



ELSEVIER

Computers in Biology and Medicine 36 (2006) 516–525

---

---

Computers in Biology  
and Medicine

---

---

[www.intl.elsevierhealth.com/journals/cobm](http://www.intl.elsevierhealth.com/journals/cobm)

# Impact of missing data in evaluating artificial neural networks trained on complete data

Mia K. Markey<sup>a,\*</sup>, Georgia D. Tourassi<sup>b</sup>, Michael Margolis<sup>a</sup>, David M. DeLong<sup>c</sup>

<sup>a</sup>*Biomedical Engineering Department, The University of Texas at Austin, 1 University Station, C0800, ENS617B, Austin, TX 78712, USA*

<sup>b</sup>*Department of Radiology, Duke University Medical Center*

<sup>c</sup>*Department of Biostatistics and Bioinformatics, Duke University Medical Center*

Received 24 November 2004; accepted 17 February 2005

---

## Abstract

This study investigated the impact of missing data in the evaluation of artificial neural network (ANN) models trained on complete data for the task of predicting whether breast lesions are benign or malignant from their mammographic Breast Imaging and Reporting Data System (BI-RADS<sup>TM</sup>) descriptors. A feed-forward, back-propagation ANN was tested with three methods for estimating the missing values. Similar results were achieved with a constraint satisfaction ANN, which can accommodate missing values without a separate estimation step. This empirical study highlights the need for additional research on developing robust clinical decision support systems for realistic environments in which key information may be unknown or inaccessible.

© 2005 Elsevier Ltd. All rights reserved.

*Keywords:* Diagnosis; Computer-assisted; Mammography; Breast neoplasms

---

## 1. Introduction

Among American women, breast cancer is the most common cancer, excluding skin cancers, and is the second leading cause of cancer deaths, after lung cancer [1]. Early detection via mammography

---

\* Corresponding author. Tel.: +1 512 471 1711; fax: +1 512 471 0616.

E-mail address: [mia.markey@mail.utexas.edu](mailto:mia.markey@mail.utexas.edu) (M.K. Markey).

improves survival [2]. However, mammography is not perfect and improvements in both the sensitivity and specificity of the exam are needed. One approach is to develop computer-aided detection and diagnosis (CAD) systems to aid radiologists in the interpretation of mammograms [3–5].

Some previous studies in breast cancer CAD have explored the use of statistical and machine learning models for predicting the pathology of breast lesions from experts' descriptions of mammographic findings [6–10]. This approach is most compelling when the Breast Imaging and Reporting Data System (BI-RADS<sup>TM</sup>) lexicon [11] is used since in principle these data are routinely collected already. However, one limitation of the work to date in this area is that carefully curated databases with minimal missing values have been collected for research purposes, yet most clinical databases in routine practice are missing some pieces of information.

The purpose of this study was to explore the impact of missing data in the evaluation of artificial neural network (ANN) models trained on complete data for the task of predicting whether breast lesions are benign or malignant based on their BI-RADS<sup>TM</sup> descriptors. We investigated two forms of ANNs that have been successfully applied in prior breast cancer CAD studies: feed-forward, back-propagation ANN (BP-ANN) and constraint satisfaction ANN (CSNN). In order to use the BP-ANN, estimates are needed for missing values. Three methods for estimating the missing values were compared: simply replacing missing values with zero, replacing missing values with the mean value from the training set, and using a multiple imputation procedure. In the multiple imputation method, several estimates are produced for the missing values to create multiple imputed versions of the data set [12]. The analysis is then performed on each imputed version of the data set and the results are subsequently combined. Prior studies have found that multiple imputation methods are frequently better than single imputation methods, such as replacing missing values with the mean. Thus, one goal of this study was to compare single and multiple imputation methods for coping with various levels of missing data processed by an ANN trained on complete data for this CAD task. In addition, we hypothesized that the non-hierarchical nature of the CSNN would enable it to effectively estimate the missing values at the same time that it predicted the biopsy outcome, avoiding the extra estimation step needed in using the BP-ANN.

## 2. Materials and methods

### 2.1. Data set

The data set consisted of 604 non-palpable, mammographically suspicious breast masses that underwent biopsy (core or excisional) at Duke University Medical Center from 1990 to 2000. The pathology outcome was coded as binary variable of benign vs. malignant. Experienced mammographers described each case using the BI-RADS<sup>TM</sup> lexicon [11]. In particular, the descriptors used in this study were mass margin, mass shape, mass size, and mass density, which were numerically encoded as described in our previous studies [13]. Patient age was also included since it was found to be an important variable in prior analyses of CAD systems that utilize BI-RADS<sup>TM</sup> descriptors [14]. Summary characteristics of the data set are shown in Fig. 1. The data were randomly partitioned into two sets of equal size, *A* and *B*. Summary properties of the cross-validation sets are provided in Table 1. Only cases for which all variable values were present were included in this study.

Mass Margin	Number	Mass Shape	Number	Mass Density	Number
Well-circumscribed	203	Round	89	Fat-containing	5
Microlobulated	28	Oval	223	Low density	19
Obscured	124	Lobulated	122	Isodense	488
Ill-defined	163	Irregular	170	High density	92
Spiculated	86				
	<b>Average</b>	<b>Standard Deviation</b>			
<b>Mass Size (mm)</b>	12.1	6.3			
<b>Age (years)</b>	54.8	12.9			

Fig. 1. Summary characteristics of the database of mammographic masses used in this study.

Table 1

Summary characteristics of the cross-validation data sets of mammographic masses used in this study

	Benign	Malignant (%)	Total
Set A	204	98 (32%)	302
Set B	205	97 (32%)	302
Total	409	195 (32%)	604

## 2.2. Missing data estimation

Alternate versions of each data set (*A*, *B*) were created in which a fraction of the variable values were missing. The amount of missing data as a fraction of the total number of variable values (302 cases  $\times$  5 = 1510) was 10%, 20%, 30%, or 40%. In the remainder of the paper, the notation “A0” will be used to indicate set *A* with 0% missing, “A10” to indicate set *A* with 10% missing, and so on.

Three methods for estimating the missing values were compared: simply replacing missing values with zero, replacing missing values with the mean value from the training set, and using a multiple imputation procedure. The multiple imputation was performed in SAS<sup>®</sup> (SAS, Cary, NC) using the MI procedure with the default parameters. Five imputed versions were created of each data set at each level of missing data. The BP-ANN outputs were averaged across the five imputations.

## 2.3. BP-ANN

A feed-forward back-propagation artificial neural network (BP-ANN) was trained to predict the biopsy outcome from the BI-RADS<sup>™</sup> descriptors and patient age. BP-ANN is common machine learning algorithm that has been described in several excellent textbooks (e.g., [15,16]). Briefly, the output of each neuron in a BP-ANN is the result of an activation function ( $y = 1/(1 + e^{-x})$ ) applied to a weighted sum of the inputs to the neuron. The weights are the parameters adjusted as the network learns a given task. The ANN is feed-forward in the sense that each neuron in one layer feeds into each neuron in the next layer. The BP-ANN is trained to minimize the mean of the sum-of-squares error (MSE) using the back-propagation algorithm. The MSE is the squared difference between the network output ( $y_i \in (0, 1)$ ) and network target ( $t_i \in \{0, 1\}$ ), averaged over all of the cases ( $N$ , indexed by  $i$ ). The back-propagation algorithm details how the error should be propagated back through the network to adjust the weights.

The number of hidden nodes, learning rate, momentum, and number of iterations were empirically optimized through leave-one-out training on sets *A0* and *B0* separately. The network parameters determined were then used for the remainder of the study. Custom software in the C language was used to implement the BP-ANN. Three methods for estimating the missing data as described above were applied prior to testing the BP-ANN.

#### 2.4. CSNN

A CSNN was also trained to predict the biopsy outcome from the BI-RADS<sup>TM</sup> descriptors and patient age. Briefly, the CSNN is a Hopfield-type network of neurons arranged in a non-hierarchical way. There are symmetric, bidirectional weights between all pairs of neurons but there are no reflexive weights. The CSNN operates as a nonlinear, dynamic system that tries to reach a globally stable state by adjusting the activation levels of the neurons under the constraints imposed by the a priori fixed weight values. The Lyapunov energy function was used as a measure of the network stability. The weights were predetermined using autoassociative back-propagation neural networks (auto-BP). Custom software in the C language was used to implement the CSNN and has been previously described [10,17]. Because of its non-hierarchical structure, the CSNN is able to analyze data for which values are missing without a prior estimation step. In essence, the CSNN imputes the missing input values in parallel with inferring the value of the biopsy node. The structures of the BP-ANN and CSNN networks are compared in Fig. 2.

#### 2.5. ROC

Receiver operating characteristic (ROC) curves can be used to show the trade-off in sensitivity and specificity achievable by a classifier by varying the threshold on the output decision variable [18,19]. Sensitivity or the true positive fraction (TPF) is the fraction of positive cases that were classified correctly as positive. The specificity, or one minus the false positive fraction (FPF), is the fraction of negative cases that were correctly classified as negative. An ROC curve is generated by applying a threshold to the output of a classification scheme and then plotting the (FPF, TPF) pairs for each threshold. The performance of classification methods can be compared in terms of indices calculated from their curves. In particular, the area under the ROC curve (AUC) is often used as a measure of classifier performance. The values for AUC range from 0.5 for chance to 1.0 for a perfect classifier.

In this study, empirical (non-parametric) ROC curves were used and the AUC was numerically integrated using the trapezoid rule. The standard errors and paired statistical comparisons were computed using bootstrap sampling on the classifier outputs. Custom software in the C language was used to implement the ROC analyses.

#### 2.6. Summary of experimental design

A BP-ANN model was trained on data *A0*, without any missing data values. For each level of missing data (10%, 20%, 30%, and 40%) in set *B*, the missing values were estimated in three ways: simply replacing missing values with zero, replacing missing values with the mean value from the training set, and using a multiple imputation procedure. The BP-ANN model trained on *A0* was then tested on set *B* with different quantities of missing data after values were estimated for the missing data (sets *B10*, *B20*, *B30*, *B40*). The same procedure was repeated for training on *B* and testing on *A*.

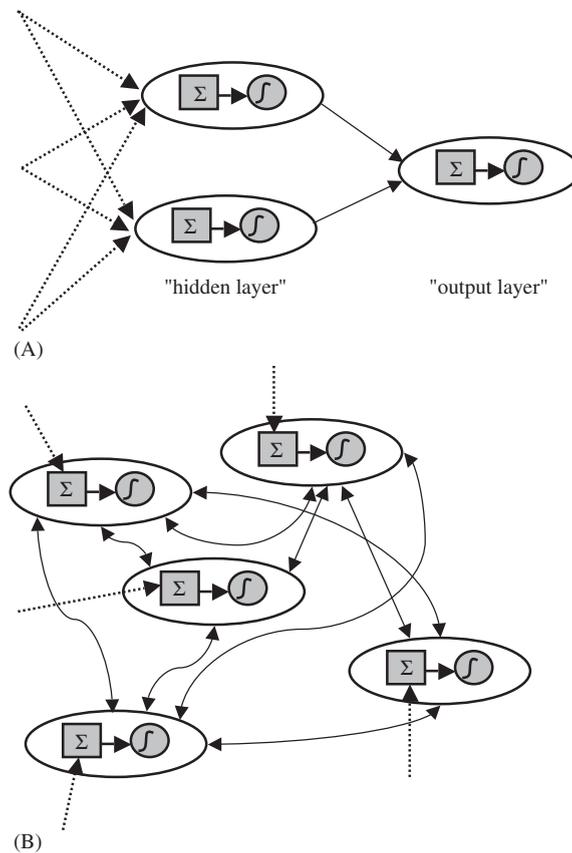


Fig. 2. In both the BP-ANN (A) and CSNN (B), the inputs to network nodes are combined by weighted summation and passed through an activation function to form the network output or activation level. External inputs (data values) are indicated by dashed arrows above. In the BP-ANN, each node in the “hidden” layer accepts multiple data inputs and the outputs of those nodes are passed to the “output” layer nodes. In the CSNN, the nodes are not organized into layers. Each node in the CSNN accepts a single data input as well as inputs from all other nodes. Moreover, the connections between nodes within the CSNN are all bidirectional while they are unidirectional in the BP-ANN. The non-hierarchical structure of the CSNN enables it to simultaneously solve for the activation levels of multiple nodes that lack external inputs.

### 3. Results

The learning parameters and network structure for a BP-ANN model were determined empirically by leave-one-out training on set *A* with no missing values, i.e., *A*<sub>0</sub>, for the task of predicting biopsy outcome of breast masses from BI-RADS™ descriptors and patient age. The selected parameters were then used to train the BP-ANN on set *A*<sub>0</sub> and test it on set *B* with no missing values, i.e., *B*<sub>0</sub>. The area under the resulting ROC curve was  $0.94 \pm 0.01$ . Similarly, network parameters were optimized for set *B*<sub>0</sub> and used to train a model which was tested on set *A*<sub>0</sub>. The AUC for training on *B*<sub>0</sub> and testing on *A*<sub>0</sub> was  $0.92 \pm 0.02$ . The difference in AUC between the two cross-validation directions was not statistically significant ( $p = 0.28$ , unpaired *z*-test).

Table 2

Performance in terms of ROC AUC for ANN models trained on complete data but tested on data sets with various percentages of missing values

Train	Test	BP-ANN (0)	BP-ANN (Ave)	BP-ANN (MI)	CSNN
A0	B0	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.95 ± 0.01
A0	B10	0.84 ± 0.03	0.94 ± 0.01	0.94 ± 0.01	0.95 ± 0.01
A0	B20	0.78 ± 0.03	0.92 ± 0.02	0.91 ± 0.02	0.92 ± 0.02
A0	B30	0.69 ± 0.03	0.90 ± 0.02	0.89 ± 0.02	0.91 ± 0.02
A0	B40	0.72 ± 0.03	0.94 ± 0.02	0.89 ± 0.02	0.90 ± 0.02
B0	A0	0.92 ± 0.02	0.92 ± 0.02	0.92 ± 0.02	0.91 ± 0.02
B0	A10	0.81 ± 0.03	0.92 ± 0.02	0.92 ± 0.02	0.88 ± 0.02
B0	A20	0.75 ± 0.03	0.89 ± 0.02	0.90 ± 0.02	0.88 ± 0.03
B0	A30	0.76 ± 0.03	0.89 ± 0.02	0.89 ± 0.02	0.85 ± 0.02
B0	A40	0.73 ± 0.03	0.84 ± 0.02	0.83 ± 0.02	0.79 ± 0.03

For the BP-ANN models, the missing values were estimated three ways: simply replacing missing values with zero (“0”), replacing missing values with the mean value from the training set (“Ave”), and using a multiple imputation procedure (“MI”).

The BP-ANNs trained on complete data sets were then evaluated on data sets with various levels of missing data. One way that the missing values were estimated was by simply replacing them with zeros. This method is sometimes used because it is very easy to do. Because of the way the BI-RADS<sup>TM</sup> variables were encoded, using a “0” value for a mass descriptor is the same as indicating that no mass was present. Thus, this approach optimistically assumes that there is no evidence for cancer when a variable value is missing. As might be expected, this method of estimating the missing values is not very effective and the performance of the BP-ANN drops of quickly as the fraction of missing data increases (Table 2). Even with only 10% of the values missing, the decrease in the AUC (B10, 0.84 ± 0.03) is statistically significantly ( $p \ll 0.01$ ) lower than that for testing on the set with no missing data (B0, 0.94 ± 0.01) for the BP-ANN trained on set A0. The difference is likewise significant for the model trained on set B0 ( $p \ll 0.01$ ).

The second way that missing values in the test set were estimated for the BP-ANN analyses was by replacing missing values with the mean of that variable in the training set. This method of estimating the missing values was effective. For the model trained on set A0, there was no statistically significant drop in performance relative to that of testing on B0 seen for 10% missing data ( $p = 0.31$ ), but the difference was significant for 20%, 30%, or 40% missing data ( $p \ll 0.01$ ). Similarly, for the BP-ANN trained on set B0, at least 10% of the values in the testing set could be missing before there was a statistically significant decrease in AUC relative to testing on A0.

The third way that missing values in the test set were estimated for the BP-ANN analyses was using a multiple imputation procedure. Five imputed versions were created of each data set at each level of missing data and then the BP-ANN outputs were averaged across the five imputations. As with the mean imputation method, no statistically significant decrease in AUC was seen for the model trained on set A0 at 10% missing data ( $p = 0.28$ ), but there was a significant decrease for higher levels of missing data. Likewise, for the BP-ANN trained on set B0, at least 10% of the values in the testing set could be missing before there was a statistically significant drop in performance relative to testing on A0.

While multiple imputation methods are better than single imputation methods in general and this has been demonstrated in other applications, for this CAD task no benefit was seen. For both cross-validation directions, there was no statistically significant difference in the AUC between the BP-ANN using mean imputation and the BP-ANN using multiple imputation for any of the levels of missing data studied.

Finally, CSNNs were trained for the task of predicting biopsy outcome of breast masses from BI-RADS<sup>TM</sup> descriptors and patient age. As with the BP-ANN analyses, CSNN models were trained for each of the two sets without any missing values (*A0*, *B0*) and then tested on the other set with various levels of missing data. However, unlike the BP-ANN, the CSNN does not require a separate estimation step. The non-hierarchical structure of the CSNN enables it to predict the values of multiple variables at once, including both the biopsy outcome and mass descriptors.

In a previous study, we demonstrated that the CSNN model was able to achieve a performance comparable to that obtained with a BP-ANN for this breast cancer CAD task using a data set containing multiple lesion types [10]. The two ANN models were likewise similar in this study using only breast masses (Table 2). The CSNN performed better than the BP-ANN when trained on *A0* and tested on *B0* ( $p=0.04$ ), but for the other cross-validation direction the BP-ANN was non-significantly better ( $p=0.24$ ). The trend was that both models performed somewhat better in the first cross-validation than the second in the case of no missing data. As discussed above, the difference was not significant for the BP-ANN ( $p=0.28$ , unpaired *z*-test). However, for the CSNN, the difference was borderline significant ( $p=0.05$ , unpaired *z*-test).

For the CSNN trained on set *A0*, there was not a statistically significant drop in performance relative to that of testing on *B0* for 10% missing data ( $p=0.63$ ), but the performance decrease was significant for higher levels of missing data. This is exactly the same trend that was seen with the BP-ANN using either mean imputation or multiple imputation. The results were similar for the CSNN trained on set *B0* (Table 2); however, the drop in performance at 10% missing was already significant ( $p < 0.01$ ).

#### 4. Discussion

The purpose of this study was to explore the impact of missing data in the evaluation of ANN models trained on complete data for the task of predicting whether breast lesions are benign or malignant based on their BI-RADS<sup>TM</sup> descriptors. We investigated two forms of ANNs that have been successfully applied in prior breast cancer CAD studies: feed-forward, BP-ANN and CSNN. For testing the BP-ANN, three methods for estimating the missing values were compared: simply replacing missing values with zero, replacing missing values with the mean value from the training set, and using a multiple imputation procedure.

As expected, the method of simply replacing missing values with zeros resulted in a dramatic drop in performance relative to that using the complete data set. The decrease in the area under the ROC curve was statistically significant even for a missing data level of only 10%. Clearly, this simple approach to dealing with missing data is not advisable, at least for this breast cancer CAD task.

Replacing missing values in the test set with the mean of that variable in the test set and using a multiple imputation procedure were found to be equally useful in this situation. Either approach enabled the BP-ANN model trained on a complete data set to maintain high performance when evaluated on a set with missing information. For levels of missing data above 10%, the decrease in the ROC AUC relative to evaluation on the complete set was statistically significant. However, the trend is that reasonable

performance levels are maintained up to about 30% missing data using either method for estimating the missing values with a BP-ANN model (Table 2).

Most interesting, the CSNN model trained on complete data was shown to maintain a high performance level when evaluated on data with missing values, without a separate estimation step. The non-hierarchical structure of the CSNN enables it to predict the values of multiple nodes in the network simultaneously, such as both the biopsy outcome and the value of a missing mass descriptor. This is a very appealing property since it allows for a one-step process that avoids the additional task of optimizing a separate estimation procedure. However, while we conclude that the CSNN and BP-ANN models have comparable performance for this task, there is some evidence that the CSNN was more sensitive to the chance differences between the data set partitions in the cross-validation analysis.

In conclusion, we have demonstrated for a breast cancer CAD task the dramatic impact that missing data can have in the performance of an ANN that was trained on a complete data set. It is encouraging that even very simple imputation methods may be adequate for some applications, such as was shown here. Finally, the comparison of different forms of ANN models suggests that ease of accommodating missing data may be an important factor to consider in selecting a classifier for a use in a CAD system. Classifiers that have equivalent performance in ideal laboratory conditions may have differing merits when the additional challenges of the clinical setting are taken into account. This empirical study highlights the need for additional research on developing robust clinical decision support systems for realistic environments in which key information may be unknown or inaccessible.

## 5. Summary

Artificial neural networks (ANN) are frequently used in the development of computer-aided diagnosis systems for breast cancer detection and diagnosis. One class of models uses descriptions of mammographic lesions encoded following the Breast Imaging and Reporting Data System (BI-RADS<sup>TM</sup>) lexicon. Data sets that have been carefully curated to ensure completeness are generally used; however, in routine practice, some information is typically missing in clinical databases. The purpose of this study was to explore the impact of missing data in the evaluation of ANN models trained on complete data for the task of predicting whether breast lesions are benign or malignant from their BI-RADS<sup>TM</sup> descriptors. A feed-forward, back-propagation ANN (BP-ANN) was tested with three methods for estimating the missing values: replacing missing values with zero, replacing missing values with the mean value, and using a multiple imputation procedure. As expected, replacing missing values with zeros results in a large drop in performance relative to the no missing data condition, as measured by the area under the receiver operating characteristic curve. Both estimation by the mean and multiple imputation enabled the BP-ANN model to maintain high performance in the presence of a small amount of missing data in the testing set. Similar results were achieved with a constraint satisfaction ANN (CSNN). Since the CSNN can accommodate missing values without a separate estimation step, it may be an advantageous ANN model. This empirical study highlights the need for additional research on developing robust clinical decision support systems for realistic environments in which key information may be unknown or inaccessible.

## Acknowledgment

The authors would like to thank Al Daniel for scientific programming assistance.

## References

- [1] American Cancer Society, Cancer Facts and Figures 2004, American Cancer Society, Atlanta, 2004.
- [2] C.H. Lee, Screening mammography: proven benefit, continued controversy, *Radiol. Clin. N. Am.* 40 (2002) 395–407.
- [3] M.L. Giger, Computer-aided diagnosis of breast lesions in medical images, *Comput. Sci. Eng.* 2 (2000) 39–45.
- [4] M.L. Giger, N. Karssemeijer, S.G. Armato, III, Computer-aided diagnosis in medical imaging, *IEEE Trans. Med. Imag.* 20 (2001) 1205–1208.
- [5] C.J. Vyborny, M.L. Giger, R.M. Nishikawa, Computer-aided detection and diagnosis of breast cancer, *Radiol. Clin. N. Am.* 38 (2000) 725–740.
- [6] Y. Wu, M.L. Giger, K. Doi, C.J. Vyborny, R.A. Schmidt, C.E. Metz, Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer, *Radiology* 187 (1993) 81–87.
- [7] C.E. Kahn Jr., L.M. Roberts, K.A. Shaffer, P. Haddawy, Construction of a Bayesian network for mammographic diagnosis of breast cancer, *Comput. Biol. Med.* 27 (1997) 19–29.
- [8] C.E. Floyd Jr., J.Y. Lo, G.D. Tourassi, Case-based reasoning computer algorithm that uses mammographic findings for breast biopsy decisions, *Am. J. Roentgenol.* 175 (2000) 1347–1352.
- [9] J.A. Baker, P.J. Kornguth, J.Y. Lo, C.E. Floyd Jr., Artificial neural network: improving the quality of breast biopsy recommendations, *Radiology* 198 (1996) 131–135.
- [10] G.D. Tourassi, M.K. Markey, J.Y. Lo, C.E. Floyd Jr., A neural network approach to breast cancer diagnosis as a constraint satisfaction problem, *Med. Phys.* 28 (2001) 804–811.
- [11] American College of Radiology, Illustrated Breast Imaging Reporting and Data System (BI-RADS<sup>TM</sup>), American College of Radiology, Reston, VA, 1998.
- [12] J.L. Schafer, Multiple imputation: a primer, *Statist. Methods Med. Res.* 8 (1999) 3–15.
- [13] M.K. Markey, J.Y. Lo, G.D. Tourassi, C.E. Floyd Jr., Self-organizing map for cluster analysis of a breast cancer database, *Artif. Intell. Med.* 27 (2003) 113–127.
- [14] J.Y. Lo, J.A. Baker, P.J. Kornguth, C.E. Floyd Jr., Effect of patient history data on the prediction of breast cancer from mammographic findings with artificial neural networks, *Acad. Radiol.* 6 (1999) 10–15.
- [15] T.M. Mitchell, *Machine Learning*, WCB/McGraw-Hill, Boston, 1997.
- [16] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, 1995.
- [17] G.D. Tourassi, M.K. Markey, J.Y. Lo, Validation of a constraint satisfaction neural network for breast cancer diagnosis: new results from 1030 cases, in: M. Sonka, J.M. Fitzpatrick (Eds.), *Medical Imaging 2003: Image Processing*, Proceedings of the SPIE 5032, San Diego, 2003, pp. 207–214.
- [18] C.E. Metz, Basic principles of ROC analysis, *Semin. Nucl. Med.* 8 (1978) 283–298.
- [19] C.E. Metz, ROC methodology in radiologic imaging, *Invest. Radiol.* 21 (1986) 720–733.

**Mia K. Markey** is currently an Assistant Professor in Biomedical Engineering at The University of Texas at Austin. She earned her B.S. in Computational Biology from Carnegie Mellon and her Ph.D. in Biomedical Engineering, as well a Certificate in Bioinformatics, from Duke University. The mission of the Biomedical Informatics Lab ([www.bme.utexas.edu/research/informatics](http://www.bme.utexas.edu/research/informatics)) is to design cost-effective, computational medical decision aids that will help physicians better diagnose, treat, and manage cancer.

**Georgia Tourassi**, is an Assistant Research Professor in the Department of Radiology at Duke University Medical Center and Adjunct Associate Professor of Computer Engineering and Computer Science at the University of Louisville. She earned a B.S. in Physics from the University of Thessaloniki in Greece and the Ph.D. in Biomedical Engineering from Duke University. Dr. Tourassi is a member of the Institute of Electrical and Electronics Engineers (IEEE), the International Society for Optical Engineering (SPIE), the American Association of Physicists in Medicine (AAPM), and the Radiological Society of North America (RSNA). She is the associate editor in Radiology and referee in several journals such as *Medical Physics*, *Academic Radiology*, and *Journal of Electronic Imaging*. Her research interests include applications of artificial intelligence in computer-aided medical diagnosis and medical image processing where she has numerous publications. Her research work has been supported by the National Institutes of Health, US, Army, and the Whitaker Foundation.

**Michael Margolis** earned his B.S. in Electrical Engineering with a Biomedical Concentration at The University of Texas at Austin. He is currently pursuing his M.D. at the University of Texas Medical Branch in Galveston.

**David M. DeLong** received a M.S. degree in statistics in 1975 and a Ph.D. in physics in 1977 from the University of North Carolina. He joined SAS Institute where he was a statistical software developer and later Director of the Applications Division. In 1993 he joined the Biostatistics department at the Duke University Medical Center where he is currently an Assistant Research Professor in the Biostatistics and Bioinformatics Department.