Mammographic Diagnosis for Breast Cancer Biopsy Predictions Using Neural Network Classification Model and Receiver Operating Characteristic (ROC) Curve Evaluation

Kathleen H. Miao, and George J. Miao, *Senior Member, IEEE*

***Abstract*—The most common method for screening and diagnosing breast cancer is mammography. However, it lacks high diagnostic accuracy. The low positive predictive values of breast biopsy outcomes using mammogram interpretations often lead to unnecessary biopsies for patients with benign outcomes. In this research report, a mammographic diagnostic method is presented in distinguishing malignant breast cancer and benign disease for biopsy outcome predictions using a neural network classification model and receiver operating characteristic (ROC) curve evaluation. The proposed model uses a two-stage back-propagation neural network approach including both linear and nonlinear components of calculations with iterative training processes and an adjustable learning rate. The iterative training processes along with the adjustable learning rate can ensure that the model has a low minimum mean-square error (MMSE) throughout the training of the model. The probability of misclassification error and performance of our model in diagnosing malignant breast cancer and benign disease for breast biopsy outcome predictions have been evaluated based on a large mammographic mass dataset using the model sensitivity, specificity, and ROC curve analysis. An estimated area of the ROC curve of our model is 0.9626±0.0069 for breast biopsy outcome predictions, which outperforms the diagnostic accuracy of previously reported methods. The sensitivity, specificity, precision, and accuracy of our model simulations are 89.33%, 89.93%, 89.33% and 89.64%, respectively. Therefore, our model along with mammography can provide highly accurate and consistent diagnoses in distinguishing malignant and benign cases for breast cancer biopsy outcome predictions, reducing the number of unnecessary biopsies for patients with benign outcomes.**

***Index Terms*—breast cancer, mammography, malignant, benign, biopsy, receiver operating characteristic (ROC) curve, neural network, classification, sensitivity, specificity, precision, minimum mean-square error (MMSE), prediction**

Manuscript received August 29, 2013.

Kathleen H. Miao is with the College of Arts and Sciences at Cornell University, Ithaca, NY 14853, USA (e-mail: khm37@cornell.edu).

George J. Miao is with Flezi, LLC, San Jose, CA 95134, USA (e-mail: g.j.miao@ieee.org).

I. Introduction

According to the World Health Organization [1], 7.6 million people worldwide die from cancer each year. Breast cancer is currently one of the top cancers diagnosed in women in both developed and developing nations. Globally, breast cancer is the principal cause of cancer death among women. Breast cancer tumors typically do not produce symptoms until they are relatively large at advanced stages. Early breast cancer diagnosis improves chances of long-term survival for patients. However, the majority of breast cancer deaths occur in low- and middle-income countries, where most of the women are diagnosed in later stages mainly due to lack of awareness and barriers to health service access [1].

According to Centers for Disease Control and Prevention [2], cancer is the second leading cause of death in the United States. American Cancer Society (ACS) projects an estimated 1,638,910 new cancer cases to occur in 2012 [3-4]. Among all cancer cases, breast cancer ranks as the second leading cause of cancer death and the leading cause of new cancer cases in women [4]. ACS also forecasts 229,060 new cases of invasive breast cancer and 39,920 breast cancer deaths in the United States in 2012.

Female breast cancer mortality rates in the United States initially increased from 1975 to 1990, and then gradually decreased annually up to the current year [5]. The recent decrease in female breast cancer mortality rates is generally attributed to greater awareness of breast cancer, earlier detections, enhanced diagnostic methods, and advanced medical treatments. Thus, diagnosing breast cancer in patients at an early stage before symptoms develop is an important factor influencing their chances of long-term survival.

Cancer is a disease that causes cells in the body to change and grow out of control. Most types of cancer eventually form lumps or masses referred to as tumors [4]. Breast cancer tumors typically do not produce symptoms until they are relatively large at advanced stages [3-4]. Most breast cancer tumors at earlier stages are benign, which are not yet malignant and life-threatening. They do not yet grow uncontrollably or spread. Thus, early detection of breast cancer masses can improve the chances of long-term survival

for patients. An accurate diagnosis highly enhances early detection, ensuring that the patient receives the most appropriate treatment and care. Therefore, diagnosing breast cancer in women at an earlier stage before symptoms appear is important.

Currently, there are four main methods of breast cancer diagnosis, which are used to distinguish malignant tumors from benign ones. They include surgical biopsy, magnetic resonance imaging (MRI), mammography, and fine needle aspiration (FNA) cytology. For surgical biopsy, the reported accuracy of breast cancer diagnosis is close to 100% [6], and for MRI, benign and malignant diagnoses are 70% and 92% accurate, respectively [7]. For mammography, the diagnostic accuracy of distinguishing malignant breast cancer and benign disease is between 68% and 79% [6]. For FNA cytology, diagnostic accuracy varies from 65% to 98% [9-10]. Of the four methods, surgical biopsy is the most accurate diagnostic method, but it is expensive, invasive, and inconvenient for the patient. The diagnostic accuracy of malignant cases via MRI screening is relatively high, but the accuracy of its benign diagnoses is comparably lower. In addition, breast MRI scans are relatively expensive. Mammography, which is a non-invasive method, lacks high diagnostic accuracy. The low positive predictions of mammogram interpretations lead to a high number of unnecessary biopsies for benign outcomes [8]. Even though FNA is the least expensive method, it is an invasive procedure. When using MRI, FNA cytology, or mammography to detect a breast cancer tumor, surgical biopsy is usually needed to confirm the state of its malignancy [11]. Increasing the diagnostic accuracy of mammography can reduce the number of unnecessary surgical biopsies for benign outcomes and provide patients with the option of a comparably inexpensive and non-invasive procedure.

Mammography is the screening process that uses low-energy x-rays to examine a woman's breast. It is used as a diagnostic and screening tool for early detection and diagnosis of breast cancer, typically through detection of characteristic masses and/or microcalcifications (such as tiny deposits of calcium).

A mammogram is an x-ray picture or image of the breast obtained by using mammography. The x-ray image makes it possible to detect tumors, which cannot be felt by patients. Currently, there are two kinds of mammograms [12]: screening and diagnostic mammograms. The screening mammogram is used to check for breast cancer in women who have no signs or symptoms of the disease. It can also be used to find microcalcifications that sometimes indicate the presence of breast cancer. The diagnostic mammogram is used to check for breast cancer after a lump or other symptom of the disease has been found. Besides a lump, signs of breast cancer can include breast pain, thickening of the breast skin, nipple discharge, and a change in breast size or shape [3, 12]. However, these signs may also be signs of benign disease symptoms. Thus, the diagnostic mammogram can also be used to evaluate changes found during a screening mammogram or to view breast tissue when it is difficult to obtain it in a screening mammogram.

Early detection of breast cancer by using mammography greatly improves the chances of survival for patients [8, 13-17]. Although mammography is a sensitive procedure for detecting breast cancer, it lacks high diagnostic accuracy. In other words, the positive predictive value of breast biopsy outcomes is low, which leads to unnecessary biopsies for benign outcomes. In addition, there are false-negative and false-positive results [12]. The false-negative results occur when mammograms appear normal even through breast cancer is present. The false-negative results may lead to delays in treatment and a false sense of security for affected women. The false-positive results occur when radiologists decide mammograms are abnormal but no cancer is actually present. In fact, only 10-34% of women, who undergo breast biopsies for mammographically detected, impalpable, and suspicious lesions, are actually found to have malignant pathology [8, 18]. Eventually, several hundreds of thousands of biopsies are performed on benign rather than malignant cases each year. The women undergo unnecessary breast biopsies for benign cases. This leads to discomfort, extra expenses, potential complications, changes in cosmetic breast appearances, anxiety, and psychological distress for affected women [18, 19]. Therefore, if a relatively objective system is created to enhance mammographic diagnosis with consistently high accuracy, patients can then bypass unnecessarily additional examinations and even the need for surgical biopsies.

A common way for radiologists to describe mammogram findings based on the breast image reporting and database system (BIRADS) is established by the American College of Radiology (ACR) [20]. BIRADS includes seven standardized categories that allow radiologists to make an assessment to support the decision of a physician to perform a breast biopsy or a follow-up diagnosis. By using BIRADS along with various characteristics, such as mass shape, obtained from a mammogram, several computer aided diagnosis (CAD) systems were developed to distinguish malignant breast cancer and benign disease. One research report used CAD approaches to predict breast cancer biopsy outcomes based on an intelligible decision process [8]. Methods in other research papers, included a case-based reasoning classifier using different similarity measures based on Euclidean and Hamming distances [19, 21-23], an artificial neural network approach based on BIRADS descriptions [24, 25], a classification based on a decision tree approach [26], and a prediction method using a distributed genetic programming approach [13]. These methods were proposed using mammogram data to predict breast cancer biopsy outcomes or to classify malignant and benign lumps.

In this research report, a different and enhanced approach is proposed for mammographic diagnosis for breast biopsy outcome predictions utilizing a neural network classification model and ROC curve evaluation. This classification model contains two components: a training model and a diagnostic model. The training model is based on a two-stage back-propagation neural network approach, along with an iterative training method and an adjustable learning rate. The diagnostic model is able to distinguish and classify malignant

breast cancers and benign diseases for breast biopsy outcome predictions. The probability of misclassification error and performance of the proposed classification model were evaluated using the model sensitivity, specificity, and ROC curve analysis.

II. Materials and Methods

In this section, mammographic dataset of breast cancer tumors are first described. Then, diagnostic methods of the neural network classification model and ROC curve evaluation are introduced in detail, including the training and diagnostic models, as well as their corresponding structures, algorithms, and approaches.

A. Mammographic Mass Dataset

The mammographic mass dataset, which was used in this research, is obtained from the Mammographic Mass Database available in the UCI Machine Learning Repository [27]. This dataset contains mammographic information of breast cancer clinical instances, contributed by the Institute of Radiology of the University Erlangen-Nuremberg in Germany. Full-field digital mammography was used to collect the clinical instances for the dataset periodically from 2003 to 2006, resulting in a constantly increasing dataset size.

The mammographic mass dataset contains 516 benign and 445 malignant cases, totaling 961 clinical instances. Each clinical instance has five attributes (referred to as BIRADS, age, shape, margin, and density) and one class attribute (referred to as severity) with a binary value of 0 or 1, indicating benign or malignant diagnoses, respectively. The physical meanings of the five attributes and the class attribute are listed in Table 1.

Among the 961 clinical instances, there are a number of instances with missing attribute values, including 2 BIRADSs, 5 ages, 31 shapes, 48 margins, and 76 densities. This results in a total of 131 clinical instances missing either one or two of the five attributes. For consistent analysis in our research, the 131 clinical instances with missing attributes were removed from this dataset, resulting in a dataset of 830 clinical instances. Among the 830 clinical instances, 427 (51.45%) instances are benign disease cases and 403 (48.55%) instances are malignant breast cancer cases.

In our research, the integer values of the BIRADS and margin attributes were normalized by their maximum values of 5, resulting in attribute values ranging from 0.2 to 1. The integer values of the shape and density attributes were normalized by their maximum values of 4, resulting in attribute values ranging from 0.25 to 1. The age attribute integer value was normalized by 100. The severity attribute was a binary value of either 0 or 1, representing benign or malignant diagnoses, respectively.

B. Diagnostic Method and Prediction Model

In this section, we establish the mammographic diagnostic method to classify malignant breast cancers and benign diseases for breast biopsy outcome predictions using the neural network classification model and ROC curve evaluation. In the following subsections, a neural network classification training model, its structure, back-propagation algorithm and approach, as well as the diagnostic model, its structure, algorithm, and method are presented in detail. The performances of the classification training model are evaluated by using the model sensitivity and specificity analyses as well as ROC curve analysis.

TABLE 1
THE MAMMOGRAPHIC MASS DATASET ATTRIBUTES

Attributes	Ranges
BIRADS	1 to 5 (ordinal, non-predictive)
Age	Age of patient in years (integer)
Shape (mass shape)	Round = 1; Oval = 2; Lobular = 3; Irregular = 4 (nominal)
Margin (mass margin)	Circumscribed = 1; Microlobulated = 2; Obscured = 3; Ill-defined = 4; Spiculated = 5 (nominal)
Density (mass density)	High = 1; Iso = 2; Low = 3; Fat-containing = 4 (ordinal)
Severity	Benign = 0; Malignant = 1 (binominal, class)

The Neural Network Training Model

The neural network training model can be considered as a processor that acquires and stores experiential knowledge through a machine learning process. In order to retain the knowledge, synaptic weights that resemble interneuron connections are used. The training process of a learning algorithm involves the modification of the synaptic weights of the model in order to obtain a desired objective.

During the training process, the training model for distinguishing and classifying malignant breast cancer and benign disease is shown in Figure 1. It contains a two-stage neural network classification unit, input vector processor, input size controller, training control center, learning rate controller, addition operation unit, target output class, weight update unit, and training weights unit.

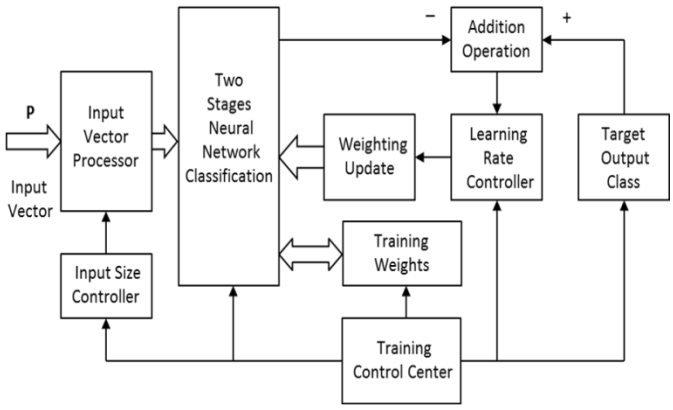

Fig. 1: A neural network training model for mammographic diagnosis to distinguish malignant breast cancer and benign disease in breast biopsy outcome predictions.

During the training process, the input vector ($P \times N$), including N clinical instances and P attributes of the mammographic data (where $P = 5$ attributes including

BIRADS, age, shape, margin, and density), were fed into the input vector processor. The input vector processor first normalized the five input attributes and then generated two additional attributes, including the combined products of "Age*BIRADS" and "Shape*BIRADS." Thus, the output data matrix of the input vector processor, which included *N* clinical instances and 7 attributes, was used for the two-stage neural network classification unit. The classification unit propagated all input patterns for determining all outputs. After comparing the outputs of the model with the target output class, an error was obtained and multiplied by a scaling parameter, which was adjusted by the learning rate controller. Next, the weights were updated after the error was minimized at each stage through the weight update unit. The process was repeated until a sum of squared error (SSE) or MMSE was less than a pre-defined error value or until the training epochs were used up. Then, the resulting weights were stored in the training weights unit.

The training control center communicated with the two-stage neural network classification unit, target output class, input size controller, learning rate controller, and training weights unit. For each iterative training process, the center regulated the input vector size *N*, the number of hidden neurons, and the initial weights. Furthermore, it adjusted the learning rates according to the learning rate controller. As the number of iterative training processes increased, the input vector size *N* was gradually increased accordingly. The final weights of each completed training process were reused as initial weights for the next training process. Consequently, the learning rate controller simultaneously decreased the learning rate. The iterative training processes were repeated until the minimum sum square of a pre-defined error was obtained. The final weights of the model at the last iterative process were stored as the final trained weights in the training weights unit, which would be used for the diagnostic model.

Iterative training processes are especially useful when the input vector size is relatively large. This is because the solution of the neural network model differs when different initial weights are used to train the model. During the iterative training processes, recycling a relatively optimal set of final weights from the previous training process as initial weights for the next training process enables the neural network training model to reach an approximately optimal solution.

Two-Stage Neural Network Classification Structure

The two-stage neural network classification unit including S_1 tan-sigmoid transfer functions, denoted by F_1, in the first neuron layer and one linear transfer function, denoted by F_2, in the second neuron layer is shown in Figure 2. R and S_1 represented the total number of attributes (including the combined attributes) and the number of hidden neurons in the first neuron layer, respectively. $\boldsymbol{W}_1$, $\boldsymbol{B}_1$ and $\boldsymbol{W}_2$, $\boldsymbol{B}_2$ represented hidden neurons of the weights and biases in the first and second neuron layers, respectively. In our research, R signified the 7 attributes (where five were independent attributes and two were combined attributes), and S_1 was set to include 40 hidden neurons. $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ were the weights of the (40×7)

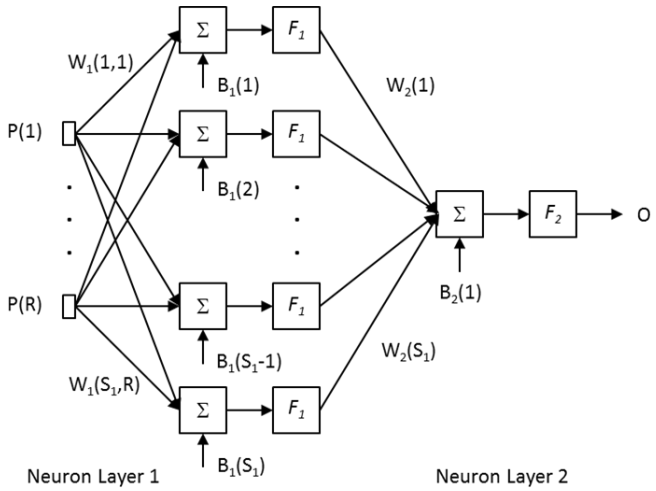

Fig. 2: A two-stage neural network classification structure, where the first and second neuron layers contained S_1 neurons and one neuron, respectively.

matrix and (1×40) matrix, respectively; and $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ were the biases of the (40×1) matrix and (1×1) matrix (which is a scalar), respectively.

In the first neuron layer, the tan-sigmoid transfer function, which was a nonlinear transfer function, generated output values between -1 to 1 as the net input of the neuron ranged from negative to positive infinity [28]. The second neuron layer contained one linear transfer function that could distinguish and designate any output values from the first neuron layer as the two output results of malignant breast cancer and benign disease in mammographic data classification. Both the linear and nonlinear transfer functions did not have minima since the linear and nonlinear transfer functions were differentiable and monotonically increasing functions. Thus, this would tend to preclude error minima from trapping the neural network classification training model when it learned.

A Back-Propagation Approach

The neural network training model used a back-propagation approach for training the two-stage neural network classification unit during the training process. The back-propagation approach based on the Widrow-Hoff learning rule [29, 30] was used to minimize the objective function for the neural network training model. The input vector and the corresponding output vector were used to train the neural network classification model until the training model appropriately approximated a function within a prior defined error value. During the training process, a learning algorithm was used to adjust weights and biases by utilizing the derivative vectors of errors back-propagated through the two-stage neural network classification unit.

The learning algorithm of the first and second neuron layers used for the neural network training model was based on the generalized delta rule [28, 30-32] described in the following steps:

1. The first step was to choose small, random values for the initial neural network weights and biases matrices, $\boldsymbol{W}_1$, $\boldsymbol{B}_1$ and $\boldsymbol{W}_2$, $\boldsymbol{B}_2$, for the first and second neuron layers,

respectively.

2. For the forward propagation calculation of the network, the equation matrix output of the first neuron layer was obtained as follows

$$\boldsymbol{a}_1 = F_1(\boldsymbol{W}_1\boldsymbol{a}_0 + \boldsymbol{B}_1) \quad (1)$$

where $\boldsymbol{a}_0 = \boldsymbol{P}$ denoted the input matrix values for the first neuron layer in the neural network; F_1 was the transfer function of the first neuron layer; and $\boldsymbol{n}_1 = \boldsymbol{W}_1\boldsymbol{a}_0 + \boldsymbol{B}_1$ was the matrix equation of the combined output of the neuron weights in the first neuron layer.

The matrix equation output of the second neuron layer was obtained by

$$\boldsymbol{a}_2 = F_2(\boldsymbol{W}_2\boldsymbol{a}_1 + \boldsymbol{B}_2) \quad (2)$$

where F_2 was the transfer function of the second neuron layer and $\boldsymbol{n}_2 = \boldsymbol{W}_2\boldsymbol{a}_1 + \boldsymbol{B}_2$ was the matrix equation of the combined output of the neuron weights in the second neuron layer.

3. The minimum SSE of the neural network was then calculated by

$$SSE = (\boldsymbol{T} - \boldsymbol{a}_2)^T(\boldsymbol{T} - \boldsymbol{a}_2) \quad (3)$$

where $\boldsymbol{T}$ is the desired pattern output or target output class.

4. This step was to propagate the sensitivities backward through the neural network from the second neuron layer into the first neuron layer. For the second neuron layer, the sensitivity matrix equation was expressed as

$$\boldsymbol{S}_2 = -2F_2'(\boldsymbol{n}_2)(\boldsymbol{T} - \boldsymbol{a}_2) \quad (4)$$

where $F_2'(\boldsymbol{n}_2)$ was the derivative of the transfer function of the second neuron layer.

The sensitivity matrix equation of the first neuron layer was obtained through the back-propagation of the sensitivity from the second neuron layer as described as follows

$$\boldsymbol{S}_1 = F_1'(\boldsymbol{n}_1)(\boldsymbol{W}_2)^T\boldsymbol{S}_2 \quad (5)$$

where $F_1'(\boldsymbol{n}_1)$ was the derivative matrix of the transfer function of the first neuron layer and was given by

$$F_1'(\boldsymbol{n}_1) = \begin{bmatrix} F_1'(n_{11}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & F_1'(n_{1N}) \end{bmatrix} \quad (6)$$

and the derivative of each transfer function in Equation (6) was obtained by

$$F_1'(n_{1j}) = \frac{\partial F_1(n_{1j})}{\partial n_{1j}} \quad (7)$$

where j (for $j=1,\ldots, N$) represented jth neuron in the first neuron layer.

5. The next step was to update the weights and biases of the neural network. For the first and second neural layers, the matrix equation for updating weights was obtained by

$$\boldsymbol{W}_i(k+1) = \boldsymbol{W}_i(k) - \beta\boldsymbol{S}_i(\boldsymbol{a}_{i-1})^T \quad (8)$$

and the matrix equation for updating biases was obtained by

$$\boldsymbol{B}_i(k+1) = \boldsymbol{B}_i(k) - \beta\boldsymbol{S}_i \quad (9)$$

where k was the number of iterations and β was the learning rate during the network training. If set to i =1, the matrix equations of (8) and (9) were used for the first neuron layer; and if set to i = 2, the matrix equations of (8) and (9) were used for the second neuron layer.

Note that in our research, each of the transfer functions in the first neuron layer was characteristically tan-sigmoid as expressed as follows

$$F_1(n_1) = \frac{1}{1+e^{-n_1}} \quad (10)$$

The corresponding derivative computation of Equation (10) was obtained by

$$F_1'(n_1) = \frac{e^{-n_1}}{(1+e^{-n_1})^2} \quad (11)$$

For the second neuron layer, the transfer function was a linear function expressed as follows

$$F_2(n_2) = n_2 \quad (12)$$

where the corresponding derivative function of Equation (12) was $F_2'(n_2) = 1$.

Thus, in order to update the weights and biases of the neural network training model using the back-propagation approach, the learning algorithm involved two phases of propagation and weight updates during the training processes. The above steps from (2) to (5) in the learning algorithm were continuously iterated until the difference between the neural network response and the target output class reached an optimal level, which was less than a pre-defined error value. In other words, the neural network training model repeated the two phases until the SSE or MMSE was less than the pre-defined error value or until the number of training epochs was used up.

The Neural Network Diagnostic Model

The neural network diagnostic model for distinguishing malignant breast cancer and benign disease for breast biopsy outcome predictions is shown in Figure 3. This diagnostic model contained an input vector processor, two-stage neural network classification unit, and two-lever hard-limit classifier. The input vector processor first normalized each input of the 5 attributes (including BIRADS, age, shape, margin, and density) by dividing each by its maximum value and then calculated two additional products of combined attributes, "Age*BIRADS" and "Shape*BIRADS." Thus, for each clinical instance, there was a total of 7 attributes resulting from the outputs of the input vector processor. The 7 attributes were simultaneously fed into the two-stage neural network classification unit. This unit had the same structure as shown in Figure 2, with the final weights loaded from the training model, where $R = 7$ attributes and $S_1 = 40$ hidden neurons in the tan-sigmoid transfer functions in the first neuron layer

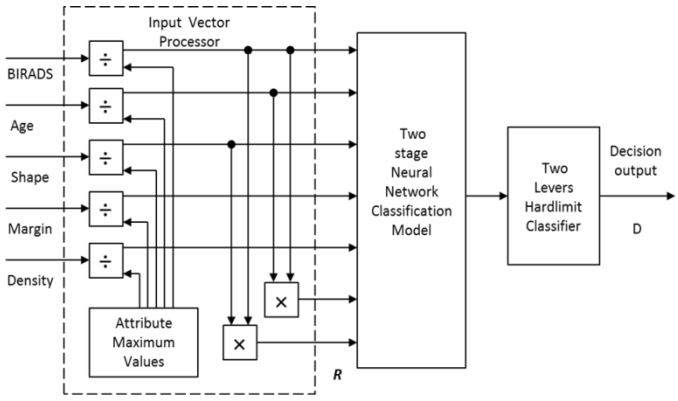

Fig. 3: A neural network diagnostic model for distinguishing malignant breast cancer and benign disease in breast biopsy outcome predictions.

and the neuron in the linear transfer function in the second neuron layer. The function of the two-lever hard-limit classifier, denoted by F_3, produced one of the binary decisions of 0 or 1, where "0" denoted benign disease and "1" denoted malignant breast cancer.

The mathematical relationship between the output vector $\boldsymbol{R}$ from the input vector processor and the binary decision output D, as shown in Figure 3, was obtained by [28]:

$$D = F_3\{F_2[\boldsymbol{W}_2 \times F_1(\boldsymbol{W}_1 \times \boldsymbol{R} + \boldsymbol{B}_1) + \boldsymbol{B}_2]\} \quad (13)$$

where D was a scalar binary decision output value of either 0 or 1, which indicated either benign disease or malignant breast cancer; $\boldsymbol{R}$ was the output vector of the (7×1) matrix of the input vector processor that represented one clinical instance; $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ were the model weights of the (40×7) and (1×40) matrices, respectively, which were referred to as connection weights in the first and second neuron layers; $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ were the model biases of the (40×1) and (1×1) matrices in the first and second neuron layers, respectively; F_1 and F_2 represented the tan-sigmoid nonlinear transfer function and linear transfer function in the first and second neuron layers, respectively; F_3 was the two-lever hard-limit classifier, having a mathematic expression as follows

$$D = F_3(x) = \begin{cases} 0, & x < T \\ 1, & x \geq T \end{cases} \quad (14)$$

where T denoted a threshold value, $0 \leq T \leq 1$. In our research, T was set to 0.5 for the neural network diagnostic model. Therefore, given a new clinical instance of mammographic mass data, the neural network diagnostic model can diagnose and distinguish malignant breast cancer and benign disease for breast biopsy outcome predictions.

The Model Sensitivity and Specificity Analyses

The performance of the neural network classification model is best presented in terms of its sensitivity and specificity, which evaluated its performance based on the number of false positive and false negative instances of the mammographic mass data. For our model, which presented a binary decision of either malignant breast cancer or benign disease, the diagnostic results in terms of positive or negative results can be summarized in Table 2. The sensitivity (referred to as the true positive rate or recall) is the probability of correctly identifying malignant breast cancers given by

$$Sensitivity = \frac{TP}{TP+FN} \quad (15)$$

and the specificity (referred to as the true negative rate) is the probability of correctly identifying benign diseases given by

$$Specificity = \frac{TN}{FP+TN} \quad (16)$$

where the difference of (1 – specificity) is referred to as the false positive rate. The precision (referred to as the positive predictive value) is then defined as

$$Precision = \frac{TP}{TP+FP} \quad (17)$$

Furthermore, for the probability of misclassification error (PME), it is obtained by

$$PME = \frac{FN+FP}{TP+FN+FP+TN} \quad (18)$$

where the difference of (1 – PME) is referred to as the model accuracy.

TABLE 2
DIAGNOSTIC TEST RESULT ACCURACY OF THE MODEL FOR MALIGNANT BREAST CANCERS AND BENIGN DISEASES

	Actual Malignant	Actual Benign	Total
Predicted Malignant	True Positive (TP)	False Positive (FP)	TP + FP
Predicted Benign	False Negative (FN)	True Negative (TN)	FN + TN
Total	TP + FN	FP + TN	TP + FP + FN + TN

The Neural Network ROC Curve Evaluation and Analysis

A neural network ROC curve evaluation is a plot generated by varying a set of trade-off (or threshold) points between the sensitivity and the difference of (1 – specificity) for cases being classified as malignant breast cancer, ranging from 0 to 1. In our simulation modeling in producing the neural network ROC curve evaluation, a set of T values was first adjusted to run discrete units from 0 to 1 in the neural network diagnostic model according to Equation (14) and a set of corresponding sensitivity and (1-specificity) values was calculated when performing the diagnostic model using the clinical instances from the mammographic mass dataset.

An area under the ROC curve is considered an effective measure of inherent validity of a diagnostic test [33] and a metric for evaluating performance of a classification model [13]. The area under the ROC curve of the neural network classification model can be determined by using a trapezoidal approximation [34]:

$$\int_0^1 f(x)dx \cong \sum_{i=0}^{N} \left(\frac{y_i + y_{i+1}}{2}\right)(x_{i+1} - x_i) \quad (19)$$

where *f(x)* denoted the function of the ROC curve, y_i and x_i represented the sensitivity and (1-specificity) at *i*th ($i = 0, 1, 2, \ldots, M$) point, respectively. Accordingly, a standard error (SE) of the area of the ROC curve in Equation (19) is given by [35]

$$SE = \sqrt{\frac{A(1-A) + (N_1-1)(Q_1-A^2) + (N_2-1)(2-A^2)}{N_1 N_2}} \quad (20)$$

where A was an estimate of the area of the ROC curve given in Equation (19); N_1 and N_2 denoted the number of clinical instances of malignant (positive) and benign (negative) results in the mammographic mass dataset, respectively; $Q_1 = A/(2-A)$ and $Q_2 = 2A^2/(1+A)$. Under the assumption that the future clinical instances are drawn from the same distribution, Equations (19) and (20) provided the method of how the neural network classification model will perform diagnoses for future clinical instances of mammographic mass data.

Generally, the area of the ROC curve is used to evaluate and rate the quality of classification models. It can be statistically interpreted as the probability of the classification model to correctly classify malignant breast cancer and benign disease. The higher the area of the ROC curve, the better the classification model and the better the diagnostic test results. For a perfect ranking of the classification model, the area of the ROC curve should be equal to 1, which means that the diagnostic accuracy of the model is 100% accurate in

distinguishing malignant breast cancer and benign disease. In other words, both the sensitivity and specificity of the model are 1 and both false positive and false negative rates are 0. If the area of the ROC curve is equal to 0.5, there is only a 50% chance that the diagnostic test will correctly discriminate between malignant breast cancer and benign disease. On the other hand, the area of the ROC curve of the classification model is useful for finding an optimal trade-off point for the neural network diagnostic model, which we will discuss in the section of conclusion and future work, thereby leading to the least probability of misclassification error for malignant and benign diagnoses of breast cancer biopsy outcome predictions.

III. Results

In our research, the neural network classification model was trained and tested using all of the available 830 clinical instances of the mammographic mass dataset after removing the 131 clinical instances that were each missing one or two attributes. Of the 830 remaining clinical instances, 427 (51.45%) instances were those of benign disease and 403 (48.55%) instances were those of malignant breast cancer.

To evaluate the performance of our neural network classification model, the probability of misclassification error and the area of the ROC curve for the model were estimated using a nonparametric approach based on a resubstitution method [36]. For the resubstitution method, the neural network classification model was trained using a pattern dataset, which included all of the available 830 clinical instances, and was tested to estimate the probability of misclassification error and the area of the ROC curve using the same pattern dataset. Generally, in this research, the resubstitution method resulted in optimistically unbiased estimates of the asymptotic probability of misclassification error and the area under the ROC curve since the pattern dataset, containing the 830 clinical instances, can be considered a relatively large dataset.

A. The Model Training Results

The 830 clinical instances of the mammographic mass dataset were used in mammographic diagnosis to distinguish malignant breast cancer and benign disease for breast biopsy outcome predictions. The pattern dataset, which contained all of the available 830 clinical instances of the mammographic mass data, was used to train the neural network training model as shown in Figure 1. The iterative training process was employed during the training processes.

For the results, the training model, which used the pattern dataset, stopped at 6×10^5 epochs with the final SSE dropping to 64.3082 throughout the last iterative training process. This is equivalent to the final MMSE dropping to 0.0775. This result implied that our training model would have reliably and highly accurately diagnosed and distinguished malignant breast cancer and benign disease for breast biopsy outcome predictions.

Figure 4 shows the 830 pairs of input and target training vectors when the final MMSE dropped to 0.0775 at 6×10^5 training epochs. Since the final MMSE was approximately close to 0, this also indicated that the neural network training model trained well for all of the pairs of input and target training vectors using the pattern dataset.

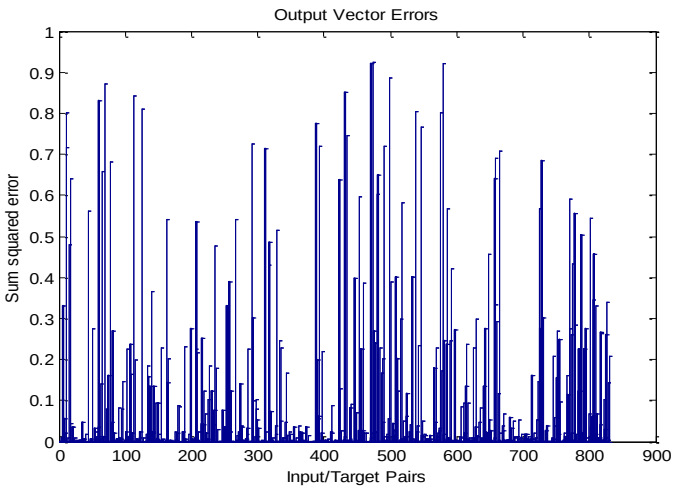

Fig. 4: A graph plot of the 830 pairs of input and target training vectors of the pattern dataset at the completion of 6×10^5 epochs with MMSE = 0.0775 during the last iterative training process of the neural network training model.

B. The Model Testing Results

In this section, we present the testing results of our neural network diagnostic model as shown in Figure 3 to estimate its probability of misclassification error for malignant breast cancer and benign disease outcomes using the resubstitution method.

When the training model completed its training in Figure 1, the final trained weights from the training model were loaded into the neural network diagnostic model. Using the same pattern dataset including all of the available 830 clinical instances of the mammographic mass data, the diagnostic model was tested to estimate its probability of misclassification error in distinguishing malignant breast cancer and benign disease for breast biopsy outcome predictions. During the testing process, the threshold T of the two-lever hard-limit classifier in Figure 3 was set to 0.5. If the output of the diagnostic model was 1, it indicated malignant breast cancer. If the output of the diagnostic model was 0, it indicated benign disease.

For the test results, the testing accuracy of the neural network diagnostic model in diagnosing and classifying malignant breast cancer and benign disease was 89.64%, details shown in Table 3. Accordingly, using the Equations (15), (16) and (17), the sensitivity, specificity, and precision results were 89.33%, 89.93%, and 89.33%, respectively.

C. The Area Under the ROC Curve Evaluation Results

The area under the ROC curve of the neural network classification model was generated by varying a set of trade-off (threshold) points between the sensitivity on the *y*-axis and the (1 – specificity) on the *x*-axis for cases being classified as malignant breast cancer and benign disease, ranging from 0 to 1 as shown in Figure 5. The area under the ROC curve estimated by using Equation (19) was 0.9626.

TABLE 3
TEST RESULT ACCURACY OF THE NEURAL NETWORK DIAGNOSTIC MODEL IN DIAGNOSING MALIGNANT BREAST CANCER AND BENIGN DISEASE

	Actual Malignant	Actual Benign	Total
Predicted Malignant	360	43	403
Predicted Benign	43	384	427
Total	403	427	830
Probability of Misclassification Error	10.67%	10.07%	10.36%

The corresponding standard error of the area under the ROC curve obtained by using Equation (20) was 0.0069. The associated asymptotic 95% confidence interval of the area under the ROC curve was from 0.9491 to 0.9761. The area of the ROC curve results implied that our neural network classification model can provide a consistently high accuracy for the diagnosis and classification of malignant breast cancer and benign disease for breast biopsy outcome predictions.

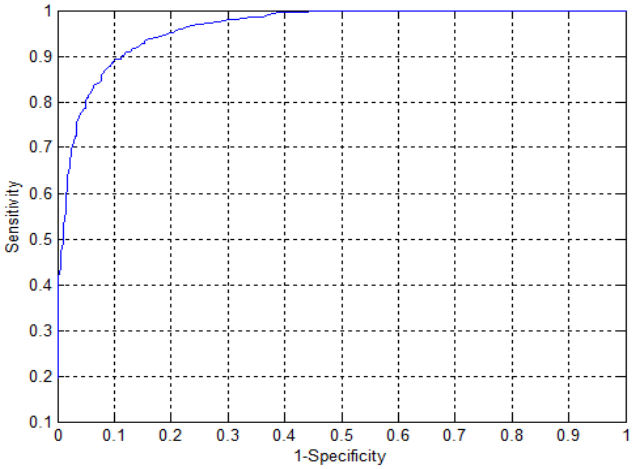

Fig. 5: A ROC curve of the neural network classification model for cases being classified as malignant breast cancer and benign disease for breast biopsy outcome predictions.

IV. DISCUSSION

Our neural network classification model, which was trained via the two-stage back-propagation approach, was used to diagnose and classify malignant breast cancer and benign disease for breast biopsy outcome predictions. The classification model was trained and tested using the pattern dataset based on the resubstitution method. The test accuracy of our classification model in distinguishing malignant breast cancer and benign disease was 89.64%. Accordingly, the sensitivity was 89.33%, specificity was 89.93%, and precision was 89.33%. The estimated area under the ROC curve was 0.9626 and its standard error was 0.0069. This is to say that if mammography was used to obtain the 5 attributes (referred to as BIRADS, age, shape, margin, and density) from a new patient, the diagnostic results via our classification model would be 89.64% accurate in diagnosing and classifying malignant breast cancer and benign disease. In addition, with the high area of the ROC curve results (0.9626±0.0069), our classification model can provide a consistently high accuracy in diagnosing malignant breast cancer and benign disease for breast biopsy outcome predictions.

In the context of related papers, there are several methods for diagnosing and classifying malignant breast cancer and benign disease for breast biopsy outcome predictions based on the mammographic mass data. These methods include a decision tree approach (DTA) and case-based reasoning classifier (CBRC) using an intelligible decision process [8], a CBRC using different similarity measures based on Euclidean and Hamming distances [19, 21-23], an artificial neural network (ANN) approach based on BIRADS descriptions [24, 25], a classification based on a DTA [26], and a prediction using a distributed genetic programming approach (DGPA) [13]. The performances of the previously reported methods and our current neural network classification method (referred to as NNCM) in terms of their estimated areas under the ROC curves and their corresponding standard errors are summarized in Table 4.

TABLE 4
THE PERFORMANCE COMPARISON OF AREAS AND STANDARD ERRORS OF ROC CURVES FOR PREVIOUSLY REPORTED METHODS AND OUR NEURAL NETWORK CLASSIFICATION METHOD (NNCM)

Methods	Area and Standard Error Under ROC Curve
NNCM	0.9626±0.0069
ANN	0.847±0.017 ~ 0.880±0.01
CBRC	0.857±0.016 ~ 0.890±0.01
DTA	0.838±0.017 ~ 0.870±0.01
DGPA	0.859±0.032 ~ 0.860±0.03

In comparison, the area under the ROC curve along with its standard error of our neural network classification model is 0.9626±0.0069, which is comparably much higher than those of the previously published papers as summarized in Table 4. In addition, since a relatively large pattern dataset was used including all of the available 830 clinical instances from the mammographic mass data, our classification model can be considered to have a reliably consistent outcome, which is an unbiased estimate of the probability of misclassification error based on the nonparametric approach of the resubstitution method. Moreover, our neural network classification model used the two-stage back-propagation approach, which is a parallel process including both linear and nonlinear components of calculations, and did not require a statistical distribution assumption for the mammographic mass data. Thus, our classification model has more flexibility and a greater percentage of accuracy in distinguishing malignant breast cancer and benign disease for breast biopsy outcome predictions, even if there are overlapping clusters between the malignant and benign cases. As can be seen, compared with

previously reported approaches, the proposed neural network curve classification model has much higher performance results for the area of the ROC curve and can provide a reliably and consistently high accuracy in diagnosing and classifying malignant breast cancer and benign disease for breast biopsy outcome predictions. Hence, using our classification model with mammography has the greater potential to reduce the number of unnecessary breast biopsies in clinical practice. Therefore, combined with mammography, the proposed neural network classification model can be implemented in hospitals and health care clinics. Furthermore, the model can aid medical professionals to diagnose malignant breast cancer and benign disease for breast biopsy outcome predictions with higher accuracy while bypassing the need for unnecessary, expensive, and invasive surgical biopsies for patients.

V. CONCLUSION AND FUTURE WORK

In this research report, we introduced the neural network classification model and ROC evaluation method to diagnose and classify malignant breast cancer and benign disease for breast biopsy outcome predictions based on the 830 clinical instances from the mammographic mass dataset. Our classification model, based on the two-stage back-propagation neural network classification approach, included both linear and nonlinear components for calculations as well as an input vector processor, adjustable learning rate controller, and training control center that allowed implementation of the iterative training processes. During each of the iterative training processes, the training model gradually increased the input data size to reuse the final trained weights from the previous iterative training stage as the initial weights for the next iterative training stage. Accordingly, the learning rate controller adjusted the learning rate. The proposed iterative training processes ensured that our model had a low SSE or MMSE. In order to obtain a highly accurate neural network diagnostic model, the proposed iterative training processes were especially useful for training our model when a large input of data and a large number of hidden neurons were present. Our research results showed that the neural network classification model had a specificity of 89.93% in diagnosing benign disease, a sensitivity of 89.33% in diagnosing malignant breast cancer, and an overall accuracy of 89.64% in diagnosing both malignant breast cancer and benign disease. An estimated area of the ROC curve for breast biopsy outcome predictions was 0.9626±0.0069. Therefore, our model along with mammography can provide highly accurate and consistent diagnoses for breast biopsy outcome predictions, allowing patients to bypass unnecessary surgical biopsies.

In future research, we can further enhance the accuracy of our neural network classification model to diagnose malignant breast cancer and benign disease by exploring additional combinations of individual attributes in different ways and increasing the number of neurons or neuron layers in our model. This enhancement would give the model more degrees of freedom, resulting in a more optimal solution. Furthermore, given the ROC curve of our model, it is also possible to determine an optimal threshold T for the neural network diagnostic model in Equation (14) by locating the point {sensitivity, (1 – specificity)} that is the shortest distance between the ROC curve and the leftmost corner point (0, 1) in Figure 5. This further improves the performances of the neural network classification model in terms of its probability of misclassification error, sensitivity, and specificity in its diagnostic accuracy of malignant breast cancer and benign disease for breast biopsy outcome predictions.

ACKNOWLEDGMENT

The authors would like to acknowledge that the mammographic mass dataset of clinical breast cancer cases was obtained from the Mammographic Mass Database available in the UCI Machine Learning Repository. This dataset contained mammographic information of breast cancer clinical instances, contributed by Dr. Rüdiger Schulz-Wendtland from the Institute of Radiology, Gynaecological Radiology, University Erlangen-Nuremberg in Germany.

REFERENCES

[1] World Health Organization, World Cancer Day, Available http://www.cdc.gov/features/worldcancerday/.

[2] Department of Health and Human Services Centers for Disease Control and Prevention, United States Cancer Statistics, Technical Notes 2007, Centers for Disease Control and Prevention, 2007, Available http://www.cdc.gov/cancer/npcr/uscs/2007/technical_notes/.

[3] American Cancer Society, *Breast Cancer Facts and Figures 2011-2012*, Atlanta, Georgia, American Cancer Society, pp. 1-32, 2011.

[4] American Cancer Society, *Cancer Facts & Figures 2012*, Atlanta, Georgia, American Cancer Society, pp. 1-63, 2012.

[5] National Cancer Institute, Cancer Trend Progress Report – 2011/2012 Update, U.S. National Institutes of Health, Available http://progressreport.cancer.gov/introduction.asp.

[6] S. W. Fletcher, W. Black, R. Harris, B. K. Rimer and S. Shapiro, "Report of the International Workshop on Screening for Breast Cancer," *Journal of the National Cancer Institute*, vol. 85, pp. 1644-1656, 1993.

[7] H. Al-Khawari, R. Athyal, A. Kovacs, M. Al-Saleh and J. P. Madda, "Accuracy of the Fischer Scoring System and the Breast Imaging Reporting and Data System in Identification of Malignant Breast Lesions," *Annals of Saudi Medicine*, vol. 29, no. 4, pp. 280-287, July-August 2009.

[8] M. Elter, R. Schulz-Wendtland, and T. Wittenberg, "The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process," *Medical Physics*, vol. 34, no. 11, pp. 4164-4172, 2007.

[9] R. W. M. Giard and J. Hermans, "The Value of Aspiration Cytologic Examination of the Breast, A Statistical Review of the Medical Literature," *Cancer*, vol. 69, pp. 2104-2110, 1992.

[10] O. L. Mangasarian, W. N. Street and W. H. Wolberg, "Breast Cancer Diagnosis and Prognosis via Linear Programming," Mathematical Programming Technical Report 94-10, University of Wisconsin, pp. 1-10, December 1994.

[11] M. S. Hung, M. Shanker and M. Y. Hu, "Estimating Breast Cancer Risks Using Neural Netwroks," *Journal of Operational Research Society*, vol. 52, pp. 1-10, 2001.

[12] National Cancer Institute, *Mammograms*, National Institute of Health, Available http://www.cancer.gov/cancertopics/factsheet/detection/mammograms.

[13] S. A. Ludwig, "Prediction of Breast Cancer Biopsy Outcomes Using a Distributed Genetic Programming Approach," *Proceedings of the 1st ACM International Health Informatics Symposium*, ACM, pp. 694-699, New York, 2010.

[14] L. L. Humphrey, M. Helfand, B. K. Chan, and S. H. Woolf, "Breast cancer screening: A summary of the evidence for the U.S. Preventive

Services Task Force," *Annals Internal Medicine*, vol. 137, pp. 347-360, 2002.
[15] L. Tabar, M. F. Yen, B. Vitak, H. H. Tony Chan, R. A. Smith, and S. W. Duffy, "Mammography service screening and mortality in breast cancer patients: 20-years follow-up before and after introduction of screening," *Lancet*, vol. 361, pp. 1405-1410, 2003.
[16] S. W. Duffy, L. Tabar, and H. H. Chen, "The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties," *Cancer*, vol. 95, pp. 458-469, 2002.
[17] L. Tabar, B. Vitak, H. H. Chen, M. F. Yen, S. W. Duffy, and R. A. Smith, "Beyond randomized controlled trials: Oraganized Mammographic screening substantially reduces breast carcinoma mortality," *Cancer*, vol. 91, pp. 1724-1731, 2001.
[18] D. B. Kopans, "The positive predictive value of mammography," *American Journal of Roentgenology*, vol. 158, pp. 521–526, 1992.
[19] C. E. Floyd, J. Y. Lo, and G. D. Tourassi, "Case-Based Reasoning Computer Algorithm that Uses Mammographic Findings for Breast Biopsy Decisions," *American Journal of Roentgenology*, vol. 175, pp. 1347-1352, November 2000.
[20] American College of Radiology, BI-RADS Atlas, Available http://www.acr.org/Quality-Safety/Resources/BIRADS
[21] A. O. Bilska-Wolak and C. E. Floyd, "Investigating different similarity measures for a case-based reasoning classifier to predict breast cancer", *Proceedings of SPIE*, vol. 4322, pp. 1862-1866, 2001.
[22] A. O. Bilska-Wolak and C. E. Floyd, "Development and evaluation of a case-based reasoning classifier for prediction of breast biopsy outcome with BI-RADS lexicon", *Medical Physics*, vol. 29, pp. 2090-2100, 2002.
[23] A. O. Bilska-Wolak, C. E. Floyd, J. Y. Lo, and J. A. Baker, "Computer aid for decision to biopsy breast masses on mammography: Validation on New Cases," *Academic Radiology*, vol. 12, pp. 671-680, 2005.
[24] J. A. Baker, P. J. Kornguth, J. Y. Lo, M. E. Williford, and C. E. Floyd, "Breast cancer: Prediction with artificial neural network based on BI-RADS standardized lexicon," *Radiology*, vol. 196, pp. 817-822, 1995.
[25] M. K. Markey, J. Y. Lo, R. Vargas-Voracek, G. D. Tourassi, and C. E. Floyd "Perception error surface analysis: A case study in breast cancer diagnosis," *Computers in Biology and Medicine*, vol. 32, pp. 99-109, 2002.
[26] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81-106, 1986.
[27] UCI Machine Learning Repository, Mammographic Mass Data Set, Available http://archive.ics.uci.edu/ml/datasets/Mammographic+Mass.
[28] G. J. Miao, K. H. Miao, and J. H. Miao, "Neural Pattern Recognition Model for Breast Cancer Diagnosis," *Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Bioinformatics (JBIO)*, August Edition, pp. 1-8, September 2012.
[29] S. Haykin, *Neural Network: A comprehensive Foundation*, Macmillan College Publishing Company, 1994.
[30] H. Demuth and M. Beale, *Neural Network Toolbox: For Use with MATLAB*, The Math Works, Inc., Third Printing, 1994.
[31] R. Schalkoff, *Pattern Recognition: Statistical, Structural and Neural Approaches*, John Wiley & Sons, Inc., 1992.
[32] M. T. Hagan, H. B. Demuth, and M. Beale, *Neural Network Design*, PWS Publishing Company, 1996.
[33] R. Kumar and A. Indrayan, "Receiver Operating Characteristic (ROC) Curve for Medical Researchers," *Indian Pediatrics*, vol. 48, pp. 277-287, April 2011.
[34] R. L. Finney, F. D. Demana, B. K. Waits, and D. Kennedy, *Calculus: A Complete Course*, Second Edition, Addison Wesley Longman, Inc., 2000.
[35] J. A. Hanley and B. J. McNeil, "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve," *Radiology*, vol. 143, no. 1, pp. 29-36, April 1982.
[36] G. J. Miao and M. A. Clements, *Digital Signal Processing and Statistical Classification*, Artech House, Inc., 2002.

Kathleen H. Miao is an undergraduate student in the College of Arts and Sciences at Cornell University, Ithaca, New York.

She is a National AP Scholar and a USA Biology Olympiad Semifinalist. She is the co-author of the peer-reviewed journal paper "*Neural Pattern Recognition Model for Breast Cancer Diagnosis*." She has received a number of science and volunteer service awards including the Intel Science Talent Search Research Report Badge Award, Synopsys Silicon Valley Science & Technology Championship Honorable Mention, IEEE Award for Best Electro-Technology, Mu Alpha Theta Winner Award, Society of Women Engineers Santa Clara Valley Section Winner Award, and Stanford University Medical Center Auxiliary Volunteer Honors.

She received the Elks National Foundation Most Valuable Student Achievement Scholarships from the California-Hawaii Elks Association Level Lodge and Sunnyvale Lodge. She is the co-founder of The Wishing Star Team for the American Cancer Society. She is interested in the biological sciences, medicine, and cancer research.

George J. Miao received a B.Eng. joint degree from Shanghai University of Science and Technology (now Shanghai University) and Shanghai Second Medical University (now Shanghai Jiao Tong University School of Medicine), China; a M.S. in Statistics from Columbia University, New York; and a Ph.D. in Electrical Engineering from the Georgia Institute of Technology, Atlanta, Georgia.

He has been a Director and Chief Scientist in Flezi, LLC since 2008. He worked and consulted for a number of U.S. Fortune 500 companies, universities, and research institutes. He is the co-author of the textbook *Digital Signal Processing and Statistical Classification* (Artech House, 2002) and the author of the textbook *Signal Processing in Digital Communications* (Artech House, 2007). He holds 16 granted U.S. patents in the areas of digital signal processing and wireless communications.

Dr. Miao is a Senior Member of the IEEE and was a Chairman of the IEEE New Jersey Coast Chapter of Signal Processing/Circuits and Systems (2003-2006). He has received a number of achievement awards, including the IEEE Region-1 Technology Award, the IEEE Chapter Distinguished Service Award, and the IEEE Signal Processing Society Certificate of Appreciation. His research interests are in digital, multirate, adaptive, and statistical signal processing, digital communication, and software-defined radio as well as data mining, machine learning, dynamic neural network, and time-series analysis, modeling, and predictions of big data analytics.