

Medical image classification using genetic-algorithm based fuzzy-logic approach

Du-Yih Tsai
Yongbum Lee
Masaru Sekiya
Masaki Ohkubo

Niigata University
Department of Radiological Technology
School of Health Sciences, Faculty of Medicine
2-746 Asahimachi-dori
Niigata City Niigata
951-8518, Japan
E-mail: tsai@clg.niigata-u.ac.jp

Abstract. In this paper we present a genetic-algorithm-based fuzzy-logic approach for computer-aided diagnosis scheme in medical imaging. The scheme is applied to discriminate myocardial heart disease from echocardiographic images and to detect and classify clustered microcalcifications from mammograms. Unlike the conventional types of membership functions such as trapezoid, triangle, S curve, and singleton used in fuzzy reasoning, Gaussian-distributed fuzzy membership functions (GDMFs) are employed in the present study. The GDMFs are initially generated using various texture-based features obtained from reference images. Subsequently the shapes of GDMFs are optimized by a genetic-algorithm learning process. After optimization, the classifier is used for disease discrimination. The results of our experiments are very promising. We achieve an average accuracy of 96% for myocardial heart disease and accuracy of 88.5% at 100% sensitivity level for microcalcification on mammograms. The results demonstrated that our proposed genetic-algorithm-based fuzzy-logic approach is an effective method for computer-aided diagnosis in disease classification. © 2004 SPIE and IS&T. [DOI: 10.1117/1.1786607]

1 Introduction

Research in computer-aided diagnosis (CAD) is a rapidly growing, dynamic field with new computer techniques, new imaging modalities, and new interpretation tasks. CAD is defined as a diagnosis made by a radiologist who uses the output from a computerized analysis of medical images as a second opinion in detecting lesions, assessing extent of disease, and making diagnostic decisions.¹ So far most CAD papers have involved either mammograms^{2–10} or chest radiographs.^{11–17} Recent reports show that CAD research has extended to other fields such as echocardiography¹⁸ and colonography.^{19,20}

In this paper, we present a generalized CAD scheme based on our previously reported CAD scheme.¹⁸ The proposed CAD scheme, containing four stages: image prepro-

cessing, feature extraction, classifier training, and classification, can be applied to various imaging modalities and diseases with minor modification. In our system, we basically employ fuzzy logic for classification. Unlike the conventional types of fuzzy membership functions such as triangle and trapezoid, Gaussian-distributed membership functions (GDMFs) are used in the system. The GDMFs are initially generated using various features obtained from image data sets. Subsequently, the shapes of the GDMFs are optimized using a genetic-algorithm (GA) learning process. After optimization, the system is used for discrimination of disease. To our knowledge, this is the first time such a CAD system has been described using the GA-based fuzzy approach. In the present study, we apply our CAD method to discriminate myocardial heart disease from echocardiographic images and to detect and classify clustered microcalcification from mammograms. The performance of our CAD method is evaluated in terms of accuracy, sensitivity, and specificity.

2 Methods

2.1 Fuzzy Membership Functions and Fuzzy Rules

The major components of the fuzzy-logic decision-making system are fuzzy sets, fuzzy membership functions, and fuzzy rules. Each fuzzy set has a corresponding fuzzy membership function. The value of the membership function ranges from 0 to 1 and can be considered a degree of truth. The current study uses simplified fuzzy rules as follows:

rule i : If x_1 is c_{i1} and \dots , and x_M is c_{iM} ,
then y is w_i , (1)

where i ($i=1,2,\dots,M$) are rule numbers, x_1,\dots,x_M are input variables to the fuzzy reasoning, y is the output, c_{i1},\dots,c_{iM} are fuzzy labels corresponding to the input vari-

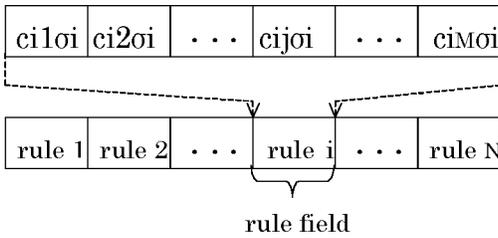


Fig. 1 Configuration of an individual consisting of N fuzzy rules. If each fuzzy rule is composed of M premise-part membership functions, a total of $M \times N$ chromosomes line up to generate an individual.

ables, and w_i is a real number of the consequent part of the fuzzy rule. While almost any type of fuzzy membership functions such as trapezoid, triangle, S curve, and singleton can be used, the GDMFs were employed in this study. The optimal shape of a membership function may vary depending on the issue being dealt with. In this study, we intended to automatically generate membership functions using various features obtained from a specific category of image data for recognition and classification. We considered that if the number of image data base is large, Gaussian distribution function can be used to appropriately describe image's features. These are the main reasons why GDMFs were used.

Consider a specific feature value x that can be measured from an image. Let the mean value of x from a set of images belonging to the same category be μ and the standard deviation of the feature values be σ . Define a fuzzy set with a GDMF having the maximum value of unity (normalized). The normalized membership function can be expressed as

$$f(x) = \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right] \quad (2)$$

When the value of x for an image is μ , the membership value should be one. When x is gradually apart from the value of μ , the membership value should become smaller. If the number of sample images is limited, the value of σ may not accurately reflect the characteristic of all images of the same category. Therefore, in this study we propose a method to use a GA at a training phase for determining the optimal membership function.

2.2 Optimization of the Fuzzy Membership Function Using Genetic Algorithms

As shown in Eq. (3), by varying the value of standard deviation σ with a coefficient c , the shape of the GDMF can be optimized using GA-based learning

$$f(x) = \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{c\sigma} \right)^2 \right]. \quad (3)$$

If N fuzzy rules are present and each fuzzy rule consists of M premise-part (antecedent) membership functions in a fuzzy system, a total of $M \times N$ chromosomes (or genes) line up to generate an individual (see Fig. 1). As shown in Fig. 2(a), it is assumed that an individual comprises two

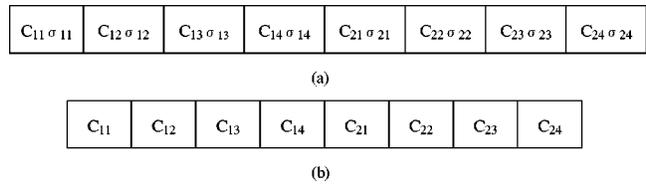


Fig. 2 (a) Chromosomes of an individual presented by two fuzzy rules. Each fuzzy rule consists of four GDMFs with various coefficients c_{ij} and standard deviations σ_{ij} ($i=1, 2$, and $j=1, 2, 3, 4$). (b) Each individual consists of eight coefficients.

fuzzy rules. Each fuzzy rule consists of four GDMFs with various coefficients c_{ij} and standard deviations σ_{ij} ($i=1, 2$, and $j=1, 2, 3, 4$). The $c_{ij}\sigma_{ij}$ are optimized by the GA. Since the values of $\sigma_{11}, \sigma_{12}, \sigma_{13}, \sigma_{14}, \sigma_{21}, \sigma_{22}, \sigma_{23}$, and σ_{24} are constant when the data set is determined, in practice only the coefficients are used in the phase of optimization. As shown in Fig. 2(b) the string having eight variables is treated as an individual. The shape of the consequent-part membership functions employed in the present study is a right isosceles triangle with the maximum value of unity (normalized).

A simple GA that has given good results in a variety of engineering problems is composed of three operators: selection (reproduction), crossover, and mutation. These operators are implemented by performing the basic tasks of copying strings, exchanging portions of strings, and generating random numbers. The GA begins by randomly generating a population of individuals (strings). Each individual is a possible solution to the optimization problem. In general, the population is initialized at random to a bit string of 0's and 1's. The selection operation mimics the natural selection process by selecting which individuals will be used to create a new generation, where the fittest individuals reproduce most often. Usually a fitness function is used to rate individuals in terms of how good they are in solving the optimization problem. The crossover operation refers to the exchange of substrings of two individuals to generate two new individuals. The third operator, mutation, enhances the ability of the GA to find near-optimal solutions. Mutation is the occasional alternation of the binary value at a particular string position.

The procedure of how to optimize the membership functions using the GA is described as follows.

1. A total of 200 individuals are randomly determined. Each individual (string) consists of eight variables as shown in Fig. 2(b).
2. The fitness values of each individual are computed.
3. The individuals are rearranged in ranking order according to the fitness value.
4. Out of 200 randomly generated initial strings, the first 140 high-ranked parent strings are selected for crossover and mutation to produce the same number of children strings. Next, the first high-ranked 60 parent strings together with the 140 children are treated as the strings of the next generation. Single-point crossover is used in this study. It should be emphasized here that the crossover is performed for each variable having eight-digit binary number, namely, the crossover are simultaneously performed at eight positions. A probability of 10% is used for mutation.

5. Operations 2–4 are treated as one generation. The number of generation employed is 50. On the completion of 50 iterations (generations), the individual having the highest ranking is finally selected, the eight variables corresponding to the eight coefficients of GDMFs are considered as optimal, and the training process is finished, followed by the classification phase.

2.3 Defuzzification

While many kinds of defuzzification approaches could be used, the MIN-MAX compositional rule of fuzzy inference is employed in this study.²¹ However, a modified “height method” rather than center-of-gravity method is employed for defuzzification. The MIN-MAX compositional height method is described as follows.

First, assuming that $\mu_{\text{nor}}(Q_1)$, $\mu_{\text{nor}}(Q_2)$, $\mu_{\text{nor}}(Q_3)$, and $\mu_{\text{nor}}(Q_4)$ are the respective fuzzy GDMF values for the normal case (or benignancy), and $\mu_{\text{ab}}(Q_1)$, $\mu_{\text{ab}}(Q_2)$, $\mu_{\text{ab}}(Q_3)$, and $\mu_{\text{ab}}(Q_4)$ are the respective fuzzy GDMF values for the abnormal case (or malignancy). Then, the minimum value of $\mu_{\text{nor}}(Q_1)$, $\mu_{\text{nor}}(Q_2)$, $\mu_{\text{nor}}(Q_3)$ and $\mu_{\text{nor}}(Q_4)$, and that of $\mu_{\text{ab}}(Q_1)$, $\mu_{\text{ab}}(Q_2)$, $\mu_{\text{ab}}(Q_3)$ and $\mu_{\text{ab}}(Q_4)$ are given by

$$\mu_{\text{nor}} = \text{MIN}[\mu_{\text{nor}}(Q_1), \mu_{\text{nor}}(Q_2), \mu_{\text{nor}}(Q_3), \mu_{\text{nor}}(Q_4)] \quad (4)$$

and

$$\mu_{\text{ab}} = \text{MIN}[\mu_{\text{ab}}(Q_1), \mu_{\text{ab}}(Q_2), \mu_{\text{ab}}(Q_3), \mu_{\text{ab}}(Q_4)]. \quad (5)$$

This is usually called MIN (minimum) operation. Finally, the larger value is taken from the two. It is called MAX (maximum) operation and is given by

$$\mu_{\text{norUab}} = \text{MAX}[\mu_{\text{nor}}, \mu_{\text{ab}}]. \quad (6)$$

The classification scheme in our applications is

$$\text{if } \mu_{\text{norUab}} = \mu_{\text{nor}} \rightarrow \text{normal case (or benignancy)}, \quad (7)$$

$$\text{if } \mu_{\text{norUab}} = \mu_{\text{ab}} \rightarrow \text{abnormal case (or malignancy)}. \quad (8)$$

In the case of $\mu_{\text{nor}} = \mu_{\text{ab}}$, the classification becomes uncertain and is regarded as misclassification in this study.

As shown in Fig. 3, suppose that the minimum values for rules 1 and 2 are 0.4 and 0.2 (absolute value), respectively. Since 0.4 (fuzzy rule 1 for normal case) is greater than 0.2 (fuzzy rule 2 for abnormal case), the output of the fuzzy inference is “it is a normal case.” The result is subsequently compared to the teacher signal, which is already known through the diagnosis made by the highly trained physician. If the result is identical to the diagnostic result made by the physician, then the case is correctly classified.

3 Applications

3.1 Classification of Myocardial Heart Disease from Ultrasonic Images

Echocardiography is one of the best tools for diagnosing cardiomyopathy. However, since the clinical interpretation process and the results largely depend on the physician’s subjective point of view and experience, the criteria of di-

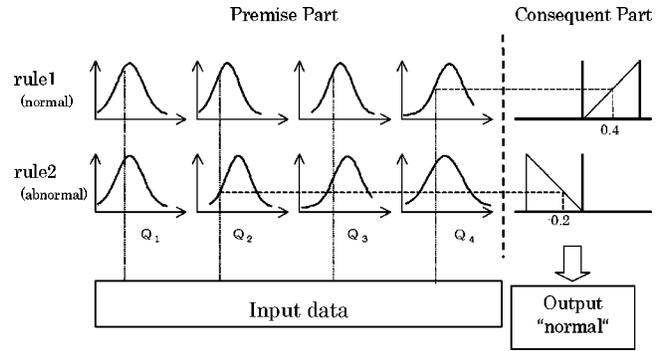


Fig. 3 Illustration of defuzzification employed in our CAD scheme. Eight Gaussian-distributed membership functions are used in the CAD scheme. The MIN-MAX compositional rule of fuzzy inference is employed in this study.

agnosis are somewhat indeterminate. If a computer-aided technique, which provides a second opinion for the physician, can be developed, then this subjectivity may be reduced and in turn the accuracy in diagnosis is expected to increase.

3.1.1 Data set

In this application, a total of 90 samples of echocardiographic images from 45 subjects (two sample images per subject: an end-diastole image and an end-systole image) were collected at the Hospital of Gifu University School of Medicine. The images were captured from a Toshiba SSH-160A device with a 2.5 MHz (central frequency) transducer. The state when the ventricles of the heart are maximally filled just before the heart contracts is called *end diastole*. The other extreme state—when the ventricles are maximally emptied—is called *end systole*. Hence, one cardiac cycle can be represented by these two extreme states to give a fair idea of the heart’s condition. Physicians usually use the end-diastole and end-systole images in the diagnostic process. Of the 45 subjects, 23 subjects were diagnosed in advance by a highly trained (15 year experienced) clinician as normal and 22 were as abnormal (dilated cardiomyopathy or hypertrophic cardiomyopathy). Each image was digitized at the resolution of 256×256 pixels. Since the original echocardiographic images have 64 gray levels, the scanned images were quantized to the same level. In our previous studies we noted that the use of composite images could provide higher recognition rate compared to that of individual images at end systole and end diastole.^{22,23} Therefore, in the present application we used composite images $h(x,y)$, which are obtained as follows:

$$h(x,y) = \max[m(x,y), n(x,y)], \quad (9)$$

where $m(x,y)$ and $n(x,y)$ refer to the images at end-systole and end-diastole states, respectively. Figure 4 shows an example of the normal case. The images at end-systole and end-diastole states are shown in Figs. 4(a) and 4(b), and the composite image is shown in Fig. 4(c).

3.1.2 Feature extraction

We generated a gray-level co-occurrence matrix from each of composite images. The gray-level co-occurrence matrix

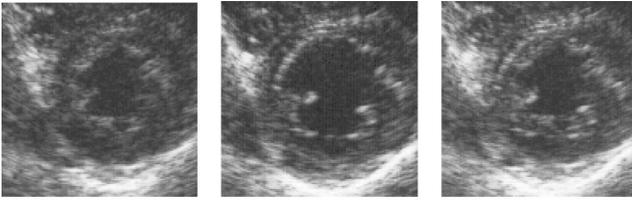


Fig. 4 Examples of echocardiograms: (a) end-systole, (b) end-diastole, and (c) composite images. The composite image $h(x,y)$ is obtained using $h(x,y)=\max[m(x,y),n(x,y)]$, where $m(x,y)$ and $n(x,y)$ refer to the images at end-systole and end-diastole states, respectively.

is a matrix used to express the correlation of spatial location and gray-level distribution of an image. From it, the local variation of gray levels on an image can be statistically investigated, which in turn enable us to know the manner of change in gray level as a whole. In the current application, we used the following conditions to generate gray-level co-occurrence matrices: (a) Number of gray levels: A co-occurrence matrix of 64×64 size can be obtained from a 64 gray-level image. In order to save computation time, we compress the gray level to 16 in this application, since our experiments showed that the matrix size of 16×16 was adequate. (b) Direction: In general the gray-level co-occurrence matrices from 0° , 45° , 90° , and 135° directions are used. Since the differences of the results from the four directions are insignificant in our preliminary investigation, only the direction of 0° was used in the study. (c) Distance: The length of five pixels was used, because we experimentally found that the value to be optimal. From the generated gray-level co-occurrence matrices, a total of 14 statistical features for each image can be calculated.²⁴ We experimentally evaluated all of these features on their ability to discriminate between normal and abnormal cases. Of the 14 features, we found that the following 4 have the most powerful discrimination ability as texture features of the composite images.

Angular second moment (Q_1):

$$Q_1 = \sum_{f_1} \sum_{f_2} P(f_1, f_2)^2, \quad (10)$$

where f_1 and f_2 are the gray-level values of two pixels at different locations, $P(f_1, f_2)$ is the probability obtained by dividing the number of (f_1, f_2) pairs in the matrix by the total number of occurrence in the co-occurrence matrix.

Contrast (Q_2):

$$Q_2 = \sum_{f_1} \sum_{f_2} (f_1 - f_2)^2 P(f_1, f_2). \quad (11)$$

Correlation (Q_3):

$$Q_3 = \frac{\sum_{f_1} \sum_{f_2} f_1 f_2 P(f_1, f_2) - \mu_1 \mu_2}{\sigma_1 \sigma_2}, \quad (12)$$

where,

$$\mu_1 = \sum_{f_1} f_1 \sum_{f_2} P(f_1, f_2),$$

$$\mu_2 = \sum_{f_2} f_2 \sum_{f_1} P(f_1, f_2),$$

$$\sigma_1^2 = \sum_{f_1} (f_1 - \mu_1)^2 \sum_{f_2} P(f_1, f_2),$$

$$\sigma_2^2 = \sum_{f_2} (f_2 - \mu_2)^2 \sum_{f_1} P(f_1, f_2).$$

Entropy (Q_4):

$$Q_4 = - \sum_{f_1} \sum_{f_2} P(f_1, f_2) \ln[P(f_1, f_2)]. \quad (13)$$

The feature Q_1 represents the uniformity of textures. The more the number of specific gray-level pairs appears, the higher the value of Q_1 . The feature Q_2 represents the mean of the gray-level differences of various gray-level pairs $(f_1 - f_2)$. The value of Q_2 increases with the increase of image contrast. The feature Q_3 represents pattern periodicity of specific directions. The feature Q_4 represents the contrary property of Q_1 . It is still not very clear that these features completely describe what kinds of physical properties of the tissue, respectively. However, in the sense that different tissue or different quality of image provides different feature values, these statistic values can be used to represent texture features of echocardiographic images.

3.1.3 Classification using GA-based fuzzy-logic approach

We randomly selected 12 normal and 12 abnormal echocardiograms from the 45 subjects and called them ‘‘set A,’’ and called the remaining (11 normal and 10 abnormal) hearts ‘‘set B.’’ We used set B as learning data for training to obtain optimal membership functions, and used set A as test data for classification, and vice versa. The classification results were then averaged. it should be noted that the GD-MFs were initially generated by using μ and σ obtained from learning data for learning and optimization phase. In other words, no learning data were used for testing in this study.

A two-step fitness function in the selection operation was employed to attempt increasing the accuracy of classification. Fitness values are computed by the fitness functions shown as follows:

$$f_1 = \frac{n - m}{n} \quad (14)$$

and

$$f_2 = \sum_{i=1}^n (d_{1i} - d_{2i})^2, \quad (15)$$

Table 1 Performance comparison for various methods applied to myocardial heart disease.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
BP-NN	82.1	83.0	80.8
GA-NN	88.7	91.7	86.4
Fuzzy	91.4	91.7	91.3
GA-fuzzy	95.9	91.7	100

where n and m refer to the number of learning data and the number of misclassification, and d_{1i} and d_{2i} represent the minimum values obtained from fuzzy rules 1 and 2, respectively. When $m=n$, that is, all of the learning data are misclassified, then $f_1=0$. On the contrary, when $m=0$, namely, all learning data are correctly classified during the training phase, then $f_1=1$. Moreover, when the difference between d_{1i} and d_{2i} increases, then the value of f_2 becomes greater. The fitness values obtained from Eqs. (14) and (15) are used for ranking the individuals. The ranking process is as follows: (a) The individuals are ranked according to the fitness value of f_1 . (b) Those individuals having the same value of f_1 are further ranked according to the value of f_2 .

After completion of learning, the individual with the highest fitness value is selected and the eight variables of this individual are considered as the optimal coefficients. By using the optimal coefficients, the width (cs) of each membership function is determined. For example, when set A was used as test data for classification, the mean value and the standard deviation of Q_1 for normal case were 0.137 and 0.037. The optimal coefficient c was 1.21. Finally the width of the corresponding membership function was 0.045.

3.1.4 Results and discussion

We evaluate the performance of the proposed method in terms of sensitivity, specificity, and overall accuracy. Sensitivity (true positive fraction) is the probability that a diagnostic test is positive, given that the person has the disease. Specificity (true negative fraction) is the probability that a diagnostic test is negative, given that the person does not have the disease. Overall accuracy is the probability that a diagnostic test is correctly performed. The three indices are defined^{13,14} as follows:

$$\text{sensitivity} = \frac{TP}{TP + FN}, \quad (16)$$

$$\text{specificity} = \frac{TN}{TN + FP}, \quad (17)$$

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (18)$$

where TP, FP, TN, and FN refer to true positive, false positive, true negative and false negative, respectively.

Table 1 shows the classification rates for performing the GA-based fuzzy method (GA-fuzzy method). In order to

demonstrate the effectiveness of the proposed method, we also provide the results of neural network with back-propagation learning method (BP-NN method), neural network with GA learning method (GA-NN method), and fuzzy method (without GA operation) for comparison. The sensitivity rates for BP-NN/GA-NN/fuzzy/GA-fuzzy methods were 83.0%/91.7%/91.7%/91.7%, respectively. Except for the BP-NN method, the sensitivity rates for the other three methods are comparable. The BP-NN method has lower sensitivity because of the number of FN cases. Specificity rates for BP-NN/GA-NN/fuzzy/GA-fuzzy methods were 80.8%/86.4%/91.3%/100%, respectively. The GA-fuzzy method provided the highest specificity, followed by fuzzy, GA-NN, and BP-NN methods. The table reveals that none of the FP cases misclassified. The results indicated that the GA-fuzzy method was effective. The overall accuracies for BP-NN/GA-NN/fuzzy/GA-fuzzy methods were 82.1%/88.7%/91.4%/95.9%, respectively. The results showed the superiority of the GA-based fuzzy method.

Regarding fuzzy-logic-based methods (fuzzy and GA-fuzzy methods), the employment of GA for optimization of GDMFs could achieve better classification rates, a 4.5% increase in accuracy. The results suggest that our proposed GA-fuzzy method for determining the GDMFs is useful, especially in the case of small number of training data available.

Furthermore, in order to reveal the merit of the use of GDMFs, we also evaluated the performance of using a triangular-type membership function. Our results showed that the average classification rate was 85%, an 11% decrease in accuracy. This inferiority may be due to the reason that those test data having feature values larger than $\mu+3\sigma$ could not be classified using the triangular-type membership function and resulted in misclassification.

To the best of our knowledge, so far, there are no other reports dealing with echocardiography classification. Therefore, we do not have references for the comparative evaluation of our results on the specific database used. However, we believe that the performance of our proposed method is satisfactory and the method is clinically useful for computer-aided diagnosis of cardiomyopathy.

3.2 Classification of Clustered Microcalcifications on Mammograms

Breast cancer is a leading cause of cancer deaths among women in many parts of the world. Mammography is known as the most effective modality for early detection of breast cancer, such as tumors and microcalcifications. Detection of microcalcifications is especially related to early detection of breast cancer because those are considered to be suspicious observations in the early stages of cancer. Thus, development of a CAD system is strongly desirable.

Several methods for classifications of microcalcifications in mammography have been reported.²⁵⁻²⁸ In these papers, methods using neural networks²⁵⁻²⁷ and using GAs²⁸ have been proposed. Recently, there has been research on the application of fuzzy logic to the CAD in medical images.^{18,29-33} The reason lies in the fact that a lot of information used for interpretation of medical images is fuzzy. In this application, we use our proposed GA-fuzzy

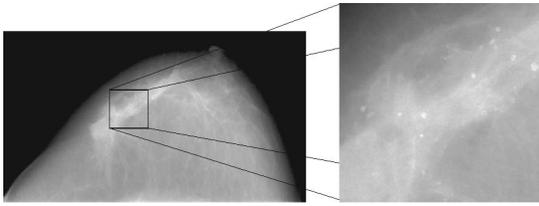


Fig. 5 An example of mammogram with microcalcifications. The mammograms used in this study was digitized at a spatial resolution of 50 μm sampling distance with an eight-bit density resolution.

method to discriminate between benign and malignant clustered microcalcifications.

3.2.1 Data set

In this application, we used the mammographic database provided by the Mammographic Image Analysis Society (MIAS) in the United Kingdom. Each of the mammograms in the MIAS database was digitized at a spatial resolution of 50 μm sampling distance with an eight-bit density resolution. Figure 5 shows an example of region of interest (ROI) with microcalcifications.

The MIAS database consists of more than 300 images. The images of the database can be categorized as normal, microcalcification, mass, architectural distortion, and asymmetry. Of the images, 25 images are obviously indicated as microcalcifications. The aim of our study is to develop a CAD system for discrimination between benign and malignant clustered microcalcifications. Therefore, we employed only 25 images including 13 benign and 12 malignant microcalcifications in the current study.

3.2.2 Feature extraction

Prior to feature extraction from the ROI images used for classification, we employed mathematical morphology to detect microcalcifications from ROIs.³⁴ After detection, four features including microcalcification number (Num), mean area (Area), mean circularity (Cir), and mean minimum distance (Dis) were used for classifying clustered microcalcifications. The main reason for using these features is that radiologists' interpretation for classifying clustered microcalcifications is generally based on the visual information distribution such as sizes and shapes of clustered microcalcifications. These features are expressed as follows:

$$\text{Num} = n, \tag{19}$$

$$\text{Area} = \frac{1}{n} \sum_{i=1}^n A_i, \tag{20}$$

$$\text{Cir} = \frac{1}{n} \sum_{i=1}^n C_i, \tag{21}$$

$$\text{Dis} = \frac{1}{n} \sum_{i=1}^n D_i, \tag{22}$$

where n is the number of isolated candidate regions of microcalcification within a ROI image. A_i and C_i are the area

Table 2 Performance comparison for various methods applied to microcalcifications.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
BP-NN	84.6	100	69.2
GA-NN	76.9	100	53.8
Fuzzy	65.4	100	30.8
GA-fuzzy	88.5	100	76.9

and the circularity of the i th candidate region, respectively. D_i is the distance from i th candidate to the nearest candidate.

3.2.3 Classification using GA-based fuzzy-logic approach

Basically, the GA-based fuzzy method used in this application is the same as that used in the preceding application. In this application, eight GDMFs are generated from four features (Num, Area, Cir, and Dis) for each of two categories (benignancy, malignancy).

As shown in Fig. 3, rule 1 (normal) and rule 2 (abnormal) correspond to benignancy rule and malignancy rule, respectively, in this application. We used GA at training phase for determining the optimal membership functions by varying the values of coefficient c . In training phase, ten benign cases and ten malignant cases were used as learning data. The remaining five cases were used for classification as unknown images. In order to obtain results of high reliability, a total of 3146 combinations for classification of malignancy and a total of 4356 combinations for classification of benignancy were used, respectively. That is, when a malignant case was selected and used for classification, the number of combinations for training and classification was ${}_{11}C_{10} \times {}_{13}C_{10} = 11 \times 286 = 3146$. Similarly, when a benign case was selected and used for classification, the number of combinations for training and classification was ${}_{12}C_{10} \times {}_{12}C_{10} = 66 \times 66 = 4356$.

3.2.4 Results and discussion

Table 2 shows the classification rates for performing our proposed method. We also compare the method with other three techniques (BP-NN, GA-NN, fuzzy). The specificity rates at 100% sensitivity level for BP-NN/GA-NN/fuzzy/GA-fuzzy methods were 69.2%/53.8%/30.8%/76.9%, respectively. The accuracies at 100% sensitivity level for BP-NN/GA-NN/fuzzy/GA-fuzzy methods were 84.6%/76.9%/65.4%/88.5%, respectively. The results showed the superiority of the proposed GA-fuzzy method. Moreover, in order to show the statistical significance of the proposed method, receiver operating characteristic (ROC) analysis was made. ROC analysis is the standard approach to evaluate the sensitivity and specificity of diagnostic procedures. The ROC analysis estimates a curve, namely, ROC curve, which describes the inherent tradeoff between sensitivity and specificity of a diagnostic test.^{35,36} Figure 6 illustrates the ROC curves. The differences in the performances among the four techniques were estimated by comparing four A_z values (the areas under the ROC curves). The A_z

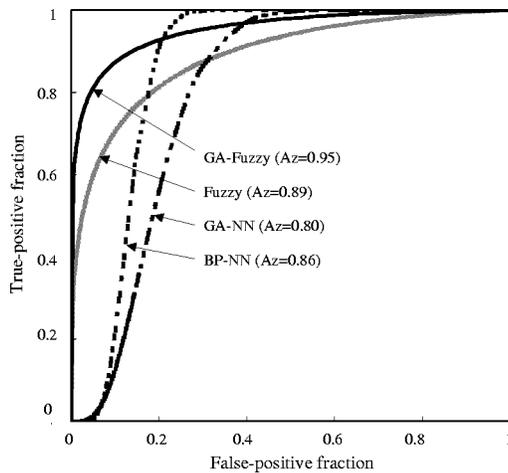


Fig. 6 ROC curves obtained from the computer analysis of four various methods, i.e., BP-NN, GA-NN, fuzzy method, and GA-fuzzy methods. A ROC curve describes the inherent tradeoff between sensitivity and specificity of a diagnostic test. The area under an ROC curve, A_z , is a quick and accepted way of comparing the performance of different classifiers. It is the average sensitivity over all possible specificities.

value has become a particularly important metric for evaluating diagnostic procedures because it is the average sensitivity over all possible specificities.³⁵ The A_z areas for BP-NN/GA-NN/fuzzy/GA-fuzzy methods were 0.86/0.80/0.89/0.95, respectively.

Paired t test was used to validate the statistical significance of the difference in A_z values in terms of p value. The p values of GA-fuzzy and GA-NN pair, GA-fuzzy and BP-NN pair, and GA-fuzzy and fuzzy pair were 0.042, 0.16, and 0.15, respectively. These results show that our proposed method is significantly superior to the GA-NN method. However, the differences in A_z values among GA-fuzzy, BP-NN, and fuzzy are statistically insignificant. The insignificance may be due to insufficient sample images. Increase of sample images for further performance assessment is needed.

Figure 7 shows four ROI images (one malignant case and three benign cases) that had low sensitivity or low specificity by using the GA-fuzzy method. The sizes of ROI images shown in this application were reduced to 35% of actual sizes of the original images. The benign cases B7, B10, and B13 were on dense-glandular mammograms as the character of background tissue. Since microcalcifications are buried in dense regions, it is difficult to correctly extract regions of microcalcifications from dense-glandular mammograms because microcalcifications are buried in dense regions. In particular, it is difficult to visualize microcalcifications in case B13 though we do not know whether microcalcifications really exist or not, because the place of each microcalcification was not indicated in the MIAS database. We suppose that detection results of microcalcifications influence classification performance. This factor may be a reason of misclassifications in this application.

In several related studies, Chan *et al.*²⁶ reported a specificity of 39% at 100% sensitivity, Hara *et al.*²⁷ reported a specificity of 94% at 79% sensitivity, and Chan *et al.*²⁸ re-

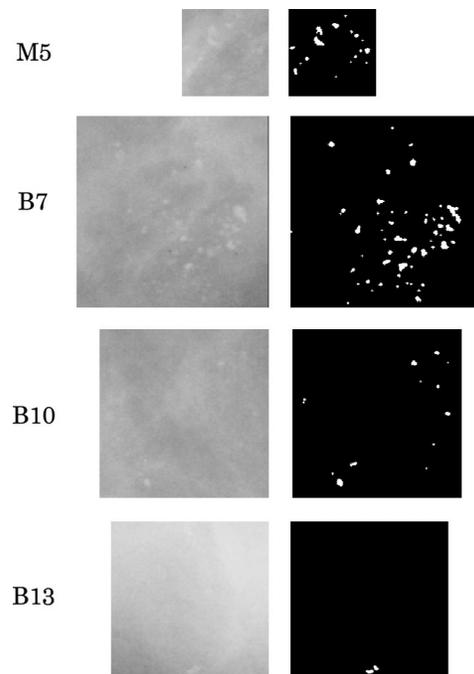


Fig. 7 Four cases that had low accuracy classified using the GA-fuzzy method. On the left are original ROI images (one malignant case M5 and three benign cases B7, B10, and B13), and on the right are the corresponding thresholded binary images. The sizes of the ROI images shown were reduced to 35% of actual sizes from the original images.

ported a specificity of 50% at 100 sensitivity. The corresponding accuracies obtained from the three reports are 70%, 86%, and 75%, respectively. The respective images used were 56, 104, and 145. Although the conditions of performance evaluation were different among these reports as well as our present study (25 images), the performance of our proposed method (an accuracy of 88.5%) is considered superior or comparable to the performance shown in the mentioned literature. The results demonstrate the usefulness of our method in the classification of clustered microcalcifications on mammograms.

However, to further improve the classification performance, it is still necessary to enhance the detection performance of microcalcifications and to investigate the existence of more powerful features and to increase sample images.

4 Conclusion

In this paper we have proposed a GA-based fuzzy approach for CAD scheme in disease classification. The proposed method was to exploit a GA-based training for optimization of membership functions. Unlike the conventional types of membership functions, Gaussian-distributed membership functions were employed. The effectiveness of our proposed method has been demonstrated through two applications, i.e., discrimination of myocardial heart disease from echocardiographic images and classification of clustered microcalcifications from mammograms. We have compared the proposed methods with other three methods, BP-NN, GA-NN, and fuzzy approaches. In the application of discrimination of myocardial heart disease, the results in terms

of accuracy have validated the superiority of the proposed method. In the application of classification of microcalcification, the ROC analysis was employed to compare the performance of the four methods and the results show the tendency for the improvement by using the proposed method. Our future works include increasing sample images for further feasibility test on the proposed method, modifying the shape of consequent-part membership function, and exploring more powerful image features.

Acknowledgments

The authors wish to thank Dr. Masaaki Tomita for providing echocardiograms with proven diagnosis and for his many useful discussions. They also thank Dr. Youichi Miyajima for his encouragement and support to this work, and the reviewers for their comments and suggestions. This work was supported in part by a Grant-In-Aid for Scientific Research (Intelligent Assistance for Diagnosis of Multi-dimensional Medical Images) from the Ministry of Education, Culture, Sports, Science and Technology.

References

- M. L. Giger, S. G. Armato III, and N. Karssemeijer, "Computer-aided diagnosis in medical imaging," *IEEE Trans. Med. Imaging* **20**, 1205–1208 (2001).
- H. Kobatake and Y. Yoshinaga, "Detection of spicules on mammogram based on skeleton analysis," *IEEE Trans. Med. Imaging* **15**, 234–245 (1996).
- C. H. Chen and G. G. Lee, "On digital mammogram segmentation and microcalcification detection using multiresolution wavelet analysis," *Graph. Models Image Process.* **59**, 349–364 (1997).
- L. M. Bruce and R. R. Adhami, "Classifying mammographic mass shapes using the wavelet transform modulus-maxima method," *IEEE Trans. Med. Imaging* **18**, 1170–1177 (1999).
- H.-P. Chan *et al.*, "Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: An ROC study," *Radiology* **212**, 817–827 (1999).
- Z. Huo, M. L. Giger, and C. E. Metz, "Effect of dominant features on neural network performance in the classification of mammographic lesions," *Phys. Med. Biol.* **44**, 2579–2595 (1999).
- Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, and C. E. Metz, "Computerized classification of benign and malignant masses on digitized mammograms: A study of robustness," *Acad. Radiol.* **7**, 1077–1084 (2000).
- N. R. Mudigonda, R. M. Rangayyan, and J. E. L. Desautels, "Detection of breast masses in mammograms by density slicing and texture flow-field analysis," *IEEE Trans. Med. Imaging* **20**, 1215–1227 (2001).
- B. Sahiner *et al.*, "Computer-aided characterization of mammographic masses: Accuracy of mass segmentation and its effects on characterization," *IEEE Trans. Med. Imaging* **20**, 1275–1284 (2001).
- S.-C. B. Lo, H. Li, Y. Wang, L. Kinnard, and M. T. Freedman, "A multiple circular path convolution neural network system for detection of mammographic masses," *IEEE Trans. Med. Imaging* **21**, 150–158 (2002).
- T. Ishida, S. Katsuragawa, T. Kobayashi, H. MacMahon, and K. Doi, "Computerized analysis of interstitial disease in chest radiographs: Improvement of geometric-pattern feature analysis," *Med. Phys.* **24**, 915–924 (1997).
- M. Carreira, D. Cabello, M. Penedo, and A. Mosquera, "Computer-aided diagnoses: Automatic detection of lung nodules," *Med. Phys.* **25**, 1998–2006 (1998).
- M. Penedo, M. Carreira, A. Mosquera, and D. Cabello, "Computer-aided diagnosis: A neural-network-based approach to lung nodule detection," *IEEE Trans. Med. Imaging* **17**, 872–880 (1998).
- A. Reeves and W. Kostis, "Computer-aided diagnosis for lung cancer," *Radiol. Clin. North Am.* **38**, 497–509 (2000).
- D. Catarious, Jr., A. Baydush, and C. Floyd, Jr., "Initial development of a computer-aided diagnosis tool for solitary pulmonary nodules," *Proc. SPIE* **4322**, 710–717 (2001).
- M. S. Brown, M. F. McNitt-Gray, J. G. Goldin, R. D. Suh, J. W. Sayre, and D. R. Aberle, "Patient-specific models for Surveillance in CT Images," *IEEE Trans. Med. Imaging* **20**, 1242–1250 (2001).
- B. V. Ginneken, S. Katsuragawa, B. M. H. Romeny, K. Doi, and M. A. Viergever, "Automatic detection of abnormalities in chest radiographs using local texture analysis," *IEEE Trans. Med. Imaging* **21**, 139–149 (2002).
- D.-Y. Tsai and Y. Lee, "Fuzzy-reasoning-based computer-aided diagnosis for automated discrimination of myocardial heart disease from ultrasonic images," *Electron. Commun. Jpn. Part 3: Fundamental Electronic Science* **85**, 1–8 (2002).
- S. B. Gokturk *et al.*, "A statistical 3-D pattern processing method for computer-aided detection of polyps in CT colonography," *IEEE Trans. Med. Imaging* **20**, 1251–1260 (2001).
- H. Yoshida and J. Nappi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," *IEEE Trans. Med. Imaging* **20**, 1261–1274 (2001).
- T. Terao, K. Asai, and M. Sugeno, *Introduction to Applied Fuzzy Systems*, Ohmsha, Tokyo (1990).
- D.-Y. Tsai and M. Tomita, "A computer-aided system for discrimination of dilated cardiomyopathy using echocardiographic images," *IEICE Trans. Fundamentals* **E-78-A**, 1649–1654 (1995).
- D.-Y. Tsai, S. Watanabe, K. Kojima, and M. Tomita, "Determination of weighting values of neural networks for discrimination of cardiomyopathy using genetic algorithms," *Med. Imaging Technol.* **15**, 557–558 (1997).
- M. Takagi and H. Shimoda, *Handbook of Image Analysis*, Tokyo University Publishing, Tokyo (1991).
- Y. C. Wu, M. T. Freedman, A. Hasegawa, R. A. Zuurbier, S. C. Lo, and S. K. Mun, "Classification of microcalcifications in radiographs of pathologic specimens for the diagnosis of breast cancer," *Acad. Radiol.* **2**, 199–204 (1995).
- H. P. Chan *et al.*, "Computerized classification of malignant and benign microcalcifications on mammograms: Texture analysis using an artificial neural network," *Phys. Med. Biol.* **42**, 549–567 (1997).
- T. Hara, A. Yamada, H. Fujita, T. Iwase, and T. Endo, "Automated classification of mammographic micro-calcifications by using artificial neural network: An ACR BI-RADS criteria," *Proc. SPIE* **4322**, 1783–1787 (2001).
- H. P. Chan *et al.*, "Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces," *Med. Phys.* **25**, 2007–2019 (1998).
- T. Katafuchi, H. Fujita, K. Asai, and T. Uehara, "Development of a computer-aided diagnosis system using fuzzy inference in ^{201}Tl exercise myocardial scintigraphy," *Jpn. J. Radiol. Technol.* **56**, 377–383 (2000).
- M. Sameti and R. K. Ward, "A fuzzy segmentation algorithm for mammogram partition," in *Digital Mammography*, International Congress Series, K. Doi, Ed., Elsevier, Amsterdam, pp. 471–474 (1996).
- M. Park, T. M. Cao, J. Jin, and L. Wilson, "Multiple classification ripple down rule and fuzzy sets in computer aided diagnosis," *Proc. of International Federation of Automatic Control (IFAC) on Modelling and Control in Biomedical Systems*, 2003 (in press).
- N. S. Iyer, A. Kandel, and M. Schneider, "Feature-based fuzzy classification for interpretation on mammograms," *Fuzzy Sets Syst.* **114**, 271–280 (2000).
- H. D. Cheng and H. Xu, "A novel fuzzy logic approach to mammogram contrast enhancement," *Inf. Sys.* **148**, 167–184 (2002).
- D.-Y. Tsai *et al.*, "A morphology-based method for automated detection of clustered microcalcifications," *Proc. of IASTED Signal Processing, Pattern Recognition, and Applications* **1**, 159–162 (2001).
- J. A. Swets, "ROC analysis applied to the evaluation of medical imaging techniques," *Invest. Radiol.* **14**, 109–121 (1979).
- C. E. Metz, "ROC methodology in radiologic imaging," *Invest. Radiol.* **21**, 720–733 (1986).



Du-Yih Tsai received his BS degree in electrical engineering from the National Cheng-Kung University, Taiwan in 1970, his MS degree in electrical engineering from Gifu University, Japan in 1979, and his PhD degree in computer engineering from Nagoya University, Japan in 1985. He is currently a full professor of the Department of Radiological Technology, School of Health Sciences, Faculty of Medicine, Niigata University, Japan. His research interests include medical image processing, pattern recognition, and applications of soft computing. He is a member of SPIE and senior member of IEEE.



Yongbum Lee received his BS, MS, and PhD degrees in information science from Gifu University, Japan in 1996, 1998, and 2001, respectively. He is currently a research associate in the Department of Radiological Technology, School of Health Sciences, Faculty of Medicine, Niigata University, Japan. His research interests include medical image processing and pattern analysis.