifying that the body is radiopaque; the less radiation that is absorbed by the tissue, the more radiation will reach the film, signifying that the body is radiolucid. This feature of X-ray images will be taken into account and will be used in

order to obtain information about the sample tissues. The input layer is formed by K input neurons; this value is obtained by considering: the regions defined for the image (image subdivisions), the amount of statistical operations applied to those regions, plus one neuron according to the position of the breast (left or right) and a set of three neurons according to the type of tissue (dense, dense-glandular, and glandular). The characteristics of each of the regions are obtained from the information provided by the tissues represented in the pixels. Data extraction in the regions is carried out via the following statistical procedures:

Mean: it is the mean data value.

Bias: it is the systemic error which frequently occurs.

Kurtosis: it measures whether the distribution values are relatively concentrated around the mean values of the sample.

Variance: it measures the existing distance between the values and the mean.

The mean is defined by the following function:

$$\bar{X} = \frac{\sum x_i}{n}$$

Bias is defined:

$$\text{bias} = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{X}}{s} \right)^3$$

Kurtosis is defined by the following formula:

$$k = \frac{n * (n-1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{X}}{s} \right)^4 - \frac{3 * (n-1)^2}{(n-2)(n-3)}$$

Lastly, variance is given by:

$$s = \frac{\sum (x_i - \bar{X})^2}{(n-1)}$$

The system offers the possibility of creating different topologies for multilayer networks ($N \geq 3$, where N is the number of layers). This feature allows the user to research new architectures. As it has been mentioned before, image information (regions) is entered in the network, for instance; for an image divided into 16 regions, there are 69 ($16 * 4 + 1 + 3 + 1$) neurons including the independent term; the output layer may have only one or two neurons.

2.5. Training

Supervised learning is characterized by a controlled training by an external agent (supervisor or teacher) who determines the response to be generated by the network

from a certain input. The supervisor verifies the network output and in case it does not match the expected output, connection weights are to be modified, in order to obtain an output as close as possible to the expected one [Hertz *et al*, 1991]. The proposed method for the system is learning by trial and error, which consists in adjusting the weights of the connections according to the distribution of the quadratic error among the expected responses r_q and the appropriate current responses O_q .

$$E_q = \frac{1}{2} \sum_{q=1}^{N_q} (r_q - O_q)^2$$

Training data are mammographies obtained by Mammographic Image Analysis Society (MIAS); they consist in 322 images from which 55 are to be analyzed in order to obtain further information, which shall be used to train the neural network.

The following information is necessary for neural network training:

Stop error: it is the acceptable output error for training, below this error, training is brought to a halt and weights are said to converge.

Number of cycles: it sets the maximum number of cycles that need to be learned by the network.

Momentum: each connection is given a certain inertia or momentum, in such a way that its tendency to change direction with a steeper descent be averaged with the “tendencies” for change of direction that were previously obtained. The momentum parameter must be between 0 and 1 [Zurada, 1995].

Learning ratio: it is a value between 0 and 1 used in error distribution (delta rule).

The training procedure is evaluated every thousand cycles, obtaining information about the last ten cycles, performing an evaluation of the mean, in a way that the speed of network convergence can be determined, thus allowing to conclude training and start with a new set of weights.

3. Experimenting with the System

The system allows for the generation of different architectures, for which reason different networks are evaluated, bearing in mind certain factors, whenever a new architecture is created. The performance (and cost) of a neural network on any given problem is critically dependent, among other things, on the network’s architecture and the particular learning algorithm used. [Fiszelew *et al*, 2003]. Too small networks are unable to adequately learn the problem well while overly large networks tend to over fit the training data and consequently result in poor generalization performance [Parekh, *et al*, 2000]. With varying numbers of hidden layers, the following configuration has yielded the best results: 69 Neurons in the input layer. 16 Neurons in the hidden layer, one for each region the image was divided into. 4 Neurons in the following hidden layer, one for each operation performed on the regions. 2 Output neurons, one for each possible direction (right and left). In addition, two networks have been set up for use depending on the location of the breast (left or right), since

the initial results obtained using one single network for both were very poor. Training times were also reduced using this modality, thereby increasing the convergence speed. The single network trained for both sides was able to classify correctly with a 30% success rate, whereas the networks trained for a specific side have a 60% success rate. In addition, the former took over 4 hours to converge, while the second did so in less than 60 minutes. Further research will be done to identify the appropriate neural network that will allow classifications to be obtained with 80% certainty.

4. Related Work

In this research the neural net architecture is trained to distinguish malignant nodules from benign ones. It differs from the CALMA project approach [Lauria *et al.*, 2003] which has tools that identify micro calcification clusters and massive lesions. The proposed architecture deals with mammography images in image formats (JPG, BMP, TIFF, GIF), which differs from standard data mining approach based on image parameters sets provided by the community [UCIMLR. 2006a.; 2006b].

5. Conclusions

This project has been an attempt to provide an environment for the continued investigation of new neural network models that will achieve better results in the classification of medical images. The project is open source, allowing access to the source code so that others may study, improve, expand and distribute it, so that healthcare institutions may have access to tools they can use to improve breast cancer detection.

The goal of the project is to improve the detection of areas suspected of containing some type of lesion. The results obtained are very different from our initial expectations, although the initial results obtained are promising. The future will bring improvements to this application as well as the possibility of inputting additional statistical data that will enhance the reading of the images.

Next research steps are: (a) to compare results of the proposed architecture in this paper with others provided by vector machines based classifiers [Fung and Mangasarian, 1999; Lee *et al.*, 2000; Fung and Mangasarian, 2003]; and (b) to study specific filters that recognize structures in mammography images.

6. Bibliography

- Antonie M., Zaïene O., Coman A. (2001). *Application of data mining techniques for medical image classification*. Proceedings of the Second International Workshop on Multimedia Data Mining. San Francisco.
- Baydush A., Cararius D., Lo J., Abbey C., Floyd C. (2001). *Computerized classification of suspicious regions in chest radiographs using subregion hotelling observers*. American Association of Physicists in Medicine. Vol 28 (12).

- Díaz, S. (2004). *Detección de microcalcificaciones en imágenes de mamografías usando diferencias de filtro gaussianos optimizados*. Universidad Nacional de Ingeniería. Facultad de Ingeniería Industrial y de Sistemas. Lima. Perú.
- Egmont-Petersen M., de Ridder D., Handels H. (2002). *Image processing with neural networks. The Journal of the pattern recognition society*. Vol. 35 (10).
- Fiszlelew, A., Britos, P., Perichisky, G. & García-Martínez, R. (2003). *Automatic Generation of Neural Networks based on Genetic Algorithms*. Revista Eletrônica de Sistemas de Informação, 2(1): 1-7.
- Fung, G. and Mangasarian, O. 1999. Semi-Supervised Support Vector Machines for Unlabeled Data Classification. *Optimization Methods and Software* Vol. 15, pp. 29-44.
- Fung, G. and Mangasarian, O. 2003. *Breast Tumor Susceptibility to Chemotherapy via Support Vector Machines*. Data Mining Institute Technical Report 03-06, November 2003. Computational Management Science.
- Gerstner W. (1998). *Supervised learning for neural networks: A tutorial with Java Exercises*. Technical Report. Laboratory of Computational Neuroscience at the Swiss Federal Institute of Technology Lausanne.
- Hertz J., A. Krogh y R. Palmer 1991. *Introduction to the Theory of Neural Computation*. Reading, MA: Addison-Wesley.
- Jankowski M., Kuska J. (2004). *Connected components labelling – algorithms in Mathematica, Java, C++ and C#*. IMS2004.
- Kung S., Hwang J. (1998). *Neural network for intelligent multimedia processing*. Proceedings of the IEEE. Vol 86 (6).
- Lauria A., Palmiero R., Forni G., Cerello P., Golosio B., Fauci F., Magro R., Raso G, Tangaro S., Indovina P. (2003). *The CALMA system: an artificial neural network for detecting masses and microcalcifications in digitalized mammograms*. 9th Pisa Meeting on Advanced Detector. La Biodola, Isola d'Elba, Italy.
- Lee, Y., Mangasarian, O. and Wolberg, W. 2000. *Breast Cancer Survival and Chemotherapy: A Support Vector Machine Analysis*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 55, pp. 1-10.
- Liew, A., Yan, H. and Yang, M. (2005). *Pattern Recognition techniques for the emerging field of bioinformatics: A review*. Pattern Recognition 38(11): 2055-2073
- Mols, F., Vingerhoets, A., Coebergh, J. and Poll-Franse, L. (2005). *Quality of life among long-term breast cancer survivors: A systematic review*. European Journal of Cancer 41(17): 2613-2619.
- Parekh R., Yang J., Honavar V. (2000) *Constructive neural-network learning algorithms for pattern classification*. IEEE Transactions on neural networks. Vol. 11 (2).
- Selman, S. (2000). *Data Mining of Digital Mammograms Will Aid in War against Cancer*. www.gatech.edu Página vigente al 17/08/2003.
- Simoff, S., Djeraba, C., y Zaiane, O. *Multimedia Data Mining between Promise and Problems*. 3rd Edition of the International Workshop on Multimedia Data Minig. Pages 118-121. SIGKDD Explortions.
- UCIMLR. 2006a. *Breast Cancer Database*. UCI Machine Learning Repository. <http://www.ics.uci.edu/~mllearn/MLSummary.html>
- UCIMLR. 2006b. *Wisconsin Breast Cancer Databases*. UCI Machine Learning Repository. <http://www.ics.uci.edu/~mllearn/MLSummary.html>
- Zurada, J. (1995). *Introduction to Artificial Neural Systems*. West Publishing Company.