Intelligent Applications: Medical and Diagnostic System

PERFORMANCE OF COMPUTER-AIDED DIAGNOSIS TECHNIQUES IN INTERPRETATION OF BREAST LESION DATA

Anatoli Nachev, Mairead Hogan, Borislav Stoyanov

Abstract*: This study explores and compares predictive abilities of six types of neural networks used as tools for computer-aided breast cancer diagnosis, namely, multilayer perceptron, cascade-correlation neural network, and four ART-based neural networks. Our experimental dataset consists of 803 patterns of 39 BI-RADS, mammographic, sonographic, and other descriptors. Using such a combination of features is not traditional in the field and we find it is better than traditional ones. The study also focuses on exploring how various feature selection techniques influence predictive abilities of the models. We found that certain feature subsets show themselves as top candidates for all the models, but each model performs differently with them. We estimated models performance by ROC analysis and metrics, such as max accuracy, area under the ROC curve, area under the convex hull, partial area under the ROC curve with sensitivity above 90%, and specificity at 98% sensitivity. We paid particular attention to the metrics with higher specificity as it reduces false positive predictions, which would allow decreasing unnecessary benign breast biopsies while minimizing the number of delayed breast cancer diagnoses. In order to validate our experiments we used 5-fold cross validation. In conclusion, out results show that among the neural networks considered here, best overall performer is the Default ARTMAP neural network.*

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

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

Introduction

Breast cancer is one of the leading causes of death for women in many countries. Mammography is currently the most widely used screening method for early detection of the disease, but it has a low negative predictive value. Many investigators have found that more than 60% of masses referred for breast biopsy on the basis of mammographic findings are actually benign [Jemal et al., 2005], [Lacey et al., 2002]. One goal of the application of computer-aided diagnosis (CAD) to mammography is to reduce the false-positive rate. Avoiding benign biopsies spares women unnecessary discomfort, anxiety, and expense. The problem is nontrivial and difficult to solve. Breast cancer diagnosis is a typical machine learning problem. It has been dealt with using various data mining techniques and tools such as linear discriminant analysis (LDA), logistic regression analysis (LRA), multilayer perceptions (MLP), support vector machines (SVM), etc. [Chen et al., 2009], [Jesneck et al., 2006].

Current CAD implementations tend to use only one information source, usually mammographic data in the form of data descriptors defined by the Breast Imaging Reporting and Data System (BI-RADS) lexicon [BI-RADS, 2003]. Recently, Jesneck et al. [2007] proposed a novel combination of BI-RADS mammographic and sonographic descriptors and some suggested by Stavros et al. [1995], which in combination with MLP show promising results. The MLP have been largely applied in the data mining tasks, but one of their major drawbacks is unclear optimal architecture, which includes number of hidden nodes, activation functions, and training algorithm to learn to predict. Another major problem is to specify optimal set of descriptors used for data mining, which effectively reduces the training and testing datasets to a dimensionality which provides best performance for the application domain. Our study was motivated by addressing those problems and particularly focusing on how reduction of dimensionality of that new combination of descriptors affects performance of not only MLPs, but also other neural network models, such as cascade-correlation nets and those based on the adaptive resonance theory (ART), introduced by Grossberg [1976].

The paper is organized as follows: Section 2 provides a brief overview of the neural networks used in this study: multilayer perceptron (MLP), cascade-correlation neural networks, fuzzy ARTMAP, distributed ARTMAP, default ARTMAP, and ic ARTMAP; Section 3 introduces the dataset and its preprocessing; Section 4 presents and discuses results from experiments; and Section 5 gives the conclusions.

Neural Networks for Data Mining

A variety of neural network models are used by practitioners and researchers for clustering and classification, ranging from very general architectures applicable to most of the learning problems, to highly specialized networks that address specific problems. Each model has a specific topology that determines the layout of the neurons (nodes) and a specific algorithm to train the network or to recall stored information.

Multilayer Perceptions (MLP)

Among the neural network models, the most common is the multilayer perceptron, which has a feed-forward topology and error-backpropagation learning algorithm [Rumelhart & McClelland, 1986]. Typically, an MLP consists of a set of input nodes that constitute the input layer, an output layer, and one or more layers sandwiched between them, called hidden layers. Nodes between subsequent layers are fully connected by weighted connections so that each signal travelling along a link is multiplied by its weight w_{ij}. Hidden and output nodes receive an extra bias signal with value 1 and weight θ. The input layer, being the first layer, has input nodes that distribute the inputs to nodes in the first hidden layer. Each hidden and output node computes its activation level by

$$s_i = \sum_j w_{ij} x_j + \theta \quad , \tag{1}$$

and then transform it to output by an activation function. The MLP we use in this study has one hidden layer with two hidden nodes and log-sigmoid activation function

$$O_i(s_i) = \frac{1}{1 + e^{-\beta s_i}} \quad , \tag{2}$$

We trained the MLP by adaptive learning rate algorithm developed by Jacob [1988], also called delta-bar-delta, or TurboProp. The adaptive learning rate method proposes more flexibility and a higher speed of convergence, compared to the classic backpropagation algorithm.

Cascade-Correlation Neural Networks (CCNN)

CCNN [Fahlman & Libiere, 1990] are supervised self-organizing networks with structure similar to backpropagation networks. Instead of adjusting the weights in a network of fixed topology, a CCNN begins with a minimal number of nodes, then automatically trains and adds new hidden nodes one by one and do not change them over the time. It creates a multi-layer structure called a 'cascade' because the output from all input and hidden nodes already in the network feed into new nodes.

A CCNN has three layers: input, hidden and output. Initially, the network begins with only input and output nodes. The output layer consists of a single neuron if the network is used for regression problems, or contains several neurons for classification problems, one per class label. The hidden layer is empty in the beginning – every input is connected to every output neuron by a connection with an adjustable weight. Such a simple cascade-correlation network has considerable predictive power and for a number of applications it provides excellent predictions. If not, however, the network adds new hidden nodes one by one as illustrated in Figure 1, until the residual error gets acceptably small or the user interrupts this process.

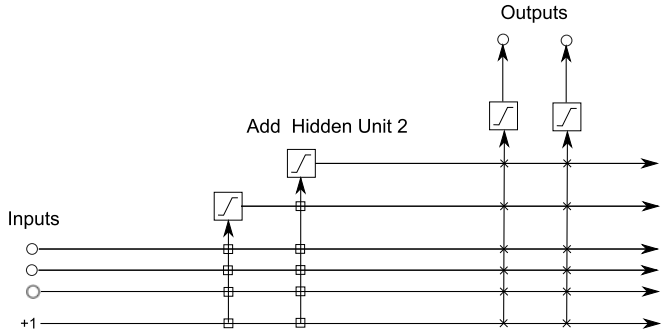

Fig.1. Cascade architecture after adding two hidden nodes (adapted from [Fahlman & Lebier, 1991]). The vertical lines sum all incoming activation. Boxed connections are frozen, 'x' connections are trained repeatedly

The cascade-correlation architecture has several advantages over the traditional backpropagation neural nets. First, as the network is self-organizing and determines its own size and topology during the growth of the hidden layer during training, there is no need to decide how many layers and neurons to use in the network. This is a major problem of the backpropagation networks which implies use of predefined network architecture and designing a near optimal network architecture is a search problem that still remains open. Secondly, cascade-correlation nets learn very quickly (often 100 times as fast as a backpropagation network) and retain the structures they have built even if the training set changes. Finally, they have less chance to get trapped in local minima compared to the backpropagation nets.

Fuzzy, Distributed, Default, and IC ARTMAP Neural Networks

The adaptive resonance theory (ART) introduced by Grossberg [1975] led to the creation of a family of self-organizing neural networks, such as the unsupervised ART1, ART2, ART2-A, ART3, fuzzy ART, distributed ART and the supervised ARTMAP, instance counting ARTMAP, fuzzy ARTMAP (FAM), distributed ARTMAP, and default ARTMAP. ARTMAP is a family of neural network that consists of two unsupervised ART modules, *ARTa* and *ARTb*, and an *inter-ART* module called map-field as shown in Figure 2. An ART module has three layers of nodes: input layer *F0*, comparison layer *F1*, and recognition layer *F2*. A set of real-valued weights W_j is associated with the *F1-to-F2* layer connections between nodes. Each *F2* node represents a recognition category that learns a binary prototype vector w_j. The *F2* layer is connected through weighted associative links to a map field F^{ab}.

The ARTMAP learning can be described by the following algorithm [Carpenter et al., 1991]:

1. *Initialization*: All *F2* nodes are uncommitted, and all weight values and network parameters are initialized.

2. *Input pattern coding*: When a training pattern is presented to the network, a process called complement coding takes place. It transforms the pattern into a form suited to the network. A network parameter called vigilance parameter (ρ) is set to its initial value. This parameter controls the network 'vigilance', that is, the level of details used by the system when it compares the input pattern with the memorized categories.

3. *Prototype selection*. The input pattern activates layer *F1* and propagates to layer *F2*, which produces a binary pattern of activity such that only the *F2* node with the greatest activation value remains active, that is, 'winner-takes-all'. If such a node does not exist, an uncommitted *F2* node becomes active and undergoes learning.

4. *Class prediction*. The class label *t* activates the F^{ab} layer in which the most active node yields the class prediction. If that node constitutes an incorrect class prediction, then another search among *F2* nodes in Step 3 takes place. This search continues until an uncommitted *F2* node becomes active (and learning directly ensues in Step 5), or a node that has previously learned the correct class prediction becomes active.

5. *Learning*. The neural network gradually updates its adaptive weights towards the presented training patterns until a convergence occur. The learning dynamic can be described by a system of ordinary differential equations.

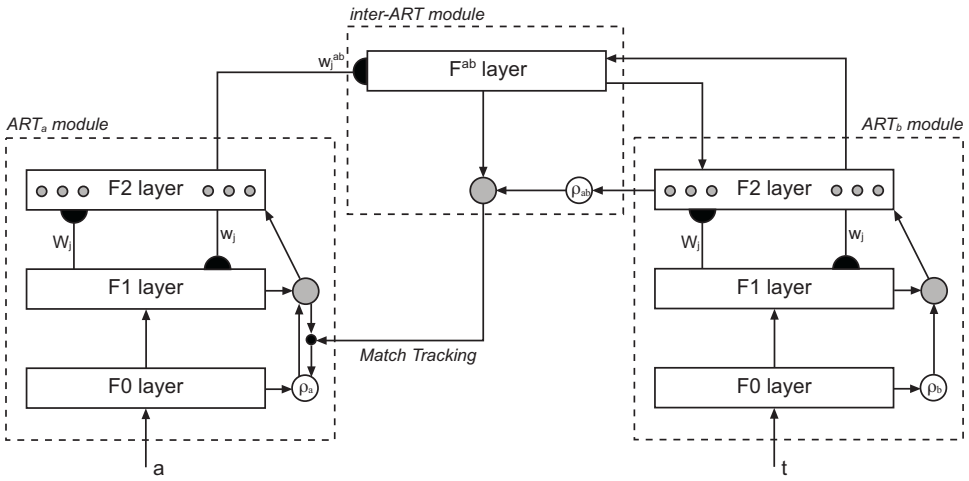

Fig. 2. Block diagram of an ARTMAP neural network adapted from [Carpenter et al, 1991]

Fuzzy ARTMAP was developed as a natural extension to the ARTMAP architecture. This is accomplished by using fuzzy ART modules instead of ART1, which in fact replaces the crisp (binary) logic embedded in the ART1 module with a fuzzy one. In fact, the intersection operator ($\cap$) that describes the ART1 dynamics is replaced by the fuzzy AND operator ($\wedge$) from the fuzzy set theory ($(p \wedge q)_I \equiv \min(p_i, q_i)$) [Carpenter et al., 1992]. This allows the fuzzy ARTMAP to learn stable categories in response to either analog or binary patterns in contrast with the basic ARTMAP, which operates with binary patterns only.

ART1 modules in an ARTMAP net map the categories into *F2* nodes according to the winner-takes-all rule, as discussed above, but this way of functioning can cause category proliferation in a noisy input environment. An explanation of that is that the system adds more and more *F2* category nodes to meet the demands of predictive accuracy, believing that the noisy patterns are samples of new categories. To address this drawback, a new distributed ART module was introduced. If the ART1 module of the basic ARTMAP is replaced by a distributed ART module, the resulting network is called **Distributed ARTMAP** [Carpenter, 1997].

Instance Counting (IC) ARTMAP adds to the basic fuzzy ARTMAP system new capabilities designed to solve computational problems that frequently arise in prediction. One such problem is inconsistent cases, where identical input vectors correspond to cases with different outcomes. A small modification of the fuzzy ARTMAP

match-tracking search algorithm allows the IC ARTMAP to encode inconsistent cases and make distributed probability estimates during testing even when training employs fast learning [Carpenter & Markuzon, 1998].

A comparative analysis of the ARTMAP modifications, including Fuzzy ARTMAP, IC ARTMAP, and Distributed ARTMAP, has led to the identification of the **Default ARTMAP** network, which combines the winner-takes-all category node activation during training, distributed activation during testing, and a set of default network parameter values that define a ready-to-use, general-purpose neural network for supervised learning and recognition [Carpenter, 2003]. The Default ARTMAP features simplicity of design and robust performance in many application domains.

Data and Preprocessing

Our tests used a dataset that contains data from physical examination of patients, including mammographic and sonographic examinations, family history of breast cancer, and personal history of breast malignancy, all 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. Data contain 803 samples, 296 of which are malignant and 507 benign. Out of 39 descriptors, 13 are mammographic BI-RADS, 13 sonographic BI-RADS, 6 sonographic suggested by Stavros et al. [1995], 4 sonographic mass descriptors, and 3 patient history features [BI-RADS, 2003], [Jesneck et al., 2007], [Nachev & Stoyanov, 2010]. There are also class label that indicates if a sample is malignant or benign.

We preprocessed the dataset in order to addresses the problem of large amplitude of variable values caused by their different nature and different units of measurements. Consistency we achieved by mapping all data values into the unit hypercube (i.e. all values between 0 and 1), using a linear transformation

$$x_i^{new} = \frac{x_i^{old} - \min_i}{\max_i - \min_i} \tag{3}$$

applied to each variable (data column) separately. This scaling down of values is essential requirement for certain types of neural networks, and particularly for the ARTMAP models we used in our study.

Another preprocessing step was feature selection. In many cases and application domains removing redundant features from the data can help to alleviate effect of curse of dimensionality, avoid overfitting, and speed up learning process. Exhaustive search approach among all possible subsets of features is not applicable in our case as the dataset has cardinality 39. Alternative approaches could be using subset selection algorithms or feature ranking techniques. The former one is preferable as it usually provides good results. We tested genetic search, best first search, subset size forward selection, race search, and scatter search. Our tests showed that the subset size forward selection, proposed by Guetlien et al. [2009] gives good results with all types neural networks we experimented with. This method output a set of 17 descriptors (s17): patient age, indication for sonography, mass margin, calcification number of particles, architectural distortion, anteroposterior diameter, mass shape, mass orientation, lesion boundary, special cases, mass shape, mass margin, thin echo pseudocapsule, mass echogenicity, edge shadow, cystic component, and mass margin. Two of these are general descriptors; three - mammographic BI-RADS; five - sonographic BI-RADS; four - Stavros'; and three - sonographic mass descriptors. The feature set is relatively balanced in representing different categories of data. In our experiments we also used a set of 14 descriptors (s14) proposed by Jesneck et al. [2007] and obtained by stepwise feature selection. We also used the original full set of 39 descriptors (s39).

Empirical Results and Discussion

We used simulators of the neural network models explored here. Series of test showed that best architecture of the multilayer perceptron is one hidden layer with two nodes. We also used a cascade-correlation neural network with Turboprop2 learning based on the Fahlman's work [Fahlman & Libiere, 1990]. Each of the four types ARTMAP neural networks were tested with 41 vigilance parameter values from 0 to 1 and step of increment 0.025. In order to avoid bias in training due to the specific order of training samples, we applied 5-fold cross validation and summarized results in four categories: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Prediction accuracy was calculated by Acc=(TP+TN)/(TP+TN+FP+FN). For the purposes of ROC analysis, we also calculated true positive rate TPR=TP/(TP+FN), and false positive rate FPR=FP/(TP+FN).

No doubts, accuracy is the most common performance estimator of a model, which is used in a vast amount of studies and applications, but in many cases and problem domains it is not sufficient, even can be misleading where important classes are underrepresented in datasets (i.e. class distribution is skewed), or if errors of type I and type II can produce different consequences and have different cost. Secondly, the accuracy depends on the classifier's operating threshold, such as threshold values of MLP or vigilance parameter of ARTMAP NN, and choosing optimal threshold can be challenging. The deficiencies of accuracy can be addressed by the Receiver Operating Characteristics (ROC) analysis [Fawcett, 2006], which plots curves between two indices: TPR and FPR.

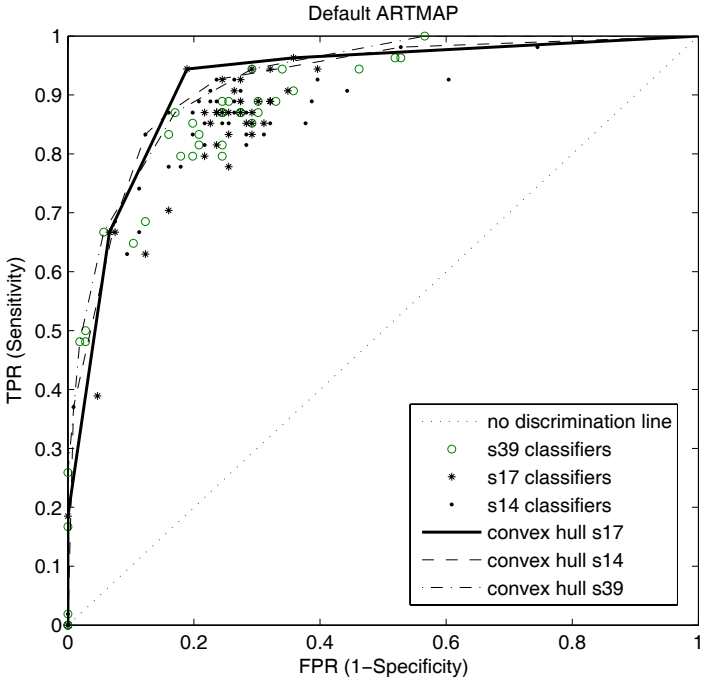

Fig. 3. ROC analysis of Default ARTMAP neural network tested with three sets of descriptors: all available (s39), a selection proposed by Jesneck et al. [2007] (s14), and a set of 17 descriptors (s17) proposed by authors. The model was tested with 41 vigilance parameter values from 0 to 1 with step of increment 0.025

ROC curves are step functions which can be used to select the optimal decision threshold by maximizing any pre-selected measure of efficacy. In general, the best possible prediction method would yield a point in the upper left corner or coordinates (0, 1) of the ROC space, representing 100% sensitivity (all true positives are found) and 100% specificity (no false positives are found). A completely random guess would give a point along a diagonal

line, also known as line of no-discrimination, which is from the left bottom to the top right corners. The optimal classifier would be represented by the most 'northwest' point on the curve, which is the most distant one from the no-discrimination line. Plotting discrete classifiers, such as ARTMAP, require additional processing. Since they do not use operational threshold to be varied in order to generate the ROC curve, the vigilance parameter variance plays the same role. Each parameter value plots one point on the ROC space, and the curve that connects the most northwest points along with the two trivial classifiers (0,0) and (1,1), called also ROC Convex Hull (ROCCH) represents the model as a whole. The ROCCH lines that link points of 'real' classifiers define a continuum of possible classifiers that can be obtained by linear combinations of the plotted ones. ROC analysis also provides additional metrics for estimation of model performance, such as Area Under the ROC curve (AUC) and partial Area Under the ROC curve (pAUC) where sensitivity is above certain value (p). The bigger the AUC / pAUC, the better the model is.

Table 1. Performance of MLP, CCNN, Fuzzy ARTMAP, Distributed ARTMAP, Default ARTMAP, and IC ARTMAP. Metrics for comparison include: area under the ROC curve (AUC), partial AUC at sensitivity above 90% ($_{0.90}$AUC), specificity at 98% sensitivity, and maximal accuracy (ACC_{max}). Models have been tested with three variable selections: s39, s17, and s14. Typical radiologist assessment values are also included.

MLP	*s39*	*s17*	*s14*	*Radiologist*
AUC	0.89	0.91	0.86	0.92
$_{0.90}$*AUC*	0.62	0.68	0.55	0.52
Spec /98% sens	0.37	0.49	0.27	0.52
ACC_{max}	0.89	0.91	0.87	n/a

CCNN	*s39*	*s17*	*s14*	*Radiologist*
AUC	0.896	0.911	0.907	0.92
$_{0.90}$*AUC*	0.648	0.68	0.731	0.52
Spec /98% sens	0.427	0.5	0.49	0.52
ACC_{max}	0.828	0.848	0.838	n/a

Fuzzy ARTMAP	*s39*	*s17*	*s14*	*Radiologist*
AUC	0.851	0.815	0.838	0.92
$_{0.90}$*AUC*	0.586	0.393	0.413	0.52
Spec /98% sens	0.099	0.082	0.091	0.52
ACC_{max}	0.838	0.813	0.819	n/a

Distributed ARTMAP	*s39*	*s17*	*s14*	*Radiologist*
AUC	0.786	0.819	0.744	0.92
$_{0.90}$*AUC*	0.226	0.272	0.187	0.52
Spec /98% sens	0.047	0.057	0.039	0.52
ACC_{max}	0.831	0.856	0.819	n/a

Default ARTMAP	*s39*	*s17*	*s14*	*Radiologist*
AUC	0.927	0.931	0.918	0.92
$_{0.90}$*AUC*	0.725	0.809	0.778	0.52
Spec /98% sens	0.686	0.649	0.690	0.52
ACC_{max}	0.85	0.856	0.863	n/a

IC ARTMAP	*s39*	*s17*	*s14*	*Radiologist*
AUC	0.776	0.821	0.860	0.92
$_{0.90}$*AUC*	0.215	0.282	0.384	0.52
Spec /98% sens	0.045	0.059	0.081	0.52
ACC_{max}	0.831	0.850	0.856	n/a

As long as AUC provides an overall estimation of the model, the partial area is more relevant to the domain of computer-aided diagnosis, and particularly where p=0.9. Another clinically relevant metric used in the application domain is sensitivity at a very high level of specificity (98%).

Table 1 summarizes results from numerous experiments where networks were trained and tested with three different sets of descriptors: s39 that contains the original 39 variables; a selection of 14 variables (s14) proposed

by Jesneck et al. [2007] as a result from stepwise feature selection technique, and our set of 17 variables (s17) we obtained by using the subset size forward selection of Guetlien et al. [2009].

Figure 4 illustrates the results. We obtained highest prediction accuracy of 91.1% by using CCNN with s17. The Default ARTMAP outperforms all other models in terms of overall performance measured by AUC. Here again, s17 is best performer. The figure also illustrates, that Default ARTMAP is the only model (among the studied here) that outperforms the average radiologist performance [Jesneck et al. 2007] in terms of AUC, but more important from a clinical viewpoint are the metrics pAUC and sensitivity at very high specificity. Figures show that again Default ARTMAP beats the others with s17 in terms of pAUC, and again is best performer in terms of sensitivity at high specificity, no matter which subset of descriptors is used.

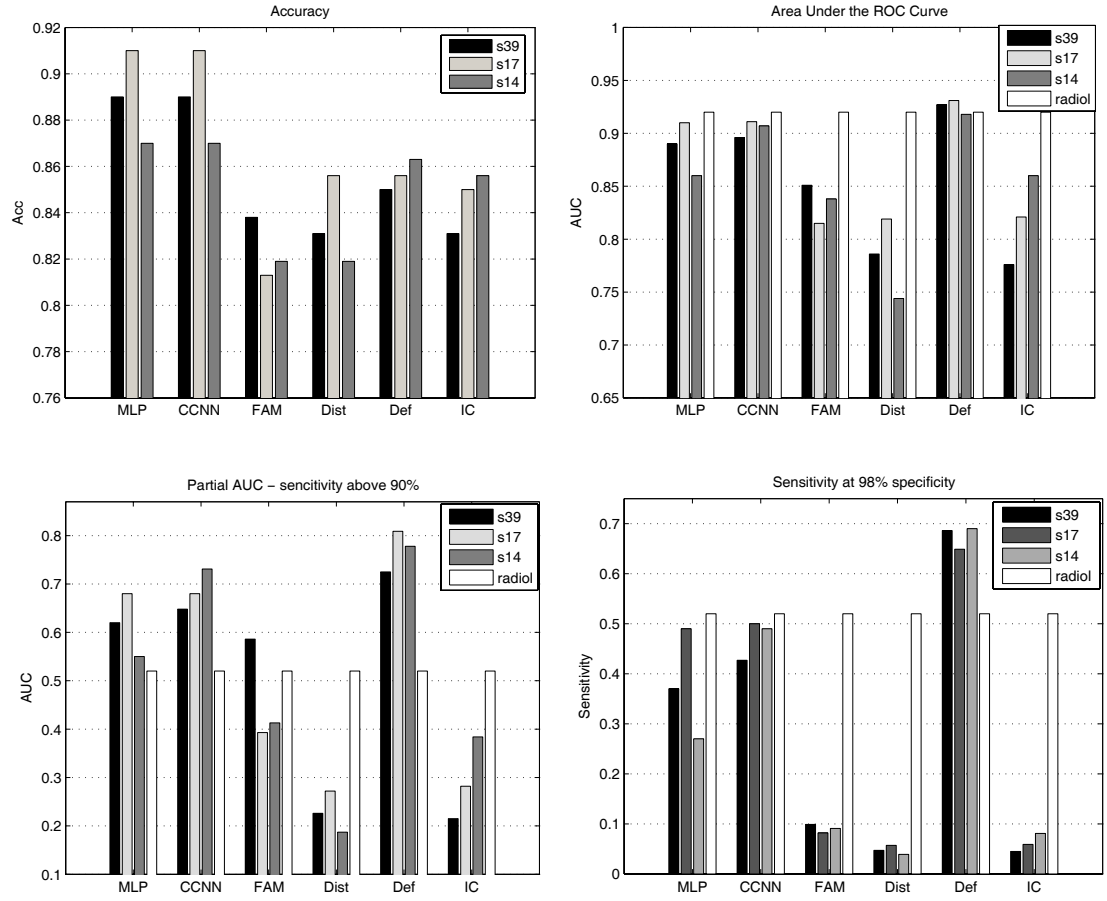

Fig. 4. Performance of three-layer perceptron (MLP), cascade-correlation NN (CCNN), Fuzzy ARTMAP (FAM), Distributer ARTMAP (Dist), Default ARTMAP (Def), and IC ATRMAP (IC), measured by max accuracy (Acc), area under the ROC curve (AUC), partial area under the ROC curve where sensitivity is above 90%, and sensitivity at 98% specificity. Models were tested with all available descriptors (s39), a selection proposed by Jesneck et al. [2007] (s14), and a new proposed set of 17 descriptors (s17). Results were also compared with a typical radiologist performance (radiol)

Figure 3 gives further details on the Default ARTMAP neural network performance obtained by ROC analysis. Bold line represents the best descriptor set. Finally, we find that Default ARTMAP is appropriate for solving the task as its clinically relevant characteristics are good, however a limitation of the model is that is that it requires a very careful tuning.

Conclusion

Many CAD systems for breast cancer screening improve lesion detection sensitivity, but improving specificity is still challenging. This study explores and compares predictive abilities of six types of neural networks: MLP, CCNN, Fuzzy, Distributed, Default, and IC ARTMAP by using a recently proposed combination of BI-RADS mammographic, sonographic, and other descriptors. We also focused our study on how various feature selection techniques influence predictive abilities of those models and found that a subset obtained by subset size forward selection provides best overall results. Our performance estimations were based on ROC analysis and metrics, such as max accuracy, area under the ROC curve and convex hull. We paid particular attention on clinically relevant metrics, such as partial area under the ROC curve with sensitivity above 90%, and specificity at 98% sensitivity, as a higher specificity reduces false positive predictions, which would allow decreasing unnecessary benign breast biopsies while minimizing the number of delayed breast cancer diagnoses. In conclusion, our results show that among all neural networks we explored for this application domain, highest prediction accuracy of 91.1% can be obtained by cascade-correlation neural network, but Default ARTMAP outperforms all other models in terms of overall performance and clinically relevant metrics. All the results confirm that the set of descriptors we propose outperforms the ones used in previous studies.

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, 1991] Carpenter, G., Grossberg, S., & Reynolds, J.: ARTMAP: Supervised Real-Time Learning and Classification of Non-stationary Data by a Self-Organizing Neural Network. Neural Networks, vol. 6, pp. 565-588 (1991)

[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.

[Carpenter, 1997] Carpenter, G.. Distributed Learning, Recognition, and Prediction by ART and ARTMAP Neural Networks. Neural Networks, vol. 10:8, 1473-1494, (1997)

[Carpenter & Markuzon, 1998] Carpenter, G. & Markuzon, N.. ARTMAP-IC and Medical Diagnosis: Instance Counting and Inconsistent Cases. Neural Networks, vol. 11:2, pp. 323-336 (1998)

[Carpenter, 2003] Carpenter, G.: Default ARTMAP. Proceedings of the International Joint Conference on Neural Networks (IJCNN'03), Portland, Oregon, pp. 1396−1401, (2003)

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

[Fahlman & Lebiere, 1990] Fahlman, S. and Lebiere C. "The Cascade-Correlation Learning Architecture" in D. S. Touretzky (ed.), Advances in Neural Information Processing Systems 2, Morgan Kaufmann, 1990.

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

[Guetlein et al., 2009] Guetlein, M., Frank, E., Hall, M., Karwath, A. "Large Scale Attribute Selection Using Wrappers"; In Proc. IEEE Symposium on CIDM, pp.332-339, 2009.

[Jacob, 1988] Jacob, R. "Increased Rates of Convergence Through Learning Rate Adaptation"; Neural Networks, Vol. 1 Issue 4, pp. 295-307, 1988.

[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.

[Rumelhart & McClelland, 1986] Rumelhart, D., McClelland, J. "Parallel Distributed Processing"; Explorations in the Microstructure of Cognition, Cambridge, MA: MIT Press, 1986.

[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*

Major Fields of Scientific Research: data mining, neural networks, support vector machines, adaptive resonance theory.

Mairead Hogan *– Business Information Systems, Cairnes Business School, National University of Ireland, Galway, Ireland; e-mail: mairead.hogan@nuigalway.ie*

Major Fields of Scientific Research: HCI, usability and accessibility in information systems, data mining.

Borislav Stoyanov *– Department of Computer Science, Shumen University, Shumen, Bulgaria; e-mail: bpstoyanov@abv.bg*

Major Fields of Scientific Research: artificial intelligence, cryptography, data mining.