ORIGINAL PAPER

Usage of Case-Based Reasoning, Neural Network and Adaptive Neuro-Fuzzy Inference System Classification Techniques in Breast Cancer Dataset Classification Diagnosis

**Mei-Ling Huang · Yung-Hsiang Hung ·
Wen-Ming Lee · R. K. Li · Tzu-Hao Wang**

Abstract Breast cancer is a common to females worldwide. Today, technological advancements in cancer treatment innovations have increased the survival rates. Many theoretical and experimental studies have shown that a multiple classifier system is an effective technique for reducing prediction errors. This study compared the particle swarm optimizer (PSO) based artificial neural network (ANN), the adaptive neuro-fuzzy inference system (ANFIS), and a case-based reasoning (CBR) classifier with a logistic regression model and decision tree model. It also applied three classification techniques to the Mammographic Mass Data Set, and measured its improvements in accuracy and classification errors. The experimental results showed that, the best CBR-based classification accuracy is 83.60%, and the classification accuracies of the PSO-based ANN classifier and ANFIS are 91.10% and 92.80%, respectively.

M.-L. Huang (✉) · Y.-H. Hung · T.-H. Wang
Department of Industrial Engineering and Management,
National Chin-Yi University of Technology,
35, Lane 215, Section 1, Chungshan Road, Taiping,
Taichung 411 Taiwan, Republic of China
e-mail: huangml@ncut.edu.tw

Y.-H. Hung
e-mail: hys502@ncut.edu.tw

W.-M. Lee · R. K. Li
Department of Industrial Engineering & Management,
National Chiao Tung University,
1001 Ta Hsueh Road,
Hsinchu, Taiwan 300, Republic of China

W.-M. Lee
e-mail: wunming@mail.hit.edu.tw

R. K. Li
e-mail: rkli@cc.nctu.edu.tw

Introduction

Breast cancer is a globally prevalent cancer among women. Currently there are three methods to diagnose breast cancer: mammography, FNA (fine needle aspirate) and surgical biopsy. Mangasarian et al. pointed out a linear-programming-based machine learning approaches to the prognosis of breast cancer. The problem considered in this research is that of prognosis, the prediction of the long-term behavior of the disease [1]. Data mining and statistics analysis is the search for valuable information in large volumes of data. It is now widely used in health care industry [2].There are many data mining technologies, and the common ones include the artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS), case-base reasoning (CBR), decision tree (DT), support vector machine (SVM), and rough set theory (RST). In medical applications, CBR and machine learning have been successfully used for decision support in tasks such as patient diagnosis [3].

Depending on problem in hand, the proportionality between data and knowledge varies. The knowledge dominant computing model, CBR is a multi-disciplinary subject that focuses on the reuse of experiences [4]. The data dominant computing models: ANN and ANFIS have been deployed in time to time and problem to problem by many researchers in medical domain [5]. CBR is a knowledge system, which can be used for special knowledge expression of cases from past experience and problems, and it possesses the technologies to develop a knowledge base system, including a case-based explanation system and a problem-solving system. In recent years, ANN has been developed to

mimic the behavior of biological neural nets, and have successfully solved problems through the generalization of limited quantities of training data and overall trends in functional relationships [6, 7]. For each liner problem and for a variety of constraint integration methods, three popular heuristic search procedures are particle swarm optimization (PSO), genetic algorithms (GA), and simulated annealing (SA), which are applied, tuned, and compared [8]. In recent years, the particle swarm optimization algorithm, which is a simple, easy to implement, and highly effective evolutionary algorithm, has been used for the purpose of ANN evolution. An ANFIS is not only a theory integrated with the fuzzy theory and neural networks, but also a fuzzy system with learning abilities. It can obtain fuzzy rules, which cannot be obtained by a common fuzzy system by regulating synaptic weights of a neural network, summarizing the relevant input and output relations, and then obtaining the fuzzy rules [9] from the synaptic weights of the network. The advantage of this system is that no accurate mathematical model is needed, therefore, if proper experimental data is given, an applicable and effective prediction model can be constructed. However, due to complementary advantage and disadvantage of CBR and ANN sometimes, in medical domain, it is difficult to solve problem independently with either. But, if their advantages are exploited and disadvantages are removed then their combination offers significant benefits. Many approaches to case based reasoning exploit feature weight setting algorithms to reduce the sensitivity to distance functions.

This paper compares the performance of three different hybrid algorithms, (1) Assigning attribute weights to be used in a CBR prediction model. The generation of the attribute weights is performed by decision tree and logistic regression (LR). (2) A hybrid PSO-Backpropagation (PSO-BP) algorithm combines the population-based evolutionary searching ability of PSO to improve the NN algorithm's capability of finding out the global optimal solution. (3) The ANFIS classifiers learned how to differentiate a new case in the domain by given a training set of such records. The proposed ANFIS model combined the neural network adaptive capabilities and the fuzzy logic qualitative approach. The study and observation would help the biomedical engineers to know the applicability of a particular method in different medical domains of practice and research.

Material and methods

Mammographic mass data set

Mammography is the most effective method for breast cancer screening currently available. In the mammogram report, includes many technical details, and one section will shows a breast imaging reporting and data system (BI-

RADS) score. BI-RADS was developed by the American College of Radiologists as a standard of comparison for rating mammograms and breast ultrasound images. It indicates the radiologist's opinion of the absence or likelihood of breast cancer [10]. The data used in this study is provided by a UCI Machine Learning Repository Mammographic Mass Data Set [11], which contains a BI-RADS assessment, the patient's age, and three BI-RADS attributes, together with the ground truth (the severity field) of 516 benign and 445 malignant masses that have been identified through full field digital mammograms collected at the Institute of Radiology of the University Erlangen-Nuremberg, from 2003 to 2006 [12]. The database contains information regarding BI-RADS, Age, Shape, Margin, and Density variables. The classification output is binary: malignant (0) or benign (1) tumors (see Table 1). The initial database was pre-processed to remove cases with missing values and to standardize the variables. After removing records with missing values the database contains 815 cases.

Case-Based Reasoning (CBR)

CBR has been successfully applied to medical diagnosis. It is an inference technique developed in the field of artificial intelligence, originating from the script model proposed by Schenk and Abelson (1977) [13]. CBR is a problem solving paradigm in artificial intelligence, where new problems are solved based on previously experienced similar problems. It is regarded as the most effective way to construct an expert system [14]. The common case-based reasoning operations mainly include 4 cyclic processes, which are retrieving similar cases, reusing cases, revising cases, and retaining [15]. There are three main retrieval techniques of case retrieval in CBR, such as nearest neighbor retrieval, inductive approach, and knowledge-guided approach.

However, how to maintain a set of appropriate feature weights applicable to solving future problems effectively and efficiently are key factors in determining the success of case-based reasoning applications [16]. Many approaches to CBR exploit feature weight setting algorithms that reduce the sensitivity of distance functions. These machine learning techniques, such as case adaptation, decision trees

Table 1 Description of the initial data set

Variables	Variable description	Value
A1	BI-RADS assessment	1–5
A2	Age	18–96
A3	Shape	1–4
A4	Margin	1–5
A5	Density	1–4
A6	Severity	malignant (0) or benign (1)

[17], and LR, can be combined with CBR to set attribute weights in a CBR prediction model. The generation of the attribute weights is performed by considering the presence, absence, and positions of the attributes in a decision tree. The provision of feature rankings provides the user with a sense of how each of the feature values contributes to a particular prediction. Nearest-neighbor retrieval is a simple approach that computes the similarity between stored cases and new input cases, based on weight features. In this study, CBR use the k-nearest neighbor algorithm for total holdout data set. Generally, the technique of the nearest neighbor uses Euclidean distance, as follows:

$$DIS_{xy} = \sqrt{\sum_{i=1}^{n} w_i (f_{xi} - f_{yi})^2} \quad i = 1, 2, \ldots, k \qquad (1)$$

where DIS_{xy} is distance x, and y, f_{xi} and f_{yi} are the values for attribute x and y of the input and retrieved cases, n is the number of attributes, w_i is the importance weighting of attributes x, y. In this study, CBR use the k-nearest neighbor algorithm for total holdout data set. This produces a formulation equivalent to [18] and the reader should consult that paper for more details.

CBR weight set ranked by decision tree

A node represents a pattern observed among instances, which is the path of the node from the root. A decision tree holds all patterns that provide better classification accuracy within an optimum height. The number of defective/good instances classified under a node can be used to measure the support and confidence of the pattern causing a defective/good case. The first step to inducing a decision tree is to rank the attributes according to their importance in classifying the data. The C4.5 algorithm ranks using an entropy measure [17]. However, if the information gain is taken as the classification rules, the attributes with many attribute values will be selected to improve the information gain by the Shannon entropy gain $Gain(A)$ [17]. The C4.5 pruning trees are the conditions for judgment based on the predicated error rate. This study adopted the Classification and Regression Trees (CART) algorithm from SPSS Answertree 3.1 software to determine the maximum gain, and executed 10-fold cross-validation.

CBR weight set ranked by Logistic Regression

Logistic Regression has been successfully applied to classifying high-dimensional cancer datasets [19]. However, we use the LR model as a feature selection (feature ranking) method. Hosmer and Lemeshow [20] provides a detailed survey and overview of the existing methods for LR model. The maximum likelihood estimation procedure is used to obtain estimates of the coefficients, and to maximize the value of a function called log-likelihood function [20]. The approach implementing LR in each node presents some peculiarities. The feature selection process is performed concomitantly with the estimation of decision functions for the classification procedure. The methodology for feature selection is known as stepwise forward, and it is based on Wald's test of significance. The equation for calculating the value of Wald is Eq. (2), as follows:

$$wald = \left(\frac{\beta_1}{S.E.} \right)^2 \qquad (2)$$

where β_1 denotes the output regression coefficient and $S.E.$ denotes the standard error.

PSO-based ANN

ANN has the special capacity to approximate dynamics of the nonlinear systems of many applications, in a black box manner. The multi-layered, feed-forward neural networks of supervised training types, trained with a BP learning algorithm, are the most popular networks applied in different fields. Similar to all neural network models, three layered, feed-forward neural networks must be trained to obtain predicting ability with a training set. Concerning the training, the BP algorithm is usually used to perform such a task, that is to minimize an average sum squared error term by performing a gradient descent in the error space. The BP neural network, as a gradient search algorithm, has some limitations associated with over fitting, local optimum problems, and sensitivity to the initial values of weights [7, 21].

However, building an optimal prediction model from nonlinear problems is complicated by the presence of many training factors. Several different attempts have been proposed by various researchers to solve this training problem, which include imposing constraints on the search space, restarting training at random points, adjusting training parameters, and restructuring the ANN structure [7]. Recently, a new evolutionary computation technique, PSO, is proposed [22]. In order to improve training performance, the PSO technique has been used to train multi-layered feed-forward neural networks to discriminate between different operating conditions. The global or local best solutions in PSO are reported only to the other particles in a swarm. Therefore, evolution only looks for the best solution, and the swarm tends to converge to the best solution quickly and efficiently, thus increasing the probability of global convergence [23, 24]. There are several articles focused on the PSO [21, 23–26]. These references indicate that NN training with PSO algorithms are successful in evolving ANN, and achieve a generalization performance comparable to or better than those of the standard BP networks. It is called PSO-Based ANN classifier.

Fig. 1 Equivalent ANFIS architecture
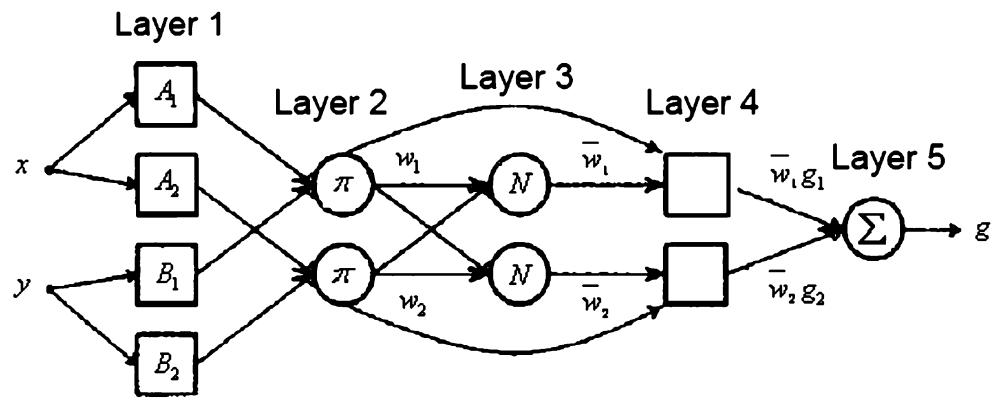
PSO-based ANN classifier

A typical three-layer feed-forward neural network consists of an input layer, a hidden layer, and an output layer. The nodes are connected by weights and output signals, which are a function of the sum of the inputs to the node, and modified by a simple nonlinear activation function. Learning involves adjustments of the synaptic connections (i.e. weights) among the neurons. Every neuron model consists of a processing element with synaptic input connections and a single output. Demuth and Beale [6] provides detailed information about the theory and applications of NN structures.

During neural network training, the weight and bias values are adjusted to minimize training errors, which are a measure of the correlation between the NN model output and training data. The PSO is a population-based stochastic optimization technique inspired by the social behavior of flocking birds [26]. The system is initialized with a population of random solutions, then searches for the optima solution by updating generations. A swarm consists of a set of particles, where each particle represents a potential solution. In particular, each particle remembers its best position from among those it occupied, which is referred to as *pbest*, as well as the best position of its neighbors [24]. The velocity vector drives the optimization process, and reflects social exchange information. Moreover, two main algorithms regularly used in PSO are a local best algorithm (*lbest*) and a global best algorithm (*gbest*). Each individual of the population has an adaptable velocity (position change), according to which it moves within the search space. In the simplest form, position $\vec{P}_i$ and velocity $\vec{v}_i$ of each particle are represented by the following equations, and consider *lbest* rather than *gbest* as the best position of the particle referred to the neighbors:

$$\vec{P}_i^t = \vec{P}_i^{t-1} + \vec{v}_i^t \qquad (3)$$

The *gbest* algorithm only maintains one single best solution, and each particle simultaneously moves towards its previous best position, and the best particle of the whole swarm. Eventually all particles will converge to this position. In our proposed method, the neural network is a neural network, the learning process is a supervised type, and the learning paradigms are the BP method and the PSO technique. Both BP and PSO training processes require bounded and differentiable activation functions, therefore, a sigmoid function is used.

Network training can be more efficient when certain preprocessing steps are performed on network inputs. The returned normalized input variables will all fall within the interval of [0, 1], Eq. (4) illustrates the use of this function, as follows.

$$F_i = \frac{X_i - \min(X_1, X_2, \cdots, X_n)}{\max(X_1, X_2, \cdots, X_n) - \min(X_1, X_2, \cdots, X_n)} \qquad (4)$$

ANFIS classifier

The adaptive neuro-fuzzy inference system is capable of approximating any real continuous function in a compact set to any degree of accuracy [27, 28]. Specifically, the ANFIS system is functionally equivalent to the Sugeno first-order
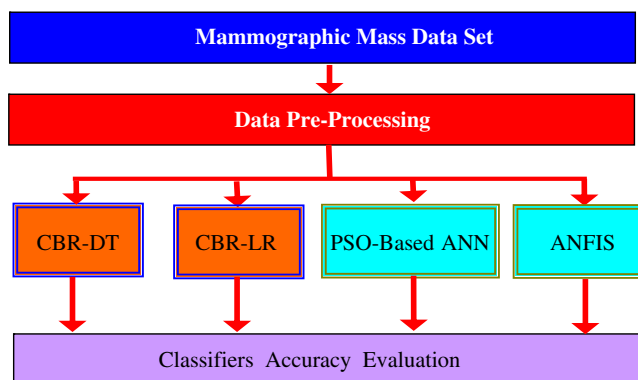

Fig. 2 Experimental process step representation

Table 2 Weight ranked by CART algorithm

Variable	A1	A2	A3	A4	A5
Weight	0.077	0.735	0.136	0	0

Table 3 Chi-square tests of coefficients results

	Chi-square	df	Sig.
Step			
Step1	515.068	5	.000
Block	515.068	5	.000
Model	515.068	5	.000

fuzzy model [9]. ANFIS and ANN models perform similarly in some cases, but the ANFIS model predicts better than the ANN model in most of the cases. A key feature of the ANFIS approach is the ability to capture knowledge from data that is inherently imprecise, and maintain a high level of performance in the presence of uncertainty. ANFIS combines the advantages of neural networks and fuzzy systems, which is important because medical tests and data are inherently imprecise due to individual differences, measurement errors, and noise [18]. In other words, ANFIS techniques with reasoning problems under uncertainty and ambiguity circumstances, and they directly exploit expert knowledge bases [29]. ANFIS has been successfully used in modeling, control, and different areas of biomedicine, including breast cancer and diabetes [30, 31].

In this study, ANFIS was used to develop a pavement performance prediction model. Using a given input/output data set, ANFIS applies a hybrid-learning rule combining back-propagation, gradient-descent, and a least-squares algorithm to identify and optimize the Sugeno system's parameters. The equivalent ANFIS architecture is depicted in Fig. 1. As can be seen, the ANFIS consists of five layers. The first layer is used to generate the membership grades for each set of input data vectors. Layer 2 implements the fuzzy AND operator, while Layer 3 acts to scale the firing strengths. The output of Layer 4 is comprised of a linear combination of the inputs, multiplied by the normalized firing strength w:

$$Y = w(pX + r) \tag{5}$$

where p and r are adaptable parameters. Layer 5 is a simple summation of the outputs of Layer 4.

Table 4 Results of the Wald significance analysis

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)						
BIRADS	2.219	.229	93.986	1	.000	9.202
Age	.047	.009	29.987	1	.000	1.048
Shape	.410	.114	12.976	1	.000	1.507
Margin	.278	.092	9.075	1	.003	1.320
Density	−.026	.289	.008	1	.929	.975
Constant	−14.255	1.332	114.520	1	.000	.000

A denotes Variable(s) entered on step 1: BIRADS, Age, Shape, Margin, Density

Table 5 Weight ranked by logistic regression algorithm

Variable	A1	A2	A3	A4	A5
Weight	0.643	0.205	0.089	0.062	0.001

Training the ANFIS is a two-pass process over a number of epochs. During each epoch, the node outputs are calculated up to Layer 4, and then, at Layer 5, the consequent parameters are calculated using a least-squares regression method. The output of the ANFIS is calculated and the errors propagated back through the layers in order to determine the premise parameter (Layer 1) updates.

The simulation for CBR, PSO-Based ANN and ANFIS classifiers was performed by MATLAB® software. The system configuration of the computer used for training both models was as follows: operating system Windows XP; processor, Intel Core 2 Duo T7500; total physical memory was1.5 GB.

Structure of experimental steps

The Mammographic Mass Data Set was used separately for CBR-DT, CBR-LR, PSO-Based ANN and ANFIS. The classification result from each method was recorded for further comparisons. *P*-value was used to demonstrate the difference between each pairwise comparison for classification performance. Fig. 2 schematically represents the experiment process steps evolving in such four ways.

Results

Assessment of predictive accuracy is a critical aspect of evaluating and comparing models, algorithms or technologies that produce the predictions. In the field of medical diagnosis, receiver operating characteristic (ROC) curves have become the standard tool for this purpose [32]. A common method is to calculate the area under the ROC curve, abbreviated AROC [33]. Since the AROC is a portion of the area of the unit square, its value will always be between 0 and 1.0.

Table 6 The control parameters used for running PSO

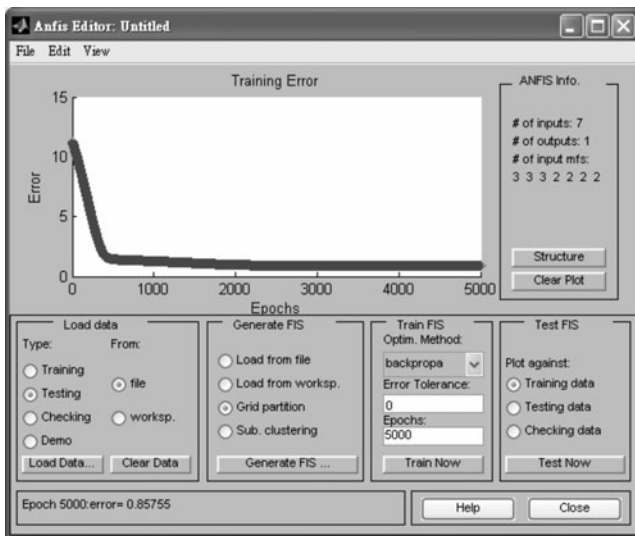| Parameter	Value
# of Population	30
# of Generation	1000
Learning factor	0.3
Inertia weights	0.45 and 0.8
Fitness	Misclassification Error

Fig. 3 Convergence curves of ANFIS

CBR systems

Decision Tree weight ranked

According to the gains of the decision tree, it screened the gains of three attributes, BI-RADS, Age, and Shape, which are 33, 336, and 58, respectively. After CART conversion, the weights, and their order are A2(0.735), A3(0.136), and A1 (0.077) respectively, as shown in Table 2. According to the experimental results, the AROC of the ROC curve is 83.60%.

Logistic Regression weight ranked

This study is based on Severity variables, divided into two categories, benign diagnosis (0) and malignant tumor (1). Table 3 is omnibus tests of model coefficients, from which we find that when $\chi^2 = 515.068$, and significance is below 0.05, it indicates that the overall adaptation of the LR model is favorable. Table 4 is the results of the Wald significance analysis, based on which, the ordering of influences regarding whether patients suffer from the diseases is: BI-RADS, Age, Shape, Margin, and Density. Finally, the value of Wald is converted into weights, according to its proportion of the total number, as the reference for CBR weights input. The converted weights

Table 7 AROC values of the proposed models

Model	Input variables	AROC
CBR-DT	A2, A3 and A1	83.6%
CBR-LR	A1, A2, A3 and A4	79.9%
PSO-Based ANN	All Variables	91.1%
ANFIS	All Variables	92.8%

Table 8 Pairwise comparison of ROC curves

P-value	ANFIS	PSO-based ANN	CBR-DT	CBR-LR
ANFIS	–	0.482	0.007	0.000
PSO-Based ANN	0.482	–	0.030	0.001
CBR-DT	0.007	0.030	–	0.305
CBR-LR	0.000	0.001	0.305	–

are as shown in Table 5. The experimental results show that the AROC of ROC curve is 79.90%.

PSO-based ANN classifier

The combinations of control parameters used to run the PSO are as shown in Table 6. The fitness function of the PSO process, also called the target function, is used to decide the degree of fitness of a chromosome under an environmental condition, namely, measuring the performance of every chromosome. The experiments were repeated ten times to ensure that any misclassification errors (ME) converged to a minimum value. The PSO-Based ANN experimental results show that the AROC of ROC curve is 91.10%.

ANFIS system

The Sugeno model is adopted by ANFIS system, using the IF-THEN rules and fuzzy reasoning. A similar function is employed by the regression function or conversion function of the nodes of the same layer to construction the main framework of ANFIS. This model takes A1, A2, A3, A4, and A5 as input variables, and A6 as output variable, with 80% (651) of the observations used for training, and the remaining 20% (164) observations used for testing; then, load the ANFIS tool kit of MATLAB by training data, select the regression function, set up the times of network training, and carry out
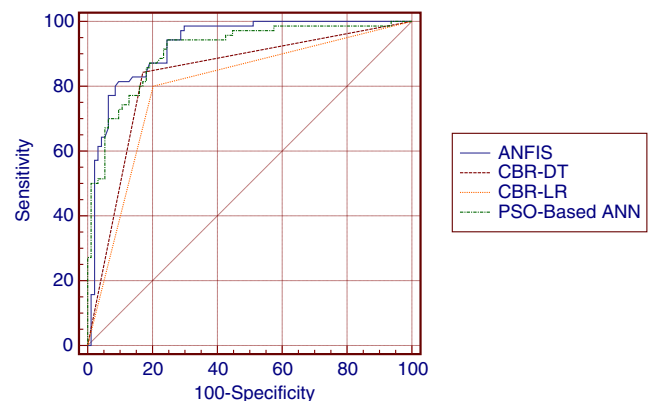

Fig. 4 Comparison of ROC curve

training until complete. After obtaining the results, load the testing data, and verify the accuracy. The network learning cycles are set at 5,000 times, mean squared error is (MSE)=0.00, and then select Gbellmf membership functions. The results of convergence training from the processes of the experiments are as shown in Fig. 3. The experimental results of ANFIS show that AROC of ROC curve is 92.80%.

Performance comparison of classifiers

The results of modeling for breast cancer prediction using a CBR-DT, CBR-LR, PSO-Based ANN, and ANFIS are computed and reported in Table 7. We found that the PSO-Based ANN and ANFIS has better performance than the CBR-based classification, and the AROC performance of ANFIS is the highest, achieving 92.80%.

The CBR-DT approach resulted in the AROC=83.60%, the CBR-LR approach in AROC=79.90%, the PSO-Based ANN in AROC=91.10% and the ANFIS approach in AROC = 92.80%. From Table 8, the classification performance from ANFIS is significantly different from both CBR-DT and CBR-LR. The P-values are 0.007 and 0.000, respectively. And the classification performance from PSO-Based ANN is also significantly different from both CBR-DT and CBR-LR. The P-values are 0.030 and 0.001, respectively. However, the difference between ANFIS and PSO-Based ANN is not significant (P-value=0.482). The comparisons of ROC curves for CBR-DT, CBR-LR, PSO-Based ANN and ANFIS are shown as Fig. 4.

Discussion

Breast cancer is one of the most common cancers in women around the world. Therefore, identifying patients within the system, that may contribute to the development of the automatic diagnostic system of breast cancer. Mammography is the most effective method for breast cancer screening available today. However, the low positive predictive value of breast biopsy resulting from mammogram interpretation leads to approximately 70% unnecessary biopsies with benign outcomes. In this paper, a CBR system with two feature ranking algorithms, PSO-Based ANN and ANFIS were proposed for breast cancer classification application.

The CBR system aims at the improvement of retrieval speed. Meanwhile, the objective and subjective factors are considered comprehensively, the similar cases can be extracted rapidly and accurately from case base. The major advantage of ANN is the flexible capability of nonlinear modeling. An ANN-Based classifier is an effective way to extract hidden knowledge from breast

cancer register, but the resulting model is not clear and could not be easily understood. The ANFIS is a developed model of ANN by combining the benefits of ANN and fuzzy inference system in a single model. The ANFIS shows characteristics of fast and accurate learning with the ability of using both linguistic information and data information and good generalization capability. ANFIS is a rule-based step to map previously extracted terms and events from files easily and can be used to support less experienced oncologists in decision-making. For the mammographic mass data set, although the result shows that the POS-Based ANN and ANFIS approach perform significantly better than the CBR approach, the application of PSO-Based ANN and ANFIS on other data sets should be cautious.

References

1. Mangasarian, O. L., Street, W. N., and Wolberg, W. H., Breast cancer diagnosis and prognosis via linear programming. *Oper. Res.* 43(4):570–577, 1995.
2. Xiong, X., Kim, Y., Baek, Y., Rhee, D. W., and Kim, S. H., Analysis of breast cancer using data mining & statistical techniques, Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005 and First ACIS International Workshop on Self-Assembling Wireless Networks. SNPD/SAWN, pp. 82–87, (2005).
3. Nilsson, M., and Sollenborn, M., Advancements and trends in medical case-based reasoning: An overview of systems and system development, International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004), Miami Beach, FL (US), 1 pp. 17–19, (2004).
4. Lieber, J., and Bresson, B., Case-based reasoning for breast cancer treatment decision helping. Proceedings of the 5th European Workshop on Case-Based Reasoning, pp. 173–185, (2000).
5. Pandey, B., and Mishra, R. B., Knowledge and intelligent computing system in medicine. *Comput. Biol. Med.* 39(3):215–230, 2009.
6. Demuth, H., and Beale, M., *Neural network toolbox user's guide for use with MATLAB*. MathWorks Inc., Natick, MA, 2006.
7. Sexton, R. S., Dorsey, R. E., and Sikander, N. A., Simultaneous optimization of neural network function and architecture algorithm. *Decis. Support Syst.* 36(3):283–296, 2004.
8. Pandey, B., and Mishra, R. B., Knowledge and intelligent computing system in medicine. *Comput. Biol. Med.* 39:215–230, 2009.
9. Jang, J.-S. R., ANFIS: Adaptive-network- based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* 23(3):665–685, 1995.
10. Riedl, C. C., Pfarl, G., and Helbich, T. H., Breast imaging reporting and data system, http://www.birads.at/.
11. Elter, M., Schulz-Wendtland, R., and Wittenberg, T., Mammographic Mass Data Set, http://archive.ics.uci.edu/ml/datasets/Mammographic+Mass, 2007.
12. Mammogram Interpretation: Categories and the ACR/BI-RADS, http://www.imaginis.com/breasthealth/acrbi.asp, (2007).
13. Schank, R. C., and Abelson, R. P., *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Erlbaum, Hillsdale, New Jersey, 1977.
14. Burke, E. K., MacCarthy, B., Petrovic, S., and Qu, R., Structured cases in case-based reasoning-re-using and adapting cases for timetabling problems. *Knowl.-Based Syst.* 13(2–3):159–165, 2000.

15. Althoff, K. D., Auriol, E., Barletta, R., and Manago, M., *A Review of Industrial case-based reasoning tools, An AI perspectives Report*. AI Intelligence, United Kingdom, pp. 3–4, 1995.

16. Zhang, Z., and Yang, Q., Feature weight maintenance in case bases using introspective learning. *J. Intell. Inf. Syst.* 16(2):95–116, 2001.

17. Quinlan, J. R., C4.5: Programs for machine learning. Morgan Kaufmann Publishers, (1993).

18. Mitra, S., and Hayashi, Y., Neuro-fuzzy rule generation: Survey in soft computing framework. *IEEE Trans. Neural Netw.* 11(3):748–768, 2000.

19. Xing, E. P., Jordan, M. I., and Karp, R. M., Feature selection for high-dimensional genomic microarray data, in: Proc. of the 18th International Conference on Machine Learning, pp. 601–608, (2001).

20. Hosmer, D. W., and Lemeshow, S., *Applied logistic regression*, 2nd edition. Wiley, New York, 2000.

21. Yu, J., Wang, S., and Xi, L., Evolving artificial neural networks using an improved PSO and DPSO. *Neurocomputing* 71(4–6):1054–1060, 2008.

22. Kennedy, J., and Eberhart, R. C., Particle swarm optimization. In: *IEEE International Conference on Neural Networks*. IEEE, New York, pp. 1942–1948, 1995.

23. Geethanjali, M., Slochanal, S. M. R., and Bhavani, R., PSO trained ANN-based differential protection scheme. *Neurocomputing* 71(1–3):904–918, 2008.

24. Parsopoulos, K. E., and Vrahatis, M. N., Recent approaches to global optimization problems through Particle Swarm Optimization. *Nat. Comput.* 1(2–3):235–306, 2002.

25. Hu, K., and Huang, S. H., Solving inverse problems using Particle Swarm Optimization: An application to aircraft fuel measurement considering sensor failure. *Intell. Data Anal.* 20(1):421–434, 2007.

26. Sadeghi, B. H. M., A BP-neural network predictor model for plastic injection molding process. *J. Mater. Process. Technol.* 103 (3):411–416, 2000.

27. Jang, J.-S. R., Sun, C. T., and Mizutani, E., Neuro-Fuzzy and soft computing: A computational approach to learning and machine intelligence, Prentice Hall; US Ed edition Prentice-Hall, (1996).

28. Papadimitriou, S., and Terzidis, K., Efficient and interpretable fuzzy classifiers from data with support vector learning. *Intell. Data Anal.* 9(6):527–550, 2005.

29. Mohammad, R., Akbarzadeh, T., and Majid, M. K., A hierarchical fuzzy rule-based approach to aphasia diagnosis. *J. Biomed. Inform.* 40(5):465–475, 2007.

30. Castellano, G., Fanelli, A. M., and Mencar, C., An empirical risk functional to improve learning in a neuro-fuzzy classifier. *IEEE Trans. Syst. Man Cybern. Part B-Cybern.* 34(1):725–731, 2004.

31. Ghazavi, S. N., and Liao, T. W., Medical data mining by fuzzy modeling with selected features. *Artif. Intell. Med.* 43(3):195–206, 2008.

32. Swets, J. A., Measuring the accuracy of diagnostic systems. *Science* 240(4857):1285–1293, 1988.

33. Bradley, A. P., The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* 30 (7):1145–1159, 1997.