

# Mass Detection and Classification in Breast Ultrasound Images Using Fuzzy SVM

Xiangjun Shi, H.D. Cheng, and Liming Hu

Department of Computer Science, Utah State University  
401B. Old Mail Hill, Logan UT 84322, USA

## Abstract

Breast cancer is the leading cause of death of women in United States. Currently the effective method for early detection and screening of breast cancers is X-ray mammography. But the high rate of false positives in mammography causes a large number of unnecessary biopsies. Sonography is an important adjunct to mammography in breast cancer detection. The accuracy rate of breast ultrasound can reach a high level in the diagnosis of simple benign cysts and reduce the number of false positives. We develop a novel CAD system for mass detection and classification in breast ultrasound images based on the fuzzy SVM. The experimental results show that the proposed CAD system greatly improves the five objective indices in comparison with other classification methods and the radiologist assessment.

**Keywords:** Breast ultrasound images; Mass detection; Feature extraction; Pattern classification; Fuzzy SVM;

## 1. Introduction

Breast cancer is the leading cause of death of women in United States. Breast cancer can be most effectively treated when it is detected at its early stage. Currently the most effective method for early detection and screening of breast cancers is X-ray mammography [1]. However, reading mammography is a demanding job for radiologists, and cannot provide consistent results from time to time. The judgments depend on training, experience, and subjective criteria. In addition, mammography produces a high false positive rate, and only about 525 of 1800 lesions that were sent to biopsy are malignant [5]. The reasons for the high miss rate and low specificity in mammography are: (1) the low conspicuity of mammographic lesions; (2) the noisy nature of the images; (3) the overlying and underlying structures that obscure features of the ultrasound image. Also, the biopsies are expensive and involve minor risks. In order to avoid unnecessary biopsies, the number of false positives in mammography has to be reduced.

It has been shown that breast sonography is superior to the mammography in two facts: (1) the ability to detect focal abnormalities in the dense breasts of adolescent women; (2) the ultrasound images are acquired in real-time, with the low health risk to the patient, and with the low cost. Sonography is an important adjunct to mammography in breast cancer detection and has been primarily useful for differentiating cysts from solid tumors. The accuracy rate of breast ultrasound can reach to a high level in the diagnosis of simple benign cysts that do not require biopsies.

In this paper, we present a novel CAD system of automatic mass detection and classification of breast ultrasound images using fuzzy support vector machines.

## 2. Preprocessing

The ultrasound images originated from an ultrasonic scanner (VIVID7) manufactured by GE Medical Systems; the frequency ranges of the ultrasonic scanner are 5-12MHz [3].

### 2.1. Enhancement

Image enhancement is one of the most important issues in low-level image processing. Its purpose is to improve the quality of low contrast images. Therefore, the underlying principle of the enhancement is to enlarge the intensity difference between objects and surroundings, and the image should be not apparently distorted. A lot of methods have been developed. They can mainly be divided into two classes: local and global methods. The proposed Multi-peak GHE (Generalized Histogram Equalization) method [4] is very effective not only in enhancing the entire image but also in enhancing the image texture. It makes the change of the order of gray levels of the original image completely controllable. The experiment results show that this method is very effective in enhancing the breast ultrasound image.

### 2.2. Segmentation

Image segmentation can be considered as the labeling problem wherein the solution is to assign a set of labels to image pixels. Labeling is a natural representation for Markov Random Fields. The algorithms based on Markov Random Field (MRF)/Gibbs Random Field (GRF) studied in [3] is adopted here to segment the ultrasound images. The Metropolis Sampler approach was used, and a new local energy is defined in this method, which makes the system more consistent with the global model. Therefore, it is more tolerant to noise, and requires less iteration to converge. It is effective to segment breast ultrasound images corrupted by speckle noise.

### 3. Feature Analysis and Extraction

A key stage of mass detection and classification by CAD (computer-aided diagnosis) schemes is feature analysis and extraction. The feature space is very large and complex due to the wide diversity of normal tissues and variety of abnormalities.

Hundreds of features might be derived from an image. But not all of the features are suitable for mass classification. Too many irrelevant features not only make the classifier complicated, but also will reduce the accuracy of the classification. The most important issue is that to select features that are able to represent the characteristics of masses in the breast ultrasound images, and based on these features; the malignant mass can be significantly discriminated from the benign masses by the classifier. In breast ultrasound images, the spiculation and the angular margins are the significant characteristics. The spiculation produced the higher positive predictive value of malignancy. The duct extension or a branch pattern might also increase the likelihood of malignancy. Also, the hyperechogenicity, well-circumscribed lobulation, ellipsoid shape and a thin capsule are the significant characteristics of benign masses in breast ultrasound images. All these four features have high sensitivities and negative predicted values [9]. Based on the characteristics of the breast ultrasound image, the following kinds of features are extracted from the ultrasound images.

#### 3.1. Textural Features

Textural features are based on the texture information [7]. All the features are derived from the spatial gray level dependence (SGLD) matrices. SGLD matrices are used to measure the texture-context information. It is a 2-D histogram. An element of the SGLD matrix  $P(i,j,d,\theta)$  is defined as the joint probability of the gray levels  $i$  and  $j$  occur separated by distance  $d$  and along direction  $\theta$  of the image.

#### 3.2. Fractal Dimensions

The fractal concept is useful to represent a statistical quality of roughness and self-similarity at different scales of most natural surfaces and/or curves. This statistical quality is most often characterized in terms of the fractal dimensions. The fractal dimensions as a geometric feature in image processing have become popular in modeling image properties. Intuitively, the degree of roughness of the image texture is proportional to the fractal dimension. The definition of the fractal dimension is similar to the Hausdorff dimension [2]. Informally, self-similar objects with parameters  $N$ , the number of similar pieces, and  $s$ , the magnification factor, are described by a power law such as  $N = s^d$ , where  $d$  is the "dimension" of the scaling law, known as the Hausdorff dimension. In this work, five features are designed and extracted based on the fractal dimensions of the ultrasound breast images:

$$fd1 = \text{fractal dimension of the ROI.}$$

$$fd2 = \text{fractal dimension of the suspicious area.}$$

$$fd3 = \text{fractal dimension of the ROI excluding the suspicious area.}$$

$$fd4 = fd2 / fd1.$$

$$fd5 = \text{fractal dimension of the edge of the suspicious area.}$$

#### 3.3. Histogram-Based Features

The Shape of the histogram provides many clues to represent the characteristic of the images. The seven statistic features extracted from the image based on the histogram is Mean, Variance, Skewness, Kurtosis, Energy, and Entropy. The mean is the average intensity level whereas the variance implies the variation of intensity around the mean. The skewness indicates whether the histogram is symmetric about the mean. The histogram is symmetrical if the skewness is zero. Otherwise, it is skewed above the mean if the skewness is positive, and skewed below the mean if the skewness is negative. The kurtosis is a measure of how sharp the histogram peaks are. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data with high kurtosis tend to have a distinct peak near the mean, declining rather rapidly, and having heavy tails. Data with low kurtosis tend to have a flat top near the mean rather than having a sharp peak. Entropy is a measure of how much disorder in a system.

#### 3.4. Feature Selection

Totally, 151 features described above are derived from the suspected mass lesion area of the ultrasound image including 140 texture features, the five fractal dimensions, and six histogram-based features. But not

all of the features are useful for mass classification. Also, the correlation may exist among the features.

Stepwise regression [6] is a statistic technique for choosing an optimal subset of explanatory variables. Based on these 151 features, stepwise regression produces an optimal subset of the features that comprised of 13 features in our experiments, including:

***The eight texture features:***

- the mean of the information measure of correlation with the distance of one pixel ( $d = 1$ );
- the mean of maximal correlation coefficient with the distance of one pixel ( $d = 1$ );
- the range of difference entropy with the distance of one pixel ( $d = 1$ );
- the range of variance with the distance of two pixels ( $d = 2$ );
- the range of the information measure of correlation with the distance of two pixels ( $d = 2$ );
- the mean of maximal correlation coefficient with the distance of four pixels ( $d = 4$ );
- the mean of the information measure of correlation with the distance of eight pixels ( $d = 8$ ); and
- the mean of inverse different moment with the distance of 16 pixels ( $d = 16$ ).

***The three fractal dimensions:***

- the fractal dimension of the ROI;
- the fractal dimension of the suspicious area; and
- the fractal dimension of the ROI not including the suspicious area.

***Two histogram-based features:***

- variance; and
- entropy.

## 4. Classifications

Computer aided diagnosis (CAD) systems may help radiologists in interpreting sonography for mass detection and classification. It is important to develop CAD systems that can distinguish benign lesions from malignant lesions. The combination of the CAD scheme and experts' knowledge would greatly improve the detection accuracy of the abnormalities.

### 4.1. Fuzzy SVM

The support vector machines (SVMs) based on statistical learning theory, are introduced by Vapnik [10]. The learning method in SVMs is motivated by statistical learning theory. SVMs are powerful in solving two-class classification problem [8], but some limitations exist in the SVM theory. Traditionally, we know each sample  $\{x_i, y_i\}$  in the training dataset belongs to either one class or the other, i.e., the value of

$y_i$  is only assigned to 1 or -1. All samples in training dataset are treated uniformly in the same class during the learning process of SVMs.

In practical classification problems, the effects of the samples in training dataset may be different. Usually, some of samples in training dataset are corrupted by noise, which is introduced during sampling. These samples are called outliers, and usually less important than others. In fact, that we care about the meaningful samples can be classified correctly.

In short, a sample in the training dataset may not completely belong to one class. For example, 90% of the sample belongs to one of the two classes and 10% is meaningless, or we say that the sample belongs to one of two classes with 90% confidence. In other words, each training sample  $\{x_i, y_i\}$  is associated with a fuzzy membership ( $0 \leq s_i \leq 1$ ). This fuzzy membership  $s_i$  indicates the certainty that the sample belongs to one of two classes is  $s_i$ , and the value  $(1 - s_i)$  can be regarded as meaningless in the classification problem.

An FSVM (fuzzy SVM) is proposed in [8]. In the FSVM, each sample  $\{x_i, y_i\}$  in the training data is weighted by using fuzzy membership function. It becomes as  $\{x_i, y_i, s_i\}$ , where  $s_i$  is the fuzzy membership, i.e., the confidence of this sample belong to one of two classes. Then the optimal hyperplane problem in SVM is reformulated as

Minimize

$$L(\alpha) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum \alpha_i$$

Subject to

$$\sum_{i=1}^n y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq s_i C, \quad i = 1, 2, \dots, n$$

The significant difference of FSVM from SVM is that the sample with smaller fuzzy member  $s_i$  is less important than all other samples in SVM during training. It indicates that the importance of the training sample can be measured by the fuzzy membership  $s_i$ .

### 4.2. Fuzzy Membership

In fuzzy logic, the fuzzy membership function is a mapping from an element to a real value within [0,1]. Logistic regression can be used to predict a discrete outcome, such as group membership, from a set of explanatory variable values. Generally, the response variable is dichotomous, such as two classification problem. That is, the response variable can take the value 1 with a probability of success  $p$ , or the value 0 with probability of failure  $1-p$ . In our two-class classification problem, the logistic regression model is used to predict the fuzzy memberships of training samples. This regression model is formulated:

$$P(y=1) = \frac{e^{\beta^T x}}{1+e^{\beta^T x}} = \frac{e^{\beta_0 + \sum_{j=1}^m \beta_j x_j}}{1+e^{\beta_0 + \sum_{j=1}^m \beta_j x_j}}$$

where  $\mathbf{x} = (1 \ x_1 \ x_2 \ \dots \ x_m)^T$ ,

and  $\boldsymbol{\beta} = (\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_m)^T$

The subscript  $m$  is the dimension of the feature space,  $\mathbf{x}$  is the explanatory variable vector, and  $\boldsymbol{\beta}$  is called the coefficient vector.

Based on the regression model, we know that the sample with higher certainty of ( $y_i = 1$ ) is having larger fuzzy membership in the class ( $y_i = 1$ ). Also, the sample with lower certainty of ( $y_i = 1$ ) larger fuzzy membership in the class ( $y_i = -1$ ). Based on this idea, we define the fuzzy membership  $s_i$  as:

$$\begin{aligned} \text{IF } (y_i = 1) \quad \text{THEN } s_i &= P(y_i = 1) \\ \text{ELSE } s_i &= 1 - P(y_i = 1) \end{aligned}$$

## 5. Experimental results and conclusions

Table 1 shows that radiologist produces a high false positive rate. It implies that about one third (18 of 52) lesions sent to biopsy are benign, even though their false negative rate is reasonably low. However, the analyses using CAD systems, including ANN, SVM, and Fuzzy SVM can greatly reduce the false positive rate while approximately keeping the same level of the false negative rate as radiologists; among the CAD systems above, the analyses using ANN and SVM are approximately same, almost produce the same positive and negative rate. The fuzzy SVM analysis produces the lowest positive and negative rate, and has the best performance.

Table 1. Summary of the Different Classifiers and the Radiologist Assessment.

	TP	FP	TN	FN
Radiologist	32	18	33	4
ANN	31	5	46	5
SVM	31	6	45	5
FSVM	33	2	49	3

By observing Table 2, we know that the three indices of accuracy, specificity, and PPV of the ANN and the SVM are much higher than that of the radiologist assessment, and the other two indices are nearly the same as those of the radiologist assessment; all five indices of the fuzzy SVM are much higher than those of the radiologist assessment. Also, all five indices of the fuzzy SVM are much better in comparison with the ANN and the SVM analyses. Therefore the proposed approach can classify the malignant solid masses more

accurately than the radiologist assessment and the ANN and the SVM analyses.

Table 2. The Comparison of the Performances among the Analyses of Different Classifiers and Radiologist.

	Radiologist	ANN	SVM	FSVM
Accuracy (%)	74.71	88.51	87.36	94.25
Sensitivity (%)	88.89	86.11	86.11	91.67
Specificity (%)	64.71	90.20	88.24	96.08
PPV (%)	64.00	86.11	83.78	94.29
NPV (%)	89.19	90.20	90.00	94.23

## 6. References

1. Bovis, K., Singh, S., Fieldsend, J. and Pinder, C., Identification of Masses in Digital Mammograms with Mlp and Rbf Nets. in *International Joint Conference on Neural Networks (IJCNN'2000)*, Jul 24-Jul 27 2000, Year.
2. Chen, C.C., Daponte, J.S. and Fox, M.D. Fractal Feature Analysis and Classification in Medical Imaging. *IEEE Transactions on Medical Imaging*, 8, 2 (1989), 133-142.
3. Cheng, H.D., Hu, L., Tian, J. and Sun, a.L. A Novel Markov Random Field Segmentation Algorithm and Its Application to Breast Ultrasound Image Analysis, in *6th International Conference on Computer Vision, Pattern Recognition and Image Processing, Proceedings, JCIS 2005*, Salt Lake City, pp.644-647, 2005.
4. Cheng, H.D. and Shi, X.J., A Simple and Effective Histogram Equalization Approach to Image Enhancement. *Digital Signal Processing*, 14, 2 (2004), 158-170.
5. Dennis, M.A., Parker, S.H., Klaus, A.J., Stavros, A.T., Kaske, T.I. and Clark, S.B. Breast Biopsy Avoidance: The Value of Normal Mammograms and Normal Sonograms in the Setting of a Palpable Lump. *Radiology*, 219 (2001), 186-191.
6. Draper, N. and Smith, H. *Applied Regression Analysis*. John Wiley and Sons, 1981.
7. Haralick, R.M., Shanmugam, H.K. and Dinstein, I. Texture Features for Image Classification. *IEEE Transaction on Systems, Man, and Cybernetics*, SMC-3, 6 (1973), 610-621.
8. Lin, C.-F. and Wang, S.-D. Fuzzy Support Vector Machines. *IEEE Trans. on Neural Networks*, 13, 2 (2002), 464-471.
9. Stavros, A.T., Thickman, D., Rapp, C.L., Dennis, M.A., Parker, S.H. and Sisney, G.A. Solid Breast Nodules - Use of Sonography to Distinguish Benign and Malignant Lesions. *Radiology*, 196, 1 (1995), 123-134.
10. Vapnik, V.N. *Statistical Learning Theory*. John Wiley & Sons, New York, 1998.