HIERARCHICAL NEURAL NETWORK SYSTEM TO BREASTS CANCER DIAGNOSIS

Mingo L.F. Castellanos J. *Giménez V.* *Vilarrasa A.*
Dept. Inteligencia Artificial Dept. Matemática Aplicada Servicio de Radiología
Facultad de Informática - Univ. Politécnica de Madrid Hospital 12 de Octubre
28660 Boadilla del Monte - Madrid - SPAIN Madrid - SPAIN
Tel: +34 91 3367440, Fax: +34 91 3524819
e-mail: jcastellanos@fi.upm.es

ABSTRACT

This paper describes a hierarchical neural network system used as a help tool in the breast cancer diagnosis. The system divides the image using a tree scheme. The preprocessing of the areas pointed by the tree takes into account the environment that the element to analyze have, so as others computations, that are done using an adaptive filter. The neural network is in charge of classifying the areas into two possible classes, a healthy class or an ill class, using the obtained descriptors from the areas. All the data employed in this research have been obtained from the digitalization of the mammographies taken to some hospitalized patients.

1 INTRODUCTION

Breast cancer is one of the frequents cancer among the female population. In the last few years, the number of cases of this cancer has been increasing, in such a way that according to the National Cancer Institute (*SEER*), the risk of suffering a breast cancer, being a woman living in USA with a medium life expectancy, has been continuously increasing between 1940 and 1995, in such a way that in 1940 the number of cases with this pathology represented only the 5% of the total cases while in 1989 that pathology represented the 11%. In 1995 the number of cases was the 12.6% of the total, being the most frequent death cause, that means that one among 25 or 30 women will be dead for this cause in the USA.

Screening breast programs are looking for a early diagnosis as soon as possible, with the purpose of finding the tumour in its earliest stage and in this way to get the opportunity for attaching the problem with more success. It has been proved, that the early detection of the sickness may avoid a great number of cases of death among women with more than 50 years old, and with big probability from the forties also.

Small masses (*about 2mm*) may be detected doing a mammography of the patient breast, the microcalsifications of 0.2mm diameter size may be also detected, that is the reason for which the mammography method has become the most reliable and sensitive way of cancer detection and diagnosis. The use of the screening method is then a fully reasonable method for attaching the sickness. Anyway the method described before has some problems that still remain to be solved, as for example: bad interpretation or bad realisation of the mammography's, no detection of the masses of small size, false diagnosis (*false positives or negatives*), poor experience of the cases, etc.

Today breast cancer is detected on the basis of four types of signs on mammogram: the morphology and characteristics of a tumour mass, certain presentations of mineral deposits as specks called microcalsifications, architectural distortion of normal tissue patterns caused by the disease, asymmetry between images of the left and right breasts.

Artificial Neural Networks techniques show excellent promise in being able to overcome the limitations of presently used computer methods to predict patient carcinoma. This is because these networks can be trained to recognize complex relationships that exist between inputs and outputs [1]. These subtle relationships in the data are automatically recognized by the network, even if they are unknown to clinicians. Because neural networks can learn any arbitrary relationship between a given set of inputs and outputs [2, 3, 4], they can normally be expected to perform at least as well as and usually better than other modelling technique.

As the complexity of the problem increases, so does the superiority of neural networks over most other methods. Importantly, neural network techniques have previously been shown to be able to handle the inaccuracy and inconsistency associate with patient histories and physical findings [5, 6, 7]. Further, the networks appears to be able to deal with the complexities of the disease states characterized by several totally differing clinical presentations [8, 9].

The disadvantage of neural network models is that, while they often have excellent overall results, they do no reveal how a given prediction was made. Physicians sometimes feel uncomfortable with this black box approach to patient management in complicated cases because it is difficult to know when to overrule the network prediction.

The main aim of this paper is to achieve the prediction of a given pathology, Infiltrating Ductal Carcinoma, using for this purpose artificial neural networks. The system could be employed as a supporting tool to the expert radiologist, who can reinforce his diagnosis with a second point of view.

2 MODEL TO DIAGNOSIS

The problem of breast cancer diagnosis is one of the most complex problems, since it is based on the identification of a given area, of the mammography, that presents some kind of anomaly, and the anomaly has not to take part into the breast structure. You do not have to focus in specific alterations since they could form part of the breast architecture. It is rather an study by areas of the mammographies, paying special attention to the environment that surrounds them.

The system implemented here, is based on the neural network techniques as the main tool. These networks are applied in a hierarchical way to the different areas of the image, in such a way that when the process ends, there exist a set of areas susceptibles of having some kind of cancer disease. Neural Networks are trained so they are able to distinguish if a given area is healthy or if it has cancer. In our case all these networks have been specialized only to detect the infiltrating ductal carcinoma.

The training process of these networks is very sensitive since environment information and punctual information must be taken into account. In order to incorporate such information, the mammographies areas are pre-processed to get a set of descriptors that are the real data that are the inputs to the neural networks.

This paper proposes a pre-processing model based on the pixel intensity, instead of others method such as the feature extraction [12], via edge detection, pattern-matching, etc. [9, 10, 11, 12], that present a bit more complex information processing.

Given an area of the original image, four descriptors are computed. The four descriptors correspond to the mean, deviation, maximum and minimum value of the pixel intensities from the given area. Such values point out some global characteristics of the area, that can tell the expert radiologist that there exists a carcinoma.

Then, other five descriptors are obtained in a adaptive way, that is, theirs values are changing till they reach the desire aim. The descriptors are the followings: the number of pixels that there are above a threshold which will be explained later on, the mean, deviation, maximum and minimum value of the pixel intensities from the given area but this area has been filtered. The filter and threshold are fixed up according to the network training, these values are changing during the training in order to get a better performance of the network, so mean square error as generalization, with a test set.

The most suitable value obtained for the threshold is: ((maximum+mean)/2.0). And the obtained filter is the one defined by the following fuzzy function, illustrated in figure 1, taking into account that the slope of the line is changing as the network is training in order to improve the training error and generalization.

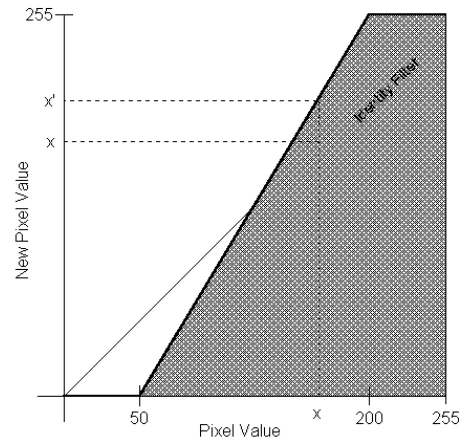

Figure 1: Filter to preprocess image areas.

This filter increases the brightness and contrast of the image. Once all the pre-processing is done, then, nine descriptors of each area are obtained -mean, deviation, maximum, minimum, the number of pixels that there are above a threshold and the mean, deviation, maximum and minimum after the filtering process-, that are the inputs to the neural network.

The hierarchy in which neural networks are apply is explained. First of all, the original image is taken, and if it has a great size then, it is divided into five sub-windows, that are continuously divided until their size is suitable in order to avoid incorrect diagnostic problems, or to avoid ignoring some areas with a very small carcinoma. In our case, only two levels of division are performed.

Once that all the windows have a correct size, all of them are pre-process to extract the information they have. The descriptors obtained with this pre-processing are the input to a network which classifies it into one of the two possible classes, with or without carcinoma. In case there are not carcinoma in the window, such window is redivided to incorporate all its subwindows to the set of windows to classify. In case there exists a carcinoma, such window is marked and the radiologist must study if the classification of the network is correct.

The most unfavourable situation is the one in which the system diagnosis is "no carcinoma" for all the windows of the mammography. In this situation the exploration tree is growing and growing till the size of the windows is small enough to considered that any tiny carcinoma could be ignored by the network. The divison of each area consists on the ampliation of the dimension of the image to its double size and the divison into five sub-areas. Such ampliation is done in order not to lose resolution in the following divisions. The training of the systems is based on a set of areas previously classified by an expert radiologist, into area with or without carninoma. Then, all the areas are pre-processed to form the training set of the neural network, the filter and the threshold are initially fixed up a priori.

2.1 Neural Network Topology

Neural networks have the ability to generalize, giving the right answer to entries not used during the training period and extracting the right tendency during contradictory and noisy events [1, 2]. Due to the type of application a backpropagation algorithm was used and several configurations were tested.

Backpropagation algorithm is based on the input to output propagation and the following modification of the connection weights using the gradient term. Mathematically [3, 4], it can be written as the multiplication of the learning rate α, by the error accumulation δ, and by the output of a given neuron o.

The task performed by neural networks is to approximate [13, 14] the function that the training set describes by means of the inputs and outputs that are fed to the network.

Cancer diagnosis can be performed in two different ways, the first one uses the whole radiological image to detect no healthy areas, and the second one only employs a zoomed fragment of the image that the radiologist thinks is not healthy. Both situations can be managed with the described system above.

3 APPLICATION TO BREAST CANCER DIAGNOSIS

The application of the model explained above has been focused on the identification of a pathology, Infiltrating Ductal Carcinoma, with the following signs: masses, asymmetry density and spiculating image, with or without microcalsifications.

The research has been performed using data collected from the records of the patients that have pass through the hospital "Hospital 12 de Octubre" at Madrid. All the mammographies have the diagnosis that the radiologist have determined and the confirmation after the biopsy is done, in case the radiologist diagnosis is carcinoma.

Taking as a sample the 161 cases of hidden diseases, in our Hospital, we found that 55 of them were of the carcinoma type, from which the 53% was the infiltrating ductal carcinoma and the 33% was the intraductal carcinoma.

The mammography data base considered has 100 mammography images, 70 of them have been diagnosed with the infiltrating ductal carcinoma, and has been check it doing the biopsy; the other 30 are diagnosed are normal mammographies. All the mammographies have been digitalized in order to be readable by a computer.

The training and testing sets have been designed choosing those areas that have some kind of anomaly and classifying then as items of the first class; later on, areas without any anomaly have been chosen and classified as item of the second class. The training set is made up of 200 patterns and the test or validation set has 50 patterns.

The pre-processing step of the areas that conform the training and validation sets is done following the guidelines established in the proposed model. Nine descriptors, extracted from each pattern (area), put the information needed to the classification of the images into the two possible diagnosis, with cancer or without cancer.

These nine descriptors are related with values of the intensity level of the pixels of the images, these are: mean, deviation, maximum, minimum, number of pixels above the threshold, and mean, deviation, maximum and minimum after the filter function.

The operation mode of the system can be treated in two different ways. In the first mode, the radiologist feeds a complete mammography into the systems and wait for the areas with some kind of anomaly, that are the second point of view that the expert radiologist will have, verifying or adding new possible areas. Such mechanism supports the expert diagnosis and even shows him some new areas that must be explore in detail, which due to their features -small microcalsifications, unlimited masses, etc.- are difficult to diagnose using screening methods. In the second mode of operation, the radiologist takes one area of the mammography, to get a second point of view or because he is not sure about his own diagnosis, and feeds it into the system waiting for the diagnosis.

4 RESULTS FROM SIMULATION

The optimum neural network architecture obtained in this particular problem has been a full-connected neural network with the back-propagation algorithm. The network has two hidden layers, the topology of this network is 9 - 5 - 5 - 1 neurons in the corresponding layers. Several architectures have also been tested to see the behaviour of the neural networks, among the possibilities tested were: one hidden layer, more hidden neurons, etc., but the better results are obtained with the topology of two hidden layers.

In the first training approach of the problem with the two-hidden layer network and with a non adaptive filter function, results were not really good, since the error in the classification is relatively high (see table 1, column fix filter). When adding the feature of an adaptive filter to the whole process, results are highly improved, since the filter is changing till it reaches the optimun slope that is suitable in order to decrease the error of the network with the pre-processed data.

Table 2 shows results obtained in terms of false negatives and false positives that the network outputs in

Table 1: Performance of the Network.

	Fix Filter	Adaptive Filter
Train	93%	99%
Test	75%	90%

Table 2: Analisys of the Network Diagnosis.

Patterns	False Positives	False Negatives
Train (200)	0	2
Test (50)	1	4

the training and testing sets. It can be seen that there is only one false positive in the test process, this is a good characteristic and shows a good behaviour of the network against new data. Since the system is a second point of view to the expert radiologist, almost all the false values can be removed by the radiologist when diagnosing a patient.

5 CONCLUSIONS

The method explained in this paper, in order to detect the infiltrating ductal carcinoma using the radiological image of the patient's breasts, has proven to have a good behaviour in the classification of some areas of the mammography with the infiltrating ductal carcinoma.

The descriptors obtained pre-processing the mammographies areas and the descriptors given by the adaptive filter, are a powerful tool that provide independence of the size of the original image. Moreover, they permit the possibility to make a zoom of the suspicious area without increasing the computational complexity of the system.

The obtained results are really good since there exists only a 10% error in the diagnosis of new patients not presented to the network during the training process. Most of these errors are false negatives, with a few false positive, this implies a robustness of the network that in the case that a cancer input it feeds to the network, it will not classify it as a normal area without cancer.

References

[1] De Villiers J., Barnard B.: *Backpropagation Neural Nets with one and two hiddden layers.* IEEE Transactions on Neural Networks. Vol. 4, No. 1, Pp. 136-141. (1993).

[2] Hertz J., Krogh A., Palmer R.: *Introduction to the Theory of Neural Computing*, Addison- Wesley, Redwood City, CA. (1991).

[3] Maren A., Harston C., Pap R.: *Handbook of Neural Computing Applications*, Academic Press, London. (1990).

[4] Muller B., Reinhardt J.: *Neural Networks: An Introduction.* Springer-Verlag, Berlin, Germany. (1990).

[5] Baxt, W.G.: *Use of an Artificial Neural Network for the Diagnosis of Myocardial Infarction.* Annals of Internal Medicine. Vol. 115, Number 11, Pp. 843. (1991).

[6] Gheorghiade M., Anderson J., Rosman H., Lakier J., et al.: *Risk Identification at the time of Admission to Coronary care unit in patients with suspected Myocardial Infarction.* American Heart Journal. Vol. 116, Pp. 1212-1217. (1987).

[7] Bridgett NA., Brandt J., Harris C.J.:*Artificial neural networks for use in the diagnosis and treatment of breast cancer.* International 4th conference on Artificial Neural Networks. Pp. 448-453. Cambridge, UK, June 1995 IEE Conf. Pub. No. 409. (1995).

[8] Dassen W. R., Mulleneers R. G., Den Dulk K.: *An artificial neural network to localize atrioventricular accessory pathways in patients suffering from the Wolff-Parkinson-White syndrome.* Pacing Clicinal Electrophysiology. Vol. 13, Pp. 1792. (1990).

[9] Cheng H.D., Chen C.H., Freimanis R.I.: *Neural Networks for Breasts Cancer Detection Using Fuzzy Entropy Approach.* International Conference on Image Processing. ICIP-95. (1995).

[10] Braal L., Ezquerra N., Schwartz E., Cooke C.D., Garcia E.: *Analyzing and Predicting Images through a Neural Network Approach.* College of Computing, Georgia Institute of Technology, Atlanta, GA 30332. (1996).

[11] Hiroyuki Y., Zhang W., Cai W., Doi K., Nishikawa M.: *Optimizing Wavelet Transform Based on Supervised Learning for Detection of Microcalcifications in Digital Mammograms.* International Conference on Image Processing. ICIP-95. (1995).

[12] Castellanos J., Mingo L.F., Rios J.; *Image Compression and Classification with Neural Networks.* VII International Symposium on Artificial Intelligence, Proceedings: Intelligent Systems Applications in Industry and Business, Monterrey, Mexico, October 18-21. pp 181- 185. (1994).

[13] Stinchcombe, M. & White, H.: *Multilayer Feedforward Networks are Universal Approximators.* Neural Networks, Vol. 2. Pp. 551-560. (1989)

[14] Hornik, K.; Stinchcombe, M. & White, H.: *Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks.* Neural Networks, Vol. 3. Pp. 625-633. (1990)