Neural Pattern Recognition Model for Breast Cancer Diagnosis

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

***Abstract*—In this article, we introduce a neural pattern recognition model for breast cancer diagnosis. This proposed model uses a two-stage back-propagation approach including both linear and nonlinear components of calculations along with iterative training processes and a learning shift controller. The iterative training processes allow the model to gradually increase the number of hidden neurons and input data size, reusing each of the iteratively final weights as new initial weights for the next iterative training stage. A learning rate is accordingly adjusted by the learning shift controller. This training approach ensures that even the local minima of the model have low enough sum-squared errors. The average testing diagnosis accuracy of our model is 98% for benign and malignant breast cancer. Therefore, our research results indicate that the proposed model can provide consistently high accuracy in the diagnosis and classification of benign and malignant breast cancer.**

***Index Terms*—breast cancer, diagnosis, classification, iterative training, neural network, pattern recognition**

I. Introduction

According to the United States Cancer Statistics in the Department of Health and Human Services Centers for Disease Control and Prevention, one out of four deaths is due to cancer diseases. Cancer is now the second-leading cause of death in the United States [1]. A total of 1,638,910 new cancer cases is projected to occur in 2012 [2-3]. American Cancer Society estimates that 39,920 breast cancer deaths (including 39,510 woman and 410 men) are expected to occur and 229,060 new cases of invasive breast cancer (including 226,870 woman and 2,190 men) will be diagnosed in the United States in 2012 [2-3]. Presently, breast cancer ranks as the second leading cause of cancer death in women after lung cancer and ranks as the first leading cause of new cancer cases in women based on 2012 estimates [3].

Female breast cancer mortality rate (number of deaths per 100,000 persons per year) slowly increased from 1975 to 1989, peaking at its maximum mortality rate in 1989, and then has steadily decreased annually up to the year 2012 [4]. The decrease in the female breast cancer mortality rate is generally attributed to a greater awareness of the disease and represents progress in earlier detections, enhanced diagnosis methods, and improved medical treatments.

When the breast cancer tumor is relatively small and most treatable, it typically does not show symptoms [2-3]. Changes to the breast, such as swelling, thickening, or skin irritation, are less common symptoms. Breast pain usually appears at the benign stage and may be a later symptom of cancer at the malignant stage. Thus, before symptoms develop, it is important for women to be able to diagnose breast cancer at an early stage.

A breast cancer victim's chances for long-term survival are improved by early detection of the disease, and early detection is in turn enhanced by an accurate diagnosis. Recently, various mathematics models have been developed to represent and simulate cancer tumor cell growth and invasion as well as quantify evaluation of treatments in the field of modeling cancer biology [5]. Chances of lengthened survival are also enhanced by correct prognosis, that is, the expected long-term behavior of the disease, which largely influences the choices of appropriate treatments immediately following surgery.

In order to diagnose breast cancer, there are currently four main methods used to distinguish benign lumps from malignant ones: surgical biopsy, mammography, magnetic resonance imaging (MRI), and fine needle aspiration (FNA) with visual interpretation. The reported accuracy of the cancer diagnosis for surgical biopsy is close to 100%, mammography ranges from 68% to 79% [6], MRI is 70% for benign diagnoses and 92% for malignant diagnoses [7], and FNA with visual interpretation varies from 65% to 98% [8-9]. Surgical biopsy is the most accurate of the four methods, but it is invasive, expensive, and time consuming. In comparison, mammography lacks accuracy. When a tumor is detected through mammography or MRI, surgical biopsy is necessary for determining the state of its malignancy [10]. Compared to surgical biopsy, MRI, and mammography, FNA is the least invasive and expensive, but the accuracy of FNA varies widely. Therefore, a relatively objective system that diagnoses FNAs with a consistently high accuracy is greatly desired. This allows FNA with an accurate diagnosis without the need for a surgical biopsy.

Utilizing characteristics of individual cells obtained from a minimally invasive FNA, several systems have been proposed and developed to distinguish between benign and malignant breast cancer lumps. An interactive computer system with machine learning techniques to diagnose breast cancer using FNA was first developed by Wolberg and Mangasarian, et al.

Manuscript received July 24, 2012; revised September 7, 2012.

G. J. Miao is with Flezi, LLC, San Jose, CA 95134 USA (e-mail: g.j.miao@ieee.org).
K. H. Miao is with Monta Vista High School, Cupertino, CA 95014 USA.
J. H. Miao is with Monta Vista High School, Cupertino, CA 95014 USA.

at the University of Wisconsin [11-14]. It is a computerized image analysis system, which enables a set of nuclear features obtained from FNA, along with a linear programming based on a classification scheme using multistage piecewise linear parallel hyperplanes [15-16]. Other research papers used an instance-based learning algorithm [17], a data-dependent upper bound estimation algorithm [18], and a hybrid fuzzy-genetic programming system [19] to classify breast cancer.

In this article, we consider an alternative approach using a neural pattern recognition model for breast cancer diagnosis and classification. This model that we propose herein is based on a two-stage back-propagation neural network approach with a training method that gradually increases the number of hidden neurons and input data size. After each of the iterative processes, the model reuses the final weights of the previous iteration as new initial weights for the next iterative training stage. Accordingly, the learning rate at each of the iterations is adjusted during the training process. The two-stage back-propagation neural network approach is a parallel process, including both linear and nonlinear components of calculations. The model is able to detect and classify benign and malignant breast cancer without the restrictions of statistical assumptions. Our research results indicate that the neural pattern recognition model for breast cancer diagnosis provides the possibility of obtaining a consistently high accuracy for the diagnosis and classification in distinguishing between benign and malignant breast cancer.

II. Breast Cancer Dataset

The breast cancer dataset that we used in this research is obtained from the Breast Cancer Wisconsin (Original) Database available in the UCI Machine Learning Repository [20]. This dataset contains information of breast cancer clinical cases, created by Dr. William H. Wolberg from University of Wisconsin Hospitals. New samples arrived periodically as Dr. Wolberg reported his clinical cases. The database therefore reflects this chronological grouping of the data and was constantly increasing in size, thereby resulting in 8 groups with different numbers of clinical instances at different reported dates as shown in Table 1.

This dataset contains a total of 699 clinical instances, with 458 benign and 241 malignant cases. Each clinical instance has 9 attributes with assigned integer values ranging from 1 to 10 and one class output with a binary value of either 2 or 4, indicating benign and malignant breast cancer diagnoses, respectively. The physical meaning of the 9 attributes are as follows: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, and mitoses.

Among the 699 clinical instances, 16 instances are each missing one of the nine attributes. Common practice is to eliminate all individuals from analysis for whom information on a variable is missing [21]. For a consistently high accuracy, the 16 instances each missing one attribute are removed from this dataset. The resulting dataset has 683 clinical instances, with 444 (65.01%) benign and 239 (34.99%) malignant diagnoses. During this research, each of the 9 attributes with original integer value in the range of 1 to 10 is normalized by the value of 10, thereby having the normalized attribute value in the range of 0.1 to 1. The class output with an original binary value 2 or 4 is changed to a binary value 0 or 1, representing benign and malignant diagnoses, respectively.

TABLE 1
Grouping of Breast Cancer Clinical Instances

Name of Group	Clinical Instances
Group 1	367
Group 2	70
Group 3	31
Group 4	17
Group 5	48
Group 6	49
Group 7	31
Group 8	86

III. Diagnosis Model and Method

This section focuses on the study of breast cancer diagnosis model and method to distinguish between benign and malignant diagnoses by utilizing 9 attributes of individual cells of each clinical instance obtained from a minimally invasive FNA. The available breast cancer dataset used in this research has 683 clinical instances from a total of 8 groups, reported on different dates. Classes for benign and malignant diagnoses are overlapping clusters, indicating that clusters are not linearly separate.

In this study, we propose an alternative method of the neural pattern recognition model for benign and malignant breast cancer diagnosis and classification. This model is based on a two-stage back-propagation approach, including both linear and nonlinear components of calculations, with an iterative training process and a learning shift controller.

A. Neural Pattern Recognition System

A neural pattern recognition system can be considered a massively parallel distributed processor that has a natural propensity for storing experiential knowledge. Knowledge is acquired by the system through a learning process. Interneuron connection strengths known as synaptic weights are used to store the knowledge. The training procedure of a learning algorithm is to modify the synaptic weights of the system to achieve a desired design objective.

The neural pattern recognition system for diagnosing breast cancer, to distinguish benign from malignant breast lumps during the training process, is shown in Fig. 1. This system contains a key computational block of a two-stage black-propagation neural pattern recognition, along with an input size controller, an iterative training control center, a target output class, a learning shift controller, a weighting update calculator, an addition operation, and a scaling parameter ε.

An input vector of $(R \times N)$ matrix, which includes N clinical instances and R attributes of breast cancer (where $R = 9$), was simultaneously fed into the two-stage back-propagation neural pattern recognition, which propagated all of the input patterns to determine all unit outputs. Comparing all unit outputs with

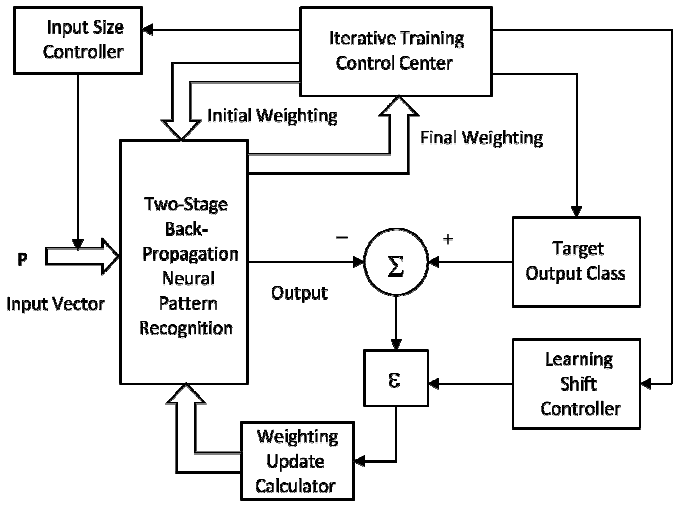

Fig. 1. A neural pattern recognition training system.

the desired pattern response from the target output class, we obtained an error that was further multiplied by the scaling parameter ε. The scaling parameter ε was adjusted by the learning shift controller. Then, weights were updated after the minimization of error at each stage through the unit weight adjustment. This process was repeated until the sum of squared errors was as small as possible and was less than a prior defined error value, or the number of set training epochs was used up.

The iterative training control center was used to control and communicate with the input size controller, the two-stage back-propagation neural pattern recognition, the target output class, and the learning shift controller. For each of the iterative training processes, this allowed the determination of the input vector size N, number of hidden neurons, initial weights, adjustment learning rates, and corresponding number of the desired pattern response. The input vector size N and the number of hidden neurons were gradually increased as the number of iterative training processes was increased. Accordingly, the final weights of each of the completed training processes were reused as initial weights for the next training process. At the same time, the learning rate was decreased, according to an instruction from the learning shift controller. The iterative training processes were repeated until the sum of squared error was obtained, which was less than a prior defined error.

The iterative training processes that we propose herein are especially useful when input vector sizes and neuron sizes are relatively larger. When a network model is trained with different initial weights, the network model solution will be different. Thus, by reusing a relatively good set of final weights obtained from the previous training as initial weights for the next training during the iterative training processes, our iterative training method enables the neural pattern recognition system to reach an approximately optimal solution while keeping the system stable during training. Fig. 2 shows a detailed iterative training process flowchart for obtaining approximately optimal weights for the neural pattern

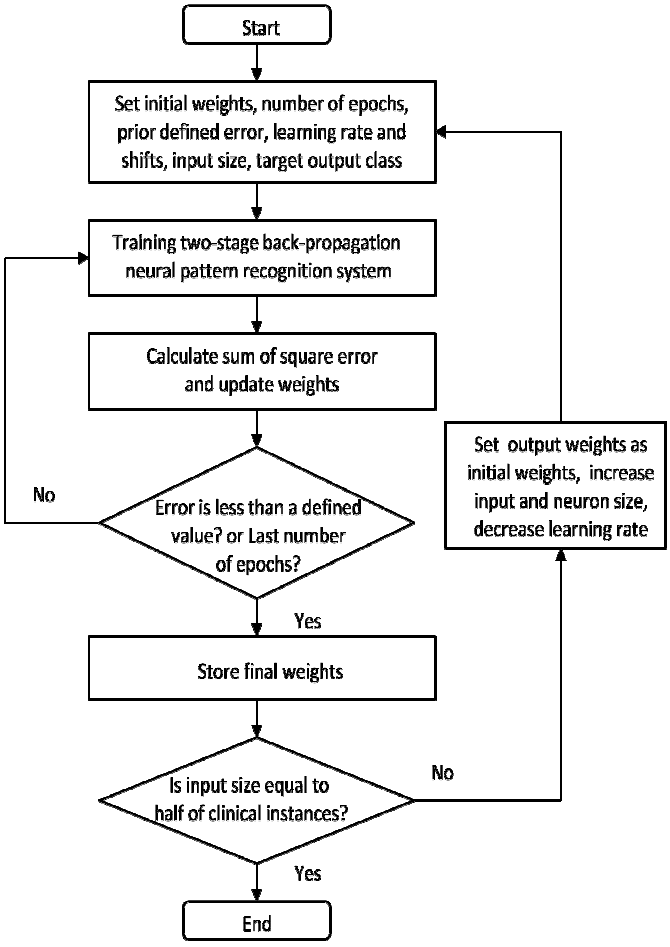

Fig. 2. An iterative training process flowchart for the neural pattern recognition system.

recognition system to diagnose benign and malignant lumps.

B. Back-propagation Neural Pattern Recognition Model

A back-propagation approach, which is a common method for training multiple-layer neural networks, was used to minimize the objective function for the neural pattern recognition system. The back-propagation approach was established by using the Widrow-Hoff learning rule in multiple-layer networks and nonlinear differentiable transfer functions [22-23]. Input sequence and the corresponding output sequence were used to train a back-propagation neural pattern recognition model until the model could appropriately approximate a function within a prior defined error value. A back-propagation learning rule was used to adjust weights and biases by the error derivative (delta) vectors back-propagated through the neural pattern recognition system. Since the desired pattern classes are known, such processing is referred to as supervised learning for pattern recognition or classification [24]. Thus, in this research, classifying the breast cancer diagnosis as benign or malignant is a supervised learning of pattern recognition.

Fig. 3 shows a two-stage back-propagation neural pattern recognition model with S_1 tan-sigmoid transfer functions F_1 in the first neuron layer and one linear transfer function F_2 in the

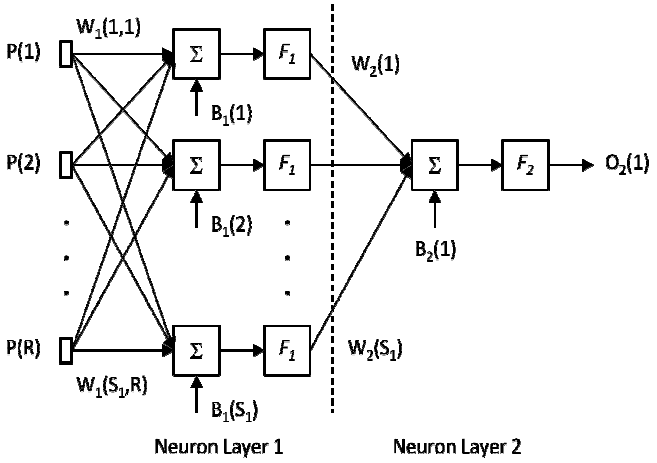

Fig. 3. A two-stage back-propagation neural pattern recognition model.

second neuron layer, where R and S_1 denote the number of input attributes and hidden neurons in the first neuron layer, respectively, and W_1, B_1 and W_2, B_2 indicate hidden neurons of the weights and biases in the first and second neuron layer, respectively. The tan-sigmoid transfer function is a nonlinear transfer function. It generates output values placed between -1 to 1 as the neuron's net input goes from negative to positive infinity. The second neuron layer has only one linear transfer function that can take on any value for breast cancer pattern recognition. Both the linear and nonlinear transfer functions are differentiable and monotonic increasing functions. The output of each transfer function increases as its input increases. Thus, the linear and nonlinear transfer functions have no minima that would not tend to cause error minima, thereby trapping the neural pattern recognition model as it has learned.

The back-propagation learning algorithm that we used for training the neural pattern recognition model is based on the generalized delta rule [25] as described in the following steps:

1. The output error based on the pth training sample is denoted by Ep and defined as

$$Ep = \frac{1}{2}\sum_{j=1}^{N}\left(T_j^p - O_j^p\right)^2 \tag{1}$$

where j is neuron j ($j = 1, 2, ..., N$), T_j is desired pattern output (or target output class), and O_j is corresponding output response of the neural pattern recognition model.

2. To move in a direction opposite the gradient, the iterative weight correction procedure using the pth training sample is denoted by $\Delta^p w_{ji}$ and given by

$$\Delta^p w_{ji} = \varepsilon \delta_j^p \tilde{O}_i^p \tag{2}$$

where i is ith input ($i = 1, 2, ..., R$), and ε is a positive constant that is referred to as the learning rate. $\tilde{O}_i^p$ is O_i^p if the input is the output of a neuron in the second neuron layer and $\tilde{O}_i^p$ is I_i if the input is a direct input to the first neuron layer. δ_j^p is the sensitivity of the pattern error on the net activation of the jth unit.

3. In the case of output units, the sensitivity of the pattern error on the net activation of the jth unit is given by

$$\delta_j^p = \left(T_j^p - O_j^p\right)\frac{\partial o_j^p}{\partial net_j^p} \tag{3}$$

Where

$$\frac{\partial o_j^p}{\partial net_j^p} = F_j'\left(net_j^p\right) \tag{4}$$

and

$$net_j^p = \sum_{i=1}^{M} w_{ji}\, I_i + bias_j \tag{5}$$

The $bias_j$ term in Equation (5) is added to net_j^p directly, and F_j in Equation (4) is a non-decreasing and differentiable transfer function for the jth neuron. In our study, the output function in the output layer is a linear transfer function, that is,

$$F_j\left(net_j^p\right) = I_j net_j + bias_j \tag{6}$$

4. In the case of internal units, the sensitivity of the pattern error on the net activation of the jth unit is given by

$$\delta_j^p = F_j'\left(net_j^p\right)\sum_n \delta_n^p w_{nj} \tag{7}$$

where δ_n^p is from the lower-numbered layer. In our study, the internal transfer function uses the tan-sigmoid characteristic as follows:

$$F_j\left(net_j\right) = \frac{1}{1+e^{-net_j}} \tag{8}$$

The corresponding derivative computation of the above equation (8) is given by

$$F_j'\left(net_j\right) = \frac{e^{-net_j}}{(1+e^{-net_j})^2} \tag{9}$$

The back-propagation learning algorithm involves two phases including propagation and weight update. During the propagation, the forward propagation of a training input passes through the neural pattern recognition model to generate the propagation's output activations. Through the neural pattern recognition model, the backward propagation of the output activations uses the training target classes to produce the deltas of all output and hidden neurons. During the weighting update, the neural pattern recognition model performs by multiplying its output delta and input activation to get the gradient of the weights. Then, the neural pattern recognition model brings the weights in the opposite direction of the gradient by subtracting a ratio of it from the weights. The neural pattern recognition model repeats the two phases until the performance of the neural pattern recognition model is satisfactory.

C. Implementation of Neural Diagnosis Model

An implementation of neural diagnosis model is shown in Fig. 4. This model contains a normalized input value, a two-stage neural pattern recognition model, a two-lever hard-limit classifier, and final trained weights. The normalized input value was used to normalize each input of attributes so that the input value ranged from 0.1 to 1. The two-stage neural pattern recognition model has the same structure as shown in Fig. 3, with the weights loaded from the final trained weights, where R = 9 attributes, S_1 = 50 hidden neurons in the first neuron layer and one neuron of the linear transfer function in the second layer. The two-lever hard-limit classifier F_3 produced one of the binary decision values 0 or 1, where the "0" represents benign and the "1" represents malignant.

The relationship between the input vector $\boldsymbol{P}$ and the decision output D, as shown in Figure 4, has the following mathematical representation:

$$D = F_3\{F_2[\boldsymbol{W}_2 \times F_1(\boldsymbol{W}_1 \times \boldsymbol{P} + \boldsymbol{B}_1) + \boldsymbol{B}_2]\} \qquad (10)$$

where D is a binary decision output value, which indicates either a benign or malignant diagnosis; $\boldsymbol{P}$ is the input vector of (9×1) matrix that represents one clinical instance; $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are the model weights of (50×9) matrix and (1×50) matrix, respectively, which are referred to as connection weights in the first and second neuron layer; $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are the model biases of (50×1) matrix and (50×1) matrix in the first and second neuron layer, respectively; F_1 and F_2 represent the tan-sigmoid nonlinear transfer functions and the linear transfer function in the first neuron layer and in the second neuron layer, respectively; F_3 is the two-lever hard-limit classifier. It has the mathematic expression as follows:

$$D(x) = \begin{cases} 0, & x < 0.5 \\ 1, & x \geq 0.5 \end{cases} \qquad (11)$$

Thus, given a new clinical instance, this implementation of neural diagnosis model can diagnose and distinguish between benign and malignant breast cancer.

IV. Results

In this study, 683 clinical instances in the Breast Cancer Wisconsin (Original) Database were used for training and testing the neural pattern recognition model after removing 16 clinical instances that were each missing one attribute. Among 683 clinical instances, there are 444 benign and 239 malignant breast cancer cases. Each of the 9 attributes with original integer value in the range of 1 to 10 was normalized by 10, thereby resulting in the data range from 0.1 to 1. The class output of an original binary value 2 or 4 was converted to a corresponding binary value 0 or 1, which indicates benign and malignant breast cancer, respectively.

Estimating a probability error on the neural pattern recognition model is to use two nonparametric approaches, including a *resubstitution* method and a *holdout* method [24]. Based on the resubstitution method, the neural pattern recognition model was trained on a pattern data and was tested

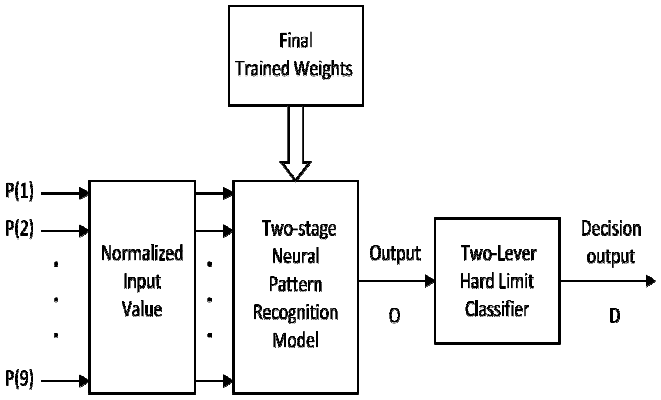

Fig. 4. An implementation of neural diagnosis model for distinguishing benign and malignant breast cancer.

on the same pattern date. Using the holdout method, the pattern data were first partitioned into two mutually exclusive datasets. We then trained the neural pattern recognition model on the first dataset (training dataset) and then tested it on the second dataset (testing dataset). The holdout method results in an unbiased estimate of the expected probability of error when the data set size is large.

A. Model Training

Note that 683 clinical instances in the dataset are distributed in 8 groups, which were reported with different dates, as shown in Table 1. We then separated the dataset into two mutually exclusive datasets from the odd and even indices, referred to as the first dataset (training dataset) and second dataset (testing dataset), respectively. This separated method resulted in two mutually exclusive datasets with relatively even distributions of benign and malignant breast cancer cases crossing over the 8 groups. The first dataset that was used for training the model has 342 clinical instances, including 221 benign and 121 malignant breast cancer cases. The second dataset that was used for testing the model has 341 clinical instances, which contains 223 benign and 118 malignant breast cancer cases.

Fig. 5 shows a graph plot result of the neural pattern recognition model error throughout the last iterative training process using the training dataset of 342 clinical instances, including 221 benign and 121 malignant breast cancer cases. The graph plot of the model error stopped at 16×10^5 epochs when the final sum of square error dropped to 0.0035 in which the minimum mean square error (MMSE) is about 0.00001. Thus, it does not matter whether the neural pattern recognition model falls into a global or local minimum. We should know that this model would have a realistically accurate solution for diagnosing benign and malignant breast cancer. In addition, this graph plot can also help show how quickly the neural pattern recognition model learned during the last iterative training process.

Fig. 6 shows a corresponding graph plot of each pair of input and target training vectors when the final sum of square error dropped to 0.0035 at the 16×10^5 training epochs. As

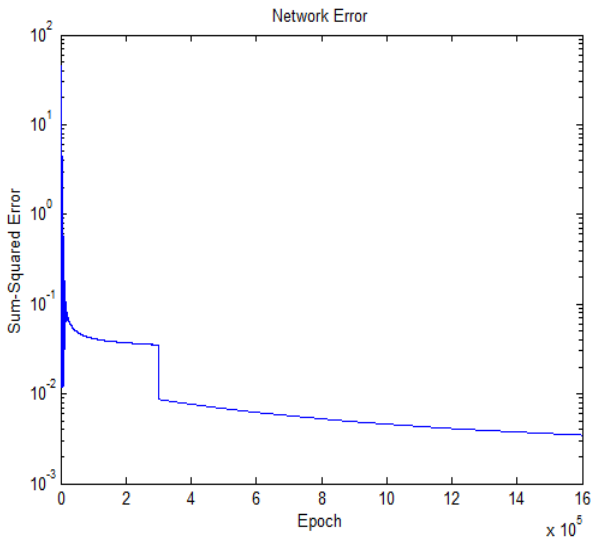

Fig. 5. A graph plot result of the neural pattern recognition model error at each of the epochs during the last iterative training process.

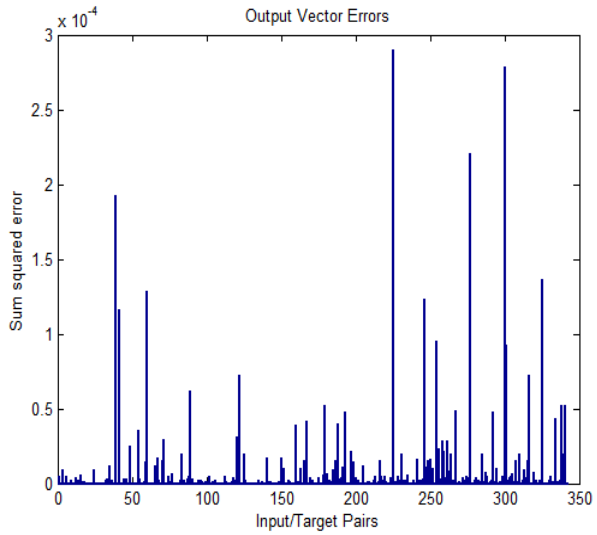

Fig. 6. A graph plot result of the final sum-squared error associated with each pair of input and target training vectors at the 16×10^5 epochs during the last iterative training process.

can be seen, the sum-squared error of each pair of input and target training vectors is also close to zero. This indicates that the neural pattern recognition model trained well.

B. Model Testing

In this section, we present the model testing results based on the nonparametric approaches. Once training was completed, the implementation of neural diagnosis model of loading the final weights, as shown in Figure 4, was used for the model testing and diagnosis of benign and malignant breast cancer. In order to estimate the probability of misclassification error of the model on how accurately it classified benign and malignant breast cancer, we tested the model by using the training dataset based on the resubstitution method and by using the testing dataset based on the holdout method.

First, the resubstitution method for estimating the probability of misclassification error on benign and malignant breast cancer was applied for testing the model. The training dataset, which contains 342 clinical instances, including 221 benign breast cancer and 121 malignant, was used to train the neural pattern recognition model. Also, the same training dataset was used to test the model. The test result of the probability of misclassification error on the model is shown in Table 2. The test accuracy of the model to diagnose and classify benign and malignant breast cancer is 100% by using the training dataset.

Second, the holdout method for estimating the probability of misclassification error on breast cancer benign and malignant was used for testing the model. We used two mutually exclusive datasets, the training dataset and the testing dataset. The training dataset has 342 clinical instances, including 221 benign and 121 malignant breast cancer cases, and was first used to train the neural pattern recognition model. The testing dataset, which contains 341 clinical instances, including 223 benign and 118 malignant breast cancer cases, was then used to estimate the probability of misclassification error for benign and malignant breast cancer.

TABLE 2
ESTIMATION OF PROBABILITY OF MISCLASSIFICATION ERROR FOR BENIGN AND MALIGNANT BREAST CANCER USING THE TRAINING DATASET

	Benign	Malignant	Total Instances
Number of Clinical Instances	221	121	342
Misclassification of Clinical Instances	0	0	0
Probability of Misclassification Error	0%	0%	**0%**

TABLE 3
ESTIMATION OF PROBABILITY OF MISCLASSIFICATION ERROR FOR BENIGN AND MALIGNANT BREAST CANCER USING THE TESTING DATASET

	Benign	Malignant	Total Instances
Number of Clinical Instances	223	118	341
Misclassification of Clinical Instances	8	6	14
Probability of Misclassification Error	3.58%	5.08%	**4.10%**

The test result of the probability of misclassification error on the model is shown in Table 3. The test accuracy of the model to diagnose and classify benign and malignant breast cancer is about 96% by using the testing dataset.

Third, in estimating the average diagnosis accuracy, we tested the model using combined training and test datasets together. As a test result, out of the 444 benign cases, there was a total of 8 misclassifications, in which the average accuracy of benign diagnoses is 98.20%. Out of the 239 malignant cases, there was a total of 6 misclassifications. Thus, the average accuracy of malignant diagnoses is 97.49%. In conclusion, the total test average accuracy of our model is 97.95% using all 683 clinical instances.

V. Discussion

In general, training a neural network obtains different network resolutions when different initial weights are used. The neural network may fall into a local minimum so that a good solution cannot be found when applying a poor set of initial weights. Adding more hidden neurons provides the neural network a better chance to solve the problem of local minimum because it gives the neural network more degrees of freedom. However, creating a good set of initial weights for training the neural network is a relatively difficult problem. It is especially difficult to train the neural network associated with large hidden neurons when a large amount of input data is present. In addition, selecting a learning rate for the neural network, especially in a nonlinear case, is a challenge. The training of the neural network becomes unstable when a learning rate is too large. This often results in extremely large weights.

In this article, our neural pattern recognition model used the two-stage back-propagation approach for diagnosing and classifying benign and malignant breast cancer. To train the model, we proposed using the iterative training processes along with the learning shift controller. During the training processes, the input vector size and the number of hidden neurons were gradually increased step-by-step as the number of iterative training processes was increased. Accordingly, at the completion of each of the iterative training processes, the final weights were reused as initial weights for the next iterative training process. At the same time, the learning rate was decreased according to a schedule instruction from the learning shift controller. The iterative training processes were repeated until we obtained a desired sum-squared of error. Our training results, as shown in Fig. 5 and 6, indicate that the iterative training processes seem to ensure that even the local minima of the neural pattern recognition model had low enough sum-squared errors. Thus, whether or not the model fell into a global or local minimum, the problem of local minimum was solved. The low-enough sum-squared errors made it more likely that our neural model found a highly accurate solution.

Our test results for estimating the accuracy in the probability to diagnose and classify benign and malignant breast cancer, using the training dataset and the test dataset based on the resubstitution method and the holdout method, are 100% and 96%, respectively. The average total test accuracy of our model for diagnosis is 98% using the combined training and test datasets together. It seems that the estimating accuracy (100%) of the resubstitution method is better than the estimating accuracy (96%) using the holdout method. However, since the testing dataset size is relatively large, the holdout method results in an unbiased estimate of expected probability of misclassification error [24]. That is, the estimating accuracy (96%) using the holdout method is more reliable for the neural pattern recognition model in diagnosing and classifying benign and malignant breast cancer. However, the average test accuracy (98%), using the combined training and test datasets together, reflects the model's actual performance.

Compared with previous papers [12, 17, 19], our test results for the probability of diagnosis accuracy are comparatively better than those of previously published test results. Furthermore, the test results of our model can be considered more reliable since we used all available clinical instances of the dataset over 8 groups in the different reported dates while most of the previous papers used only the clinical instances of the dataset from the first group.

VI. Conclusion

In this article, we introduced an alternative approach using the neural pattern recognition model for diagnosing and classifying benign and malignant breast cancer. This model was based on the two-stage back-propagation approach including both linear and nonlinear components of calculations along with the iterative training processes and the learning shift controller. The iterative training processes that we proposed in this article allowed the model to gradually increase the number of hidden neurons and the input data size, reusing each of the iteratively final weights, obtained from the previous iterative training stage as new initial weights, for the next iterative training stage during each of the training processes. In addition, a learning rate was adjusted according to an instruction from the learning shift controller. The proposed iterative training processes ensured that even the local minima of the neural pattern recognition model had low enough sum-squared errors. This method is especially useful for obtaining a highly accurate diagnosis model to train a neural network associated with large hidden neurons when a large amount of input data is present. Our research results also indicate that the neural pattern recognition model can provide a consistently high accuracy in the diagnosis and classification of benign and malignant breast cancer. Therefore, this allows FNA with a highly accurate diagnosis percentage rate without the need for a surgical biopsy.

Although a surgical biopsy results in almost 100% accuracy in diagnosing benign and malignant breast cancer, it is invasive, expensive, and inconvenient for the breast cancer patient. FNA with visual interpretation varies from 65% to 98% in accuracy [7-9]. However, utilizing FNA with the neural pattern recognition model described in this paper ensures a consistently high accuracy of 98% in diagnosing benign and malignant breast cancer while eliminating the need for an invasive and expensive surgical biopsy.

For future research, it is possible to further enhance the model accuracy of benign and malignant breast cancer diagnosis by increasing the number of neuron layers or neurons in the model. This provides the model with more degrees of freedom, thereby leading to an approximately optimal solution. On the other hand, in dealing with multidimensional attribute observations, visualization and understanding can be aided by representing the observations in a two-dimensional space via a linear transformation such as a principal components transform, canonical discriminant transform, and optimal declustering transform [24]. A transform-based plot is helpful in exploring relationships between the cluster groups of benign and malignant breast

cancer, and in identifying a typical attribute observation. Moreover, an optimal declustering transform-based neural pattern recognition model may be capable of further improving the diagnosis accuracy of benign and malignant breast cancer.

ACKNOWLEDGMENT

The authors would like to acknowledge that the breast cancer dataset is obtained from the Breast Cancer Wisconsin (Original) Database available in the UCI Machine Learning Repository. This dataset contains clinical information of breast cancer cases contributed by Dr. William H. Wolberg from the University of Wisconsin Hospitals.

REFERENCES

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

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

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

[4] K. H. Miao, J. H. Miao, and G. J. Miao, "Cancer Mortality Rate Trend Identification and Prediction Using Iterative Backpropagation Neural Network Approach," Submitted for publication, under reviewing process, April 2012.

[5] K. Cobb, Modeling Cancer Biology, *Biomedical Computation Review*, pp. 16-24, Spring 2007.

[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] 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.

[9] 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.

[10] 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.

[11] O. L. Mangasarian and W. H. Wolberg, "Cancer diagnosis via linear programming," *SIAM News*, vol. 23, no. 5, pp. 1-18, September 1990.

[12] W. H. Wolberg and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," in *Proc. of the National Academy of Sciences*, U.S.A., vol. 87, pp. 9193-9196, December 1990.

[13] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, "Machine learning techniques to diagnose breast cancer from fine-needle aspirates," *Cancer Letters*, vol. 77, pp. 163-171, 1994.

[14] W. H. Wolberg, W. N. Street, D. M. Heisey, and O. L. Mangasarian, "Computerized breast cancer diagnosis and prognosis from fine needle aspirates," *Archives of Surgery*, vol. 130, pp. 511-516, 1995.

[15] O. L. Mangasarian, R. Setiono, and W. H. Wolberg, "Pattern recognition via linear programming: Theory and application to medical diagnosis," in *Large-scale numerical optimization*, T. F. Coleman and Y. Li, Ed, SIAM Publications, pp 22-30, Philadelphia 1990.

[16] K. P. Bennett and O. L. Mangasarian, "Robust linear programming discrimination of two linearly inseparable sets," *Optimization Methods and Software 1*, Gordon & Breach Science Publishers, pp. 23-34, 1992.

[17] J. Zhang, "Selecting typical instances in instance-based learning," in *Proc. of the 9th Int. Machine Learning Conf.*, Aberdeen, Scotland, Morgan Kaufmann, pp. 470-479, 1992.

[18] A. Antos, B. Kegl, T. Linder, and G. Lugosi, "Data-dependent margin-based generalization bounds for classification," *Journal of Machine Learning Research*, vol. 3, 2002.

[19] H. Mallinson and P. Bentley, "Evolving Fuzzy Rules for Pattern Classification," in *Proc. of the Int. Conf. on Computational Intelligence for Modelling, Control and Automation,* 1999.

[20] UCI Machine Learning Repository, Breast Cancer Wisconsin (Original) Data Set, http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29.

[21] N. E. Breslow, N. E. Day, and W. Davis, "Statistical Methods in Cancer Research," *The Analysis of Case-Control Studies*, World Health Organization International Agency for Research on Cancer, vol. 1, Third Reimpression 1989.

[22] S. Haykin, *Neural Network: A comprehensive Foundation*, Macmillan College Publishing Company, 1994.

[23] H. Demuth and M. Beale, *Neural Network Toolbox: For Use with MATLAB*, The Math Works, Inc., Third Printing, 1994.

[24] G. J. Miao and M. A. Clements, *Digital Signal Processing and Statistical Classification*, Artech House Publisher, 2002.

[25] R. Schalkoff, *Pattern Recognition: Statistical, Structural and Neural Approaches*, John Wiley & Sons, Inc., 1992.

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. from Columbia University, New York, and a Ph.D. in Electrical Engineering from the Georgia Institute of Technology, Atlanta, Georgia, USA.

He has been a 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 Publisher, 2002) and the author of the textbook *Signal Processing in Digital Communications* (Artech House Publisher, 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 Award "for innovative work in digital signal processing leading to applications in wireless communications," and the IEEE Chapter Distinguish Service Award. His research interests are in digital, multirate, adaptive and statistical signal processing, digital communication and software radio, pattern recognition, data mining, and time-series analysis and forecasting.

Kathleen H. Miao is a Monta Vista High School student in Cupertino, California, USA. She is the Co-President of Monta Vista Science National Honor Society and Co-President of Monta Vista Future Practicing Physicians Network Clubs. She has been a Junior Auxiliary Volunteer at Stanford University Hospital and Medical Center since June 2010. She has received the AP Scholar with Honor Award. She is interested in the biological sciences and medicine, especially in the area of cancer research.

Julia H. Miao is a Monta Vista High School student in Cupertino, California. She is a member of Monta Vista Biotechnology Club, Monta Vista Future Practicing Physicians Network Club, and Monta Vista Science National Honor Society. She is a volunteer at Santa Clara Kaiser Permanente Hospital. She is interested in pursuing biotechnology, cancer research, and medicine.