

Tumor detection in colonoscopic images using hybrid methods for on-line neural network training

V.P. Plagianakos^{1,3}, G.D. Magoulas^{2,3} and M.N. Vrahatis^{1,3}

¹Department of Mathematics,
University of Patras,
GR-26110 Patras, Greece.

²Centre for Health Informatics and
Computing, Department of Information
Systems and Computing, Brunel University,
Uxbridge UB8 3PH, UK.

³University of Patras Artificial Intelligence Research Center (UPAIRC).

ABSTRACT: In this paper the effectiveness of a new Hybrid Evolutionary Algorithm in on-line Neural Network training for tumor detection is investigated. To this end, a Lamarck-inspired combination of Evolutionary Algorithms and Stochastic Gradient Descent is proposed. The Evolutionary Algorithm works on the termination point of the Stochastic Gradient Descent. Thus, the method consists in a Stochastic Gradient Descent-based on-line training stage and an Evolutionary Algorithm-based retraining stage. On-line training is considered eminently suitable for large (or even redundant) training sets and/or networks; it also helps escaping local minima and provides a more natural approach for learning nonstationary tasks. Furthermore, the notion of retraining aids the hybrid method to exhibit reliable and stable performance, and increases the generalization capability of the trained neural network. Experimental results suggest that the proposed hybrid strategy is capable to train on-line, efficiently and effectively. Here, an artificial neural network architecture has been successfully used for detecting abnormalities in colonoscopic video images.

1. INTRODUCTION

Artificial neural networks (ANNs) provide to computing an alternative algorithmic model, which is biologically motivated: the computation is massively distributed and parallel, and the learning replaces a priori program development, i.e. ANNs develop their functionality based on training (sampled) data. The ANN approach to medical information processing has several benefits, such as:

- Training by examples instead of rules
- Learning from experience
- Generalization to new test data
- Reduction of the number of false alarms, without increasing significantly the number of false negatives
- Automation of the learning process
- Eliminating issues associated with human fatigue and habituation
- Rapid identification of problematic cases
- Analysis of conditions and diagnosis in real time.

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. 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. 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.

In this paper, a new hybrid Evolutionary Algorithm (EA) combined with on-line ANN training is investigated. The proposed method is used for tumor detection in colonoscopic video images. In the first stage of the proposed methodology, a recently proposed on-line learning algorithm [11] is employed to train the ANN. In the second stage, EAs [14] are used for retraining the network. The usage of EAs is based on the fact that the first stage has produced a population of potential solutions and hence the EA is less sensitive to the nonstationaries of the task.

The rest of the paper is organized as follows: batch and on-line ANN training are discussed in Section 2, while a short introduction to EAs is given in Section 3. Section 4 describes the special task of interpreting endoscopic images, while Section 5 presents details on the application of the new hybrid method in training ANNs in interpreting colonoscopic images and outlines the implementation results. Finally, in Section 6, conclusions and a short discussion of future work are presented.

2. BATCH vs. ON-LINE NEURAL NETWORK TRAINING

Learning in ANNs is usually achieved by minimizing the network's error, which is a measure of its performance and is defined as the difference between the actual output vector of the network and the desired one. This approach is very popular for training artificial neural networks and includes training algorithms that can be divided in two categories: *stochastic*, also called *on-line*, and *batch*, also called *off-line*.

The batch training of ANNs is considered as the classical machine learning approach: a set of examples is used for learning a good approximating function, i.e. train the ANN, before the network is used in the application. Batch training is consistent with the theory of unconstrained optimization, since the information from all the training set is used. Thus, the aim is to find a minimizer $w^* = (w_1^*, w_2^*, \dots, w_n^*) \in \mathbb{R}^n$, such that:

$$w^* = \min_{w \in \mathbb{R}^n} E(w)$$

where E is the batch error measure of an ANN, whose l -th layer ($l = 1, \dots, M$) contains N_l neurons

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

In the above relation, $(y_{j,p}^M - t_{j,p})^2$ is the squared difference between the actual output value at the j -th output layer neuron for pattern p and the target output value; p is an index over input-output pairs.

The rapid computation of such a minimizer is a rather difficult task since, in general, the dimension of parameter space is high and the error function generates a complicated surface in this space, possessing multitudes of local minima and having broad flat regions adjoined to narrow steep ones that need to be searched to locate an "optimal" weight set.

On the other hand, in on-line training, the ANN parameters are updated after the presentation of each training example, which may be sampled with or without repetition. On-line training may be the appropriate choice for learning a task either because of the very large (or even redundant) training set, or because of the slowly time-varying nature of the task. Although batch training seems faster for small-size training sets and networks, on-line training is probably more efficient for large training sets and ANNs. It helps escaping local minima and provides a more natural approach for learning non-stationary tasks. On-line methods seem to be more robust than batch methods as errors, omissions or redundant data in the training set can be corrected or ejected during the training phase. Additionally, training data can often be generated easily and in great quantities when the system is in operation, whereas they are usually scarce and precious before. Lastly, on-line training is necessary in order to learn and track time varying functions and continuously adapt in a changing environment. As Sutton pointed out [22], "on-line learning is essential if we want to obtain *learning* systems as opposed to merely *learned* ones".

Given the inherent efficiency of stochastic gradient descent, various schemes have been recently proposed [1,18,19,20,22]. However, these schemes suffer from several drawbacks such as sensitivity to learning parameters [18]. Note that in this

framework it is not possible to use advanced optimization methods, such as conjugate gradient, variable metric, simulated annealing etc., as these methods rely on a fixed error surface [18].

So, despite the abundance of methods for learning from examples, there are only few that can be used effectively for on-line learning. For example, the classic batch training algorithms cannot straightforwardly handle nonstationary data. Even when some of them are used in on-line training there exists the problem of "catastrophic interference", in which training on new examples interferes excessively with previously learned examples leading to saturation and slow convergence [23]. Methods suited to on-line learning are those that can handle nonstationary (time-varying) data, while at the same time, require relatively little additional memory and computation in order to process one additional example.

3. EVOLUTIONARY ALGORITHMS

In the second stage of the proposed hybrid algorithm, an evolutionary algorithm is employed on the termination point of the on-line training method. *Evolutionary Algorithms* (EAs) are stochastic search methods that mimic the metaphor of natural biological evolution. EAs operate on a population of potential solutions applying the principle of survival of the fittest to produce better and better approximations to a solution. At each generation, a new set of approximations is created by the process of selecting individuals according to their level of fitness in the problem domain and breeding them together using operators borrowed from natural genetics.

To demonstrate the efficiency of the EAs, we have used *Differential Evolution* (DE) strategies to train ANNs [12,21]. DE strategies can handle non differentiable, nonlinear and multimodal objective functions efficiently, and require few easily chosen control parameters. Experimental results have shown that DE strategies have good convergence properties and outperform other evolutionary algorithms [14]. To apply DE strategies to ANN training we start with a specific number (NP) of n -dimensional weight vectors, as an initial weight population, and evolve them over time; NP is fixed throughout the training process and the weight population is initialized randomly following a uniform probability distribution.

At each iteration, called *generation*, new weight vectors are generated by the combination of weight vectors randomly chosen from the population. This operation is called *mutation*. The outgoing weight vectors are then mixed with another predetermined weight vector - the target vector - and this operation is called *crossover*. This operation yields the so-called trial vector. The trial vector is accepted for the next generation if and only if it reduces the value of the error function E . This last operation is called *selection*. We now briefly review the two basic DE operators.

The first DE operator used is the *mutation operator*. Specifically, for each weight vector, a new vector called mutant vector is generated according to the following relation:

$$\text{Mutant_Vector} = w_i^p + \xi(w_{best} - w_i^p) + \xi(w_{r_1} - w_{r_2}),$$

where w_{best} is the best member of the previous generation, $\xi > 0$ is a real parameter called mutation constant and controls the amplification of the difference between two weight vectors,

and w_1, w_2 are two randomly chosen weight vectors, different from w_i^p . To increase further the diversity of the mutant weight vector, the *crossover operator* is applied. Specifically, for each component j , ($j=1,2,\dots, n$), of the mutant weight vector, we randomly choose a real number r from the interval $[0,1]$. Then, we compare this number with $\rho > 0$ (crossover constant), and if $r \leq \rho$ we select, as the j -th component of the trial vector, the corresponding component j of the mutant vector. Otherwise, we pick the j -th component of the target vector.

The above operators introduce diversity in the population and are used to help the algorithm escape the local minima in the weight space. The combined action of mutation and crossover is responsible for much of the effectiveness of DE's search, and allows them to act as parallel, noise-tolerant hill-climbing algorithms, which efficiently search the whole weight space.

4. INTERPRETATION OF ENDOSCOPIC IMAGES

Neural networks have been increasingly used in medicine and especially in the development of neural expert systems for intelligent medical image interpretation [3,5,6,8,13,24]. In most cases, the development of such systems is considered an attempt to emulate the doctor's expertise in the identification of malignant regions in minimally invasive imaging procedures (for example, computed tomography, ultrasonography, endoscopy, confocal microscopy, computed radiography, or magnetic resonance imaging). The objective is to increase the expert's ability to identify cancer regions while decreasing the need for intervention and maintaining the ability for accurate diagnosis. Furthermore, it may be possible to examine a larger area, studying living tissue *in vivo* - possibly at a distance [2] - and thus minimize the shortcomings of biopsies, such as a limited number of tissue samples, a delay in diagnosis, and discomfort for the patient. The need for more effective methods of early detection - such as those that computer assisted medical diagnosis systems aim to provide - is obvious.

In technical terms, the problem in automatic image interpretation is to associate sets of pixels (structures) in an image with the unknown objects that are present in the scene from which the image has been drawn. The difficulty increases when several objects of different kinds, related by a set of spatial-temporal relations, are present in the observed scene. In medical practice, endoscopic approaches and other minimally invasive techniques (for example, computed tomography and magnetic resonance imaging) are now permitting visualization of previously inaccessible regions of the body. In diagnostic endoscopy, the medical expert, based on a distributed percept of local changes, interprets the physical surface properties of the tissue - such as the roughness or the smoothness, the regularity, and the shape - to detect abnormalities. Adjacent surfaces showing different surface properties are distinguished on the basis of the texture differences.

It is important to note, however, the vast difficulties in physical attributes of the organs. For example, in colonoscopy, no two colons are alike. Even within the same colon, one section may have very different characteristics from another. This fact introduces severe limitations in the use of computer-assisted endoscopy for interpreting colonoscopic images [7]. Given a medical image, the 'true' features associated with the physical

surface properties of the tissue are not exactly known to the image-interpretation system developer. Usually, one or more feature-extraction models [10] are used to provide values for each feature's parameters. The findings are then used to infer the correct interpretation. On this same task of interpretation on the basis of local changes on the properties of the tissue under examination, the performance of human perception is considered outstanding. Furthermore, medical experts have the ability to either add or remove components from an image to give meaning to what they see. Medical experts can also adapt to changes to the extent that even a distorted image can be recognized.

Neural network methodologies present some human-like qualities, such as learning from experience, generalization, and handling uncertainty and ambiguity in distorted or noisy images. Thus, such methods provide human experts with significant assistance in medical diagnosis [6,8,13,24].

5. IMPLEMENTATION AND RESULTS

The interpretation of diagnostic medical images is usually quite sophisticated and involves multiple levels of processing. To provide a common platform for studying the various problems of medical-image-based diagnostics, a three-level model is employed as shown in Figure 1 (adapted from [9,p.95]).

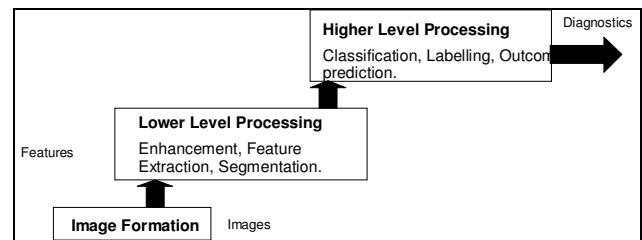


Figure 1: Model for diagnostic system using medical images.



Figure 2: A frame of the video sequence illustrating a polypoid tumor of the colon.

Hybrid On-line Training Algorithm	
Stage 1	<p>Step 0: Initialize the weights w^0, η^0, and the meta-stepsize K.</p> <p>Step 1: Repeat for each input pattern p</p> <p>Step 2: Calculate $E(w^p)$ and then $\nabla E(w^p)$.</p> <p>Step 3: Update the weights: $w^{p+1} = w^p - \eta^p \nabla E(w^p).$</p> <p>Step 4: Calculate the stepsize to be used with the next pattern $p + 1$: $\eta^{p+1} = \eta^p + K \langle \nabla E(w^{p-1}), \nabla E(w^p) \rangle.$</p> <p>Step 5: Until the termination condition is met.</p> <p>Step 6: Return the final weights w^{p+1} to the Stage 2.</p>
Stage 2	<p>Step 0: Initialize the DE population, in the neighborhood of w^{p+1}.</p> <p>Step 1: Repeat for each input pattern p</p> <p>Step 2: For i = 1 to NP</p> <p>Step 3: MUTATION(w_i^p) → Mutant_Vector.</p> <p>Step 4: Crossover(Mutant_Vector) → Trial_Vector.</p> <p>Step 5: If $E(\text{Trial_Vector}) \leq E(w_i^p)$, accept Trial_Vector for the next generation.</p> <p>Step 6: EndFor</p> <p>Step 7: Until the termination condition is met.</p>

Algorithm 1

Image processing is separated into two levels: the lower-level processing and the higher-level processing. The lower-level processing takes image pixels as input and performs various tasks such as image enhancement, feature extraction and image segmentation. The higher-level processing takes the output from the lower-level processing as input and generates output related to medical diagnostics. Tasks accomplished in the higher-level processing include classification of features, detection of specific lesions and diagnosis for various abnormalities.

In this application example we will focus on computer-assisted endoscopy for interpreting colonoscopic images. As already mentioned, colonoscopy is a minimal invasive technique for the production of medical images. A narrow pipe like structure, an endoscope, is passed into the patient's body. Video endoscopes have small cameras in their tips, when passed into a body, what the camera observes is displayed on a television monitor (see Figure 2 for a sample frame of the video sequence). The physician controls the endoscope's direction using wheels and buttons.

An important stage of the implementation is the feature extraction process. In our experiments we have used a technique known as co-occurrence matrices to generate features, as illustrated in Figure 3. More specifically, the endoscopic image was separated into windows of size 16 pixels by 16 pixels. Then the co-occurrence matrices algorithm was used to gather information regarding each pixel in an image window. Cooccurrence matrices [4] represent the spatial distribution and the dependence of the gray levels within a local area. Based on these matrices, sets of statistical measures are computed for different angles. Four angles were considered, as well as a predefined distance of one pixel. The following four statistical measures, provide high discrimination accuracy [4]: *Energy-Angular Second Moment, Correlation, Inverse Difference Moment, Entropy* and were used to extract the feature vectors. The elements of these feature vectors are the data that is to be

presented to the ANN in order to train it. For a full explanation see [8].

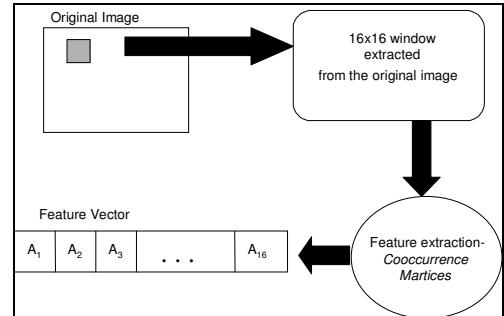


Figure 3: Feature extraction process.

So in our experiments, the feature vectors contain sixteen elements each and therefore the first layer of the ANNs will consist of sixteen neurons.

A high-level description of the proposed algorithm, is given in Algorithm 1. First, the stochastic gradient descent is outlined in the Stage 1 of Algorithm 1, where η is the stepsize, K is the meta-stepsize and $\langle \cdot, \cdot \rangle$ stands for the usual inner product in \Re^n . The memory-based calculation of the stepsize, in Step 4, takes into consideration previously computed pieces of information to adapt the stepsize for the next pattern presentation. This seems to provide some kind of stabilization in the calculated values of the stepsize, and helps the stochastic gradient descent to exhibit fast convergence and high success rate. Note that the classification error or an upper limit to the error function evaluations can be used as the termination condition in Step 5. The key features of the on-line method are the low storage requirements and the inexpensive computations. Moreover, in order to calculate the stepsize to be used at the next iteration, this on-line algorithm uses information from the current, as well as the previous iteration.

In Stage 2 of Algorithm 1, the DE algorithm, responsible for the retraining is outlined. No operation for tuning the mutation and crossover constants was carried out; the fixed values $\xi = 0.5$ and $\rho = 0.7$ have been used.

In the first stage of the algorithm, an ANN having 16 inputs, 30 hidden and 2 output nodes was initially trained to discriminate between normal and abnormal image regions using 300 randomly selected patterns from the first frame. The method used is the stochastic training method introduced in [11]. The training procedure stopped when the ANN exhibited 3% misclassifications on the training set. At this moment, the generalization capability of the ANN at the whole first frame was 83.77%. It must be noted that the first stage was extremely fast; approximately 40 training epochs were needed.

In the second stage, the same ANN architecture has been retrained using the DE algorithm [14]. The DE population has been initialized with weight vectors in the neighborhood of the weight vector found after the first stage had been completed. The new training set consisted of 1200 patterns; the 300 patterns that were selected from the first frame, plus 900 patterns randomly selected from other three video frames of the same sequence. The DE algorithm is allowed to perform only two iteration with each pattern. This was necessary to prevent the “catastrophic interference” between the patterns of the different training sets.

To test the performance of the trained ANN approximately 4000 patterns have been created from each frame. These test sets constitute the whole image region in the frame and contain normal and abnormal samples. After the retraining phase, it is easy to visually pinpoint the actual location of the tumor. The generalization results without and without retraining, are exhibited in Table 1.

As shown in Table 1, by training the ANN with on-line backpropagation using data extracted from the frame 3 and testing it on the whole frame a recognition success of 82.84% was achieved. On the other hand, the hybrid method by applying retraining provides a percentage of 93.09%.

	Without Retraining	With Retraining
Frame 1	83.77%	91.91%
Frame 2	77.18%	83.57%
Frame 3	82.84%	93.09%
Frame 4	87.60%	89.24%

Table 1: Generalization Results

6. DISCUSSION AND CONCLUSIONS

Research in computer-assisted interpretation of endoscopic images to-date is centred on technological issues and is mostly application driven.

Towards this direction, a new hybrid method for on-line neural network training has been developed, tested and applied to tumor detection in colonoscopic video images. Simulation results suggest that the new method exhibits fast and stable

learning, good generalization and therefore a great possibility of good performance. The proposed algorithm is able to train large networks using on-line data, and is better suited for tasks with large, redundant or slowly time-varying training sets, such those of medical imaging. In general, the results obtained indicate that the proposed hybrid algorithm is capable of detecting successfully tumors in single images, as well as, in sequences of frames with an acceptable recognition success rate. Further work is needed to optimize the hybrid algorithm performance, as well as to test it on even bigger training sets from different endoscopic video images.

Our ultimate aim is to incorporate this learning mechanism to a prototype intelligent system for interpreting endoscopic images. At this point it is useful to mention that previous research and experience suggests that the successful implementation of computerised systems [16], and decision support systems in particular [17], in the area of healthcare relies on the successful integration of the technology with the organisational and social context within which it is applied. Therefore, the successful implementation of intelligent medical image interpretation systems should not only rely on their technical feasibility and effectiveness but also on organisational and social aspects that may arise from their applications, as clinical information is acquired, processed, used and exchanged between professionals [15]. All these issues are critical in healthcare applications because they ultimately reflect on the quality of care provided.

ACKNOWLEDGEMENTS

The authors greatly acknowledge the contribution of Dr. S. Karkanis, Dr. D. Maroulis and Mr. D. Iakovidis of the *Image Processing Lab* at the Department of Informatics and Telecommunications, University of Athens, Greece, in the acquisition and extraction of the medical data.

REFERENCES

1. Almeida, L.B., Langlois, T., Amaral, J.D., Plankhov, A., Parameter adaptation in Stochastic Optimization, In: *Online Learning in Neural Networks*, Saad, D. (Ed.), 111-134, Cambridge University Press, (1998).
2. Delaney, P.M, Papworth, G.D. and King, R.G., Fibre optic confocal imaging (FOCI) for in vivo subsurface microscopy of the colon, in: *Methods in disease: Investigating the gastrointestinal tract*, Preedy V.R. and Watson R.R. (Ed.), Greenwich Medical Media, London, UK, (1998).
3. Hanka, R., Harte, T.P., Dixon, A.K., Lomas, D.J. and Britton, P.D., Neural networks in the interpretation of contrast-enhanced magnetic resonance images of the breast, in: *Proceedings of Healthcare Computing 1996*, Harrogate, UK, 275-283, (1996).
4. Haralick RM., Statistical and structural approaches to texture. *IEEE Proc.*, **67**, 786-804, (1979).
5. Ifeachor, E.C. and Rosen, K.G., *Proceedings of the international conference on neural networks and expert systems in medicine and healthcare*, Plymouth, UK, (1994).

6. Innocent, P.R., Barnes, M. and John, R., Application of the fuzzy ART/MAP and MinMax/MAP neural network models to radiographic image classification, *Artificial Intelligence in Medicine*, **11**, 241–263, (1997).
7. Kwoh, C.K., *Probabilistic reasoning from correlated objective data*, Ph.D. Thesis, Imperial College, London, UK, (1995).
8. Karkanis, S., Magoulas, G.D. and Theofanous N., Image recognition and neuronal networks: Intelligent systems for the improvement of imaging information, *Minimal Invasive Therapy & Allied Technologies*, **9**, 225–230, (2000).
9. Leondes, C.T., *Image Processing and Pattern Recognition*, Neural Network Systems Techniques and Applications Series, 5 , Academic Press, (1998).
10. Looney, C.G., *Pattern recognition using neural networks*, Oxford University Press, Oxford, UK, (1997).
11. Magoulas, G.D., Plagianakos, V.P., and Vrahatis, M.N., Adaptive stepsize algorithms for on-line training of neural networks, *Journal of Nonlinear Analysis Theory and Applications*, in press, (2001).
12. Michalewicz, Z. and Fogel, D.B., *How to solve it: Modern Heuristics*, Springer, (2000).
13. Phee, S.J., Ng, W.S., Chen, I.M., Seow-Choen, F. and Davies, B.L., Automation of colonoscopy part II: Visual-control aspects, *IEEE Engineering in Medicine & Biology*, 81–88, (1998).
14. Plagianakos, V.P. and Vrahatis, M.N., Training Neural Networks with Threshold Activation Functions and Constrained Integer Weights, *In Proc. of the IEEE Int. Joint Conf. on Neural Networks (IJCNN'2000)*, Italy, (2000).
15. Pouloudi, A. and Magoulas, G.D., Neural Expert Systems in Medical Image Interpretation: Development, Use and Ethical Issues, *Journal of Intelligent Systems*, 10, **5-6**, 451–471, (2000).
16. Pouloudi A., Information technology for collaborative advantage in health care revisited, *Information & Management* 1999, **35**, 345-357, (1999).
17. Ridderikhoff, J., van Herk B., Who is afraid of the system? Doctors' attitude towards diagnostic systems, *International Journal of Medical Informatics* 1999, **53**, 91-100, (1999).
18. Saad, D., *On-line Learning in Neural Networks*, Cambridge University Press, (1998).
19. Schraudolph, N.N., *Online Local Gain Adaptation for Multi-layer perceptrons*, Technical Report, IDSIA-09-98, IDSIA, Lugano, Switzerland, (1998).
20. Schraudolph, N.N., Local Gain Adaptation in Stochastic Gradient Descend, Technical Report, IDSIA-09-99, IDSIA, Lugano, Switzerland, (1999).
21. Storn, R., and Price, K., Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous spaces, *Journal of Global Optimization*, **11**, 341-359, (1997).
22. Sutton, R.S., Adapting Bias by Gradient Descent: an Incremental Version of Delta-Bar-Delta, *In: Proc. of the Tenth National Conference on Artificial Intelligence*, MIT Press, 171-176, (1992).
23. Sutton, R.S. and Whitehead, S.D., Online Learning with Random Representations, *In: Proc. of the Tenth International Conference on Machine Learning*, Morgan Kaufmann, 314-321, (1993).
24. Zhu, Y. and Yan, H., Computerized tumor boundary detection using a Hopfield neural network, *IEEE Transactions in Medical Imaging*, **16**, 55–67, (1997).