A Study of Genetic Neural Network as Classifiers and its Application in Breast Cancer Diagnosis

[a] Yaoying Huang，[b,]* Wanggen Li and [c] Xiaojiao Ye
Department of Computer Science，Anhui Normal University，Wuhu， China，241003

***Abstract*—In this paper, a genetic neural network classification model is developed. The proposed model can not only optimize the weights and thresholds of neural network, but also reduce network size by identifying the feature subset effectively using genetic algorithm. This technique is aimed at finding out a network that is large enough to capture the accurate class attributes of the data as much as possible, while retaining the generalization capability of neural network. Breast cancer datasets (BCD) in UCI Machine Learning Repository are utilized to evaluate the proposed genetic neural network approach in this paper. Simulation results show that the developed model achieved dimensional reduction and the identification of benign and malignant tumors. Accordingly, it improved classification accuracy and demonstrated excellent classification efficiency.**

***Index Terms*— Breast Cancer Diagnosis, Neural Networks, Classifier, Genetic Algorithm, Feature Selection**

I. INTRODUCTION

Breast cancer is one of the most common causes of death among women worldwide. Identifying the breast cancer tumor quickly and accurately, either benign or malignant, is critical for taking the right medical treatment timely. Owing to the complex nonlinear relationship among the main morphological features of breast cancer cells, it is difficult to describe by the traditional linear regression methods. Among a great variety of classification techniques suggested so far for medical diagnosis neural network (NN) has been one of the most popular methods that consistently demonstrated its strengths and potentials in solving practical classification problems[1]-[4], [7]-[12].

As is commonly known, a great deal of information needs to be collected for the purpose of breast cancer diagnosis (BCD), which inevitably leads to a high dimensional representation of the sample in the mode space. Therefore, we have every reason to doubt whether there exists a correlation among the various features used for quantifying each sample. That is, whether there exist redundant or useless information in these features. In this case, it is essential to carry out the task of feature selection, a technique commonly used in machine learning for selecting a subset of relevant features for building a robust learning model. Feature selection can not only discover the optimum feature subset which is rich in relevant information but also reduce the space and time complexity in computations [5]. Above all, it can improve the accuracy and efficiency of classification.

Artificial neural networks, when equipped with the ability of self-organizing and self-learning with high stability can process the nonlinear problem in the identification of the benign and malignant cells [7]. However, NNs often converge too slowly and become trapped at a local minimum. In this research, we propose a machine learning technique where neural network (NN) as classifier is combined with genetic algorithms (GA) to improve the performance of NNs in solving practical problems such as breast cancer diagnosis. In general, the structure of neural network is sensitive to the initial weights, bias and hidden nodes etc. [6], so optimizing the network is of particular importance to improve the classification accuracy. Compared with other search methods, the following aspects shed lights on why genetic algorithm is adopted: (1) GA isn't prone to get stuck in local optimum in the search process, and hence they are capable of finding the global optimal solution with great probability. (2) The inherent parallelism of genetic algorithms makes it very suitable for large-scale parallel distributed processing.

Using neural network for breast cancer diagnosis has recently received a good deal of attention [1], [7]-[12]. It has been reported that in [8] that both the evolutionary and the ensemble approach can deal with the diagnosis problem well in terms of good generalization of neural networks. Yi et al. [11] proposed a model of breast cancer diagnosis with wavelet neural network (WNN) based on genetic algorithm. In addition, two diagnosis models based on BPNN and RBFNN in [12] is constructed and compared with each other. It has been shown that NNs are feasible for solving classified problem occurred in BCD. In spite of this, there remain some issues which need to be investigated further. One of these issues is that the classification precision is not known in general. Some researchers have also proposed a variety of non-linear feature extraction methods [13], like wavelet transform, discrete Fourier transform, discrete Fermat transform and such. However, these transformations can obtain sub-optimal results to some extent.

This paper investigated GA-NN approach to feature selection (GANNFS). Our primary interest is not only to achieve the performance of feature subset, but also to

* Corresponding author. Tel: 0553-5910719. Fax: 0553-5910719.
E-mail addresses: [a] hyy091103@163.com; [b] xchen@mail.ahnu.edu.cn; [c] smzxyxj@163.com.

doi:10.4304/jcp.6.7.1319-1324

optimize the weights and biases of neural networks. For illustration, breast cancer datasets in UCI Machine Learning are taken in this study. Firstly, breast features are extracted by GA-NN strategy to avoid over-fitting. Then, this algorithm proceeds with the resulting features. Meanwhile, the better weights and bias can be obtained to greatly reduce subjective choice of structural parameters. Last but not least, the accuracy rate of classification in BCD has been improved considerably. In other words, our GANNFS technique solved the problem of diagnostic classification of breast cancer effectively.

The rest of the paper is organized as follows: We start with short descriptions of neural networks as classifier and genetic algorithms including a study of feature selection in Section II and our theoretical framework is put forward simultaneously. Section III introduces our proposed model in details. Experiment results are presented in Section IV, together with some comparisons with other training techniques. In the end, Section V concludes with a brief summary of the study.

II. GANN APPROACH TO FEATURE SELECTION

A. Neural Network as Breast Classifier

Neural networks (NN) have been widely used in various fields as an intelligent tool in recent years，such as artificial intelligence, pattern recognition, medical diagnosis, machine learning and so on [15]. Among them, pattern recognition is a class of problem that neural network is particularly suitable for solving. BP neural network can be applied to all aspects of pattern recognition: feature extraction, data compression, cluster analysis, classification and discrimination and so forth.

In fact, NN can be viewed as the mapping from input to output. If each different input is regarded as a kind of input mode, the mapping to the output is considered as output response model, the mapping from input to output is undoubtedly the issue of pattern classification. Nevertheless, learning is the first step to design classifier, that is, ascertain the requirements for the classification error rate and choose appropriate discrimination rule.

Strictly speaking, the learning algorithm of NN is a supervised learning method by training feed forward neural network using error back propagation technique to determine the parameters of neural network. Its unique advantages lie in the greatest tolerance of the noisy data, as well as the ability to classify untrained data pattern.

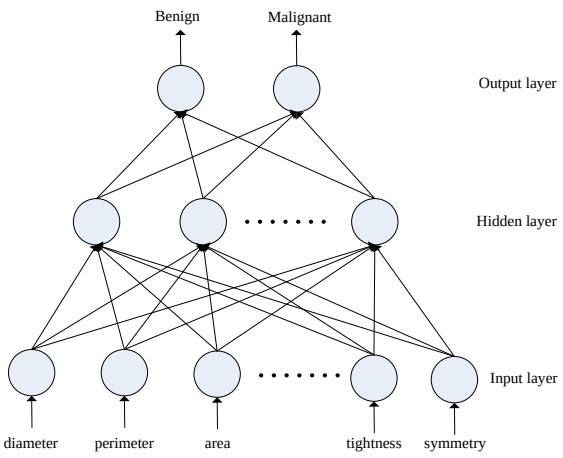

Figure 1. BPNN topology

When it comes to the breast cancer data classification, the major steps of using neural network learning algorithm can be summarized as follows: to begin with, through the provision of training samples and the class of sample, the network prediction of each sample is compared with the actual known class label, and then the weight of each training sample is adjusted to achieve the purpose of classifying other sample data.

The use of neural network to classify breast cancer data is illustrated in Fig.1. Each node in the network corresponds to the output node of a network unit, while the real lines from the bottom into node are regarded as its input. Intermediate cell is called the hidden layer units, whose output are only in the internal network, not a part of all the network output. The output of the hidden layer is considered as the input of two output units, corresponding to a result of the diagnosis of breast cancer, benign or malignant tumor.

B. Processing Procedure of Genetic Algorithm

Genetic algorithm (GA) is a global optimization algorithm drawn from the evolutionary ideas and inspiration [14], [15]. In essence, it is a direct search method which is independent of the concrete problems. GA has gained extensive application in image processing, biological science, neural networks, pattern recognition, machine learning and the like. The major steps of GA are as follows:

- GA resolves the solution of the problem into a chromosome, namely, a binary encoding string in the algorithm.
- Before the implementation of genetic algorithm, a group of chromosome is given that is assumed solution normally and these hypothetic solutions are put into the problems.
- Especially, an excellent chromosome is chosen to copy, crossover, mutation so as to produce a new generation group according to the principle of survival of the fittest.
- As a result, they evolve from generation to generation, and finally converge to the most acceptable chromosome, that is, the optimal solution.

In summary, genetic algorithm proceeds from an initial population, through a series of genetic operation, such as, selection, crossover and mutation, to search a better space step by step until reach the optimal solution. It is obvious that genetic algorithm is an optimization methodology. Here, genetic algorithm with global optimization strategies is integrated into the neural network model to improve the classification rate of breast cancer diagnosis.

C. GANN to Feature Selection

As a rule, it is generally acknowledged that breast cancer is more likely to be correctly diagnosed if there is more learning sample. Therefore, when people collect the

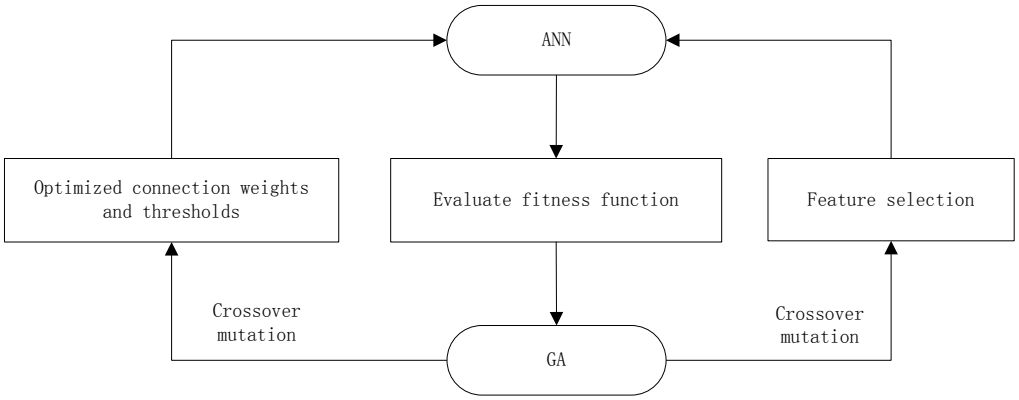

Figure 2. Proposed Model

data of breast cancer diagnosis (BCD), measurement data is always gathered as much as possible. On the other hand, although the sample data with high dimension contains abundant information about objective things in machine learning, this will give rise to a series of problems, such as rapid growth in the time of processing data together with storage space of samples, and some redundant information, even sometimes it is impossible to be directly used for classification.

That refers to the so-called curse of dimensionality problem. More importantly, feature dimension affects both classification accuracy and the computational complexity of designing classifier. Accordingly, we hope to figure out the most essential information and some of the characteristics which best describe sample data by compressing the dimension of mode. In general, such techniques as mapping or transformation can change the high-dimensional feature in the high-dimensional space U_h into the low-dimensional feature in the low-dimensional space U_l to some extent. We denote the prototypes as $U_h \Rightarrow U_l \quad h \geq l$. Note that the feature dimensions are compressed on the premise that the classification property of the sample remains unchanged.

Considering feature subset selection problem of the n class. In the original feature set U_h, the instance space $S=\{s_1,s_2,\ldots,s_q\}$, $s_i=(s_{i1},s_{i2},\ldots,s_{ip})$, where the number of attributes is p, the total number of instances is q, and $S = S_1 \cup S_2 \cup \cdots \cup S_n$. q_m represents the number of various class inclusive instances, where $m = 1,2,...,n; \sum_{m=1}^{n} q_m = q$. A subset of n class instances is consistent with the same feature set U_h. In this study, we not merely take full advantage of genetic algorithms to implement optimal feature subset selection, but optimize weights and thresholds of neural network as well. The entire structure of the proposed model GANNFS is described in Fig.2.

III. Implementation Details

In order to design splendid neural network architecture and carry through feature selection using genetic algorithm, the first step is to solve the network coding problem; then get the optimal structure after selection, crossover and mutation operators.

A. Encoding

Genetic algorithm needs to adopt a certain encoding scheme which is able to map the solution space to code space. Coding theory is an important factor when the efficiency of GA is taken into consideration. To improve the computational efficiency of time and space, we use a binary encoding scheme. Suppose that a binary bit of each chromosome p_i corresponds to a characteristic set of quantitative feature. For every chromosome p_i,

$$p_i = (p_{i1}, p_{i2}, \cdots, p_{in}), \quad p_{ij} = \begin{Bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{Nn} \end{Bmatrix} \tag{1}$$

where N, n represent the population size and the length of chromosome, respectively. If a certain genes p_{ij} is 1, it means that the corresponding feature participates in the modeling, otherwise, the bit is 0, which indicates that the corresponding feature does not involve in modeling. Take the problem of ten variables as an example, as Table I shows, "0100011010"denotes that the second feature as well as the sixth, seventh, ninth one are selected to participate in modeling, others don't involve in modeling.

TABLE I.
AN EXAMPLE OF FEATURE SUBSET SELECTION USING GA

The bit string of chromosome	Feature subset
100110010001	s1,s4,s5,s8,s12
01101000100000001100	s2,s3,s5,s9,s17,s18

B. Initial Population

In general, the binary strings of length n constitute initial solution of genetic algorithm, commonly known as initial population. Each bit is just the gene of individual chromosomes in each string. If multiple-initialization is introduced in genetic algorithm, it will have a stronger search ability of the global best solution. There is no doubt that large groups contain more models to provide sufficient sample capacity for genetic algorithm, improving the searching quality of GA and preventing premature convergence. However, it will increase the computation of large groups of individual fitness evaluation, so that the convergence rate decreases. Let the initial population P be defined as

$$\mathrm{P} = \{p_1, p_2, \cdots, p_N\} \tag{2}$$

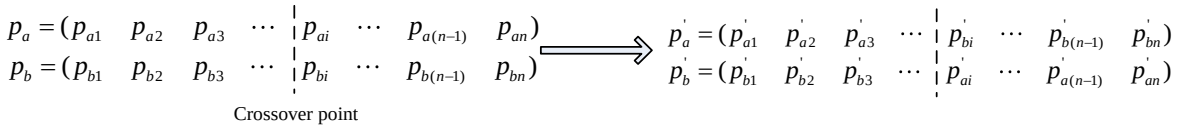

Figure 3. An example of the crossover operation

where p_i is defined in (1). Here the population size N is set to 40.

C. Fitness

The fitness of one chromosome is regarded as the evaluation of viability, which determines the probability of being selected of the chromosomes in the selection operation. The higher fitness the individuals obtain, the more generation it has. In this study, the error function E of neural network classifier can be derived by the following equation:

$$\mathrm{E}=\sum_{k=1}^{n}\left[t_k(\delta)-o_k(\delta)\right]^2 \tag{3}$$

where t, o denote the predicted output value, the actual output value, respectively. Then the inverse of the sum squared error is selected as the fitness function f in the test dataset.

$$f(p_i)=\frac{1}{E} \tag{4}$$

The weights and thresholds built by BP neural network are all optimized using genetic algorithm in respect of calculating every individual fitness function value.

D. Selection

The main purpose of selection operation is to avoid the loss of useful genetic information and to improve the global convergence and computational efficiency. The choice is based upon the size of the fitness of each chromosome. Just as mentioned above, the higher fitness function value the chromosome possesses, the larger the likelihood to be selected is. In brief, it spreads more widely in the next generation, with a large number of offspring.

In this paper, we employed roulette wheel selection method. For one thing, the fitness value of each chromosome is calculated. For another, the proportion of the fitness in total fitness value is computed as well, which implies the individual probability of being selected in the selection process. In fact, for a given population size and individual fitness value, generally, the selected probability P_S is computed by

$$P_S(\mathrm{p_i})=\frac{f(p_i)}{\sum_{k=1}^{N}f(p_k)}=\frac{f(p_i)}{N\overline{f}},\quad i=1,2,\cdots,N \tag{5}$$

with

$$\overline{f}=\frac{\sum_{j=1}^{N}f_j}{N} \tag{6}$$

where $\overline{f}$ denotes the average fitness value of the group.

E. Crossover

Generally speaking, the selection operator is to pass the original fine genes on to the next generation, while the crossover operator can generate new individual containing more excellent genes. It plays an important role in searching optimal solutions [16]. To establish a close correlation between the design of crossover mode and the structure of coding bit string, we designed two kinds of crossover operator in this article, respectively.

As for feature extraction of the input feature set, one-point crossover is adopted. It is safe to draw the conclusion that the number of one-point crossover is up to 29, concerning two individuals of 30-bit string length in this research. The operation procedure is described in Fig. 3: Firstly, two chromosomes are randomly selected in the new copy of population, and each chromosome is composed of multi-bits (genes); and then one position is stochastically taken among the genes of two chromosomes, what's more, both of them swap the end part of the genes from this point.

With regard to optimizing the weights and thresholds of neural network using the genetic algorithm, the arithmetic crossover operator [17] is utilized. The offspring chromosomes $p_a^{'}$ and $p_b^{'}$ are obtained according to

$$p_a^{'}=\theta p_b+(1-\theta)p_a,\ p_b^{'}=\theta p_a+(1-\theta)p_b \tag{7}$$

where θ is called random variable, which lies in the range $0\le\theta\le1$,and p_a and p_b are parents' strings, respectively.

F. Mutation

Mutation operation can ensure the diversity of the population to some extent, so that the search covers space as large as possible to avoid falling into local solutions due to losing useful information in the search. We made good use of simple mutation when feature extraction of training sets is taken into account. It will change a gene at a certain probability in the range of 1 and 0 at random to produce a new individual.

Since uniform mutation tends to result in weak local search performance, we take full advantage of non-uniform mutation operator to optimize the weights and biases. This technique creates a random disturbance to the original gene variation. Meanwhile, the result of perturbation after mutation is considered as new gene value.

Crossover probability P_c is set to 0.2 in this research. By and large, the higher crossover rate is, the faster the loss speed of fine genetic structure is. Conversely, if crossover rate is too low, it may lead to stop searching. Similarly, a type of premature genetic information could not be recovered when mutation probability is too small, whereas the genetic search will turn into a random search if the mutation rate is too high. Therefore, mutation rate P_m is set to 0.05 to avoid such a dilemma.

G. Termination criterion

Determine GANNFS whether continues to run or not. If it reaches the stopping condition, that is the maximum number of iterations *T*, end the algorithm and output the optimal feature subset, together with weights and thresholds of neural networks.

IV. EXPERIMENTAL DESIGN

A. Data and variables

Our experimental dataset is obtained from the UCI machine learning database, which was generated by Dr. William at hospital at the University of Wisconsin Madison [18]. This Wisconsin breast cancer dataset contains 2 classes and 569 instances, of which 357 cases were benign, 212 cases were malignant. The number of attributes is 30, including ten quantifying feature, such as diameter, perimeter, symmetry etc. and their mean value, standard deviation, the worst value. It is obvious that there is a close relationship between these characteristics and identification of benign and malignant cells.

We divided the dataset into two parts, of which one is the training set and the other is the testing set. It should be noted that the training set is used for training the parameter of BPNN using a BP algorithm. On the other hand, the testing set is introduced for the purpose of testing the generalization ability of neural network. The first 400 samples are adopted for training, and the rest 169 samples for testing.

Fig. 1 shows our designed NN model and Fig. 2 shows our theoretical framework including neural network classifier in combination with genetic algorithm to feature selection. At first, BP neural network is trained through 20 epochs, a learning rate of 0.1. What's more, we design the structure of BPNN with 30 nodes in the input layer, 2 nodes in the output layer. Our experimental studies showed that 31 nodes in the hidden layer gave the best results because a higher number caused over-training. The parameter settings on GANNFS algorithm are as follows:

- Population size (*M*):40
- Generation number (*T*):80
- Crossover rate (P_c):0.2
- Mutation rate (P_m):0.05

B. Experimental results and performance analysis

As far as the classification performance of the model is concerned, the classification rate (*C*) denotes the percentage of correctly classified samples, which is computed by the following formula.

$$C = \frac{n_c}{n_t} * 100\%, \; n_c \le n_t \qquad (8)$$

where n_c, n_t represent the number of correctly classified samples and the total number of the samples, respectively.

GANNFS was started with 40 randomly generated chromosomes, and then the process of calculating the fitness value, selection, crossover and mutation was iterated through 80 generations. The average and the best fitness function of our model with nine features are given in Fig. 4, which vary with genetic algebras.

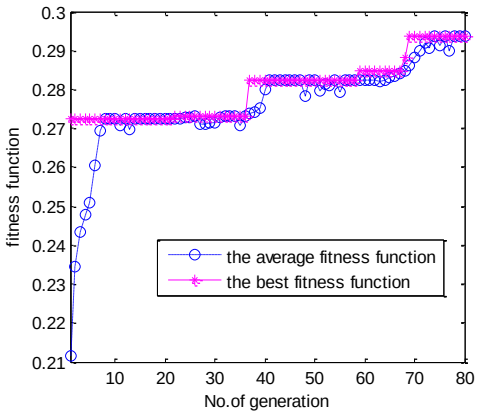

Figure 4. Fitness function of GANNFS model

To verify the effectiveness of our model, Back Propagation (BP) Neural Network classifier, Radial Basis Function (RBF) Neural Network classifier, Learning Vector Quantization (LVQ) Neural Network classifier, Genetic Algorithm BP neural network (GABP) and our designed classifier are adopted in the same experimental conditions to study classification performance of the samples for the testing sets.

The comparison results of five classifier concerning classification precision are depicted in Table Ⅱ. It can be obviously seen that the recognition rate in terms of the right classification percentage has distinctly increased, which is measured by our GANNFS model. For one thing, our GABPFS model has the highest classification accuracy (99.1%), containing nine features. For another, the modeling time using GABPFS technique is also greatly reduced, which further demonstrates that our algorithm is feasible.

TABLE II.
COMPARISON BETWEEN OUR MODEL AND OTHERS

	BP	RBF	LVQ	GABP	GABPFS
C /%	89.3	90.5	97.8	96.3	99.1
modeling time /s	25.6	24.3	16.5	8.8	3.2

TABLE III.
THE CLASSIFICATION RESULTS OF DIVERSE FEATURE SUBSETS

Feature numbers	Feature subsets	C/% (benign)	C/% (malignant)
18	s1,s2,s4,s6,s7,s12,s13,s14,s15,s18, s19,s21,s23,s25,s27,s28,s29,s30	99.02	91.37
14	s5,s8,s9,s11,s12,s13,s14,s17, s21,s24,s25,s27,s28,s29	97.12	91.88
12	s1,s2,s9,s16,s17,s18,s21,s23, s24,s25,s27,s28	98.26	93.75
9	s2,s3,s6,s8,s13,s17,s22,s26, s29	100	97.10

Numerous experiments using our proposed model GANNFS are conducted, most of which have got high recognition rate. Table III shows four corresponding experimental results. For instance, s1 denotes the first feature to be selected in modeling, to this analogizes, as has been noted in Table I. As is reflected in Table III, we can easily deduce from the results that our model with nine features achieved the highest classification rate, regardless of benign or malignant. In a word, genetic neural network classifier to feature extraction enjoys overwhelming superiority in medical diagnosis to solve the issue of breast cancer.

V. CONCLUSION

In this research, a classification framework is put forward using neural network and genetic algorithm. In this study, feature subsets are firstly extracted from the datasets by genetic algorithm, and then the structure parameters of feature subset and neural network were optimized by our presented technique, namely, GANNFS. The proposed model has been implemented to identify benign and malignant breast cell. Most importantly, the experiment results showed that the technique proposed in this paper is superior to those popularly recognized approaches. Compared with the previously reported results [12], our model is much more effective in detecting breast cell.

ACKNOWLEDGMENT

The authors would like to thank the National Natural Science Foundation of China under Grant No.61070060 and the Key Project of Natural Science Foundation from the Educational Committee of Anhui Province of China under Grant No. 2010KJ140A for supporting this work. The Research of the Ministry of Science and Technology Program of Wuhu also provided additional funding.

REFERENCES

[1] H. A. Abbass. "An evolutionary artificial neural networks approach for breast cancer diagnosis," *Artificial Intelligence in Medicine*, vol.25, no.3, Jul., 2002, pp. 265–281.

[2] M. G. Penedo, M. J. Carreira, A. Mosquera, et al. "Computer-aided diagnosis: a neural-network-based approach to lung nodule detection," *IEEE Trans. on Medical Imaging*, vol.17, no.6, 1998, pp. 872-880.

[3] D. Gila, M. Johnsson, J. M. Garcia Chamizo, A. S. Paya and D. R. Fernandez. "Application of Artificial Neural Networks in the Diagnosis of Urological Dysfunctions", *Expert Systems with Applications*, vol.36, no.3, Apr., 2009, pp. 5754-5760.

[4] S. C. Tan, C. P. Lim, K. S. Tan and J. C. Navarro. "An Evolutionary Artificial Neural Network for Medical Pattern Classification," In *Neural Information Processing (LNCS)*, vol.5864, 2009, Springer, pp. 475-482.

[5] N. Abe, M. Kudo, "Non-parametric classifier-independent feature selection," *Pattern Recognition*, vol.39, 2006, pp. 737–746.

[6] X. Zeng, D. S. Yeung. "Sensitivity analysis of multilayer perceptron to input and weight perturbations," *IEEE Trans. on Neural Networks*, vol.12, no.6, Nov., 2001, pp. 1358-1366.

[7] D. Delen, G. Walker, A. Kadam. "Predicting breast cancer survivability: a comparison of three data mining methods," *Artificial Intelligence in Medicine*, vol.34, no.2, Jun., 2005, pp.113-127.

[8] X. Yao, Y. Liu. "Neural Networks for Breast Cancer Diagnosis," In *Proc. of the 1999 Congress on Evolutionary Computation*, vol.3, Jul., 1999, IEEE Press, pp.1760-1767.

[9] M. A. Mazurowski, P. A. Habas, J. M. Zurada, et al. "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," *Neural Networks*, vol.21, no.2-3, Mar.-Apr., 2008, pp. 427-436.

[10] R. Wang, B. Wan, X. Cao. "Research on Computer Image Processing of Breast Cancer," *Journal of Tianjin University*, vol.35, no.4, Jul., 2002, pp. 497-500.

[11] X. Yi, P. Wu, J. Li, L. Liu. "Breast Cancer Diagnosis Using WNN Based on GA," *Life System Modeling and Intelligent Computing* (*LNCS*), vol.6330, 2010, pp.367-374.

[12] Q. Liu, L. He. "Diagnosis Model Based Neural-network in Galactophore Cancer Cell Identification," *Journal of Chongqing University*, vol.26, no.4, Apr., 2003, pp.70-72.

[13] R. N. Strickland, H. I. Hahn. "Wavelet transforms for detecting micro-calcifications in mammograms," *IEEE Trans. on Medical Imaging*, vol.15, no.2, 2002, pp.218-229.

[14] K. Hornik, M. Stinchcombe, H. White. "Multilayer Feedforward Networks are Universal Approximators," *Neural networks*, vol. 2, no. 5, 1989, pp. 359-366.

[15] D. E. Goldberg. "Genetic Algorithms in Search, Optimization and Machine Learning," *Addison Wesley*, Reading, MA, 1989.

[16] H. Chen, D.W. Cui, X. Li, H. L. Wei. "Optimization crossover scale for improving performance of crossover operator," *Journal of Software*, vol.20, no.4, Apr., 2009, pp.890-901. (in Chinese)

[17] C. Shen, L. Wang, Q. Li. "Optimization of injection molding process parameters using combination of artificial neural network and genetic algorithm method," *Journal of Materials Processing Technology*, vol.183, no.2-3, Mar., 2007, pp. 412-418.

[18] C. L. Blake, C. J. Merz, D. J. Newman, et al. UCI Repository of Machine Learning Databases, 2002. http://www.ics.uci.edu/`mlearn/MLRepository.html.

Yaoying Huang is currently a M.S candidate at the Department of Computer Science, Anhui Normal University, Wuhu, China. She received her BS degree in computer science from Anhui Normal University, in 2008. Her main research areas include intelligent algorithm, neural networks and machine learning.

Wanggen Li was born in Anhui Province, China, in 1973. He received the Ph.D. degree from Donghua University. He is currently an associate professor of the Department of Computer Science and Technology at Anhui Normal University. His major research fields are intelligent computing and biological computing. His research activities have been supported by the National Natural Science Foundation of China and the Key Project of Natural Science Foundation from the Educational Committee of Anhui Province of China.

Xiaojiao Ye received the bachelor's degree of computer science from Ningbo University in 2006. She is now a M.S candidate at the Department of Computer Science. Her research interests include intelligent algorithm and its application.