NEURAL NETWORK BASED MICROCALCIFICATION DETECTION IN A MAMMOGRAPHIC CAD SYSTEM

László Lasztovicza, Béla Pataki, Nóra Székely, Norbert Tóth

Budapest University of Technology and Economics,
Budapest, pf. 91, H-1521, Hungary,
{pataki, szekelyn}@mit.bme.hu, laszto@hotmail.com
http://home.mit.bme.hu/~pataki, http://home.mit.bme.hu/~szekelyn

Abstract: *This document presents the computer aided diagnosis system being developed to help experts in screening mammography. It is a very important project because about 8 % of women develop breast cancer in her lifetime therefore global screening is necessary. It means that reliable diagnosis of huge number of images must be solved. The basic architecture of the system and the information processing needed is presented. One of the most important tasks in mammographic diagnosis systems is microcalcification detection. It is solved by a hierarchical neural architecture. The original suggestion of that structure was improved by two ways. The image features used as inputs to the neural networks were analyzed and the feature set was extended. The neural architecture was embedded in a neural ensemble context for improving the quality of the solution further. Results of the tests of that detection procedure show that the false detection ratios are acceptable.*

Keywords: *microcalcification detection, neural networks, mammography, computer-aided diagnosis*

1. INTRODUCTION

Breast cancer is the most common form of cancer among women. According to statistics, 8% of women will develop it in her lifetime. Thanks to recent advances in medicine, there are effective methods in the treatment. The sooner the illness is detected, the more effective the treatment is. If detected early, the five-year survival rate exceeds 95%. Mammography is one of the most effective ways for early detection. In a mammographic session four x-ray pictures of the two breasts are taken (typically craniocaudal (CC) and mediolateral (ML) views) [1]. It is used to detect abnormalities and judge their severity: to differentiate benign and malignant cases.

Screening mammography is an x-ray examination when the patient is asymptomatic. Most of the mammograms are normal, but an enormous number of images must be evaluated. This would mean around one million images per year in Hungary, if global screening were done (women over 40). The diagnosis of this amount of images takes very long time and can produce human errors due to the length and monotony of the process. Therefore screening mammography could be improved with the aid of a computer system.

A system that could support the efforts of screening mammography – filter out normal images or draw attention to the suspicious ones – would be extremely useful. It could save time and help to avoid false diagnoses. Section 2 goes through with the introduction of such a system being developed by a consortium of doctors, researchers from the Budapest University of Technology and software engineers.

The rest of the paper deals with the description of a part of this CAD system, a neural based method for microcalcification detection, which is an extension of the work presented in [2, 3].

2. ARCHITECTURE OF A COMPUTER-AIDED DIAGNOSIS (CAD) SYSTEM

The task of computer-aided diagnosis is a very hard and complex one. In several cases the pictures taken are of poor quality, the abnormalities are not well defined. Even the best human specialist cannot diagnose 100% of the images correctly. Several methods are published, some results are achieved, but no final, satisfying solution is known for every problem. In Fig.1 the basic architecture of our CAD system is shown.
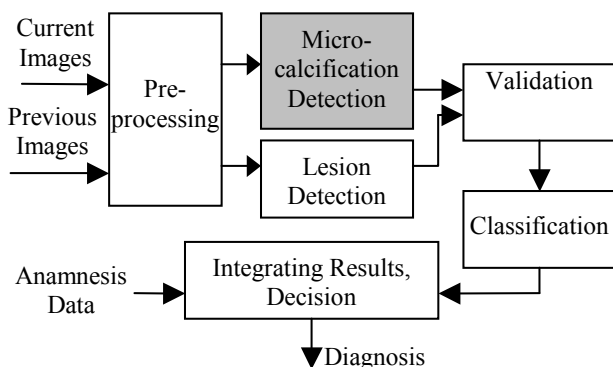
Fig. 1 – A CAD system for mammography.

Current Images: In each mammographic session four pictures are taken. In the first processing steps each image is investigated separately. After some steps the localization and classification of the suspected abnormalities are improved using the two views (CC and ML) of the same breast. Asymmetry of the left and right breasts is investigated at the end using e.g. texture analysis.

Previous Images: If the images of a previous mammographic examination of the same patient are available, then the change of suspected abnormalities is checked.

Anamnesis Data: Several personal parameters of the patient influence the probability of breast cancer development e.g. the family records (were similar diseases in the family or not), the profession etc. Therefore in the final evaluation the diagnosis depends on that data as well.

Preprocessing: There are several important tasks before the diagnostic evaluation of the images can start. Some of the pictures are of poor quality therefore image enhancement methods should be applied. The parenchymal patterns are patient dependant and probably different methods are optimal for the diagnosis of different basic patterns. Therefore in the preprocessing step the basic parenchymal type is determined.

Detection: Several methods of determining the location of suspected abnormalities were suggested, but none of them gives solution for every problem in every tissue. Two types of abnormalities are of primer importance; reliable detection of microcalcifications and lesions must be solved. Several methods were suggested especially for microcalcification detection e.g. neural network based methods, texture analysis, heuristic methods, wavelet transform based methods etc. [4-8]. Some of these procedures can be used for lesion detection as well e.g. neural nets. Plenty of algorithms are available and no single algorithm solves even one of the detection problems perfectly, therefore in the final system several methods will be applied parallel

and the results of them are integrated in a final judgment.

Validation: When suspicious locations are detected in a mammogram, some different algorithms are used to justify or reject the hypothesis. This step is needed because the detection algorithms tend to give 2-5 suspicious locations even in an absolutely normal mammogram. Several parallel methods are used in that step - heuristic methods, some shape description and evaluation procedures for lesions, integration of the findings on the 4 images of the two breasts, comparison of the current images to older ones etc.

Classification: In the classification step a diagnosis is established based on the images only. The diagnosis consists of the findings, and some probability like measure estimating the reliability of the diagnosis.

Decision: In this final step the anamnesis data are integrated to the information extracted from the images and the best possible diagnosis is determined.

3. MICROCALCIFICATION DETECTION WITH A HIERARCHICAL NEURAL NETWORK

One of the most important symptoms is the presence of microcalcifications. These appear as small bright spots (0.1 to 5 mm) in the mammogram.

Fig.2 – Microcalcifications in a mammogram.

Several methods were suggested for microcalcification detection but one of the most promising candidates is the use of neural nets. Since the size, position and shape of the calcifications are different from case to case, analytical mathematical description is not possible. On the other hand lots of data are available for training.

In the proposed CAD system we used a modification of a hierarchical neural architecture, presented in [2], for microcalcification detection. The architecture is based on a multiscale decomposition of an image and a hierarchy of neural networks. The main advantage of the multiscale

approach is the simultaneous examination of small details and large scale context, while neural networks have been proven useful in many image processing applications, especially in detection and classification of various objects. Fig.3 shows the hierarchical architecture and some processing details.

First the input image is sampled and an image pyramid is built, the resolution is halved at each level. Next a number of features are extracted from each image. It turned out that the relevance and quality of the features have a very important role in detection. In the original suggestion [2] only one type of features are used, while in our implementation we tried and finally added several new attributes. The originally used features were direction independent (the position and direction of the calcifications are neither important nor known). In [2] steerable filters were used for feature extraction, which compute local oriented gradient, and can be used to measure orientation having maximum energy at each pixel locations.

We examined and added the following features. Variance:

$$avg(W) = \frac{1}{N_W} \sum_{(m,n)\in W} u(m,n) \qquad (1)$$

$$var(W) = \frac{1}{N_W} \sum_{(m,n)\in W} [u(m,n) - avg(W)]^2 \qquad (2)$$

where u is the input image, W is the processing window with size N_W, *avg* is the average of the pixel intensity values within the window, and *var* is the variance.

Standard deviation:

$$std(W) = \sqrt{\frac{1}{N_W} \sum_{(m,n)\in W} [u(m,n) - avg(W)]^2} \qquad (3)$$

Variance helps to detect strong edges, while standard deviation emphasizes the finer edge structures too. The local difference of maximum and minimum intensity is similar to standard deviation and is defined as:

$$diff_1(W) = \max_{(m,n)\in W} [u(m,n)] - \min_{(m,n)\in W} [u(m,n)] \qquad (4)$$

Difference from local average has proved to be the most useful feature. This feature is concentrates directly on microcalcifications while the others only on edges:

$$diff_2(m,n,W) = u(m,n) - avg(W) \qquad (5)$$

Block activity is a spectral feature [4], where $U(m,n)$ is the Fourier transform of the input $u(m,n)$:

$$A(W) = \sum_{(m,n)\in W} |U(m,n)| - U(0,0) \qquad (6)$$

The extracted features then serve as inputs to the neural networks.

According to [2] the direction of processing in the hierarchy can be chosen in two ways. The fine-to-coarse approach is capable of detecting large scale objects, as lesions, while the coarse-to-fine approach is applicable for the detection of small scale objects like microcalcifications, so we used the latter in our implementation. In Fig.3 the arrow from top to down shows the path in the hierarchy that the coarse-to-fine processing scheme follows.
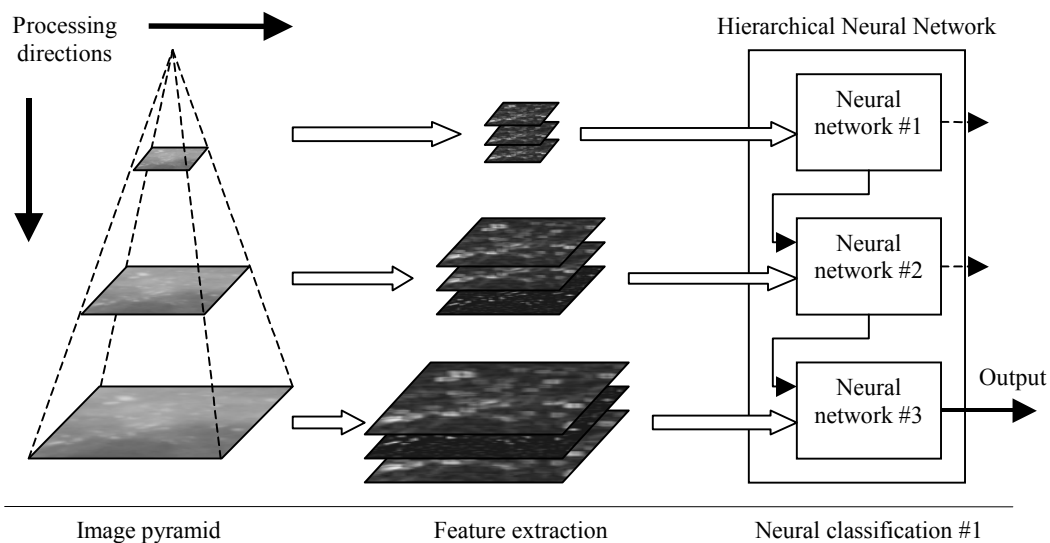

Fig.3 – Hierarchical neural network architecture.

Simple feedforward multilayer perceptrons were used with only one hidden layer. The outputs of the neurons in the hidden layer of each network are connected to the input of the next network in the hierarchy. (see Fig.4)

The output of the neural network at the top level (network#2 in Fig. 4) is the output of the system. The outputs of the lower level networks in the hierarchy are only used in training. The lowest level network (network#1 in Fig. 4) is trained first. The training is based on the input features of that level and the desired outputcompared to output#1. After that the second level network is trained. The training uses the features at second level, the hidden layer outputs of the first level network and the desired output. If more than two levels are used, then this process goes on. Throughout the whole system, the training and processing works pixel by pixel.
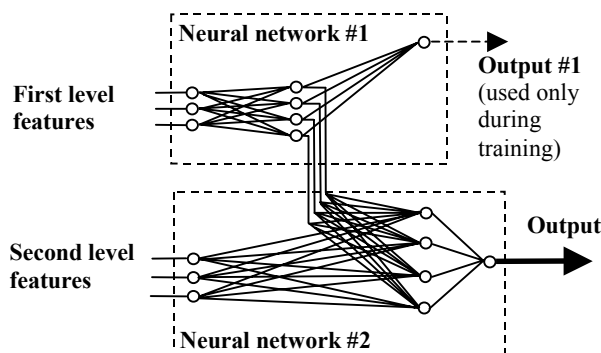

Fig.4 - Hierarchical network structure (shown for 2-level case).

4. A MODULAR ARCHITECTURE

The goal in [2] was to optimize a given CAD system, because it gave too high false positive ratio. The hierarchical neural structure was only used to improve that aspect. We had different motivations. In medical screening practice experts have to evaluate all images, both the positive and negative ones. Our goal is to reliably filter out as much negative case as possible (to decrease the burden of the expert), and highlight the suspicious locations of the remaining pictures (to help the expert). Therefore in our case both the false negative and false positive ratios are extremely important.

On the other hand in our system the method is used as an internal step not as an output filter. Therefore not only the final positive/negative decision about the image is essential, but the locations of the individual calcifications are important as well. The algorithm in our system has to make a positive or negative decision on the individual pixels of the input images too.

The design of our system was guided by some pre-examination of the hierarchical neural network architecture. Our experiments showed that the results are influenced by the composition of the training set, the number and relevance of the features used and the normalization of the network inputs. For example when different training sets or input features were used the resulting network became different. This suggests that we could combine the outputs of some differently trained networks to improve the performance. It has been shown by many researchers [9] that it could be useful to combine the outputs of several networks instead of using only one. The ensemble gives more accurate solution if the networks themselves provide precise output, and they are diverse in such a way that they make errors in different parts of the input space.

In the final system 5 modules of hierarchical networks were used. Each module gets different features of the above-mentioned set providing diversity of the solutions. Good features are independent from each other but correlated to the desired output. The resulting modular system can be seen on Fig.5.

The output of the hierarchical networks provides information pixel by pixel. To decide whether the mammogram is abnormal or not this information must be interpreted. To do this decision trees are used, evaluating for example the number of pixels classified as calcification, the number of connected areas, etc. A voting block integrates the results of the five modules.

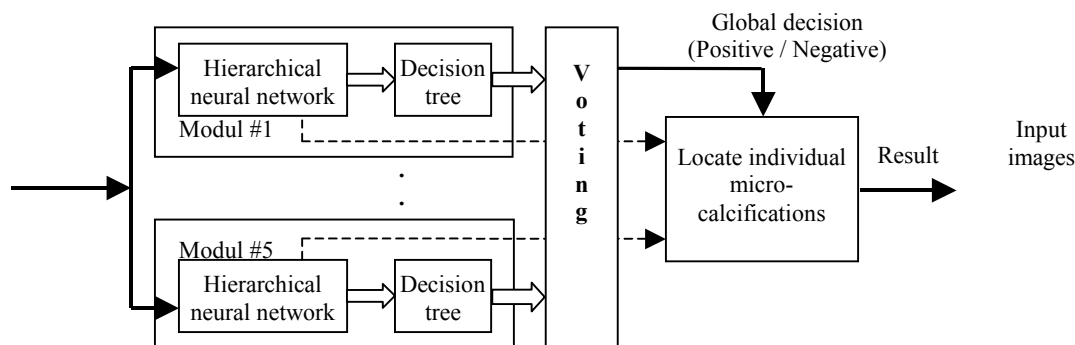The location of the individual calcifications is determined only in cases classified as positive ones.

Fig. 5 – A modular, hierarchical, neural system for microcalcification detection.

If the majority of the modules mark a pixel as calcification in an image decided to be positive then that location is finally marked.

5. EXPERIMENTS AND RESULTS

In our experiments we mainly used images from the MIAS digital mammography database [10]. This database contains 322 images. We used the lower resolution version of 1024x1024 pixels and 8-bit depth (256 gray levels). The database also includes a description of the locations and types of the abnormalities that are present. 25 images contain microcalcifications.

We extracted 430 regions of interest (ROI) from the mammograms each had a size of 64x64 pixel. We used 100 ROIs for training, 74 normal and 26 with microcalcifications. 330 of the ROIs were only used for testing the system, 32 that contain microcalcifications, and 298 normal.

The goal of the first experiment was to determine the performance of the global decision method, that tries to make a decision on the input ROI whether it is positive or negative, whether it contains microcalcification or not. The second experiment focused on the examination of the performance of the system in detecting individual microcalcifications in a pixel level.

Nash voting [11] was used; therefore the first experiment had different outcomes according to the applied threshold in the voting block. (In Nash voting each vote is in the [0.5 1.5] interval, and the final value is the product of the votes.) In our case the minimal value of the product is about 0.03 and the maximal is about 7.5. Table 1. contains the number of false positive and false negative decisions in case of different thresholds.

The best overall result can be measured at a threshold of 2.5, when only 8 ROIs from the 330 was improperly classified. However it means a problem that 5 of the 8 cases are false negative, therefore it would be better to choose 1.85 as a threshold. In this case we have 3 false negatives what is 9.375% of ROIs containing microcalcifications, and 6 false positives that is only 2% of the ROIs. By choosing 0.0625 for the threshold we get a 20% false positive rate at 0 false negative.

Table 1. Results of the global decision method

Threshold	False positive	False negative
0.03125	114	0
0.0625	69	0
0.5	21	2
1	13	2
1.85	6	3
2.5	3	5
6	0	18

The detection of individual calcifications is important from the aspect of the classification task performed by a later processing stage in our CAD system. The classification is based upon the shape of individual calcifications and clusters formed by them. Therefore in the second experiment we examined the number of correctly detected calcifications as well as the number of correctly detected pixels they consist of.

In order to make a reliable assessment of the real system performance, we only ran the detection algorithm on images considered by the global decision method as positive. We have chosen 1.85 for the threshold value; this is the trade-off between false positive and false negative decisions.

The true positive and false positive detection ratios for the individual calcifications are defined as:

$$TPC_C = \frac{\#\,\text{true positive calcifications detected}}{\#\,\text{true positive calcifications to be detected}}$$

$$FPC_C = \frac{\#\,\text{false positive calcifications detected}}{\#\,\text{images}}$$

A detected object is considered as false positive calcification if it does not correspond to any desired output. The FROC (free-response receiver operating characteristic) curve can be seen in Fig. 6.
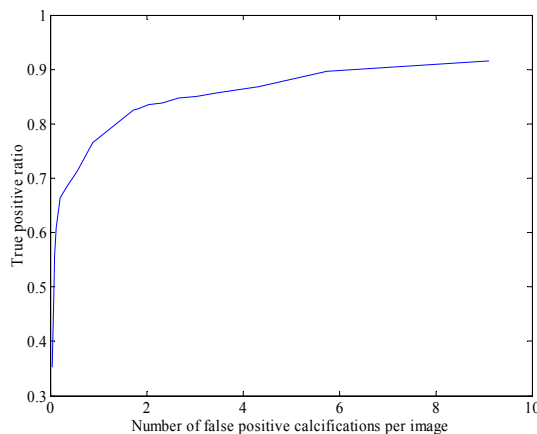

Fig. 6 – FROC curve of individual calcifications detection.

The above stands for pixel detection too. A pixel marked as positive by the algorithm is considered as false positive, if it is not part of any true calcification.

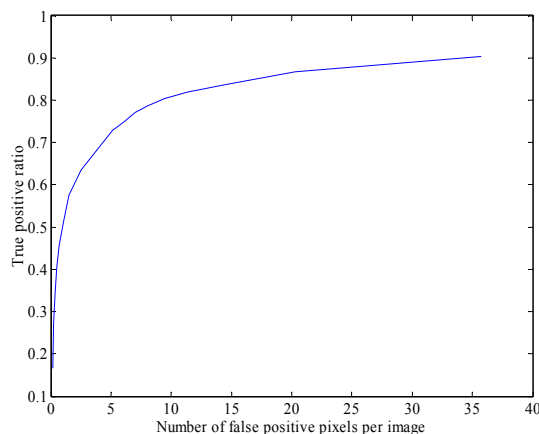The FROC curve for pixel detection is show on Fig. 7.

Fig.7 – FROC curve of pixel detection.

6. CONCLUSION

A complex hybrid system to support the efforts of screening mammography is being developed. One of the key tasks of that system is to detect the presence and location of microcalcifications in the mammographic images. Neural networks seem to be good candidates for that job but several problems must be solved.

Some of the problems were solved by a promising hierarchical neural architecture. In our work the original suggestion was improved in two ways: the feature extraction was extended to use several new attributes, and the neural architecture was embedded in a network ensemble context.

Results achieved show that better false negative, false positive ratio can be realized by the new architecture. The performance of the detection system is acceptable. (No perfect method is known yet; even the best human experts fail to detect all the diagnostically important cases.)

7. ACKNOWLEDGEMENT

8. REFERENCES

[1] R. Highnam, M. Brady. *Mammographic Image Analysis.* Kluwer Academic Publishers R, 1999.

[2] P. Sajda, C. Spence, J. Pearson. A Hierarchical Neural Network Architecture that Learns Target Context: Applications to Digital Mammography. *Proc. 2nd IEEE Intern. Conference on Image Processing*, Washington, D.C., 1995, pp. 149-155.

[3] P. Sajda, C. Spence. Learning Contextual Relationships in Mammograms using a Hierarchical Pyramid Neural Network, *IEEE Transactions on Medical Imaging* 21 (3) (2002).

[4] B. Zheng, W. Qian, L. P. Clarke. Digital mammography: Mixed feature neural network with spectral-entropy decision for detection of microcalcifications, *IEEE Transactions on Medical Imaging*, Vol. 15 (Oct. 1996). pp. 589-597.

[5] N. Karssemeijer. Adaptive noise equalization and recognition of micro-calcification clusters in mammograms, *International Journal of Pattern Recognition and Artificial Intelligence* 7. pp. 1357-1376.

[6] Moti Melloul, Leo Joskowicz, *Segmentation of microcalcification in X-ray mammograms using entropy thresholding*, Technical Report, May 2002, Hebrew University, Leibniz Center.

[7] T. Shubhachint, A.P. Dhawan. Wavelet Based Multi-resolution Algorithm for Detection of Mammographic microcalcifications. *Proceedings IEEE 21th International Conference of Engineering in Medicine & Biology*, October 13-16, 1999.

[8] H. D. Cheng, J. R. Chen, R. I. Freimanis, X. H. Jiang. A novel fuzzy logic approach to microcalcification detection, *Information Sciences*, Vol. 111, Issues 1-4 (November 1998), pp. 189-205.

[9] L. K. Hansen, P. Salamon. Neural network ensembles, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (1990). pp. 993-1001.

[10] J. Suckling et al. The Mammographic Image Analysis Society Digital Mammogram Database, *Exerpta Medica, International Congress Series* 1069, 1994, www.wiau.man.ac.uk/services/MIAS/MIASweb.html pp. 375-378.

[11] H. Normi. *Comparing Voting Systems*. D. Riedel Publishing Company, 1987.

László Lasztovicza received his M.Sc. degree in electrical engineering from Budapest University of Technology and Economics (BUTE) in 2003. He has been a PhD student at BUTE in computer sciences since 2003. In the past 2 years he has carried out research in the field of intelligent systems and methods for diagnostic purposes. He has taken part in a research project at BUTE dealing with the development of a complete diagnostic system for mammography. His main research interests are classification methods for pattern recognition including neural networks, decision trees and subspace methods.

__Béla Pataki__ received his M.Sc. and PhD degree in electrical engineering from Budapest University of Technology and Economics (BUTE) in 1978 and 1997 respectively. Currently he is an associate professor at the Department of Measurement and Information Systems of BUTE. He has taken part in several research projects at BUTE dealing with intelligent modeling and signal processing in complex industrial and medical systems His main research interests are theory and application of neural networks and hybrid intelligent systems.

__Nóra Székely__ received her M.Sc. degree in computer science from Budapest University of Technology and Economics (BUTE) in 2002. She has been a PhD student at BUTE, Dept. of Measurement and Intelligent Systems in computer sciences since 2002. In the past 3 years she has carried out research in the field of intelligent systems and methods for diagnostic purposes. She has taken part in a research project at BUTE dealing with the development of a complete diagnostic system for mammography. Her main research interests are intelligent systems and medical image processing, especially pattern recognition.

__Norbert Tóth__ received his M.Sc. degree in electrical engineering from Budapest University of Technology and Economics (BUTE) in 2003. He has been a PhD student at BUTE, Dept. of Measurement and Intelligent Systems in computer sciences since 2003. In the past 2 years he has carried out research in the field of intelligent systems and methods for diagnostic purposes. He has taken part in a research project at BUTE dealing with the development of a complete diagnostic system for mammography. His main research interests are image processing, texture analysis, pattern recognition, and hybrid intelligent systems for medical diagnosis.