Applied Mathematical Sciences, Vol. 8, 2014, no. 24, 1193 - 1200
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ams.2014.4151

Mammographic Image Recognition Using Rough Sets and Support Vector Machines

Jiming Lan

School of Computer Science
Sichuan University of Science and Engineering
Zigong, 643000, P. R. China

Abstract

Medical image recognition is one of the most important unsolved problems in medicine. Taking the mammogram as the object for research, this paper proposes a method for mammographic image recognition using rough sets and support vector machines (SVMs). Firstly, reduce mammographic noise. Secondly, extract texture and shape features to consist of feature vector that can represent the mammogram accurately. Next, the features are normalized. Finally ,attribute reduction by rough sets and classification recognition by SVMs is completed. The experimental results show that this method for mammographic recognition can achieve a satisfactory effect.

Keywords: mammographic image recognition, image preprocessing, rough sets, SVMs

1. Introduction

In the United States, an estimated 40,030 breast cancer deaths (39,620 women, 410 men) are expected in 2013. Breast cancer ranks second as a cause of cancer death in women (after lung cancer)[1]. Early detection and diagnosis of breast cancer increases the survival rate and increases the treatment options.

Screening mammography, x-ray imaging of the breast, is currently the most effective tool for early detection of breast cancer. A method for mammographic recognition using rough sets and support vector machines (SVMs) is proposed in this paper. This method is based on the modern image processing technology, artificial-intelligence algorithms, and pattern classification algorithms. It can effectively recognize the malignant breast tumor in the mammogram. After studying the various methods to process images, we use the following methods in sequence: Firstly, extracting the region of interest (ROI) in the images by median filtering technique, fuzzy enhancement algorithms, and region growing (RG) algorithm. Secondly, reducing the number of image features by the rough set approach. Finally, using reduced features as input vectors for SVMs to distinguish the benign tumor and malignant tumor. The basic principle of the method is illustrated briefly in Figure 1.

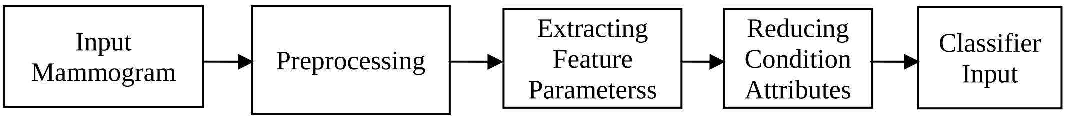
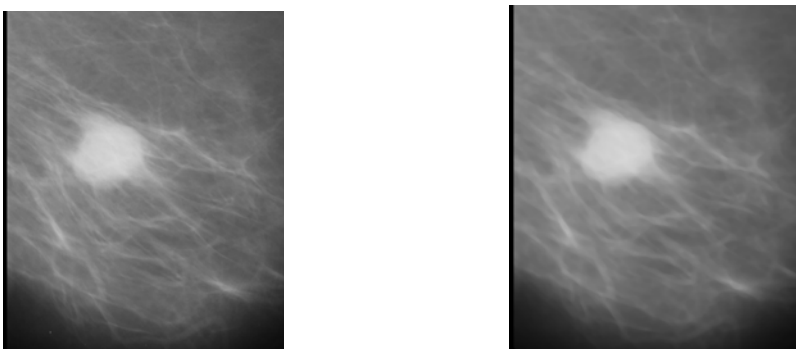


Figure 1: Flow diagram of mammographic image recognition by rough sets and SVMs

2. Mammographic Image Preprocessing

All the mammograms processed in this article come from Mammographic Image Analysis Society (MIAS) database. The pictures in the mini-MIAS database of mammograms are taken with a safe, low-dose X-ray machine. The principles of X-ray imaging lead inevitably to the noisy mammograms. Image denoising is one of the key steps in image processing and the basis of image subsequent processing [2]. In this paper, the mammograms are preprocessed by the median filtering technique, fuzzy enhancement algorithms, and region growing algorithm.

Difference between the mammograms before and after filtering is shown in Figure 2. The smooth image contours and clear boundaries are preserved while the image noise is reduced through the medium filtering.



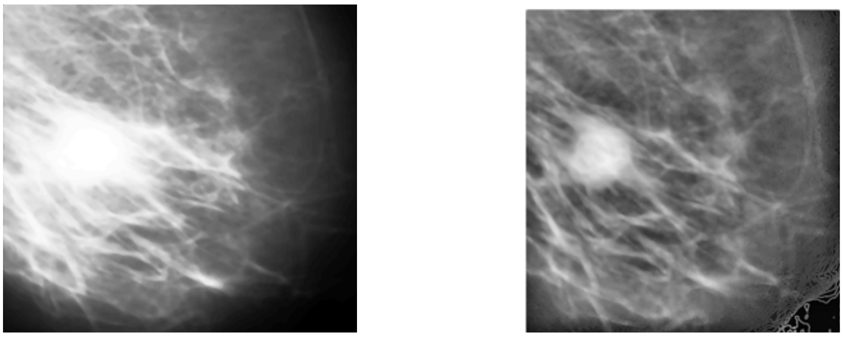
a- Mammogram before filtering

b- Mammogram after filtering

Figure 2: Part of the mammograms before and after using medium filtering

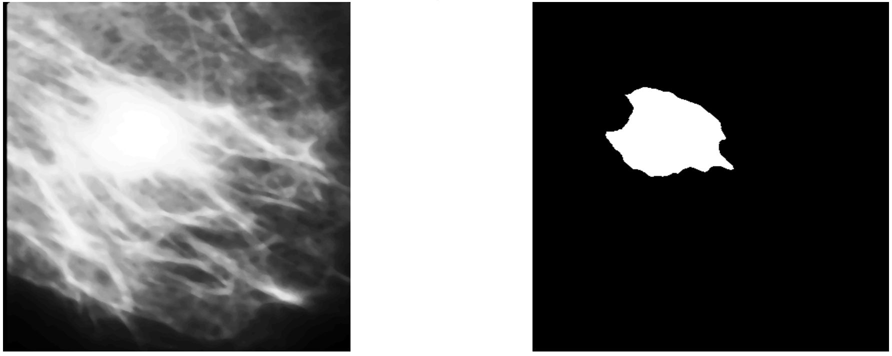
Fig. 3 compares the filtered mammogram using the fuzzy enhancement algorithms with the previous one. It is easy to see that the contrast between the breast tumor and its background is enhanced, and boundary processing is also fairly satisfactory [3].

Fig. 4 is the mammograms segmented by region growing algorithm. In the segmentation stage, the mass is segmented accurately from the background normal tissue in Fig. 4 b.



a- Mammogram before enhancement b- Mammogram after enhancement

Figure 3: Part of the mammograms before and after using fuzzy enhancement algorithms



a- Mammogram before segmentation b- Mammogram after segmentation

Figure 4: Part of the mammograms before and after using region growing algorithms

3. Feature Extraction

Hongxin Zhang et al. used fractal geometry to extract the breast tumor features, and measure the fractal dimensions of the nuclear boundary in 69 cases of breast carcinoma and 38 cases of benign breast fiber adenoma [4]. In this paper, we use shape and texture features based on gray-level co-occurrence matrices (GLCM) to extract the breast tumor features [5, 6].

3.1. Shape Features

The breast tumors are classified into the benign breast tumors and the malignant breast tumors. The benign breast tumors are usually round or oval, but the malignant ones can be lobular, nodular, stellate, or irregular. Tumors with spiculate or indistinct margins have a higher probability of malignancy than tumors with circumscribed margins. We may summarize the shape features of the breast tumors as following: the contour perimeter, the area, the circularity, and the elliptical compactness of breast tumors, etc.

The contour perimeter p and the area s can be directly calculated from the pixels in the segmented image. The circularity of breast tumors c is defined as

$$c = 4\pi s / p^2$$

where c ranges from 0 to 1. The rounder the breast tumor is, the closer its circularity is to 1. The elliptical compactness EC is defined as

$$EC = (m+n)\pi / p$$

where m and n are respectively the major and minor axis of the ellipse used for fitting a breast tumor. The smaller EC is, more serious the spiculate extent, and the greater the likelihood of the malignant breast tumor.

3.2. Texture Features

In this paper, gray-level co-occurrence matrices are used for extracting the texture features of the breast tumor. 9 gray features are respectively selected in four directions (namely 0^0, 45^0, 90^0, and 135^0)[7]. They are

$$T_1 = \sqrt{\sum_{i,j=1}^{L} p(i,j)^2}\ , \quad T_2 = \sum_{i,j=1}^{L} i * p(i,j)\ , \quad T_3 = \sum_{i,j=1}^{L} j * p(i,j)\ ,$$

$$T_4 = \sum_{i,j=1}^{L} \frac{p(i,j)(i-T_2)(j-T_3)}{\sigma_i \sigma_j}\ , \quad T_5 = \sum_{m=0}^{L-1} m^2 \{\sum_{i,j=1}^{L} p(i,j)\} \quad (m = |i-j|)\ ,$$

$$T_6 = \sum_{i,j=1}^{L} p(i,j)m\ ,$$

$$T_7 = -\sum_{i,j=1}^{L} p(i,j)\log(p(i,j))\ , \quad T_8 = \sum_{i,j=1}^{L} p(i,j)(i-T_2)\ , \text{ and } \quad T_9 = \sum_{i,j=1}^{L} \frac{p(i,j)}{1+m^2}\ .$$

To avoid the influence on these feature values for reasons of image rotation and shift, we take the average of these feature values respectively generated from four co-occurrence matrices, and get the final gray feature values.

Taken together, we select 13 features, which are 4 shape features and 9 texture features, as the recognition criterion of malignant and benign breast tumors. To avoid the problem of the curse of dimensionality, it is necessary to reduce these features before recognition.

4. Attribute Reduction Based on Rough Set Theory

The usual attribute reduction algorithms include the genetic algorithm, Johnson algorithm, Holte's IR algorithm, and so on. In this paper, the genetic algorithm is used for attribute reduction through the software Rosetta. Specific steps are as follows:

1) Input the decision table s.

2) Solve the $core(C)$ of the condition attribute C to the decision attribute D. Let $core(C)=\varnothing$, remove the attribute $r \in C$ one by one. If $pos_{C-\{r\}}(D) \neq pos_{c}(D)$, $core(C)=core(C) \bigcup \{r\}$.

3) Randomly generate M binary strings, whose length is n (namely the number of the condition attributes in the decision table s), to establish initial population. The corresponding values to the attributes of $core(C)$ are assigned 1, and others are randomly assigned 0 or 1.

4) Calculate the fitness of each individual by the formula $f = e^{k-\frac{card(x)}{n}}$, where k is the dependability of the decision attribute in the individual to the conditional attribute. The roulette selection is applied, and the best individuals are kept to next generation.

5) Select the parent individuals by the single-point crossover and simple mutation.

6) Eliminate the similar individuals, supplement with the new individuals, and rebirth into the next population.

7) Terminate the algorithm and output the optimal individuals if the fitness of the optimal individuals is no longer improved over successive generations, otherwise go to step 4.

8) Output the attributes reduction table.

After the above steps, 13 features are reduced to 9 features (namely the perimeter, circularity, elliptical compactness of breast tumors, and the energy, average, correlation, inertial, entropy, variance of the GLCM).

5. Mammographic Image Recognition by SVMs

Through preprocessing, feature extraction, and reduction based on rough sets for the mammograms from the mini-MIAS database, the training set and the testing set are inputted to SVMs for training and testing, as shown in Figure 5.

The size of the mammographic images in the mini-MIAS database is 1024 pixels x 1024 pixels. There are 322 images, which come from 161 patients, in this database. They are classified into 3 types: normal mammograms without mass, normal mammograms with masses, and abnormal mammograms. The location and category of the mass is specified by experienced radiologists in the mammogram with masses. The breast tumors in this database are classified into 3 types: well-defined or circumscribed masses, ill-defined masses, and speculated masses. To test our approach we used 44 mammographic images: circumscribed masses are 14 cases (10 cases are benign, and 4 cases are malignant), ill-defined masses are 15 cases (8 cases are benign, and 7 cases are malignant), and speculated masses are 15 cases (7 cases are benign, and 8 cases are malignant). We consider the proportion not only of positive and negative samples but of different samples that can be gotten from the database while selecting these samples.

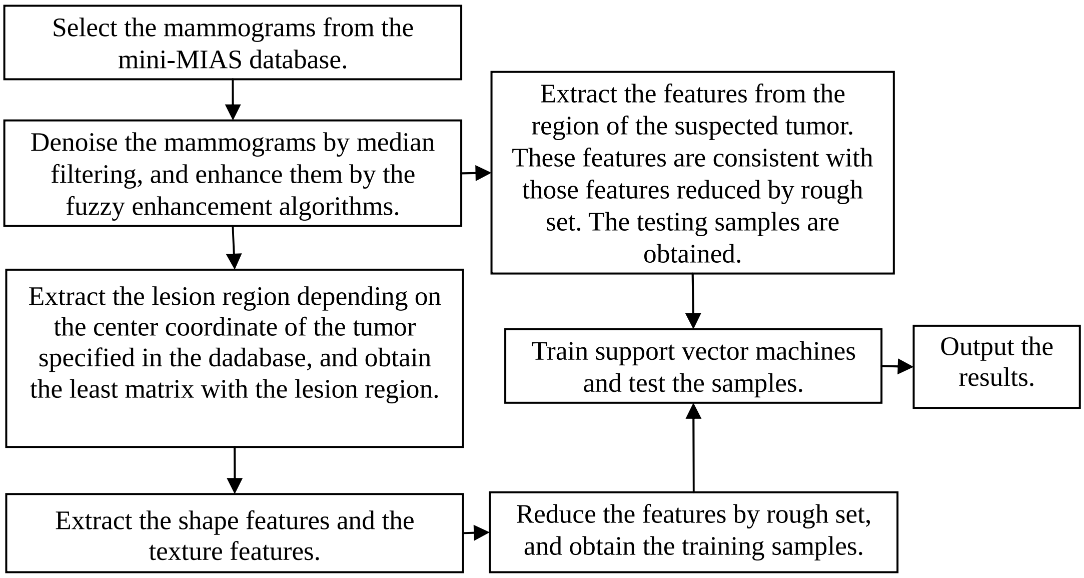

Select the mammograms from the mini-MIAS database.
Denoise the mammograms by median filtering, and enhance them by the fuzzy enhancement algorithms.
Extract the lesion region depending on the center coordinate of the tumor specified in the dadabase, and obtain the least matrix with the lesion region.
Extract the shape features and the texture features.
Extract the features from the region of the suspected tumor. These features are consistent with those features reduced by rough set. The testing samples are obtained.
Train support vector machines and test the samples.
Output the results.
Reduce the features by rough set, and obtain the training samples.

Figure 5: Flow diagram of mammographic image recognition by SVMs

44 training samples are processed in turn according to the flow diagram in Fig. 5. Cross-training is used during training. All the training samples are divided into 4 groups. 3 groups of them are used as the training set, and the rest is used as the testing set. A 4-Class Support Vector Machine is constructed in this work[8]. The Gaussian radial basis function (RBF) is adopted as the kernel function of this model. Through experiments the values of the parameters in this model are taken as $c = 100, \delta = 4$.

All the prediction results by cross-training are summarized, and the final prediction results and prediction accuracy are obtained, as listed in Table 1.

Table 1: recognition rate of the model

Name	Total number	Number correctly recognized	Recognition rate
Benign masses	25	21	84.00%
Circumscribed malignant masses	4	4	100.00%
Ill-defined masses	7	6	85.71%
Speculated masses	8	8	100.00%
Malignant masses	19	17	89.473%

Overall, the recognition rate for malignant masses is high in this work. However, there is a problem in this work. All the samples in the mini-MIAS database are included in the training samples according to the proportion of positive and negative samples. The number of the training samples is limited. In future work, the numbers of different samples should be expanded.

6. Conclusion

Mining features of tumors in the mammogram, is a process of discovering knowledge from mass data, which is inaccurate, incomplete, and uncertain. Rough set theory and SVMs show the endless charm in this respect. SVMs can realize supervised and unsupervised learning, and solve some problems of traditional machine learning algorithms, such as the over study and local minimization of artificial neural network (ANN). But SVMs cannot determine what knowledge is redundant, and which attributes are useful. Rough set theory can describe importance of the different attributes in knowledge representation. But rough sets are worse than SVMs in the adaptability to the changing environment and their own fault tolerance. In this work, the two are combined to give play to their respective advantage. Rough sets are used as the front system of SVMs, which reduce the number of input space dimension, remove the redundant information in the training sets, shorten training time, and improve the recognition accuracy. Experiments show that the mammographic image recognition method using rough sets and SVMs is effective in this work.

Acknowledgment

The author thanks the following persons for their encouragement, discussions, suggestions, and comments: Ms. Haiyan Zhang and Dr. Jinshan Zhang at School of Science, Sichuan University of Science & Engineering, Zigong, China.

This work was supported by the Zigong key science and technology plan project (2013ZC11).

References

[1] American Cancer Society, *Cancer Facts & Figures 2013*, Atlanta: American Cancer Society, 2013.

[2] M. S. El-Bashir, *Face Recognition Using Multi-Classifier*, Applied Mathematical Sciences, 6 (2012), pp. 2235-2244.

[3] M. P. Sampat, G. J. Whitman, A. C. Bovik and M. K. Markey, *Comparison of algorithms to enhance spicules of spiculated masses on mammography*, Journal of digital imaging, 21 (2008), pp. 9-17.

[4] Hongxin Zhang, Peirong Zhao, Dongling Gao, Yuhui Yin and Ahong Zhao, *Fractal study on nuclear boundary of breast tumor*, Journal of Zhengzhou University(Medical Sciences) 4(2002), pp. 431-433.

[5] M. P. Sampat, M. K. Markey, A. C. Bovik and Others, *Computer-aided detection and diagnosis in mammography*, Handbook of image and video processing, 2 (2005), pp. 1195-1217.

[6] P. Kamencay, R. Hudec, M. Benco and M. Zachariasova, *Feature extraction for object recognition using PCA-KNN with application to medical image analysis,* 2013, pp. 830-834.

[7] F. Eddaoudi, F. Regragui, A. Mahmoudi and N. Lamouri, *Masses detection using SVM classifier based on textures analysis*, Applied Mathematical Sciences, 5 (2011), pp. 367-379.

[8] G. Guo, S. Z. Li and K. Chan, *Face recognition by support vector machines,* 2000, pp. 196-201.

Received: January 21, 2014