Tumor Recognition in Endoscopic Video Images using Artificial Neural Network Architectures

S.A. Karkanis[1], D.K. Iakovidis[2], D.E. Maroulis[3], G.D. Magoulas[*] and N.G. Theofanous[4]
Dept. of Informatics, University of Athens, TYPA Build. Panepistimiopolis,15784 Athens, GREECE
{sk[1],dmarou[3],optel[4]}@di.uoa.gr, [2]digod@yahoo.com
[*]*Dept of Information Systems and Computing, Brunel University, Uxbridge, UB8 3PH, U K*
George.Magoulas@brunel.ac.uk

Abstract

This paper focuses on a scheme for automated tumor recognition using images acquired during endoscopic sessions. The proposed recognition system is based on multi-layer feed forward neural networks (MFNNs) and uses texture information encoded with corresponding statistical measures that are fed as input to the MFNN. Experiments were performed for recognition of different types of tumors in various images and also a number of sequentially acquired frames. The recognition of a polypoid tumor of the colon in the original image, which were used for training was very high. The trained network was also able to recognize satisfactory the tumor in a sequence of video frames. The results of the proposed approach were very promising and seem that it can be efficiently applied for tumor recognition.

1. Introduction

Medical imaging covers a major application area providing significant assistance in medical diagnosis. The development of these systems leads to valuable diagnostic tools that may largely assist physicians in the identification of tumors or malignant formations.

Systems capable to discriminate among various tumor categories, aim to improve expert's ability to identify abnormal (e.g. cancerous regions) in tissue while decreasing the need for aggressive intervention and enhancing the capability to make accurate diagnosis. Furthermore, with these techniques it is possible to examine a larger area, studying living tissue *in vivo*, possibly at a distance [1], and thus minimize the shortcomings of biopsies, such as discomfort for the patient, delay in diagnosis, and limited number of tissue samples. In this context, the potentials of new imaging principles, such as fluorescence imaging or laser scanning microscopy, are very high.

The main clinical idea behind these developments is early detection of malignant lesions, particularly in stages were local endoscopic therapy is possible. The need for more effective methods of early detection such as those using intelligent systems for medical imaging is obvious. Although advanced technical developments in this field are in progress and seem very promising, however, as yet, clinical results are still pending and ongoing and upgraded research is indispensable to promote the technologies in question and clarify their real potential for clinical use.

In this paper, we present an approach that will result in detection of tumors during an endoscopic procedure. The presented approach is based on the texture information that is estimated for different image regions. This kind of information is represented using corresponding textual features, which are then fed to the MFNN for recognition and characterization purposes. The recognition capability of the proposed approach has been extensively tested in still images and also in a sequence of frames.

In section 2 of this paper, there is a description of the textual analysis used. The recognition system is described in section 3. In section 4, we present the results of experiments performed for various still endoscopic images. In section 5, we present the results of experiments performed with a sequence of frames and finally in section 6 the main conclusions are summarized.

2. Texture analysis

2.1 Texture analysis using statistical descriptors

Co-occurrence matrices [2][3], represent the spatial distribution dependence of the gray levels within an area. Each (i,j)th entry of the matrices, represents the probability of going from one pixel with gray level (i) to another with a gray level (j) under a predefined distance and angle. More matrices are formed for specific spatial

distances and predefined angles. From these matrices, sets of statistical measures are computed (called feature vectors) for building different texture models. We have considered four angles, namely 0^o, 45^o, 90^o, 135^o as well as a predefined distance of one pixel in the formation of the co-occurrence matrices. Therefore, we have formed four co-occurrence matrices. Among the 14 statistical measures, originally proposed by Haralick [2][4], that are derived from each co-occurrence matrix we have considered only four. Namely, angular second moment, correlation, inverse difference, moment and entropy.

- ***Energy - Angular Second Moment***

$$f_1 = \sum_i \sum_j p(i,j)^2$$

- ***Correlation***

$$f_2 = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i*j)p(i,j) - m_x m_y}{\sigma_x \sigma_y}$$

- ***Inverse Difference Moment***

$$f_3 = \sum_i \sum_j \frac{1}{1+(i-j)} p(i,j)$$

- ***Entropy***

$$f_4 = -\sum_i \sum_j p(i,j)\log\left(p(i,j)\right).$$

We have experimentally found that these measures provide high discrimination accuracy, which can be only marginally increased by adding more measures in the feature vector. Thus, using the above mentioned four co-occurrence matrices we have obtained 16 features describing spatial distribution in each window corresponding to a region in which an original image is divided in order to apply the proposed image indexing scheme. The image is raster scanned with sliding windows of M x M dimensions. For each such window we perform analysis based on the co-occurrence matrices.

2.2 Texture analysis based on Discrete Wavelet Transform

The problem of texture classification, aiming at discriminating among various texture classes, is considered in both the time and the wavelet domain, since it has been demonstrated that discrete wavelet transform (DWT) can lead to better texture modeling [17]. Thus, we can better exploit the well-known local information extraction properties of the wavelet signal decomposition as well as features of wavelet denoising procedures [10]. It is expected that this kind of information considered in the wavelet domain should be smooth due to the time-frequency localization properties of the wavelet transform. It is interesting, that only the 2-D Haar wavelet transform, which is considered as a simple one compared with the other wavelet bases, exhibited the expected and desired properties. We have performed a one-level wavelet decomposition of the images, thus resulting in four wavelet channels. As already mentioned, among the three channels 2, 3, 4 (frequency index) the one whose histogram presents the maximum variance, which is the channel that represents the most clear appearance of the changes between the different textures, has been selected for further processing.

The subsequent step in the proposed methodology is to obtain image windows from the selected wavelet channel and the original image of dimensions M×M and 2M×2M respectively. Feature extraction is conducted by using the information that comes from the co-occurrence matrices [2]. Among the 14 statistical measures, originally proposed by Haralick [3], that are derived from each co-occurrence matrix we have considered only four of them. Namely, angular second moment, correlation, inverse difference moment and entropy as these have been described in a previous paragraph.

These measures, as experiments indicated, provide high discrimination accuracy that can be only marginally increased by adding more measures in the feature vector. Using the above mentioned co-occurrence matrices 16 features describing spatial distribution in each window in the wavelet domain have been obtained. For each window in the image of the selected wavelet channel, a feature vector containing 16 features that uniquely characterizes it in the wavelet domain has been formed. For each such window a set of four features has been obtained by calculating the above four mentioned statistical measures. Finally, these 48 component feature vectors form the input vector of the neural classifier.

3. Texture discrimination using artificial neural networks

Scientific interest in models of neuronal networks or *artificial neural networks (ANNs)* mainly arises from their potential ability to perform interesting computational tasks. Nodes, or *artificial neurons*, in neuronal network models are usually considered as simplified models of biological neurons, i.e. real nerve cells, and the connection weights between nodes resemble to synapses between neurons [5].

Advances in ANNs may contribute to the design and development of new computational tools to analyze

multidimensional and multimodal medical images. This holds also in the case of images obtained through minimally invasive imaging procedures, especially when therapy is guided by these images (video-surgery, interventional radiology, guided radiotherapy, etc.).

In medical imaging, ANNs learning from data sets encounters several difficulties, since these sets may be characterized by incompleteness (missing parameter values), incorrectness (systematic or random noise in the data), sparseness (few and/or non-representable records available from the patient), and inexactness (inappropriate selection of parameters for the given task). In principle, ANNs are able to handle these data sets and are mostly used for their pattern matching capabilities and their human-like characteristics (generalization, robustness to noise), in order to assist medical decision-making [11,14]. Furthermore, it is acknowledged that ANNs contribute to the improvement of imaging information and to the development and spread of intelligent systems in medical imaging [6-10], [12-16]. ANN-based intelligent systems strongly depend on the existence of technology that provides computers with high computing performance for processing large amount of information in reasonable time.

The most popular ANN is the so-called multi-layer feed-forward neural network (MFNN). In a MFNN, whose l-th layer contains N_l nodes, (l = 1,...,M), artificial neurons operate according to the following equations:

$$net_j^l = \sum_{i=1}^{N_{l-1}} w_{ij}^{l-1,l} y_i^{l-1},$$
$$y_j^l = f(net_j^l),$$

where net_j^l is, for the j-th neuron in the l-th layer (j = 1,...,N_l), the sum of its weighted inputs. The weights for connections from the i-th neuron at the (l-1) layer to the j-th neuron at the l-th layer are denoted by $w^{l-1,l}_{ij}$, y^l_j is the output of the j-th neuron that belongs to the l-th layer, and the logistic function $f(net^l_j)=(1+\exp(-net^l_j))^{-1}$ is the j-th's neuron non-linear activation function.

Training a MFNN to recognize abnormalities in image regions is typically realized by adjusting the network weights through a gradient descent method following an error correction strategy. In a MFNN this operation corresponds to minimizing the network's learning error:

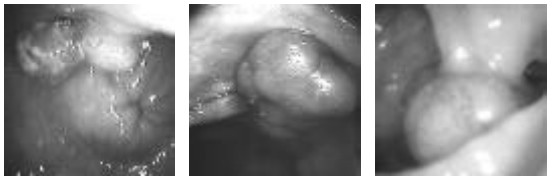
$$E = \frac{1}{2}\sum_{p=1}^{P}\sum_{j=1}^{N_M}(y_{j,p}^M - t_{j,p})^2,$$

where $(y^M_{j,p} - t_{j,p})^2$ is the squared difference between the actual output value at the j-th output layer neuron, for an input sample p, and the target output value; p is an index over input-output patterns. After training, the ANN is able to discriminate between normal and abnormal texture regions by forming hyperplane decision boundaries in the pattern space.

A three-layer FNN with 16 linear inputs and 2 nonlinear outputs were used for the experiments described in this paper. Several variations on hidden layer's non linear neurons, ranging from 10 to 50, were tested. Best results were achieved with a 20-neuron configuration, which is used for the experiments that follow.

4. Experiments on endoscopic images

A number of experiments were conducted using the images illustrated in Fig. 1A, 1B and 1C. Feature extraction was performed using windows of different size 32x32 and 64x64 pixels on each image. The feature set for different experiments was chosen using co-occurrence matrices (16 features) and the DWT (48 features). Figure 2, shows the proportion of normal tissue compared to abnormal tissue data for the images in Fig. 1A, 1B and 1C respectively.



A B C

Fig. 1. Endoscopic images presenting different tumors.

A number of vectors (approx. 400), randomly acquired from each image, were used for the training of various configurations of a three-layer FNN. These variations was of the form I-X-2, where I = {16, 48} inputs and the number of neurons of the hidden layer X ∈ [10, 40]. Test data from each image acquired using a raster scanning sliding window technique. Tables 1, 2a, 2b, 3a, 3b, 4a and 4b contain some of the most interesting results of the experiments. Maximum number of training epochs was set to 40000.

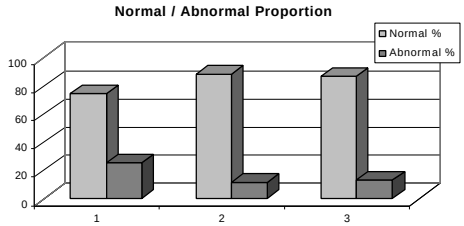


Fig.2. Distribution of normal and possibly abnormal tissues illustrated in fig. 1A, 1B and 1C.

The results show that the 3-layer FNN configuration that performs better, contains 25-30 neurons in the hidden layer. Also, the last image (Fig.1C) seems to be

recognized better than the second (Fig.1B) and the first (Fig.1A). The reconstructed images built using the FNN's output according to the corresponding success rates described in Table 4a, are illustrated in Fig.3.

A B C

Fig. 3. The images of Fig.1 reconstructed, given the output of the FNN in Table 4a.

Comparing the feature sets and the corresponding success rates (Fig.4) achieved using DWT and the co-occurrence features for 64x64 windows is worth noticing that the wavelet transformation results with slightly better performance. It is also worth to notice that the FNN can be trained by one order of magnitude faster using the wavelets (Fig.5). In general, features based on 32x32 windows can possibly perform worse. So, the rest of our experiments are focused on the use of 64x64 window-size.

Co-occurrence - Window 32x32	
Average Recognition Success (%)	89.55
Target Classification Error (%)	**4.5**
X (neurons)	**30**
Training Epochs	36989
Function Evaluations	75303
Image A Success (%)	85.9
Image B Success (%)	89.9
Image C Success (%)	92.8

Table 1. Success rates using co-occurrence matrices and 32x32 windows.

Co-occurrence - Window 64x64	
Average Recognition Success (%)	92.4
Target Classification Error (%)	2.5
X (neurons)	25
Training Epochs	40000
Function Evaluations	80000
Image A Success (%)	92.18
Image B Success (%)	92.18
Image C Success (%)	93.07

Table 2a. Success rates using co-occurrence matrices and 64x64 windows.

Co-occurrence - Window 64x64	
Average Recognition Success (%)	91.02
Target Classification Error (%)	**4.5**
X (neurons)	**30**
Training Epochs	23779
Function Evaluations	46861
Image A Success (%)	90.7
Image B Success (%)	90.18
Image C Success (%)	92.13

Table 2b. Success rates using co-occurrence matrices and 64x64 windows, for Table-1's settings.

DWT - Window 32x32	
Average Recognition Success (%)	86.57
Target Classification Error (%)	8.86
X (neurons)	30
Training Epochs	18131
Function Evaluations	35453
Image A Success (%)	82.16
Image B Success (%)	88.49
Image C Success (%)	89.07

Table 3a. Success rates using DWT and 32x32 windows.

DWT - Window 32x32	
Average Recognition Success (%)	86.28
Target Classification Error (%)	**4.5**
X (neurons)	**30**
Training Epochs	18149
Function Evaluations	35478
Image A Success (%)	82.79
Image B Success (%)	87.87
Image C Success (%)	88.18

Table 3b. Success rates using DWT and 32x32 windows, for Table-1's settings.

DWT - Window 64x64	
Average Recognition Success (%)	93.3
Target Classification Error (%)	1
X (neurons)	30
Training Epochs	2936
Function Evaluations	5776
Image A Success (%)	92.95
Image B Success (%)	92.89
Image C Success (%)	94.07

Table 4a. Success rates using DWT and 64x64 windows.

DWT - Window 64x64	
Average Recognition Success (%)	91.46
Target Classification Error (%)	**4.5**
X (neurons)	**30**
Training Epochs	866
Function Evaluations	1768
Image A Success (%)	91.91
Image B Success (%)	91.01
Image C Success (%)	91.48

Table 4b. Success rates using DWT and 64x64 windows, for Table-1's settings.

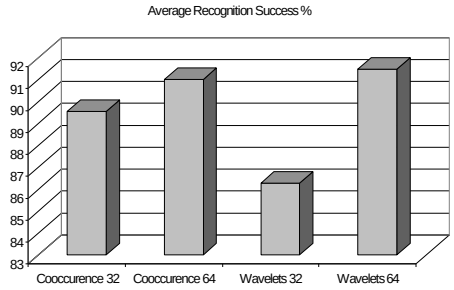


Fig.4. Comparison between the success rate of different feature sets for the FNN in table 1.

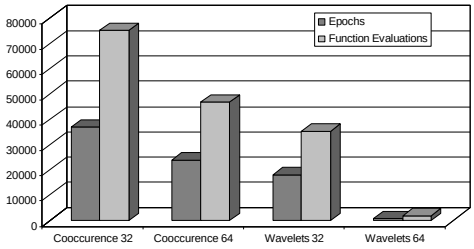


Fig.5. The number of epochs and function evaluations for needed for the training of the FNN in table 1, using different feature sets.

5. Experiments on sequences of endoscopic images

VHS-type videotapes exhibiting colonoscopic examinations, were used for the acquisition of the sequences of frames as shown in Figs.6 and 7. The window size was defined at 64x64 pixels, as this has been concluded by previous study. It is important to notice that these frames have not been through any pre-processing procedure, in order to study MFNN's performance in conditions closer to reality.

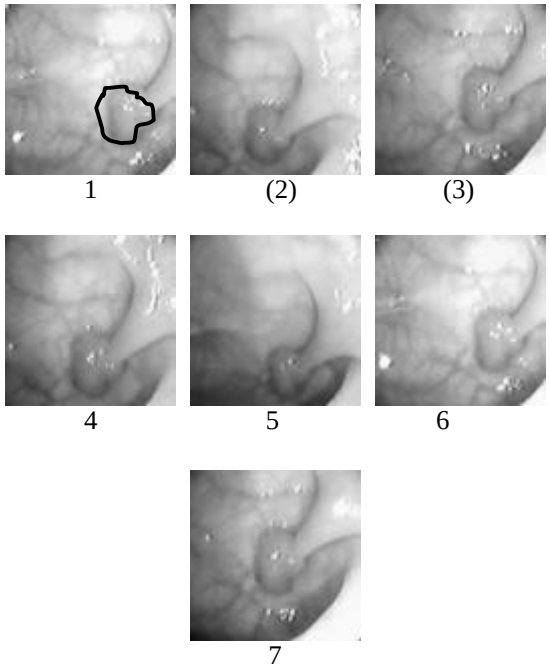


Fig.6. Sequence of 7 frames illustrating a polypoid tumor of the colon. Frames 2 and 3 where used for training the MFNN.

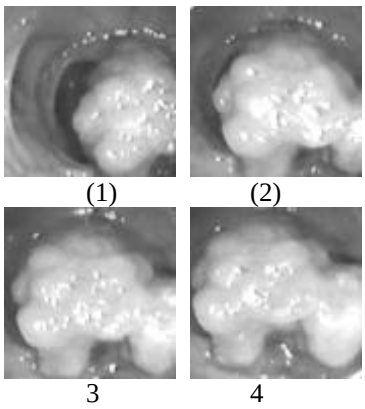


Fig.7. Sequence of frames illustrating another polypoid tumor of the colon. Frames 1 and 2 where used for training the MFNN.

The first experiment was performed using co-occurrence features on the frames illustrated in Fig.6. Abnormal to normal proportion in the in the training sample was approximately 1/10. The architecture of the network that tried was comprised by a 3-layer FNN with 20 neurons in the hidden layer. The area defined as tumor to the FNN had a round shape as outlined in frame-1 (Fig.6). Tests were performed using all frames from 1 to 7 and these corresponded to different training classification error goals of the MFNN as these are shown in Fig.8 (upper). When the network was trained using the patterns

existed in second frame (Fig. 6.2), then the MFNN resulted to percentage of 97.5% of success, but the success percentage recognizing the rest of frames was tolerable. The overall performance exceeded 89%. Interesting results also occurred using a training set consisted of textures from both frames 2 and 3. As it can be seen in Fig.8 (lower), success rate distribution has been significantly differentiated.

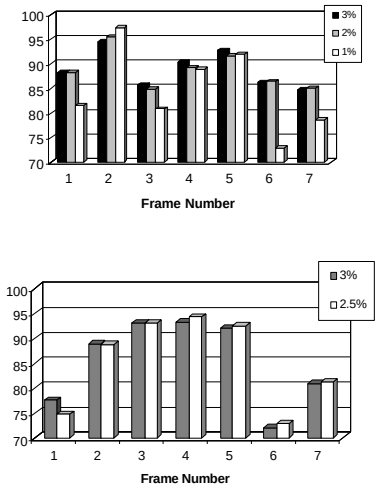


Fig. 8. Percentage of successfully recognized patterns per frame, after training the MFNN with frame-2 (upper) and with frame 2 and 3 (lower), for various target classification errors.

The overall performance achieved can be evaluated as satisfactory by the experts since that tumor's size was very small compared to the window size used for feature extraction (approximately double compared to window size). The statistical information that acquired from abnormal areas is possibly insufficient. Additionally, illumination conditions could significantly influence the statistical information extracted from each image and degrading MFNN's ability to recognize the tumor in different frames.

Another experiment was performed using the wavelet transformation for the frames illustrated in Fig.7. Once again the 48-20-2 MFNN's structure has given the highest success rates. Training with patterns coming from frame-1 (Fig. 1.1), and setting the target classification error at 1.2%, the recognition success for each frame of Fig.7 respectively, is illustrated in Fig.9. The same distribution of recognition success was being repeated in almost all the experiments for various MFNN structures and training classification errors. It is also worth to notice that the success rates increased uniformly not only for frame-1 but also for the rest of the frames, as the target classification error during training decreased.

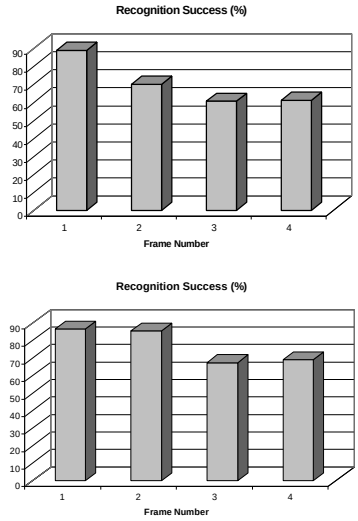


Fig. 9. Percentage of successfully recognized patterns per frame, after training the MFNN with frame-1 (upper) and with frame 1 and 2 for (lower), for various target classification errors.

The behavior of the 3-layer FNN configurations that have been trained with both frames 1 and 2, was almost the same, and the network was able to recognize those frames better. The highest results achieved using 25 hidden neurons, for a target classification error of 0.8%, as illustrated in Fig.9 (lower).

6. Conclusions

In general, the results obtained indicate that the proposed scheme is capable of detecting successfully various types of tumors in single images and in sequences of frames with a success rate in recognition that is accepted by the experts. Current results appear to be promising and several pre-processing techniques and MFNN configurations are being tested to optimize overall recognition performance for sequences of frames. Major improvements on the techniques will be focused on the preprocessing stages in order to eliminate influences coming from the illumination conditions during the endoscopic procedure. The different appearance of the normal tissues as well as the variations on the possible abnormal tissues will be studied by extensive experimentation using data from different sources. A database is being built as a reference tool in which the user will be able to maintain the various cases. Such cases will then be used for the training of the system. A major target is to setup a system that will be guided by the expert during the endoscopy. It will be capable to warn the doctors for small regions that correspond to suspicious tissues.

7. References

[1] Delaney, PM, Papworth, GD, King, RG. Fibre optic confocal imaging (FOCI) for in vivo subsurface microscopy of the colon. In: Preedy VR and Watson RR, eds. *Methods in disease: Investigating the Gastrointestinal Tract*. London: Greenwich Medical Media, 1998.

[2] Haralick, R. M., Shanmugam, K. and Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Trans. Systems, Man and Cybernetics*, 3, 6, 610-621.

[3] Gotlieb C.C., Kreyszig. Texture descriptors based on co-occurrence matrices. *Comp. Vision, Graph. and Image Proc*. 1990; 51: 70-86.

[4] Haralick, R. M. (1979). Statistical and structural approaches to texture. *IEEE Proceedings*, 67, 786-804.

[5] Durbin R, Miall C, Mitchison G. *The Computing Neuron.* Reading: Addison-Wesley, 1989.

[6] Coppini G, Poli R, Valli G. Recovery of the 3-D shape of the left ventricle from echocardiographic images. *IEEE Transactions on Medical Imaging* 1995; 14: 301-317.

[7] Hanka R, Harte TP, Dixon AK, Lomas DJ, Britton PD. Neural networks in the interpretation of contrast-enhanced magnetic resonance images of the breast. In: *Proceedings of Healthcare Computing.* Harrogate: UK, 1996: 275-283.

[8] Ifeachor EC, Rosen KG, eds. *Proceedings of the International Conference on Neural Networks and Expert Systems in Medicine and Healthcare.* Plymouth: UK, 1994.

[9] Innocent PR, Barnes M, John R. Application of the fuzzy ART/MAP and MinMax/MAP neural network models to radiographic image classification. *Artif. Intell. in Med.* 1997; 11: 241-263.

[10] Karkanis S, Magoulas GD, Grigoriadou M, Schurr M. Detecting abnormalities in colonoscopic images by textural description and neural networks. In: *Proc. of Work. on Mach. Learn. in Med. Appl., Advance Course in Artif. Intell.-ACAI99.* Chania: Greece, 1999: 59-62.

[11] Lim CP, Harrison RF, Kennedy RL. Application of autonomous neural network systems to medical pattern classification tasks. *Artificial Intelligence in Medicine* 1997; 11: 215-239.

[12] Miller AS, Blott BH, Hames TK. Review of neural network applications in medical imaging and signal processing. *Medical and Biological Engineering and Computing* 1992; 30: 449-464.

[13] Phee SJ, Ng WS, Chen IM, Seow-Choen F, Davies BL. Automation of colonoscopy part II: visual-control aspects. *IEEE Engineering in Medicine and Biology* May/June 1998: 81-88.

[14] Reategui EB, Campbell JA, Leao BF. Combining a neural network with case-based reasoning in a diagnostic system. *Artificial Intelligence in Medicine* 1996; 9: 5-27.

[15] Veropoulos K, Campbell C, Learmonth G. Image processing and neural computing used in the diagnosis of tuberculosis. In: *Colloq. Intelligent Meth. in Health. and Med. Appl.* York: UK, 1998.

[16] Zhu Y, Yan H. Computerized tumor boundary detection using a Hopfield neural network. *IEEE Tr. on Medical Imaging* 1997; 16: 55-67.

[17] Meyer Y., *Wavelets: Algorithms and Applications*, Philadelphia: SIAM, 1993.