Support Vector Machine Approach for Classification of Cancerous Prostate Regions

Metehan Makinacı

Abstract—The objective of this paper, is to apply support vector machine (SVM) approach for the classification of cancerous and normal regions of prostate images. Three kinds of textural features are extracted and used for the analysis: parameters of the Gauss-Markov random field (GMRF), correlation function and relative entropy. Prostate images are acquired by the system consisting of a microscope, video camera and a digitizing board. Cross-validated classification over a database of 46 images is implemented to evaluate the performance. In SVM classification, sensitivity and specificity of 96.2% and 97.0% are achieved for the 32x32 pixel block sized data, respectively, with an overall accuracy of 96.6%. Classification performance is compared with artificial neural network and k-nearest neighbor classifiers. Experimental results demonstrate that the SVM approach gives the best performance.

Keywords—Computer-aided diagnosis, support vector machines, Gauss-Markov random fields, texture classification.

I. INTRODUCTION

EARLY detection of cancer is very important for saving lives. Prostate cancer is one of the commonly diagnosed cancer in men. Diagnosis of prostate cancer requires the tissue and cell specimens. These specimens (as given in Fig. 1) are screened and analyzed by a pathologist using a microscope. Optimum medical treatment is decided according to this information reported by the pathologist. Wrong diagnosis of cancer cases is a major problem in pathology. In some cases, correct diagnosis is very hard and there can be 30-40% difference between pathologists' decisions [1]. Examples of diagnostic errors of biopsy slides are given in [2]. For the solution of this problem, computer aided diagnosis (CAD) systems can be used. A CAD system's goal is to increase the pathologists' performance and accuracy by indicating the regions of abnormalities (e.g. cancerous regions). Quantitative analysis results of prostate images can reduce the number of misclassifications and improve the diagnosis performance.

In this work, support vector machine (SVM) learning and classification is investigated for the detection of cancerous prostate regions. It is shown that SVM provides good performance among the tested methods. SVM training is based on the statistical learning theory [3]. In recent years, SVM learning is widely used in biomedical applications, including microcalcification detection [4], rheumatoid joint inflammation classification [5], fetal lung maturity analysis [6], and polyp detection [7]. The use of SVM in classification tasks has some advantages. Compared with the neural network (NN) approaches, finding a general NN classifier model to fit any data is very difficult. This difficulty is not a concern when the SVM classifier is used.

In this study, classification of cancerous and normal prostate regions is demonstrated as a two class pattern recognition task. Supervised learning is used to train the nonlinear SVM classifier. In Section II image acquisition system and histology of the specimens are presented. Feature extraction and classification procedures are given in Section III and IV, respectively. In Section V, performance analysis methods are summarized, and classification results are evaluated in Section VI. Some conclusions and further research directions are provided in Section VII.

II. MATERIALS

A. Image Acquisition

The specimen images are x100 magnified by the Leica microscope. An oil immersion objective is used. The analog image signal is acquired with a color camera and with a s-video connection the signal is transmitted to the computer. The images are acquired at 480x360 pixel 24bit/pixel format and saved.

B. Histology

Microscopic slices are the radical prostatectomy materials belonging to the patients who had a surgical operation because of prostate cancer. The radical prostatectomy materials are fixed in 10% formalin for 24-48 h. to preserve the biologic structure. Routine tissue inspection is applied to the entire structure. Paraffin blocks prepared from the tissue cut into 5-μm thick slices and stained with heamatoxylin and eosin.

M. Makinacı is with the Electrical and Electronics Engineering Department, Dokuz Eylül University, İzmir, Turkey. (phone: +90 232-412-7160; fax: +90 232-453-1085; e-mail: makinaci@eee.deu.edu.tr).

World Academy of Science, Engineering and Technology
International Journal of Medical and Health Sciences
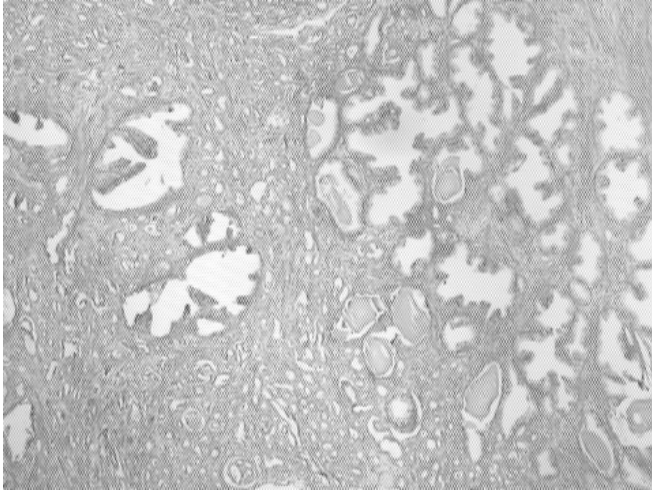Vol:1, No:7, 2007

Fig. 1 Prostate image. At the right normal prostatic glands and at the left prostatic adenocarcinoma (except the 2 glands found in the middle left of the image)

III. FEATURE EXTRACTION

46 images with labeled regions are used for the analysis. Non-overlapping 32x32 pixel blocks are taken from the labeled regions. The methods utilized for the extraction of prostate texture features are:

1) Gauss-Markov random field (GMRF) model [8]-[9].
2) Sample correlation function [8].
3) Relative entropy method [10].

GMRF model parameters are estimated using maximum likelihood method. Neighborhood systems between 1 and 9 are used for the extraction of feature set. Two dimensional time series and random field models probably are the most common class of image models. The different class of these models, such as auto models, multi-level logistic model, and hierarchical Gibbs model [11], characterizes the local statistical information in the image. In computer vision, image is defined as a 2D spatially discrete array of intensity image samples, and from this point of view, the usefulness of modeling the image with a 2D random field is obvious [12]. Sample correlation function on the lattice, defined as the lossless feature set. 5x5, 7x7, and 9x9 lattices are used on the image blocks. In relative entropy method, the set of relative entropy parameters of the normalized energy spectrum is calculated from discrete Fourier transform of the image.

A 50 element feature vector is calculated for each of the prostate region image. The feature vector consists of 16 elements from GMRF, 7 from Fourier entropy, and 27 from correlation function. Statistically significant ($p<0.001$) features are selected using t-test, giving a 27 element vector. After all, sequential forward selection method is used to select best features and to reduce the dimension of the feature vector. The resultant feature vector has 9 elements, 2 from GMRF, 4 from relative entropy, and 2 from correlation function model.

IV. SUPPORT VECTOR MACHINE

Support vector machine is a learning technique which is well-founded in modern statistical learning theory [3]. Support vector machines use the training data to crate the optimal separating hyperplane between the two classes. The optimal hyperplane maximizes the margin of the closest data points. In this way the SVM minimizes the misclassification probability of new cases. The optimal separating hyperplane is computed as a decision surface of the form:

$$g(\mathbf{x}) = \mathrm{sgn}\left(\sum_{i=1}^{l_s} d_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \right) \qquad (1)$$

where $\mathbf{x}_i$ are support vectors which are determined from the training data, $K(\ ,\ )$ is the inner product kernel which must satisfy Mercer's theorem [3], and is used to map the data from its original dimension to higher dimension so that the data is linearly separable in the mapped dimension, l_s is the number of support vectors, d_i is the class indicator $\left(d_i \in \{-1,+1\}\right)$ of $\mathbf{x}_i$, and b is bias. The coefficients α_i are calculated by solving the quadratic programming problem:

Maximize $$\sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} d_i d_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \qquad (2)$$

subject to $$\sum_{i=1}^{l} \alpha_i d_i = 0, \ \ 0 \le \alpha_i \le C \ \ \text{for } i=1,\ldots,l \qquad (3)$$

where C is a user specified positive regularization parameter used to control the amount of allowed overlap between classes. Given the expression $g(\mathbf{x})$, the decision is based on the sign of $g(\mathbf{x})$ as:

$$\text{Decision} = \begin{cases} \text{cancer} & \text{if } g(\mathbf{x}) \le 0 \\ \text{normal} & \text{if } g(\mathbf{x}) > 0 \end{cases} \qquad (4)$$

In this work a radial basis function (RBF) is chosen as an inner product kernel, which is defined as:

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2} \right) \qquad (5)$$

where $\sigma > 0$ is a user specified constant which defines the kernel width. In RBF kernel support vector machine, number and value properties of support vectors determine the number of kernels and their centers [8]. Using RBF as an inner product kernel provides classification of a non-linear set of data, which means perfect discrimination of the prostate texture features.

World Academy of Science, Engineering and Technology
International Journal of Medical and Health Sciences
Vol:1, No:7, 2007

		SVM		MLP		k-NN		
		N	C	N	C	N	C	Total
%	N	97.00	3.00	96.00	4.00	95.17	4.83	100.0
	C	3.80	96.20	6.80	93.20	7.09	92.91	100.0
Overall %		**96.60**		**94.60**		**94.04**		

N: Normal, C: Cancer

V. SYSTEM PERFORMANCE EVALUATION

As described in section III, feature vectors of 250 cancerous and 250 normal blocks from 46 prostate specimen images are used for the classification applications. Classifier performances are calculated empirically using leave-*m*-out cross validation. In this approach, total of *m* (=100) samples are selected randomly from each class with equal amounts. Rest of the data is used for training. Held out subset is used to measure the classification accuracy. Selection, training, and testing steps are repeated and the performance results are averaged for the estimation of classification rates. The procedure is used to test the SVM with various parameter settings. Performance evaluation is based on the following criteria:

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \qquad (6)$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \qquad (7)$$

$$Overall\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \qquad (8)$$

where,

TP (True Positive): Number of cancerous regions classified as cancer.
FP (False Positive): Number of normal regions classified as cancer.
TN (True Negative): Number of normal regions classified as normal.
FN (False Negative): Number of cancerous regions classified as normal.

Sensitivity is the correct classification rate of cancerous block, specificity is the correct classification rate of normal blocks, and overall performance is the correct classification rate of both normal and cancerous blocks. Another important method of performance evaluation method is receiver operating characteristic (ROC) curve [13]. ROC curve demonstrates the sensitivity and specificity performance of the classifier for the continuum of the decision threshold.

VI. RESULTS

For comparison, multilayer perceptron (MLP) and k-nearest-neighbor (k-NN) classifiers [13] are utilized. Single hidden layer MLP is trained and tested with 2 to 15 hidden nodes. Maximum average classification rates are achieved with 5 and 10 hidden neurons. k-NN is tested with number of neighbors from 1 to 20. Calculated sensitivity, specificity, and overall accuracy rates of the classifiers are summarized in Table I. SVM has the highest sensitivity, specificity, and overall accuracy rates of 96.2%, 97.0%, and 96.6%, respectively.

A series of parameter values are used for SVM training. Plot of the overall misclassification rate for various parameter values is given in Fig. 2. Minimum misclassification rate of 3.4% is achieved for certain range of parameter values.

Quantitative analysis is performed about the detection performances of support vector machine, multilayer perceptron, and k-nearest-neighbor classifiers using receiver operating characteristic curves. Test data of prostate regions are used for the ROC curves to evaluate the results objectively. ROC curves are given in Fig. 3. Area under the ROC curve is an important quantity. Closer to 1 means better classifier performance. Area under ROC curve values of 0.9752, 0.9870, and 0.9779 are calculated for SMV, k-NN, and MLP classifiers, respectively.
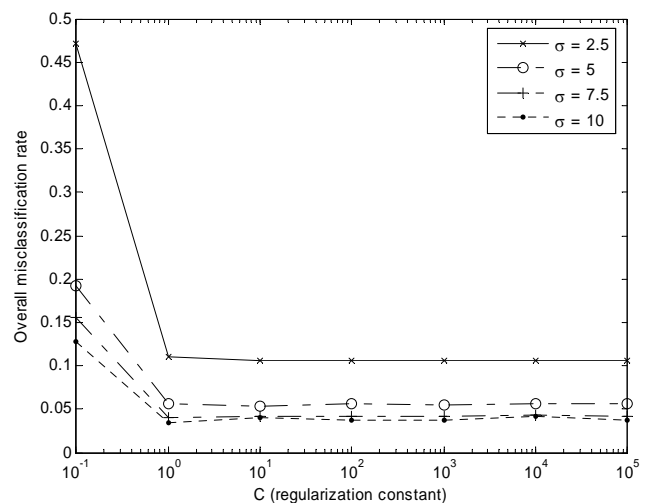

Fig. 2 Plot of SVM misclassification rate versus regularization parameter for given σ values
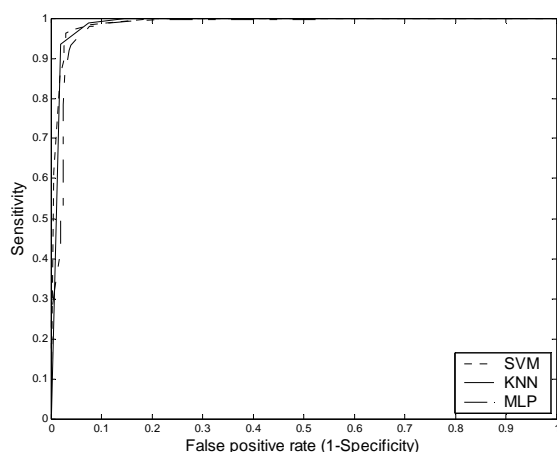
Fig. 3 ROC curves of classifiers

VII. CONCLUSION

In this work, support vector machine based pattern recognition methods are demonstrated for classification of normal and cancerous prostate regions. Experimental results show that the proposed method can be used (with additional improvements) as a CAD system in a pathology laboratory. A high performance, accurate and real time diagnosis system will aid to the pathologist and to the patients.

ACKNOWLEDGMENT

REFERENCES

[1] U. Schenck, and W. Planding, "Quantitation of visual screening technique in cytology," in *Proc. Image Analysis in Medicine, II. National Symposium*, pp. 7-14, 1998.
[2] D. Kopec, M.H. Kabir, D. Reinharth, O. Rothschild, and J.A. Castiglione, "Human Errors in Medical Practice: Systematic Classification and Reduction with Automated Information Systems," *J. of Medical Systems*, vol. 27, no. 4, pp. 297-313, Aug. 2003.
[3] Vapnik, V., *Statistical Learning Theory*, Wiley, New York, 1998.
[4] I. El-Naqa, Y. Yang, M.N. Wernick, N.P. Galatsanos, R.M. Nishikawa, "A Support Vector Machine Approach for Detection of Microcalcifications", IEEE Tran. Med. Imaging, vol. 21, no. 12, pp. 1552-1563, Dec. 2002.
[5] A. Schwaighofer, P. Mayer, A. Krause, J. Beuthan, H. Rost, G. Metzger, G.A. Müller, A.K. Scheel, "Classification Of Rheumatoid Inflammation Based on Laser Imaging", IEEE Tran. on Biomed. Eng., vol. 50, no. 3, pp. 375-382, March 2003.
[6] K.N. Prakash, A.G. Ramakrishnan, S. Suresh, T.W.P. Chow, "Fetal Lung Maturity Analysis Using Ultrasound Inage Features", IEEE Tran. on Inf. Tech. in Biomed., vol. 6, no. 1, pp. 38-45, March 2002.
[7] S.B. Göktürk, C. Tomasi, B. Acar, C.F. Beaulieu, D.S. Paik, R.B. Jeffrey, J. Yee, S. Napel, "A Statistical 3-D Pattern Processing Method for Computer-Aided Detection of Polyps in CT Colonography", IEEE Tran. on Med. Imag., vol. 20, no. 12, pp. 1251-1260, Dec. 2001.
[8] S. Chatterjee, "Classification of natural textures using Gaussian Markov random field models", in *Markov Random Fields, Theory and Applications*, R. Chellappa, A. Jain, Ed., Academic Press, 1993, pp. 159-177.
[9] B.S. Manjunath, R. Chellappa, "Unsupervised texture segmentation using Markov random field models," *IEEE Tran. Patt. Anal. Machine Intel.*, vol. 13, pp. 478-482, 1991.
[10] M.E. Jernigan, F. D'Astous, "Entropy-based texture analysis in the spatial frequency domain," *IEEE Tran. Patt. Anal. Machine Intel.*, vol. 6, pp. 237-243, 1984.
[11] S.Z. Li, *Markov Random Field Modelling in Computer Vision*, Springer-Verlag, 1995.
[12] A. Rosenfeld, "Image modelling during the 1980s: A brief overview", in *Markov Random Fields, Theory and Applications*, R. Chellappa, A. Jain, Ed., Academic Press, 1993, pp. 1-10.
[13] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, 2nd ed., John Wiley & Sons, Inc., 2001, ch. 2,4,6.