

An Artificial Immune System-Based Support Vector Machine Approach for Classifying Ultrasound Breast Tumor Images

Wen-Jie Wu · Shih-Wei Lin · Woo Kyung Moon

Published online: 6 January 2015
© Society for Imaging Informatics in Medicine 2014

Abstract A rapid and highly accurate diagnostic tool for distinguishing benign tumors from malignant ones is required owing to the high incidence of breast cancer. Although various computer-aided diagnosis (CAD) systems have been developed to interpret ultrasound images of breast tumors, feature selection and the setting of parameters are still essential to classification accuracy and the minimization of computational complexity. This work develops a highly accurate CAD system that is based on a support vector machine (SVM) and the artificial immune system (AIS) algorithm for evaluating breast tumors. Experiments demonstrate that the accuracy of the proposed CAD system for classifying breast tumors is 96.67 %. The sensitivity, specificity, PPV, and NPV of the proposed CAD system are 96.67, 96.67, 95.60, and 97.48 %, respectively. The receiver operator characteristic (ROC) area index A_z is 0.9827. Hence, the proposed CAD system can reduce the number of biopsies and yield useful results that assist physicians in diagnosing breast tumors.

Keywords Breast tumors · Textural feature · Morphological feature · Artificial immune system algorithm · Support vector machine

Introduction

Breast cancer is the most common cancer among women globally. In 2013, it was the second cause of death from cancer

in women. In the United States, approximately 232,340 women were diagnosed with breast cancer that year, of which 39,620 died [1]. Breast cancer can be treated effectively when detected in its early stage [2]. Therefore, the early detection and diagnosis of breast cancer are critical in reducing the death rate and to extending a patient's life.

Among the many methods for detecting breast cancer, biopsy is the most accurate in identifying various kinds of tumor. However, biopsy is invasive and much more expensive than other methods of detection [3]. Mammography and ultrasound (US) imaging are popular techniques for detecting breast cancer. Although US usually plays an adjunctive role to mammography in detecting breast tumors, it is more convenient and safer than mammography [3]. Accordingly, in recent years, the ultrasound examination has gradually become the preferred method for the diagnosing breast lesions. However, when diagnosing breast cancer, the performance of US depends on the size, number, location, and properties of the lesions; the skill of the operator of the US device; and the specifications of the system [4].

Computer-aided diagnosis (CAD) has become a major research topic in the field medical imaging and diagnosis [5]. Using a CAD system, physicians can obtain much information that is useful in identifying breast tumors [6]. Many researchers have developed CAD systems that are based on ultrasound imaging. For example, Chang et al. [7] used six morphological ultrasound image features to classify breast tumors using a support vector machine (SVM). The accuracy and the area under the receiver operating characteristic (ROC) curve of their CAD system for classifying breast tumors is 90.95 % (191/210) and 0.9467, respectively. Chen et al. [8] developed a CAD system for breast tumors that was based on image retrieval techniques that involve principal component analysis (PCA) of the textural features. The area under the ROC curve for their CAD system with six practical textural features was 0.925. Moon et al. [9] utilized a fuzzy clustering

W.-J. Wu · S.-W. Lin (✉)
Department of Information Management, Chang Gung University,
Tao-Yuan, Taiwan 333, Republic of China
e-mail: swlin@mail.cgu.edu.tw

W. K. Moon
Department of Diagnostic Radiology, Seoul National University
Hospital, Seoul, South Korea

method to classify breast tumors using information about their elasticity. In their CAD system, the accuracy and the area under the ROC curve for distinguishing breast tumors is 83.5 % (81/97) and 0.9017, respectively. Chen et al. [10] performed a comparative analysis of three statistical models—logistic regression analysis (LRA), SVM, and a neural network (NN)—for diagnosis using breast ultrasonography. They demonstrated that the three models (LRA, SVM, and NN) exhibit the same diagnostic performance, as demonstrated by curve analysis. The areas under the ROC curve of the three models in their CAD system are 0.9341, 0.9151, and 0.9086, respectively. Tseng et al. [11] proved that the diagnostic performance of a CAD system that was based on ultrasonically determined breast tumor morphology is unaffected by speckles. According to their experiments, the areas under the ROC curve of classifying speckle reduced tumor images and non-speckle reduced ones are 0.82 and 0.81, respectively. Because only one kind of features is used to classify breast tumors, the accuracy of above CAD system was not too good. Tan et al. [12] proved that spiculation patterns can be used to distinguished malignant breast tumors from benign ones. They computed spiculation patterns and combined them with echotexture, echogenicity, shape, posterior acoustic behavior, and margins features to classify breast tumors. In their CAD system, the area under the ROC curve for classifying breast tumors is 0.93. Although they combined many kinds of features to classify breast tumors, they did not use any feature selection method to identify the dominant features. Wu et al. [13] developed a diagnostic system for evaluating breast tumors using a GASVM approach. The accuracy and the area under the ROC curve of their CAD system is 95.24 % and 0.9615, respectively. Because GA algorithm has randomness and its specific characteristic, it may have the possibility to improve the accuracy of the CAD with the GASVM approach.

Classification is a common task of most CAD systems. A supervised learning technique is used to classify a set of dependent variables based on the features of the input data. However, many classification problems that involve a large set of potential features can only be solved by identifying the dominant features for classification purposes. Finding “good” features by feature selection is important and the usefulness of each feature must be determined. Parameter tuning is another important problem associated with classification algorithms [14, 15]. Parameter values must be set carefully to avoid obtaining poor classification results. Unfortunately, the optimal values of the parameters of classification algorithms vary among problems. No clear rules for finding the optimal values of parameters and features have been established. Trial and error is frequently used to perform these tasks but it is time-consuming. Most heuristic methods use a hill-climbing method that begins with

initial values of “good” parameters or selected subset of features. Although heuristic methods yield results rapidly, they tend to fall into sub-optimal solutions.

The artificial immune system (AIS) emerged in the 1990s, since when many interdisciplinary researchers have used it in artificially intelligent (AI) applications [16]. In recent years, numerous researchers have applied AISs to improve the performance of classification algorithms. For example, Latifoğlu et al. [17] developed an AIS-based method for diagnosing atherosclerosis based on Doppler signals from the carotid artery. Lin and Chen [18] applied the AIS algorithm to enhance the classifying capacity of the case-based reasoning algorithm. Kuo et al. [19] combined the AIS algorithm with a fuzzy neural network (FNN) to increase the accuracy of an RFID-based positioning system.

This work develops an efficient SVM [20–25] method that is based on the AIS algorithm (AISSVM) for diagnosing ultrasound images of breast tumors. The proposed CAD system simultaneously performs parameter tuning and feature selection, and thus achieves high classification accuracy.

Materials and Methods

Data Acquisition

The CAD system in this work was evaluated using 210 ultrasound images of breast tumors. The dataset comprised 120 benign and 90 malignant cases. Each case is that of a patient with an age of between 18 and 64 years. All cases are verified by both biopsy and lumpectomy. The local ethics committee approved this study and informed consent was waived.

All of the images in the dataset were obtained using an ATL HDI 3000 system with a L10-5 small-part transducer. One of the authors (Dr. Moon) supplied all of the images. No acoustic standoff pad was used in the acquisition of any image. Notably, all solid nodules according to the categorization of the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) exceeded C3 and the algorithm was applied to all of the images without selection bias.

Image Segmentation

Image segmentation was an important step before the extraction of the features of each tumor because segmentation may influence the extraction of features from tumor images. In this work, a series of steps that have been proven to be effective [7] are used to segment the ultrasound images of tumors. First, the original images were processed using an anisotropic diffusion filter [26], the stick method [27], and

the thresholding method [28]. The above methods yield a binary image, which can be combined with the original image. The binary image is combined with the original one with an equal weighting. Finally, the level set method [29] was utilized to segment the tumor in the combined image. Figure 1 presents the result of each segmentation step.

Feature Extraction

Many features in the ultrasound images of breast tumors are used to distinguish benign tumors from malignant ones. The two most commonly used kinds of feature are texture and shape-related. Textural features are extracted easily but depend on the region of interest (ROI) that is delineated by physicians. Shape features (also called morphological features) are also effective for identifying breast tumors. Although the ROI does not affect shape features, extracting such features generally involves considerable computation. In this work,

textural and morphological features are used to identify breast tumors. The following subsection describes textural features and a later subsection will introduce morphological features.

Textural Features

An ultrasound image is composed of many pixels with various gray levels. Since different tissues have different textures, textural features provide useful information for classifying breast tumors as benign or malignant. The auto-covariance matrix can specify the inter-pixel relationships in an image and it has been used to classify breast tumors [8, 21, 22]. In the method that was proposed by Chang et al. [22], the modified auto-covariance coefficients inside a tumor are defined as follows.

$$\gamma(\Delta m, \Delta n) = \frac{A(\Delta m, \Delta n)}{A(0, 0)} \quad (1)$$

$$A(\Delta m, \Delta n) = \frac{1}{\text{PixelCount}_{(\Delta m, \Delta n)}} \cdot \sum_{x=0}^{M-1-\Delta m} \sum_{y=0}^{N-1-\Delta n} [f_{in}(x, y) - \bar{f}_{in}] [f_{in}(x + \Delta m, y + \Delta n) - \bar{f}_{in}] \quad (2)$$

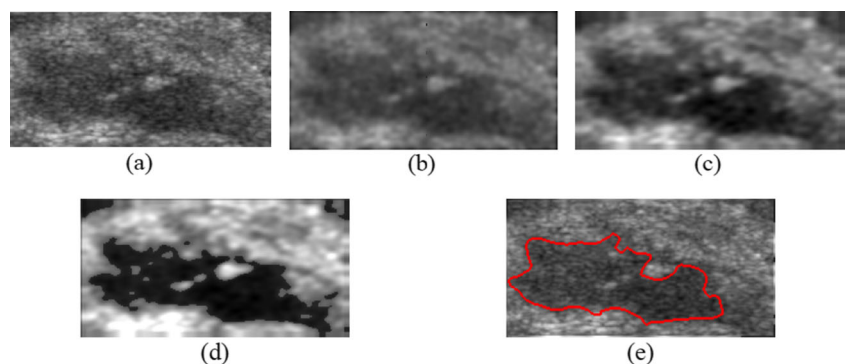
where $f_{in}(x, y)$ and $f_{in}(x + \Delta m, y + \Delta n)$ are the gray levels of two pixels (x, y) and $(x + \Delta m, y + \Delta n)$; \bar{f}_{in} denotes the mean value of $f_{in}(x, y)$; $M \times N$ is the size of the image, and $\text{PixelCount}_{(\Delta m, \Delta n)}$ is the number of pairs of pixels inside a tumor that are separated in the x direction by Δm and in the y direction by Δn .

In this work, the auto-covariance coefficients for each tumor were initially calculated using a 5×5 auto-covariance matrix (with 25 auto-covariance coefficients, with both Δm and $\Delta n = 4$). The 25 auto-covariance coefficients, except $\gamma(0, 0)$, were used to generate a 24-dimensional textural feature vector for a tumor.

Morphological Features

Since the shape of benign tumors typically differs from that of malignant ones (Fig. 2), tumors can be distinguished using information about their morphology. In this work, six morphological features of each tumor were calculated and used as the morphological information for that tumor. These morphological features have been proven to be useful in diagnosing breast tumors [7] and they are defined as follows.

Fig. 1 Tumor segmentation using proposed approach. **a** Original image. **b** Anisotropic diffusion filter applied to image (a). **c** Stick method applied to image (b). **d** Automatic threshold method used to combine image (c) with image (a). **e** Level set method applied to image (d)



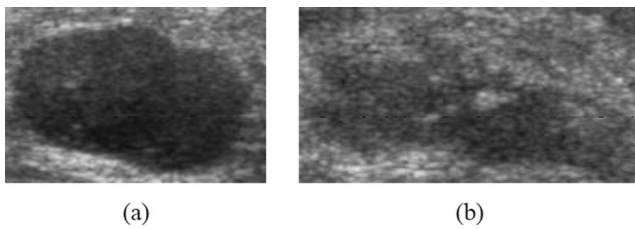


Fig. 2 Ultrasound images of two types of breast tumor: **a** benign tumor and **b** malignant tumor

1.

$$\text{Form_Factor} = \frac{4\pi \cdot \text{Area}}{\text{Perimeter}^2}, \quad (3)$$

2.

$$\text{Roundness} = \frac{4 \cdot \text{Area}}{\pi \cdot \text{Diameter}_{\max}^2}, \quad (4)$$

3.

$$\text{Aspect_Ratio} = \frac{\text{Diameter}_{\max}}{\text{Diameter}_{\min}}, \quad (5)$$

4.

$$\text{Convexity} = \frac{\text{ConvexPerimeter}}{\text{Perimeter}}, \quad (6)$$

5.

$$\text{Solidity} = \frac{\text{ConvexArea} - \text{Area}}{\sum_{i=1}^{N_t} \text{ConvexArea}_i - \text{Area}_i / N_t}, \quad (7)$$

6.

$$\text{Extent} = \frac{\text{Area}}{\text{Bounding_Rectangle}}, \quad (8)$$

where *Area* and *Perimeter* represent the area and perimeter of a tumor, respectively, and *Diameter_{max}* and *Diameter_{min}* are the maximal and minimal dimensions of a tumor at various angles of projection, respectively. *ConvexPerimeter* and *ConvexArea* are the perimeter and area of the convex hull of a tumor, respectively. *N_t* is the number of tumors in the database, and *Bounding_Rectangle* is the area of the minimal rectangle that contains a tumor. Figures 3 and 4 display examples of convex hull and bounding rectangle of a tumor, respectively.

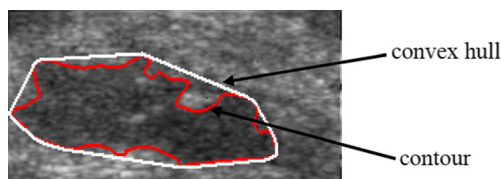


Fig. 3 Convex hull and contour of a tumor

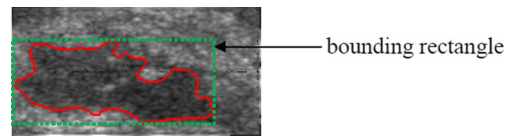


Fig. 4 Bounding rectangle of a tumor

Classification by SVM

In recent years, SVMs have been demonstrated to be effective in many fields, owing to their high classification accuracy and application to high-dimensional data [30]. SVMs can map input vectors into a high-dimensional feature space by applying non-linear mapping functions and find a computationally efficient way of searching a separate hyperplane in the space. Given a set of points $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_N, y_N)$, where $x_i \in \mathbb{R}^N$ represents the *i*th input vector and $y_i \in \{-1, 1\}$ is the corresponding designed output, the objective of an SVM is to find a hyperplane $w: wx + b = 0$ that separates the data with a maximal margin. Points on the border of the margin are called support vectors. The classification is carried out by applying the following decision function:

$$f(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i k(s_i, x) + b \right), \quad (9)$$

where α_i is the positive Lagrange multiplier; s_i is the support vector, and $k(s_i, x)$ is the kernel function that transforms input vectors into a high-dimensional feature space.

The most frequently used kernel functions are the polynomial kernel, the sigmoid kernel, and the radial basis kernel functions (RBF) [31–33]. Since the RBF is usually used in classification using support vector machines [34] and was found by Huang et al. [21] to exhibit the best performance, it is used in this work. The RBF is defined as follows.

$$k(x, y) = \exp \left(-\gamma (x - y)^2 \right) \quad (10)$$

where $\gamma \in \mathbb{R}$ is a non-zero parameter.

The performance of an SVM may be poor because the input space has many dimensions and is non-clean. Such a situation arises when some features are redundant or contain false correlations among each other. Such features may increase the computing time and reduce classification accuracy. To eliminate noisy, irrelevant, and redundant data without harming the classifying power of the SVM, an efficient and robust feature selection method is required.

Parameter tuning importantly influences the performance of an SVM. The SVM with the RBF kernel has two major parameter values, *C* and γ , that must be set [35]. Parameter *C* controls the tradeoff between the complexity of the model and

the training error and parameter γ is used in the RBF kernel. Although the grid search [36] process is usually used in the selection of parameters in the construction of SVM models, it requires much computation time. Therefore, a quick and robust method for setting the parameters of the SVM is required. In this work, the AIS algorithm is used to determine the values of the parameters of the SVM and simultaneously to select useful features for classifying breast tumors.

Parameter Tuning and Feature Selection Using AIS

AIS is an efficient meta-heuristic algorithm that is used in various optimizations [37–40]. AIS searches a solution space using a population of antibodies that each represents an encoded solution. Each antibody has an affinity value that is based on its performance. Better antibody has a higher affinity. The population evolves by performing a series of operations until some stopping criterion is met.

To increase the accuracy of classification of breast tumors in ultrasound images, this work proposes an AIS-based approach for parameter tuning and feature selection for use with the SVM (called AISSVM). The following subsection describes the elements of the AIS.

Antibody Representation and Initial Solutions

An antibody is a solution that includes the values of the parameters of the SVM that are turned and the subset of selected features. Two parameters of the SVM, C and γ , must be determined, and N values specify whether the N features are selected. Consequently, the antibody representation includes 2 plus N numbers. Each number in the antibody representation is encoded using 8 bits, which are scaled into certain ranges according to the following normalization formula.

$$V' = MIN_a + (MAX_a - MIN_a) * \left(\frac{V - MIN_a}{MAX_a - MIN_a} \right) \quad (11)$$

where V represents the original value, V' denotes the scaled value, MAX_a represents the upper bound on the interval, and MIN_a represents the lower bound on the interval.

The values of the N variables range from zero to one. If a variable has a value of ≤ 0.5 , then its corresponding feature is not chosen; otherwise, it is chosen.

Calculation of Affinity

The affinity of each n_{pop} antibody in the population of antibodies AB is calculated using the affinity function, where n_{pop} is the number of antibodies. The affinity function is related to the accuracy of classification. A higher accuracy of classification is associated with a higher affinity of the antibody.

Selection and Cloning

The “ n_s ” antibodies with the highest affinities in AB form a new set of high-affinity antibodies to be cloned. The number of clones of each of the n_s selected antibodies is calculated as $round(\beta \times (n_s - i + 1) / n_s)$, where i represents the antibody in the AB with the i th highest affinity, β represents a multiplicative factor, and $round(\bullet)$ is an operator that rounds its argument. Accordingly, the total number of generated clones is as follows:

$$\sum_{i=1}^{n_s} round\left(\beta \times (n_s - i + 1) / n_s\right) \quad (12)$$

Mutation

Each clone is then mutated by changing the value of some bits to find its neighboring solutions. Each number in the antibody representation can be encoded using 8 bits, so the chosen bit(s) is (are) changed from one to zero or from zero to one.

Updating Antibody Population

To maintain the diversity of antibody population, the similarity (f_{ij}) of antibody pair i (ab_i) and antibody j (ab_j) is calculated as $\sum_{m=1}^L ab_{ij}^m$, where L represents the length of the antibody representation. If the antibodies at the m locus of ab_i and ab_j are identical, then $ab_{ij}^m = 1$; otherwise, $ab_{ij}^m = 0$. If the similarity

Table 1 Experimental results in five runs

Number	Parameters of SVM	Selected features	Accuracy rate (%)
1	$C=4057 \gamma=0.0508$	f[1], f[5], f[6], f[15], f[22], f[25], f[26], f[29]	96.67
2	$C=8561 \gamma=0.048$	f[10], f[16], f[23], f[25], f[29], f[30]	95.72
3	$C=8488 \gamma=0.064$	f[10], f[16], f[22], f[25], f[26], f[29]	96.19
4	$C=29,371 \gamma=0.038$	f[1], f[5], f[6], f[10], f[27], f[29]	95.72
5	$C=12791 \gamma=0.022$	f[5], f[6], f[15], f[20], f[27], f[28], f[29]	95.24

Table 2 Classification of breast tumors by proposed approach

Sonographic classification	Benign*	Malignant*
Benign	TN 116	FN 3
Malignant	FP 4	TP 87
Total	120	90

* Histological finding

of an antibody is large than a predefined threshold d , then the antibody will be discarded. Finally, antibodies with the n_s highest affinities in the mutated clone set and the antibody population AB are cloned in the next generation.

Parameters and Procedure in AISSVM Approach

Five parameters in the AISSVM approach must be set. The parameter n_{pop} denotes the number of antibodies in the population, n_s is the number of antibodies that are selected for cloning, β is a multiplicative factor that determines the number of clones of each chosen antibody, d is the threshold for affecting the diversity of antibodies, and S_{max} is the maximum number of solutions evaluated.

Figure 5 presents the pseudo-code of the proposed AISSVM approach. In the initialization phase, AIS generates the initial population of n_{pop} antibodies at random, and the affinity of each antibody is then calculated. Following the initialization phase, the following sequence of steps is applied.

1. The antibodies in AB with the “ n_s ” highest affinities are selected.
2. The selected “ n_s ” antibodies are cloned.
3. The clones are mutated. Each mutated clone is a solution, which includes parameter values and selected features, and is fed into the SVM to train the classification model.
4. The affinity of each mutated clones is calculated.
5. AB is updated by applying antibody population update method to the AB and mutated clones.
6. The antibody with the highest affinity in the AB is recorded.
7. The total number of evaluated solutions, Sol , is updated.
8. The termination condition is applied. If Sol exceeds S_{max} , then the AISSVM procedure terminates; otherwise, it continues to the next generation.

Table 3 Objective indices obtained using various approaches: (a) proposed approach, (b) GASVM approach, (c) PCA approach, (d) SFFS approach, (e) approach using all features, (f) approach using textural features, and (g) approach using morphological features

Item	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Accuracy	96.67 %	95.24 %	92.38 %	92.86 %	92.38 %	83.33 %	90.95 %
Sensitivity	96.67 %	97.78 %	95.56 %	94.44 %	93.33 %	85.56 %	88.89 %
Specificity	96.67 %	93.33 %	90.00 %	91.67 %	91.67 %	81.67 %	92.50 %
PPV	95.60 %	91.67 %	87.76 %	89.47 %	89.36 %	77.78 %	89.89 %
NPV	97.48 %	98.25 %	96.43 %	95.65 %	94.83 %	88.29 %	91.47 %

Results

The proposed AISSVM approach was implemented using a PC with an Intel Core i7-2600 CPU, 8 GB of RAM, the Microsoft Windows 7 operating system, and the Visual C++ 6.0 development environment. The SVM is based on the libSVM [41]. The image processing and level set methods were implemented using MATLAB from MathWorks (Natick, MA, USA). After testing with several combinations of parameter values, the following parameter values were used in the AISSVM: n_{pop} , n_s , β , d , and S_{max} were 20, 5, 6, 3, and 10,000, respectively. For the SVM, the range of parameter C was 1 to 30,000 and the range of parameter γ was 0.001 to 1.

To test the performance of the proposed CAD system, 210 ultrasound images of breast tumors were used. These images presented 120 pathologically proven benign cases and 90 malignant ones. The k -fold cross-validation method [42] was used to estimate the performance of the proposed CAD system. All experimental images were divided into k groups, and one group was chosen as the testing group and the remaining $(k-1)$ groups were used to train the AISSVM. This process was repeated until all k groups had been used for testing. In this work, k was five and each group comprised 42 ultrasound images of breast tumors.

In extracting features, the textural and morphological features of each tumor were extracted from the images to generate a 30-dimensional feature vector as the input to the AISSVM. The textural features comprised 24 autocovariance coefficients whose Δm and Δn were 0 to 4 (excluding $\Delta m = \Delta n = 0$) and the six morphological features were defined by formulas (3)–(8). For convenience of notation, each component $f[i]$ of the feature vector was defined as follows:

$$f[i] = \begin{cases} \gamma[j][k] & | i = j*5 + k & \text{if } 1 \leq i \leq 24 \\ \text{Form_Factor} & & \text{if } i = 25 \\ \text{Roundness} & & \text{if } i = 26 \\ \text{Aspect_Ratio} & & \text{if } i = 27 \\ \text{Convexity} & & \text{if } i = 28 \\ \text{Solidity} & & \text{if } i = 29 \\ \text{Extent} & & \text{if } i = 30 \end{cases}$$

where $0 \leq j, k \leq 4$ (excluding $j=k=0$).

Table 4 Computational time required to obtain their best models (feature extraction time is not included)

Approach	AISSVM feature selection	GASVM feature selection	PCA feature selection (using SVM in grid search)	SFFS feature selection (using SVM in grid search)	All features (using SVM in grid search)	Only texture features (using SVM in grid search)	Only morphologic features (using SVM in grid search)
Time (s)	1890	1897	>86,400	>86,400	>86,400	>86,400	>86,400

To confirm the reliability of the AIS algorithm, five runs of the proposed system were executed. Table 1 presents the selected parameter values, features, and classification accuracy in each run. The best run (with the highest classification accuracy rate) has eight features ($i=1, 5, 6, 15, 22, 26, 28$, and 29) to identify the type of breast tumors in the SVM with the RBF kernel whose parameters C and γ determined by the AIS algorithm are 4057 and 0.051, respectively. Table 2 presents the confusion matrix for the best run. Table 3 presents the experimental accuracy and the other four objective indices of the performance of the proposed approach for the best run. These indices are sensitivity, specificity, positive predictive value, and negative predictive value. The accuracy of the proposed approach is 96.67 %; its sensitivity is 96.67 %, specificity is 96.67 %, positive predictive value (PPV) is 95.60 %, and negative predictive value (NPV) is 97.48 %.

The proposed approach is compared to other approaches to prove its effectiveness. These other approaches are the GASVM-based feature selection approach; the PCA-based feature selection approach; the sequential floating forward selection (SFFS)-based feature selection approach; and classification with textural features, morphological features, and all features. The results of the GASVM-based feature selection approach are taken from the original report of Wu et al. [13]. In the PCA-based feature selection approach, the original

30-dimensional feature vector of each tumor is analyzed by PCA and the principal components that explain over 95 % (six features) are selected to form new six-dimensional feature vectors. In the SFFS-based feature selection approach, the original feature vector of each tumor is analyzed by SFFS algorithm to form new seven-dimensional feature vectors. Finally, all of the above approaches determine the features that are input to the SVM to train it and to calculate its accuracy of classification of the breast tumors. Table 3 presents the performance of each approach. We also show the computational time required for various approaches to obtain their best models (time of feature extraction is not included) in Table 4. According to Table 4, AISSVM-based and GASVM-based approaches took less time than other approaches to obtain their best models. This is because other approaches obtain their best models by using grid search method but AISSVM-based and GASVM-based approaches are not.

In addition to the accuracy and the other four objective indices, the receiver operating characteristic (ROC) curves are plotted using ROCKIT software (C. Metz; University of Chicago, Chicago, IL, USA) and the area (A_z) under each ROC curve is determined (Fig. 6). The area A_z is an index of the

```

AISSVM( $n_{pop}$ ,  $n_s$ ,  $\beta$ ,  $d$ ,  $S_{max}$ )
Initialization
Begin
    Generating the  $n_{pop}$  antibodies randomly for initial population;
     $affinities \leftarrow \text{calcAffinities}(AB)$ ;
End
Loop
Begin
     $selectedAntibodies \leftarrow \text{Select}(AB, Affinities, n_s)$ ;
     $clones \leftarrow \text{clone}(selectedAntibodies, Affinities, \beta)$ ;
     $clones \leftarrow \text{mutate}(clones)$ ;
     $cloneAffinities \leftarrow \text{calcAffinities}(clones)$ ;
     $update(AB, clones, affinities, cloneAffinities, n_s, d)$ ;
     $bestAffinity \leftarrow \text{getBestAffinity}(AB)$ ;
     $Sol \leftarrow \text{noSolEval}()$ ;
    If ( $Sol > S_{max}$ )
        Terminate the AISSVM procedure;
    End
End

```

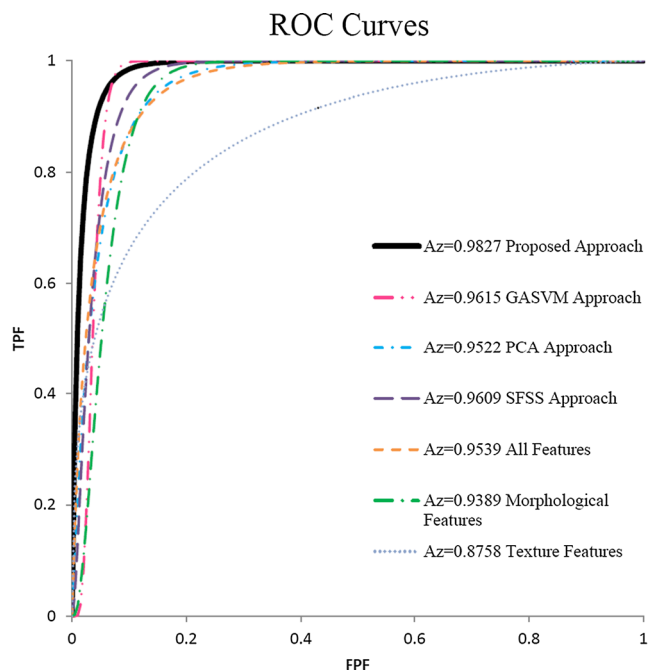
Fig. 5 Pseudo-code of proposed AISSVM-based approach**Fig. 6** ROC analysis of various approaches

Table 5 Paired *t* tests on average classification accuracy for different approaches

AISSVM v.s.	GASVM feature selection	PCA feature selection (using SVM in grid search)	SFFS feature selection (using SVM in grid search)	All features (using SVM in grid search)	Only texture features (using SVM in grid search)	Only morphologic features (using SVM in grid search)
<i>p</i> value	<0.02	<0.02	<0.05	<0.0005	<0.007	<0.015

overall performance of a diagnosis and a diagnostic system can be evaluated using this value. The A_z value of the AISSVM-based approach is 0.9827. Figure 6 displays the A_z values of the other approaches.

To confirm the computational results, we performed paired *t* tests on average classification accuracy to compare the AISSVM-based approach with other approaches. We show the results in Table 5 (confidence level $\alpha=0.05$). In addition, an analysis of variance (ANOVA) was conducted to assess the statistical significance of the differences observed. Figure 7 indicates that the means plot with HSD Tukey intervals ($\alpha=0.05$) for the proposed approach does not overlap that for another approach. Hence, the proposed approach has a higher classification accuracy rate than the other approaches. Therefore, the proposed approach significantly outperforms the GASVM-based approach, the PCA-based approach, SFFS-based approach, the approach that uses only textural features, the approach that uses only morphological features, and the approach that uses all features. According to Figs. 6 and 7 and Tables 3, 4, and 5, the proposed CAD system for diagnosing ultrasound images of breast tumors is very accurate.

Conclusion

To increase the rate of survival of breast cancer patients, early detection and treatment are very important. However, early detection depends on a rapid and highly accurate diagnostic tool. Following the rapid development of ultrasound technology in recent years, ultrasound has become one of the major imaging modalities for diagnosing breast

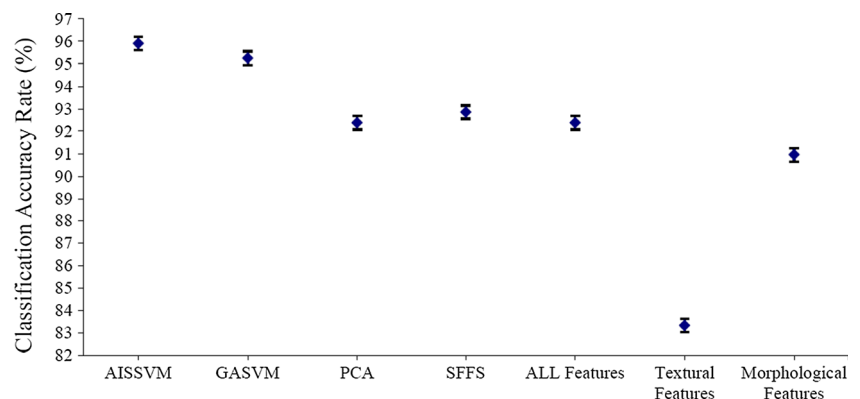
tumors. Although the accuracy of the ultrasound diagnostic method is controversial owing to the overlap between benignancy and malignancy in ultrasound images, it is becoming increasingly popular today because of its safety, convenience, and low cost.

In this study, textural and morphological features of images of breast tumors were used with SVM to distinguish benign tumors from malignant ones. Initially, breast tumors were classified using all 30 features of the images of the tumors. However, some features may be unnecessary for the classification process, increasing feature extraction time and even reducing classification accuracy. Therefore, feature selection is an important step in maximizing classification accuracy and reducing computational complexity.

The tuning of the parameters of an SVM is critical to increasing the accuracy of the proposed method. In many investigations, the grid search process has been used to find the good parameters of an SVM. However, the approach is time-consuming and may not find favorable results. Accordingly, to increase the accuracy of the proposed CAD system, an AIS algorithm was used herein to find the good parameters in the SVM and the significant features of the images of breast tumors.

In this study, eight significant features and near-optimal parameters of the SVM are found by using the AIS process to classify breast tumors in ultrasound images. From the experimental results, the accuracy, sensitivity, specificity, PPV, and NPV of the proposed CAD system are 96.67, 96.67, 96.67, 95.60, and 97.48 %, respectively. The ROC area index, A_z , is 0.9827. The high sensitivity of the proposed CAD system results in its high probability of detecting malignant tumors.

Fig. 7 Means plot and Tukey–Kramer HSD confidence intervals of proposed approach and other approaches



Moreover, the high PPV and NPV reveal that the CAD system can reduce the number of biopsies required to identify benign lesions. Restated, the proposed CAD system is an accurate diagnostic tool for identifying breast tumors. It can provide information that helps physicians prevent misdiagnosis. Although the AISSVM approach increased the capability of the proposed CAD system to distinguish between benign and malignant breast lesions, some lesions still have similar characteristics that cause the CAD system to fail. In the future, we may combine other kinds of features to decrease the FNs and FPs.

Acknowledgements The first author of this paper is grateful to the Ministry of Science and Technology of the Republic of China (Taiwan) for financially supporting this research [grant number NSC99-2221-E-182-041]. The second author is grateful to the Ministry of Science and Technology of the Republic of China (Taiwan) and the Linkou Chang Gung Memorial Hospital for financially supporting this research [grant numbers MOST103-2410-H-182-006 and CARPD3B0012].

References

1. Siegel R, Naishadham D, Jemal A: Cancer statistics, 2013. *CA Cancer J Clin* 63(1):11–30, 2013
2. Sivaramakrishna R, Gordon R: Detection of breast cancer at a smaller size can reduce the likelihood of metastatic spread: a quantitative analysis. *Acad Radiol* 4(1):8–12, 1997
3. Chen DR, Hsiao YH: Computer-aided diagnosis in breast ultrasound. *J Med Ultrasound* 16(1):46–56, 2008
4. Zhou Y: Ultrasound diagnosis of breast cancer. *J Med Imaging Health Inform* 3(2):157–170, 2013
5. Lodwick GS: Computer diagnosis in radiology. *J Mich State Med Soc* 61:1239–1242, 1962
6. Drukker K, Sennett CA, Giger ML: Automated method for improving system performance of computer-aided diagnosis in breast ultrasound. *IEEE Trans Med Imaging* 28(1):122–128, 2009
7. Chang RF, Wu WJ, Moon WK, Chen DR: Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. *Breast Cancer Res Treat* 89(2):179–185, 2005
8. Chen DR, Huang YL, Lin SH: Computer-aided diagnosis with textural features for breast lesions in sonograms. *Comput Med Imaging Graph* 35(3):220–226, 2011
9. Moon WK, Chang SC, Huang CS, Chang RF: Breast tumor classification using fuzzy clustering for breast elastography. *Ultrasound Med Biol* 37(5):700–708, 2011
10. Chen ST, Hsiao YH, Huang YL, Kuo SJ, Tseng HS, Wu HK, Chen DR: Comparative analysis of logistic regression, support vector machine and artificial neural network for the differential diagnosis of benign and malignant solid breast tumors by the use of three-dimensional power Doppler imaging. *Korean J Radiol* 10(5):464–471, 2009
11. Tseng HS, Wu HK, Chen ST, Kuo SJ, Huang YL, Chen DR: Speckle reduction imaging of breast ultrasound does not improve the diagnostic performance in morphology-based CAD System. *J Clin Ultrasound* 40(1):1–6, 2012
12. Tan T, Platel B, Huismam H, Sánchez CI, Mus R, Karssemeijer N: Computer-aided lesion diagnosis in automated 3-D breast ultrasound using coronal spiculation. *IEEE Trans Med Imaging* 31(5):1034–1042, 2012
13. Wu WJ, Lin SW, Moon WK: Combining support vector machine with genetic algorithm to classify ultrasound breast tumor images. *Comput Med Imaging Graph* 36(8):627–633, 2012
14. Han J, Kamber M: *Data Mining: Concepts and Techniques*. Morgan Kaufmann, San Francisco, 2007
15. Montani S: Exploring new roles for case based reasoning in heterogeneous AI system for medical decision support. *Appl Intell* 28(3):275–285, 2008
16. Andrews PS, Timmis J: *Inspiration for the next generation of artificial immune systems*. Artificial Immune Systems. Springer, Berlin, 2005
17. Latifoğlu F, Kodaz H, Kara S, Güneş S: Medical application of artificial immune recognition system (AIRS): diagnosis of atherosclerosis from carotid artery Doppler signals. *Comput Biol Med* 37(8):1092–1099, 2007
18. Lin SW, Chen SC: Parameter tuning, feature selection and weight assignment of features for case-based reasoning by artificial immune system. *Appl Soft Comput* 11(8):5042–5052, 2011
19. Kuo RJ, Tseng WL, Tien FC, Warren L: Application of an artificial immune system-based fuzzy neural network to a RFID-based positioning system. *Comput Ind Eng* 63(4):943–956, 2012
20. Chen GY, Xie WF: Pattern recognition with SVM and dual-tree complex wavelets. *Image Vis Comput* 25(6):960–966, 2007
21. Huang YL, Wang KL, Chen DR: Diagnosis of breast tumors with ultrasonic analysis using support vector machines. *Neural Comput & Applic* 15(2):164–169, 2006
22. Chang RF, Wu WJ, Moon WK, Chou YH, Chen DR: Support vector machines for diagnosis of breast tumors on US images. *Acad Radiol* 10(2):189–197, 2003
23. Wang XY, Yang HY, Cui CY: An SVM-based robust digital image watermarking against desynchronization attacks. *Signal Process* 88(9):2193–2205, 2008
24. Wong WT, Hsu SH: Application of SVM and ANN for image retrieval. *Eur J Oper Res* 173(3):938–950, 2006
25. Vapnik VN: *The Nature of Statistical Learning Theory*. Springer, New York, 1995
26. Deguchi K, Izumitani T, Hontani H: Detection and enhancement of line structures in an image by anisotropic diffusion. *Pattern Recogn Lett* 23(12):1399–1405, 2002
27. Czerwinski RN, Jones DL, O'Brien WD: Detection of lines and boundaries in speckle images—application to medical ultrasound. *IEEE Trans Med Imaging* 18(2):126–136, 1999
28. Blayvas I, Bruckstein A, Kimmel R: Efficient computation of adaptive threshold surfaces for image binarization. *Pattern Recognit* 39(1):89–101, 2006
29. Zhou H, Yuan Y, Lin F, Liu: Level set image segmentation with Bayesian analysis. *Neurocomputing* 71(10):1994–2000, 2008
30. Fabiszewska E, Grabska I, Jankowska K, Wesolowska E, Bulski W: Comparison of results from quality control of physical parameters and results from clinical evaluation of mammographic images for the mammography screening facilities in Poland. *Radiat Prot Dosim* 147(1–2):206–209, 2011
31. Liao Y, Fang SC, Nuttle HLW: A neural network model with bounded-weights for pattern classification. *Comput Oper Res* 31(9):1411–1426, 2004
32. Lin HT, Lin CJ: A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. *Neural Comput* 3:1–32, 2003
33. Müller KR, Mike S, Rätsch G, Tsuda K, Schölkopf B: An introduction to kernel-based learning algorithms. *IEEE Trans Neural Netw* 12(2):181–201, 2001
34. Chang YW, Hsieh CJ, Chang KW, Ringgaard M, Lin CJ: Training and testing low-degree polynomial data mappings via linear SVM. *J Mach Learn Res* 11:1471–1490, 2010
35. Pardo M, Sberveglieri G: Classification of electronic nose data with support vector machines. *Sensors Actuators B Chem* 107(2):730–737, 2005
36. Bergstra J, Bengio Y: Random search for hyper-parameter optimization. *J Mach Learn Res* 13(1):281–305, 2012

37. Li Z, He C: Optimal scheduling-based RFID reader-to-reader collision avoidance method using artificial immune system. *Appl Soft Comput* 13(5):2557–2568, 2013
38. Lin SW, Ying KC: Minimizing makespan in a blocking flowshop using a revised artificial immune system algorithm. *Omega* 41(2):383–389, 2013
39. Shang R, Qi L, Jiao L, Stolkin R, Li Y: Change detection in SAR images by artificial immune multi-objective clustering. *Eng Appl Artif Intell* 31:53–67, 2014
40. Ying KC, Lin SW: Efficient wafer sorting scheduling using a hybrid artificial immune system. *J Oper Res Soc* 65(2):169–179, 2014
41. Chang CC, Lin CJ: LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2(3):1–27, 2011
42. Salzberg SL: On comparing classifiers: pitfalls to avoid and a recommended approach. *Data Min Knowl Disc* 1(3):317–328, 1997