Hybrid Intelligent Systems

FUZZY ARTMAP NEURAL NETWORKS FOR COMPUTER AIDED DIAGNOSIS

Anatoli Nachev

Abstract*: The economic and social values of breast cancer diagnosis are very high. This study explores the predictive abilities of Fuzzy ARTMAP neural networks for breast cancer diagnosis. The data used is a combination of 39 mammographic, sonographic, and other descriptors, which is novel for the field. By using feature selection techniques we propose a subset of 21 descriptors that outperform the full feature set and outperforms the prediction model based on the most popular MLP neural networks. We also explored the model performance by ROC analysis and used metrics, such as max accuracy, area under the ROC curve, and area under the convex hull. Due to lack of specificity, many diagnosis tools entail unnecessary surgical biopsies, which motivated us to explore the clinically relevant metrics partial area under the ROC curve where sensitivity is above 90% and specificity at 98% sensitivity. In conclusion we find that the Fuzzy ARTMAP neural network is a promising prediction tool for breast cancer diagnosis. To the best of our knowledge, the Fuzzy ARTMAP neural networks have not been studied in that area until now.*

Keywords*: data mining, neural networks, Fuzzy ARTMAP, heterogeneous data; breast cancer diagnosis, computer aided diagnosis*

ACM Classification Keywords*: I.5.1- Computing Methodologies - Pattern Recognition – Models* - Neural Nets

Introduction

Breast cancer ranks first in the causes of cancer deaths among women in developed countries and is second in developing countries [Parker, 1997], [Lacey et al., 2002]. The best way to reduce deaths due to breast cancer is to treat the disease at an earlier stage. Earlier treatment requires early diagnosis, and early diagnosis requires an accurate and reliable diagnostic procedure that allows physicians to differentiate benign from malignant lesions. The economic and social values of breast cancer diagnosis are very high. Some studies show that only a third of suspicious masses are determined to be malignant and many surgical biopsies are unnecessary [Jemal et al., 2005]. Breast cancer diagnosis is a typical machine learning problem for many years. It has been dealt with using various machine learning algorithms and computer aided detection/diagnosis (CAD) tools. The problem is nontrivial and difficult to solve as the data set is noisy and relatively small. Techniques which rely on a large training data set would not work well for this problem.

A considerable amount of research in the area has been done based on rich variety of modalities and sources of medical information, such as: digitized screen-film mammograms, sonograms, magnetic resonance imaging (MRI) images, and gene expression profiles, etc. [Jesneck et al. 2006]. Current computerized breast cancer diagnosis tend to use only one information source, usually mammographic data in the form of descriptors defined by the Breast Imaging Reporting and Data System (BI-RADS) lexicon [BI-RADS, 2003]. Initially, BI-RADS was developed for standardization of mammographic descriptors only, but recently the American College of Radiology included a breast sonography extension, which standardizes various sonographic descriptors of lesions. Jesneck et al. [2007] have used a specific combination of BI-RADS mammographic and sonographic descriptors and

some proposed by Stavros et al. [1995] to build a predictive model. Their study was pioneering in using such a combination and they report that predictive abilities of the linear discriminant analysis (LDA) and multi-layer preceptrons (MPL) are similar in the context of using either all 39 descriptors or a suggested subset of 14 descriptors. MLP have been largely applied to breast cancer diagnosis applications, but they have a drawback: the model assumes predefined network architecture, including connectivity and node activation functions, and training algorithm to learn to predict. The issue of designing a near optimal network architecture can be formulated as a search problem and still remains open.

This paper describes an alternative approach based on Fuzzy ARTMAP neural networks which feature well established architecture and fast one-pass online learning. To the best of our knowledge Fuzzy ARTMAP neural networks has not been used to date for breast cancer diagnosis with a combination of mammographic and sonographic descriptors.

The paper is organized as follows: Section 2 provides an overview of the Fuzzy ARTMAP neural network architecture used build a predictive model; Section 3 introduces the dataset used in the study, its features, and preprocessing of data; Section 4 presents and discuses the experimental results; and Section 5 gives the conclusions.

Fuzzy ARTMAP Neural Networks

ARTMAP systems are based on the Adaptive Resonance Theory (ART) for neural network modeling [Grossberg, 1976]. ARTMAP architectures are neural networks that develop stable recognition codes in real time in response to arbitrary sequences of input patterns. They were designed to solve the stability-plasticity dilemma that every intelligent machine learning system has to face: how to keep learning from new events without forgetting previously learned information. Fuzzy ARTMAP networks were designed to accept real-valued input patterns [Carpenter et al., 1992]. They learn by either simultaneously establishing suitable categories in both input and output space (tasks carried out within the so-called ARTa and ARTb modules respectively) or linking input and output categories according to joint occurrence and predictive success (the linkages being stored in a special unit called the map field or ARTab (see Figure 1). Modules are made of fields, which consist of neurons (nodes). Each ART module a comparison layer (F_1), and a recognition layer (F_2) with m and n neurons, respectively. All categorization and learning are achieved by sequentially modifying the connection weights, between the layers (black circles in the figure). In Fuzzy ARTMAP systems data can be processed with either natural or complement coding [Carpenter et al., 1992]: if natural coding is used, a data item is processed as it is, otherwise, it is augmented with its complement to 1. Thus, if a data sample *d* is an *n*-dimensional vector, system actually works on a *2n*-dimensional vector.

The numbers of weights in the ARTa and ARTb modules are system parameters determining the number and dimension of weights in the ARTab module. During training, a sample and the category label are provided as input to the ARTa and ARTb modules, which causes an activation to flow from the excited neurons (categories) in ARTa and ARTb into ARTab and then eventually back to ARTa. During testing a given input vector activates (predicts) a single category in the ARTb and ARTab modules. Whenever a pattern *A* activates a layer F1, it propagates through weighted connections w_{ij} to layer F2. Activation of each node *j* in the F2 layer is determined by the function:

$$T_j(A) = \frac{|A \wedge w_j|}{\alpha + |w_j|} \tag{1}$$

where $\wedge$ is the fuzzy AND operator: $(A \wedge w_j) \equiv \min(A_i, w_{ij})$. The F2 layer produces a winner-take-all (WTA) pattern of activity such that only the node $j=J$ with the greates t activation value remains active. Node J propagates its top-down expectation, or prototype vector w_J, back onto F1 and the vigilance test is performed. This test compares the degree of match between w_J and A against the dimensionless vigilance parameter ρ (rhobar). Within a given ARTa or ARTb module, the system decision to commit new neurons, as opposed to using previously commited neurons, is controlled by the vigilance parameter. When ρ is large, the system tends to commit neurons more easily: otherwise, relatively fewer and therefore larger categories are constructed.

The system learns an input sample a by updating the vector of weights w_J associated with the prototype:

$$w'_J = \beta(A \wedge w_J) + (1-\beta)w_J \tag{2}$$

where β is a fixed learning rate parameter. Then a new association between the F2 node J and ARTab field takes place.

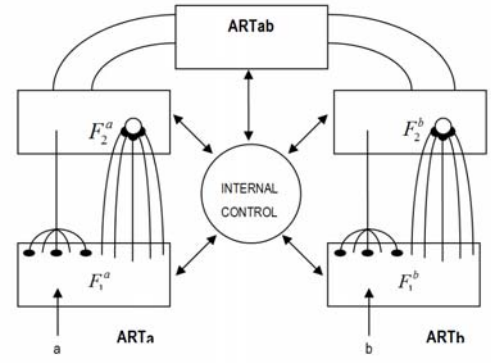

Figure 1. Simplified structurel of a Fuzzy ARTMAP neural network

Data, Features, and Preprocessing

The data used in this study consists of mammographic and sonographic examinations collected from 2000 to 2005 at Duke University Medical Centre [Jesneck et al., 2007]. Samples included in the dataset are those selected for biopsy only if the lesions corresponded to solid masses on sonograms and if both mammographic and sonographic images taken before the biopsy were available for review. The data set contains 803 samples of which 296 malignant and 507 benign. Information about patient physical examination findings, family history of breast cancer, and personal history of breast malignancy has been available to each radiologist to reproduce a realistic clinical situation.

Out of 39 features in total, 13 are mammographic BI-RADS features, 13 are sonographic BI-RADS features, six are sonographic features suggested by Stavros et al. [1995], four are other sonographic features, and three were patient history features [BI-RADS, 2003], [Jesneck et al., 2007], [Nachev & Stoyanov, 2010]. The features are as follows: mass size, parenchyma density, mass margin, mass shape, mass density, calcification number of particles, calcification distribution, calcification description, architectural distortion, associated findings, special cases (as defined by the BI-RADS lexicon: asymmetric tubular structure, intramammary lymph node, global asymmetry, and focal asymmetry), comparison with findings at prior examination, and change in mass size. The sonographic features are radial diameter, antiradial diameter, anteroposterior diameter, background tissue echo texture, mass shape, mass orientation, mass margin, lesion boundary, echo pattern, posterior acoustic features, calcifications within mass, special cases (as defined by the BI-RADS lexicon: clustered microcysts, complicated

cysts, mass in or on skin, foreign body, intramammary lymph node, and axillary lymph node), and vascularity. The six features suggested by Stavros [Stavros et al., 1995] are mass shape, mass margin, acoustic transmission, thin echo pseudocapsule, mass echogenicity, and calcifications. The four other sonographic mass descriptors are edge shadow, cystic component, and two mammographic BI-RADS descriptors applied to sonography—mass shape (oval and lobulated are separate descriptors) and mass margin (replaces sonographic descriptor angular with obscured). The three patient history features were family history, patient age, and indication for sonography.

The Fuzzy ARTMAP neural network we use require a specific input format, which presumes preprocessing of the original dataset. We applied two linear transformations: normalization and rescaling. The normalization addresses a problem of the input variables – they differ significantly in their values due to different units in which they are expressed. This difference can lead to poor classification as some variable dominate others. By calculating the deviation of each variable value from the variable mean, normalized by its standard deviation, we obtained the new values of the dataset using (3) and (4).

$$\tilde{x}_i^n = \frac{x_i^n - \bar{x}_i}{\sigma_i} \tag{3}$$

where $\tilde{x}_i^n$ is the new value, x_i^n is the original one, and

$$\bar{x}_i = \frac{1}{N}\sum_{n=1}^{N} x_i^n , \quad \sigma_i^2 = \frac{1}{N-1}\sum_{n=1}^{N} (x_i^n - \bar{x}_i)^2 \tag{4}$$

The second transformation, rescaling, maps the dataset values into [0, 1] using (5) as this is a requirement of the NN.

$$\hat{x}_i^n = \frac{(\tilde{x}_i^n - \tilde{x}_i^{\min})}{(\tilde{x}_i^{\max} - \tilde{x}_i^{\min})} \tag{5}$$

where $x_i^{\max}$ and $x_i^{\min}$ are the max, and min values of the variable x_i , respectively.

Experimental Results and Discussion

A series of tests was carried out in order to investigate how a trained Fuzzy ARTMAP NN predicts breast cancer decease, based on a set of descriptors outlined above. In order to validate our experiments we used 5-fold cross-validation (CV) technique which avoids bias in selection of the training and test sets. It does so by creating 5 copies of the classifier, and testing each on 20% (1/5) of the data set, after training it on the remainder. The classification error estimate is computed as the average of the values obtained for each test set.

The experiments were focused to two aspects of the model functioning: reduction of dimensionality of the data and optimal values of the network parameters.

We considered and experimented with various feature selection techniques, such as best first, genetic search, subset size forward selection, race search, and scatter search, and sults showed that best feature selection technique is genetic search [Goldberg, 1989] combined with a set evaluation technique that considers individual predictive ability of each feature along with the degree of redundancy between them [Hall, 1998]. The feature set we obtained as best consists of the following 21 descriptors: patient age, family history, mass margin, architectural distortion, associated findings, comparison with prior examinations, anteroposterior diameter, mass shape, mass orientation, mass margin, lesion boundary, calcification within mass, special cases, mass shape, mass margin, thin echo pseudocapsule, mass echogenicity, edge shadow, cystic component, mass shape, and

mass margin. We also tested the model with three other feature sets: the full set of 39 descriptors (s39), the set of 14 features obtained by [Jesneck et al., 2007] by stepwise feature selection method, and the set of 17 features (s17) proposed by [Nachev & Stoyanov, 2010] for MLP neural networks.

In order to estimate the Fuzzy ARTMAP performance with different network parameters, we explored each of them individually. Results show that the parameters baseline vigilance, signal rule, and the learning fraction provide the model with best discriminatory power by values $\rho_{test} = 0$, $\alpha = 0.01$, $\beta = 1.0$, and regardless of the feature selection. The vigilance parameter ρ (rhobar), however, shows dependency to each selected feature set.

For each set the model was trained and tested with 42 vigilance parameter values from 0 to 1 with step of increment 0.025. Figure 2 shows the prediction accuracy of the four sets with all values of the vigilance parameter. The sets show similar performance, but certain values of the parameter cause picks of accuracy with best result of 84.4% achieved by s21 at $\rho = 0.225$.

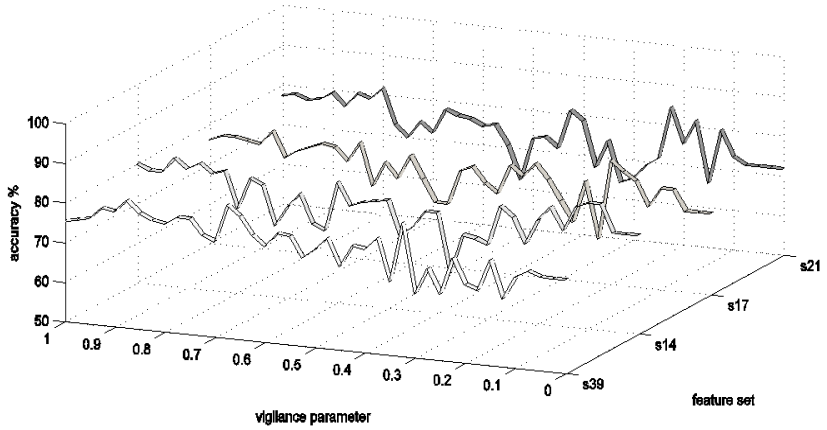

Figure 2. Accuracy of Fuzzy ARTMAP with four feature sets and vigilance parameter (rhobar) values from 0 to 1 with step of increment 0.025

Accuracy is the most common estimator used to date, but it can be misleading if the distribution of the classes is skewed, or if errors of type I and type II have different clinical implications and different cost. Taking into account those drawbacks we did Receiver Operating Characteristics (ROC) analysis of the results. ROC curves describe the relation between true positive rate (TPR) and false positive rate (FPR) [Fawcett, 2006]. In the case of crisp classifiers, such as Fuzzy ARTMAP, each classifier is represented by one point on the ROC space. By varying a parameter, such as the vigilance one, we generate an aggregation of points (Figure 3). The solid line represents the ROC convex hull (ROCCH), a line made up by connecting the most northwest points and the two trivial classifiers (0,0) and (1,1). All the candidates for optimal classifier lie on the convex hull as these are the most northwest point with minimum FPR and maximum TPR. All other classifiers that are 'capped' by the ROCCH can be ignored as they cannot be optimal. Each ROCCH line section between two adjacent corner points represents a continuum of possible intermediate classifiers that can be constructed by randomly weighting both corner classifiers giving more or less weight to one or the other. The optimal classifiers in terms of ROC are the most 'northwest' or most distant from the no-discrimination line. These are the square points on Figure 3 and they are the same that give maximal accuracy, which confirms that best in terms of accuracy is best in terms of ROC.

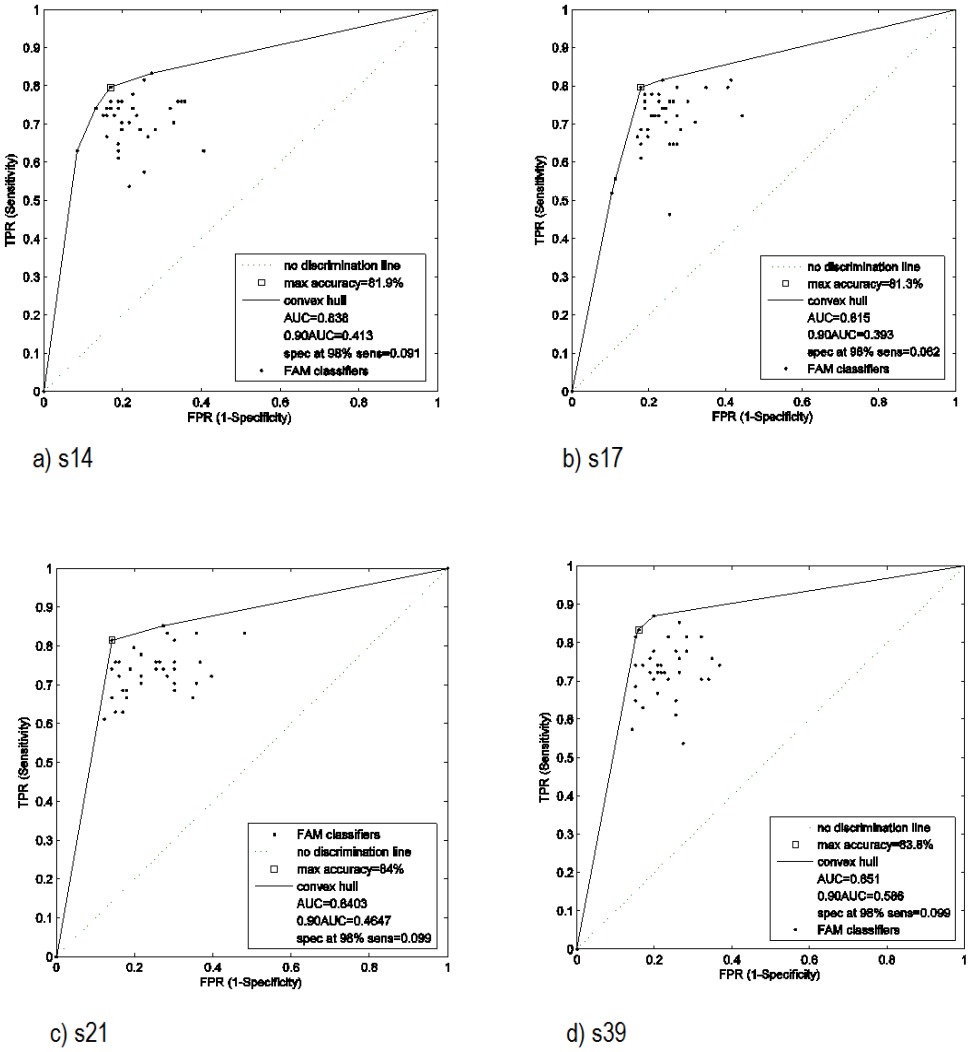

Figure 3. ROC analysis of Fuzzy ARTMAP with feature set of: a) 14 descriptors [Jesneck, 2007]; b) 17 descriptors [Nachev & Stoyanov, 2010]; c) 21 descriptors proposed in this study; d) all 39 descriptors. Each point represents a classifier determined by a value of the vigilance parameter that varies from 0 to 1 with step of increment 0.025

ROC analysis also allows to calculate metrics that estimate a model performance, such as area under the ROC curve (AUC), partial AUC ($_{0.90}$AUC of sensitivity above 0.90), and specificity at 98% sensitivity. AUC represents the overall model performance regardless the choice of vigilance; $_{0.90}$AUC shows the model performance at high values of sensitivity, which is important from a clinical viewpoint; specificity at 98% sensitivity is also important from a clinical perspective. Figure 3 shows the ROC space and the performance metrics for each of the four sets.

The figures are compared in Table 1, that shows that Fuzzy ARTMAP has max accuracy 84.4% with the 21 descriptor set and vigilance parameter value 0.225. The model also outperforms predictors based on the most popular neural networks – MLP, which yield 82.5% accuracy [Nachev & Stoyanov, 2010]. Second best is the set of 39 descriptors, which means that collecting and processing all the data is time consuming and not necessary. Lowest accuracy is obtained by using the 17 and 14 feature sets, which suggests that despite those sets give good results with LDA and MLP, they don't work well with another model, such as Fuzzy ARTMAP. AUC and

$_{0.90}$AUC show that the full feature set outperforms the selection of 21 descriptors and all features should be used if sensitivity above 90% is required.

Table 1. Accuracy of Fuzzy ARTMAP with four feature sets and vigilance parameter (rhobar) values from 0 to 1 with step of increment 0.025.

Performance Metric	*39 attr. Set*	*21 attr Set*	*17 attr. set*	*14 attr set.*
AUC	0.851	0.840	0.815	0.838
$_{0.90}$AUC	0.586	0.465	0.393	0.413
Spec at 98% sens	0.099	0.099	0.082	0.091
Acc_{max}	0.838	0.844	0.813	0.819

Conclusion

Computer-aided breast cancer diagnosis is a typical classification problem which was approached over the years by many techniques and methods, and algorithms.

This paper explores the discriminatory power of Fuzzy ARTMAP neural networks in differentiation between malignant and benign lesions based on data from mammographic and sonographic examinations. We used a data collected at Duke University Medical Centre which contains 39 descriptors. In order to improve the model performance we did reduction of dimensionality by applying various feature selection techniques.

We found that a subset of 21 descriptors outperforms the full descriptor set, as well as two other subsets used with the same dataset in other studies. A careful adjusted Fuzzy ARTMAP neural network outputs 84.4% prediction accuracy of the dataset versus 82.5% of the MLP neural network with the same dataset.

The model performance was also estimated by ROC analysis and the metrics such as area under the ROC curve, partial area under the ROC curve above 90% sensitivity, and specificity at 98% sensitivity.

In conclusion we find that the Fuzzy ARTMAP neural network is a promising technique for diagnosis, but when used it requires a careful reduction of dimensionality and well tuned network parameters. The model also provides additional benefits such as one-pass online learning that retains already acquired knowledge, in contrast to the widely used MLP neural networks. To the best of our knowledge Fuzzy ARTMAP neural networks has not been used to date for breast cancer diagnosis in combination with both mammographic and sonographic descriptors.

Bibliography

[BI-RADS, 2003] American College of Radiology. BI-RADS: ultrasound, 1st ed. In: Breast imaging reporting and data system: BI-RADS atlas, 4th ed. Reston, VA: American College of Radiology, 2003

[Carpenter et al., 1992] Carpenter, G.A., Grossberg, S., Markuzon, N., Reynorlds, J.H. and Rosen, D.B. “Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps”, IEEE Transaction on Neural Networks, 3(5), pp. 698-713, 1992.

[Goldberg, 1989] Goldberg, D. “Genetic Algorithms in Search, Optimization, and Machine Learning”, Addison-Wesley: Reading, MA, 1989.

[Grossberg, 1976] S. Grossberg, “Adaptive pattern classification and universal recoding. II: Feedback, expectation, olfaction, and illusions.” Biological Cybernetics, 23, 1976.

[Fawcett, 2006] Fawcett, T. "An introduction to ROC analysis"; Pattern Recognition Letters, Vol. 27 Issue 8, pp. 861-874, 2006.

[Hall, 1999] Hall, M. Hall,M. "Correlation based feature selection for machine learning", Doctoral dissertation, Department of Computer Science, The University of Waikato, Hamilton, New Zealand, 1999.

[Jemal et al., 2005] Jemal, A., Murray, T, Ward, E., Samuels, A., Tiwari, R., Ghafoor, A., Feuer, E., Thun, M.. "Cancer statistics", Ca-Cancer J. Clin 2005;55:10–30, 2005.

[Jesneck et al., 2006] Jesneck, J., Nolte, L., Baker, J., Floyd, C., Lo, J. "Optimized Approach to Decision Fusion of Heterogeneous Data for Breast Cancer Diagnosis."; Medical Physics, Vol. 33, Issue 8, pp.2945-2954, 2006

[Jesneck et al., 2007] Jesneck, J., Lo, J., Baker, J. "Breast Mass Lesions: Computer-Aided Diagnosis Models with Mamographic and Sonographic Descriptors"; Radiology, vol.244, Issue 2, pp 390-398, 2007.

[Lacey et al., 2002] Lacey, J., Devesa, S., Brinton, L. "Recent Trends in Breast Cancer Incidence and Mortality."; Environmental and Molecular Mutagenesis, Vol. 39, pp. 82–88, 2002.

[Nachev & Stoyanov, 2010] Nachev, A. and Stoyanov, B., "An Approach to Computer Aided Diagnosis by Multi-Layer Preceptrons", In Proceedings of International Conference Artificial Intelligence (IC-AI'10), Las Vegas, 2010.

[Parker, 1997] Parker S., Tong T., Bolden S., Wingo P. "Cancer Statistics.", CA Cancer J Clin 1997; 47:5-27.

[Stavros et al., 1995] Stavros, A., Thickman, D., Rapp, C., Dennis, M., Parker, S., Sisney, G. "Solid Breast Modules: Use of Sonography to Destinguish between Benign and Malignant Lesions"; Radiology, Vol. 196, pp. 123-134, 1995.

Authors' Information

Anatoli Nachev *– Business Information Systems, Cairnes Business School, National University of Ireland, Galway, Ireland; e-mail: anatoli.nachev@nuigalway.ie*