

# Multiple Support Vector-based approach to Breast tumor recognition

**Abstract.** Based on Multi-Support Vector Machine (MSVM), this paper provides a method of MSVM for breast tumor recognition to solve unfixed size and individual difference with the breast tumor. Support Vector Machine (SVM) on the eight direction of bump area is taken to generate vector classifier and to select Gauss kernel function as kernel function. The system application and test shows that the MSVM in breast tumor recognition achieved good result, and provide the reliable basis for further medical diagnosis. The breast tumor recognition accuracy achieved 97.3% when  $\sigma=30$ .

**Streszczenie.** Opisano wykorzystanie metody MSVM (multi-support vector machine) do wykrywania guza piersi. Metoda generuje klasyfikator wektorowy w ośmiu kierunkach a selekcję cech przeprowadza się wykorzystując funkcje Gaussa jako kernel. (Wykorzystanie metody MSVM do wykrywania guza piersi).

**Keywords:** Multi- Support Vector Machines; Kernel Function; Image Segmentation; Edge Detection; Tumor recognition.

**Słowa kluczowe:** metoda MSVB – multi support vector machine, guz piersi, rozpoznawanie wzoru

## Introduction

Breast cancer is one of the most common malignant tumor for the women, the main way to decrease mortality caused by breast cancer is the inchoate diagnosis and treatment. Currently, there has been a lot of methods for breast tumor and calcified point detection and recognition, among them the main methods are the neural network, the wavelet transform, fuzzy sets, mathematical morphology, image enhancement technique [1], etc. In order to identify tumors, paper [2] proposes to establish neural network for the morphological characters of breast tumor ultrasound images. But the neural network faced a huge application limit in high dimensional space. Paper [3] proposes a new technique to extract micro calcified point ROI (region of interest) based on wavelet transform. Paper [4] proposes a method to extract breast micro calcified point based on morphology gray reconstruction, but breast cancer of calcified point type is just a small part of breast cancer, breast tumor is the basic feature of most breast cancer. Meanwhile, being sensitive to noise, mathematical morphology can't get good diversion result. Liu<sup>[5]</sup> proposes multidimensional methods to diagnose acicular tumor. Because of the huge amount of calculation in result of extracting feature from each pixel put forward by Liu, it is hard to use in practice.

In recent years, the Support Vector machine (SVM) is more and more widely used in medical image processing area. Support Vector machine (SVM) is a new method in data mining [6], it is a kind of pattern recognition method of statistical theory, which can explore a best compromise between pattern complexity and learning ability in accordance with limited sample information so that generalization capability could be acquired. Previous research of breast tumors and calcified point mostly based on man-machine interaction, rely on expert experience and the precision of image preprocessing and dividing, this paper makes support vector machine (SVM) to apply to the recognition of breast tumor, proposes a method of extracting outline of tumor automatically.

## Support vector machine (SVM) method

### A. Optimal Hyperplane

SVM is developed on the basis of statistics as a new pattern of recognition method. It studies how to construct learning machine and prior pattern classification. Support vector machine has a good performance in solving problems such as small sample learning, nonlinear high

dimensional learning.

SVM is developed from linearly separable optimal hyperplane, the basic idea could be illustrated by figure 1 in bidimensional condition.

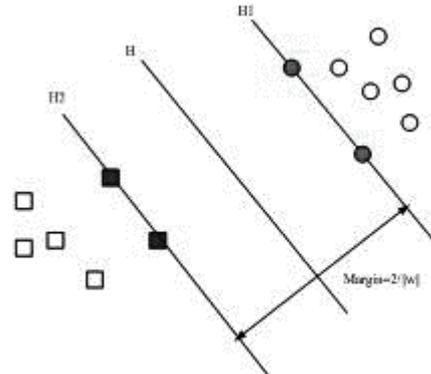


Fig1.Optimal classifier

Presume linear sample set  $(X_i, Y_i), i=1, \dots, n, X \in R^d, Y \in \{+1, -1\}$ . Linear discriminant function in d-dimensional space  $g(x) = \omega x + b$ , classification plane function is  $\omega x + b = 0$ , normalise the discriminant function in order to make all the samples from the two classes satisfy  $|g(x)| \geq 1$ , the closest sample to classification plane is  $|g(x)| = 1$ , thus, the classification interval equals to  $2/|\omega|$ , therefor, to request a max classification interval, is to request a minimum  $|\omega|$  ( $|\omega| \omega^2$ ), to request that the classification plane could classify all the samples correctly, is to request it satisfy:

$$(1) \quad y_i [\omega x_i + b] - 1 \geq 0 \quad i=1, \dots, n$$

Therefor, the classification plane which satisfy the formula above and minimum  $\|\omega\|_2$  is optimal hyperplane. The training samples which line the nearest point to classification plane and are from hyperplane H1、H2 which parallel to the optimal hyperplane, namely the samples satisfy the equation are support vectors.

### B. Nonlinear problems

For nonlinear problem, we could transform it into a linear problem in some high dimensional space through nonlinear conversions, optimal hyperplane could be found in conversion space. According to the functional theory, a kernel function  $K(x_i, x_j)$  corresponds to a conversion space inner product as long as it satisfies Mercer condition<sup>[6,7]</sup>. Therefor, to choose an appropriate inner product

function could achieve a linear classification after nonlinear conversion, the calculation complexity ,however, doesn't increase. In table 1 several common nonlinear transform kernel functions are presented:

C. multiple support vector machine

The actual problems are usually not simple two-class problems, support vector machine can be extending to multiple class problems and we could acquire multiple classifier. The common methods to construct SVM multi-value classifier are: one-to-many, one-to-one and decision tree, etc.

Table 1. Kernel functions and expression

kernel function	expression
Linear kernel function	$K(x, x_i) = x^T x_i$
Polynomial kernel function	$[x^T x_i + 1]^q$
Gaussian kernel function	$K(x, x_i) = x^T x_i$
S form kernel function	$K(x, x_i) = \tanh[v(x \bullet x_i) + a]$

Paper [8]proposes a method to detect and divide nuclear magnetic resonance images of ventriculus sinister automatically using multiple support vector machine. Because of the features of breast tumors such as that it has a relatively complete center lesion, it appears as a bright spot in the image, the gray value of image of center area is much higher than normal tissue, gray value of edge area is the transition from tumor center to normal tissue. This paper puts forward to judge edge of tumor using multiple support vector machine in order to increase speed and rate of recognition. Then find out points with highest gradient in edge area, line them up, and the outline of tumor could be divided so that shape of tumor could be distinguished.

**Tumor identification algorithm**

A. Algorithm

1) The preprocessing of the mammary gland pictures, filter out all sorts of noise. This paper uses traditional 3×3 median filter, the edge of lesion area maintain clear after median filter.

2) Extract region of interest. In recognition of breast tumor, it is an important step to divide lesion area accurately which is a preparation for extracting feature vector. This paper extracts region of interest by threshold technique.

3) Feature extraction. Extract characteristics parameter of lesion area and prepare for tumor recognition. Mudigond[10]extracts feature by using method based on pixels belt of tumor edge. Extract 5 textural features as character vectors based on gray level paragenesis matrix from tumor area divided. Calculate gray level paragenesis matrix adopting method of gray level paragenesis matrix texture analysis, then extract feature. The features extracted form vector as the input of classifier and proceed with classification.

4) Design of multiple SVM classifier. Multiple SVM structure presented by figure 2, the input is the 8×8(pixel) area to be detected in purpose round area, what is in the judgement layer, is a combination of several vectors, assume that edge was mainly divided into 8 directions ,there are 8 SVMs which are in charge of detecting edge of direction respectively, consider each input as the input of every SVM in judgement layer, if all the 8 detection results are positive, the sample turns out not to be the edge of

tumor , if only one SVM result is negative, the sample turns out to be the edge of tumor in direction which this SVM is responsible for, then record it , it also appears that there are two SVM judged to be positive, the edge direction is between theses two directions, samples in this condition should be recorded as edge area as well. Speed and recognition of multiple SVM have been improved.

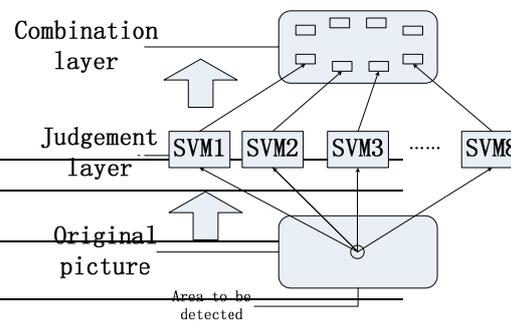


Fig2.The logic unit of eight support vector combination method

Train SVM1, so that it can divide tumor region from mammary grand picture. Train SVM2, so that it can recognize edge region of tumor from tumor region recognized by SVM1 and record it. Extract edge point from recorded edge region using gradient information , join edge point and the shape of tumor is acquired.

In the training sample set is pictures of mammary grand picked up from MIAS Dataset. We pick up 218 positive samples as representative of no tumor existing, 114 negative samples as representative of tumor existing. Among them, negative samples contain the conditions of benign and malignancy tumor, trained SVM can recognize various types of tumors. The amount of samples and the choosing of sample region relate directly to training results, therefore, positive and negative samples should be contained in conditions and regions as many as possible. In the experiment, we adopt multiple SVM to detect edge of tumor. If we adopt simple SVM, sample is the whole edge region, SVM recognition rate is low because of the irregularity of tumor edge, we can not get accurate segmentation result. This paper adopts multiple SVM to recognize tumor edge, each SVM is in charge of a fixed direction of recognition of edge region, it avail recognition of tumor in star and fuzzy type.

After detecting edge region, we only have to find out points with highest gradient in the region, and join them together so that relatively accurate outline of tumor appears.

**Analysis study**

All the pictures that showed in this paper are from Mammographic Image Analysis (MIAS) Dataset, the size of picture in MAIS Dataset is 1024\* 1024, picture format is. pgm. Every pixel represented by 8 bits. Among the 322 pictures, 208 pictures are normal, 63 pictures show that tumor existed, 51 pictures show that malignancy tumor existed.

During SVM training, although kernel functions which satisfy Mercer condition are available theoretically, different kernel functions have totally different performances. It is very important to identify kernel function and parameters reasonable and efficient. This paper identifies kernel function using method of cross validation. Gaussian kernel function chosen in this experiment and segmentation accuracy rates according to different parameters are presented below.

Table 2. Comparison rate of  $\sigma$  in the Gaussian kernel function

$\sigma$	0.5	20	30	50	200
accuracy	65.7%	84%	97.3%	82.7%	82.3%

From table 2 we can conclude that under Gaussian kernel function, the segmentation accuracy rate is highest when  $\sigma$  is 30, so we choose Gaussian kernel function,  $\sigma=30$  when utilize SVM to design classifier.

Figure 3 is recognition of mammary grand tumor, (a) is the original image with tumor, (b) is the identification of tumor region utilizing SVM1, (c) is the edge extraction by multiple SVM, (d) is the final rendering of tumor segmentation.

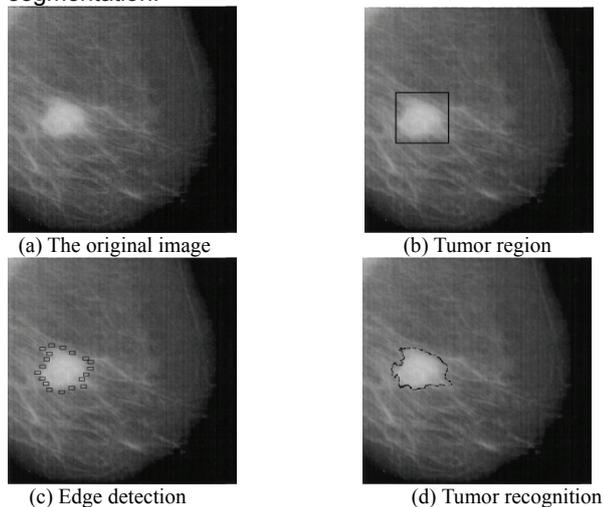


Fig3. Breast tumor recognition

### Conclusions

As a new machine learning method, support vector machine has advantages such as overall optimization, strong flexibility, complete theory, short training time, good generalization performance. The experiment result shows that not only speed but also accuracy rate of segmentation can be improved if use multiple SVM to extract tumor edge. According to feature of breast tumor, apply multiple SVM to segmentation, choose appropriate kernel function through

experiment, finally we can acquire good segmentation result. In this paper, choosing kernel function is based on abundant of experiments, if we can move forward to choose optimal kernel function according to feature of mammary grand, the segmentation effect will be more precise.

### REFERENCES

- [1] LIN-Yao, TIAN-Jie, A Survey On Medical Image Segmentation Methods[J]. Pattern Recognition and Artificial Intelligence, 2002, 15(2):192-204. (in Chinese).
- [2] Wang Yuan, Shen Jia-lin. Breast tumor classification based on shape features of ultrasonic images [J]. Optics Precision Engineering, 2006, 14(2):333-340. (in Chinese).
- [3] LI Shu-nan, WAN Bai-kun, et al. A Novel ROI Extracting Technique Based on Wavelet Transform for the Detection of Micro-calcifications in Mammograms[J]. J Biomed Eng, 2005, 2(22):360-362. (in Chinese).
- [4] WEN Hao, Ma Jin-sheng, et al. On Microcalcifications Detection in Mammograms Based on Morphological Grayscale Reconstruction[J]. CT Theory and Applications, 2006, 15(2):33-37. (in Chinese).
- [5] Liu, C.F. Babbs, E.J. Delp. Multiresolution detection of speculated lesions in digital mammograms[J]. IEEE Transactions on Image Processing, 2001, 10(6):874-884.
- [6] DENG Nai-yang, TIAN Ying-jie. The New Method in data mining: Support vector machine [M]. Beijing: Science press, 2004. (in Chinese).
- [7] V.N. Vapnik. The essence of statistical learning theory (ZHANG Xue-gong translation) [M]. Beijing: Tsinghua University Press, 2000. (in Chinese).
- [8] ZHU Jia-qun. Support vector machine and application of support vector machine technology in image segmentation of medical image visualization [D]. Nanjing University of Science & Technology, 2007. (in Chinese) Infor [8].
- [9] DUAN Rui, GUAN Yi-hong. Multi-threshold value segmentation approach for medical images [J]. Journal of Computer Applications, 2008, 28(S2):196-197. (in Chinese).
- [10] GAO Ni. Research and application of support vector machine technology in computer-aided medical diagnosing system for breast cancer [D]. Northwest University, 2009. (in Chinese).

**Authors:** lecturer. Hong Jiang, School of Computer Science and Engineering, Xi'an Technological University, Xi'an, 710032, China, E-mail: jhong519@sina.com.