Breast Cancer Detection Based on Statistical Textural Features Classification

Al Mutaz M. Abdalla[1], Safaai Deris[2], Nazar Zaki[3] and Doaa M. Ghoneim[4]
[1, 2]*Faculty of Computer Science and Info. Systems , Universiti Teknologi Malaysia, Malaysia*
[3]*College of Info. Technology, United Arab Emirates University, UAE*
[4]*College of Science, United Arab Emirates University, UAE*
[1]*alabdullah@tawam-hosp.gov.ae;* [3]*nzaki@uaeu.ac.ae*

Abstract

Localized textural analysis of breast tissue on mammograms has recently gained considerable attention by researchers studying breast cancer detection. Despite the research progress to solve the problem, detecting breast cancer based on textural features has not been investigated in depth. In this paper we study the breast cancer detection based on statistical texture features using Support Vector Machine (SVM). A set of textural features was applied to a set of 120 digital mammographic images, from the Digital Database for Screening Mammography. These features are then used in conjunction with SVMs to detect the breast cancer. Other linear and non-linear classifiers were also employed to be compared to the SVM performance. SVM was able to achieve better classification accuracy of 82.5%.

1. Introduction

Breast cancer is the second causes of death for women around the world. Early detection of breast cancer by means of screening mammography is the most effective way to reduce the mortality rate resulting from breast cancer [1]-[2]. Despite significant recent progress, the recognition of suspicious abnormalities in digital mammograms still remains a difficult task. Recent use of localized textural and machine learning (ML) classifiers has established a new research direction to detect breast cancer. Many studies have been focused on general issue of textural analysis on mammographic images, in the context of detection of the boundary of tumors and microcalcifications [3]-[4]. However, none of these studies has taken in consideration the classification of normal, benign, and malignant cases. In this paper, we study the classification of a total of 120 digital mammographic images contain 60 normal, 30 benign, and 30 malignant cases.

2. Materials and Method

A set of statistical texture feature functions was applied to set of 120 digitized mammograms in specified regions of interest. The measurements are made on co-occurrence matrices in two different directions given a total of 1000 features.

In the first step the digitized sample consists of 120 mammographic images originating from the Digital Database for Screening Mammography (DDSM) [5] has been randomly selected (similar data will be used to test the performance of different ML techniques). The database contains 60 normal, 30 benign, and 30 malignant digitized cases at 50 micron meter and 12 bit gray level. In the second stage the region of interest has been selected which contain the suspicious Region of Interest (ROI) as shown in Figure 1. In the third stage, the feature selected from the ROI and statistical texture features are calculated for each ROI.

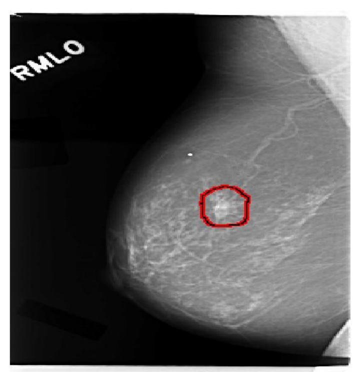

Fig. 1: Digitized mammogram showing one manually segmented malignant mass.

978-1-4244-1841-1/08/$25.00 ©2008 IEEE

In Figure 2, we illustrate the main steps in texture extraction and classification.

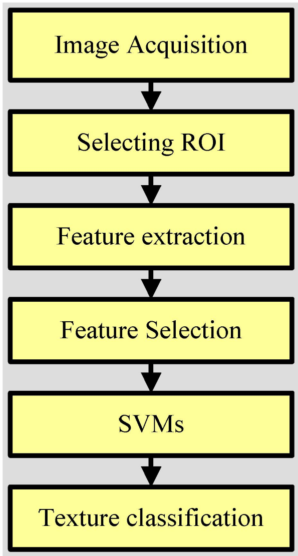

Fig. 2: Diagram showing the main steps in texture classification

3. Feature Extraction

The implementation of feature extraction procedure relies on the texture, which is the main descriptor for all the mammograms. In this paper, we concentrate on statistical descriptors that include variance, skewness, and Spatial Gray Level Dependence Matrix (SGLD) or co-occurrence matrix for texture description. These features are then used in conjunction with SVM to separate the three classes from each other.

4. Support Vector Machine

SVM is a powerful classification algorithm and well suited the given task [6]-[7]. It addresses the general problem of learning to discriminate between positive and negative members of a given class of n-dimensional vectors. The algorithm operates by mapping the given training set into a possibly high-dimensional feature space and attempting to learn a separating hyperplane between the positive and the negative examples for possible maximization of the margin between them [8]. The margin corresponds to the distance between the points residing on the two edges of the hyperplane as shown in Figure 3.

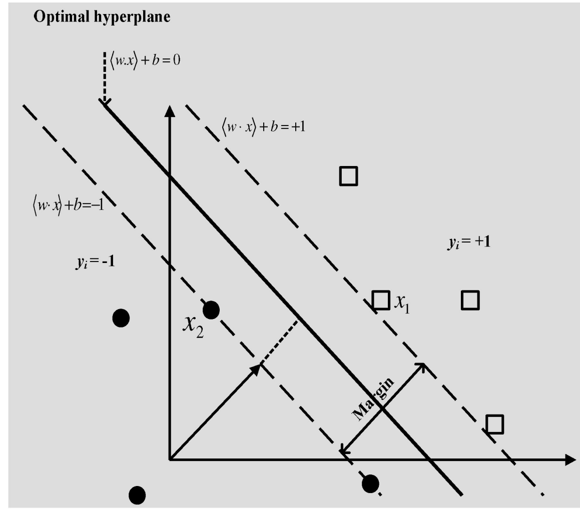

Fig. 3: Illustration of the hyperplane separation between the positive and the negative examples in Support Vector Machine

Having found such a plane, the SVM can then predict the classification of an unlabeled example. The formulation of the SVM is described as follows:

Suppose our training set S consists of labeled input vectors (x_i, y_i), $i = 1...m$ where $x_i \in \Re^n$ and $y_i \in \{\pm 1\}$. We can specify a linear classification rule f by a pair (w,b), where the normal vector $w \in \Re^n$ and the bias $b \in \Re$, via

$$f(x) = (w,b) + b \quad \text{... (1)}$$

where a point x is classified as positive if $f(x) > 0$. Geometrically, the decision boundary is the hyperplane

$$\{x \in \Re^n : (w,x) + b = 0\} \quad \text{... (2)}$$

The idea makes it possible to efficiently deal with vary high dimensional futures spaces is the use of kernels:

$$K(x,z) = \langle \phi(x) \cdot \phi(z) \rangle \quad \text{for all } x,z \in X \quad \text{... (3)}$$

where ϕ is the mapping from X to an inner product feature space. We thus get the following optimization problem:

$$\max_{\lambda} \sum_{i=1}^{m} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{m} \lambda_i \lambda_j y_i y_j K(x_i, x_j) \quad \dots (4)$$

subject to the constraints $\lambda_i \geq 0$

$$\sum_{i=1}^{m} \lambda_i y_i = 0 \qquad \dots (5)$$

Given the labeled feature vectors, we can learn SVM classifier to separate the normal, benign, and malignant classes from each other. The appeal of SVMs is twofold. First they do not require any complex tuning of parameters, and second they exhibit a great ability to generalize given small training samples. They are particularly amenable for learning in high dimensional spaces. In this particular implementation, we used the LibSVM software implemented by Chih-Chung Chang and Chih-Jen Lin. The software is available for download at http://www.csie.ntu.edu.tw/~cjlin/libsvm-tools/

5. Results

The overall results of classification obtained for the 120 image dataset are summaries in Figure 4. Performances of other ML techniques such as Linear Discriminant Analysis (LDA), Non-linear Discriminant Analysis (NDA), Principal Component Analysis (PCA) and Artificial Neural Network (ANN) were also shown in Fig 4. The results show that, SVM was able to achieve a better accuracy of 82.5%.

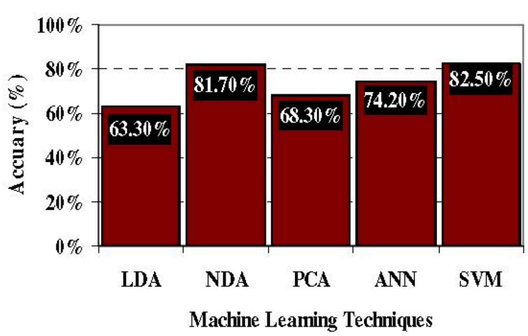

Fig. 4: Comparison of different ML techniques

6. Conclusion

Texture analysis is a promising tool for clinical decision making and one of the most valuable and promising area in breast tissue analysis. Several factors affect its performance and are still not completely understood. In this study we analyzed the region of interest on screening mammograms in order to provide radiologist an aid for estimation of tumors. Texture analysis was performed on small ROIs. Five COM measures were calculated from each ROI. The use of the statistical textural features in conjunction with SVM delivered more accurate results of 82.5%. In conclusion we can suggest that texture analysis can contribute to computer aided diagnosis of breast cancer. Completion of the proposed method should include a larger dataset and investigation of additional classification schemes.

7. Acknowledgment

The authors would like to acknowledge the help provided by the AI & Bioinformatics Laboratory at the Universiti Teknologi Malaysia and the College of Information Technology at the UAE University.

8. References

[1] R. A. Smith, "Screening women aged 40-49: where are we today?", *J Natl Cancer Inst*, 1995, pp. 1198-1199.

[2] L. Tabar, G. Fagerber and R. H. Chen, "Efficacy of breast screening by age: new results from the swedish two country trial", *Cancer*;, 1995, pp. 1412-1419.

[3] P. G. Lisboa, "A review of evidence of health benefit form artificial neural networks", *Neural Networks*, 2002, pp. 11-39.

[4] S. Arivazhagan and L. Ganesan, "Textural classification using wavelet transform", *Pat Rec Lett* , 2003, pp. 1513-21.

[5] J. T. Kwok, H. Zhu, Y. Wang, "Texture classification using the support vector machines", *Pat Re,* 2003, pp. 2883-93.

[6] Cristianini, N., and J. Shawe-Taylor, *An introduction to Support Vector Machines*, Cambridge, UK: Cambridge University Press. 2000.

[7] Vapnik, V. N., *Statistical Learning Theory*," Wiley, 1998.

[8] N. M. Zaki, S. Deris and H. Alashwal, "Protein Protein Interaction Detection Based on Substring Sensitivity Measure", *Inter. J. of Biomedical Sciences*, 2006, pp. 148-154.