

# Tumor Classification in Histological Images of Prostate Using Color Texture

Ali Tabesh and Mikhail Teverovskiy  
Aureon Laboratories, Inc.  
Yonkers, NY 10701 USA  
{ali.tabesh, mikhail.teverovskiy}@aureon.com

**Abstract**— We present a wavelet-based color texture approach to tumor classification in the histological images of prostate. We extend our previous work on intensity images to incorporate color information and rotational invariance. Our results on a set of 367 images stained using hematoxylin and eosin indicate that incorporating color and rotational invariance into the features significantly reduces the classification error. We obtained a 5-fold cross-validation error of 8.7% for intensity images and no rotational invariance. Incorporation of color and rotational invariance lowered the error to 4.4%, using the CIELAB space. Both results were obtained using support vector machine classifiers along with the linear kernel. The improvement achieved in classification accuracy corresponds to a significance level of 0.0093.

## I. INTRODUCTION

Prostate cancer is the most prevalent form of cancer and the second most common cause of cancer deaths among men in the United States [1]. One of the most reliable ways for diagnosing prostate cancer is the examination of the histological specimens under the microscope by a pathologist.

Prostate tissue, depicted in Fig. 1, consists of major building blocks referred to as gland units surrounded by fibromuscular tissue, called stroma, which holds the gland units together. Each gland unit is composed of rows of epithelial cells located around a duct in the tissue, named the lumen. When cancer occurs (Fig. 1(b,c)), epithelial cells replicate in an uncontrolled way, disrupting the regular arrangement of gland units.

It is well-known that pathologists are prone to subjectivity and limited intra- and inter-observer reproducibility in their labeling of histological images. This phenomenon has been extensively studied in the context of the Gleason grading of prostate tissue [2]. Moreover, high-throughput applications require rapid and low-cost image processing not achievable by human observers.

A few research efforts have been directed towards automating prostate cancer diagnosis in histological images [3, 4, 5]. Diamond *et al.* [3] used shape and intensity texture features to identify stroma, normal, and cancerous image blocks. Normal tissue is identified as epithelial cells in the vicinity of lumens. The remaining tissue is classified into stroma and cancerous regions based on a Haralick texture feature. The authors report an accuracy of 79.3% in classifying image regions.

Roula *et al.* [4] presented a system for classifying tissue image blocks into the stroma, benign prostatic hyperplasia, prostatic intraepithelial neoplasia (PIN), and prostate cancer classes. The images are captured in multiple spectral bands. For each image block, texture features and the area occupied by nuclei and lumens relative to the area of the image block are computed. These features are then subjected to the Karhunen-Loeve transform (KLT) and a few of the resulting principal components are used to classify the image block using the linear Gaussian classifier. The authors report a cross-validation (CV) classification accuracy of 94%.

In [5, 6], we proposed the use of wavelet texture features

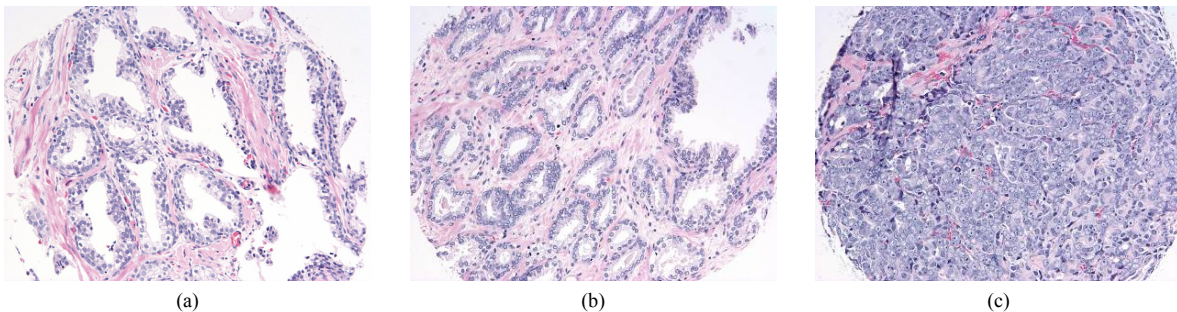


Fig. 1. Samples of hematoxylin and eosin (H&E)-stained prostate tissue; (a) normal; (b) well-differentiated, and (c) poorly differentiated cancer.

among other features for tumor classification and Gleason grading. Color images were first converted to intensity images. The wavelet transform of the intensity image was then computed, and variances of subband coefficients were obtained. The wavelet transform features resulted in a 5-fold CV accuracy of 91.3% using the linear Gaussian classifier.

In this study, we extend our previous work to incorporate color information and rotational invariance into wavelet texture features. In the next section, we describe the features used in this study. In Section III, feature selection and classification algorithms are described. Classification results are presented in Section IV. Finally, the conclusions are described in Section V.

## II. COLOR TEXTURE FEATURES

### A. Subband Variances as Texture Descriptors

Consider the tissue images in Fig. 1. It can be seen that normal tissue contains fewer epithelial cells which are stained blue, more lumens which are not stained and stay white, and larger stroma regions which are stained red. As the tissue transforms from normal to high-grade tumor, its color and texture properties change significantly. Proliferation of epithelial cells alters the color composition of the image from predominantly red to blue. Moreover, the image texture shifts from heterogeneous and coarse-grain to homogeneous and fine-grain.

The above changes in color and texture may be captured via wavelet transform features, which characterize texture at different color components, scales, and orientations [7]. The wavelet representation we used was based on an  $L$ -level, dyadic transform and the symlet 4 wavelet filter [8]. Let  $w_{bc}(i, j)$ ,  $i = 1, \dots, M$ ,  $j = 1, \dots, N$ ,  $b = 1, \dots, B$ ,  $c = 1, \dots, C$ , denote a wavelet coefficient in subband  $b$  of color component  $c$ , where  $MN$  is the number of coefficients in the subband,  $B = 3L + 1$  is the number of bands, and  $C$  is the number of color components. Variance  $\sigma_{bc}^2$  given by

$$\sigma_{bc}^2 = \frac{1}{MN} \sum_i \sum_j w_{bc}(i, j) - \left( \frac{1}{MN} \sum_i \sum_j w_{bc}(i, j) \right)^2, \quad (1)$$

was used to represent texture across color components and subbands. The total number of features is given by  $BC$ . Via univariate analysis, we found that the classification accuracy for individual features does not improve for  $L > 4$ . Thus, we set  $L = 4$ . Thus,  $4 \times 3 + 1 = 13$  and  $3 \times (4 \times 3 + 1) = 39$  features were obtained for intensity and color images, respectively.

### B. Color Spaces

Typical image capture devices such as CCD cameras produce raw image data in the RGB color space. However, this space is not the optimal choice for representing color information in texture analysis. For instance, in [9] the perceptually uniform CIELAB space, as well as the approximately perceptually uniform HSV space have been shown to outperform the RGB representation in color texture classification based on Gabor filters. Motivated by this and

similar results, we considered the RGB, CIELAB, and HSV spaces in our experiments. The conversion equations between these spaces are given in [9].

We also considered the space obtained by applying the KLT to the RGB pixel values. Given a vector of RGB values  $\mathbf{I}$ , rows of the KLT  $\mathbf{A}$  consist of the eigenvectors of  $\Sigma_{\mathbf{I}}$ , the covariance matrix of  $\mathbf{I}$ . The matrix  $\Sigma_{\mathbf{I}}$  is estimated separately for each input image. Our experience showed that despite the inter-image variations,  $\mathbf{A}$  remains fairly consistent across images. A typical  $\mathbf{A}$  was

$$\mathbf{A} = \begin{bmatrix} 0.60 & 0.67 & 0.44 \\ -0.79 & 0.41 & 0.45 \\ -0.12 & 0.62 & -0.78 \end{bmatrix}.$$

For each of the four color spaces mentioned above, we obtained a set of wavelet features. As a baseline for comparison, we also considered the wavelet features obtained from intensity images as well. Intensity images were obtained as the Y component after converting the RGB images into the YCbCr space [10].

### C. Rotational Invariance

Variance features described in Section II.A are not invariant to rotations of the input image. However, we can make them invariant to rotations by multiples of 90 degrees via simple transformations of the features. We shall refer to this as approximate rotational invariance. First, we note that  $\sigma_{\text{HH}}^2$  are invariant to such rotations, where HH denotes the diagonal detail subband. For horizontal and vertical detail subbands, denoted as HL and LH, respectively, let  $f_i$ ,  $i = 1, 2$ , be symmetric functions such that

$$f_i(\sigma_{\text{HL}}^2, \sigma_{\text{LH}}^2) = f_i(\sigma_{\text{LH}}^2, \sigma_{\text{HL}}^2). \quad (2)$$

Features  $(f_1, f_2)$  are approximately rotationally invariant. In our experiments, we considered the transformation pair

$$(f_1, f_2) = (\min(\sigma_{\text{HL}}^2, \sigma_{\text{LH}}^2), \max(\sigma_{\text{HL}}^2, \sigma_{\text{LH}}^2)). \quad (3)$$

## III. FEATURE SELECTION AND CLASSIFICATION

### A. Feature Selection

Feature selection algorithms are characterized by the optimality criterion with respect to which the quality of a feature subset is evaluated, and the search strategy for finding the optimal feature combination. In our experiments, the classification accuracy estimated using the leave-one-out (LOO) procedure [11] served as the optimality criterion  $J$  and the sequential forward search (SFS) algorithm [11] was used as the search strategy. The SFS algorithm begins with selecting the individual feature that maximizes  $J$ . Each consequent stage of the algorithm consists of augmenting the set of already selected features with a new feature such that the resulting

feature subset maximizes  $J$ . The process of adding new features is continued until  $J$  reaches a maximum.

In our experiments, it was noted that the accuracy of feature subsets is not always a unimodal function of the number of features selected. That is, the feature combination corresponding to the first peak in the classification accuracy is not necessarily the optimal subset. To alleviate this problem, the SFS algorithm was continued until fifteen features were selected. Then, among the feature combinations obtained, the optimal subset was chosen as the one that maximized the accuracy.

### B. Classification

Statistical classifiers fall into two broad categories of parametric and non-parametric methods. Parametric methods rely on the assumption that the functional forms of the class-conditional distributions are known (e.g., Gaussian), whereas non-parametric methods make minimal assumptions about the form of the distributions. The choice of the classifier type depends on the sample size and prior knowledge about the class-conditional distributions. In this study, we used both parametric (linear and quadratic Gaussian) and non-parametric ( $k$ NN and support vector machine (SVM)) classifiers.

Let  $\mathbf{X} = [\sigma_{11}^2, \dots, \sigma_{bc}^2]^T$  denote the vector of variance features. Given the discriminant function  $f(\mathbf{X})$ , the decision rule is as follows. If  $f(\mathbf{X}) < 0$ , classify  $\mathbf{X}$  into class 0; if  $f(\mathbf{X}) > 0$ , assign  $\mathbf{X}$  to class 1; and if  $f(\mathbf{X}) = 0$ , pick a class at random.

1) *Gaussian classifiers*: The quadratic Gaussian classifier is given by [11]

$$f(\mathbf{X}) = (\mathbf{X} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{X} - \boldsymbol{\mu}_0) - (\mathbf{X} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1} (\mathbf{X} - \boldsymbol{\mu}_1) + \ln \frac{|\boldsymbol{\Sigma}_0|}{|\boldsymbol{\Sigma}_1|} - \ln \frac{P_0}{P_1}. \quad (4)$$

where  $P_i$ ,  $\boldsymbol{\mu}_i$ ,  $\boldsymbol{\Sigma}_i$  are the prior probability, mean, and covariance matrix of class  $c_i$ ,  $i = 0, 1$ , respectively, and  $|\cdot|$  denotes matrix determinant. Parameters  $P_i$ ,  $\boldsymbol{\mu}_i$ ,  $\boldsymbol{\Sigma}_i$  were obtained from the training set via maximum likelihood estimation [11].

For  $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}$ , (4) simplifies to the linear Gaussian classifier given by

$$f(\mathbf{X}) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1} \mathbf{X} + \frac{1}{2} (\boldsymbol{\mu}_0^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1) - \ln \frac{P_0}{P_1} \quad (5)$$

The combination of SFS and Gaussian classifiers was realized using Tooldiag [12], which in turn was wrapped in a MATLAB [13] script to implement CV.

2) *kNN classifier*: The  $k$ NN classifier is given by [11]

$$f(\mathbf{X}) = \ln \left( \frac{n_1}{n_0} \right), \quad (6)$$

where  $n_i$  is the number of samples belonging to class  $c_i$  that fall within the set of  $k$  nearest samples to  $\mathbf{X}$ , i.e.,  $n_0 + n_1 = k$ . Nearness is defined based on a distance measure  $d(\cdot, \cdot)$  between  $\mathbf{X}$  and the samples  $\mathbf{Y}_i$  in the training set. Here, we use the weighted Euclidean distance given by

$$d(\mathbf{X}, \mathbf{Y}_i) = \sum_{j=1}^N \frac{(X_j - Y_{ij})^2}{s_j^2}, \quad (7)$$

where  $s_j^2$  denotes the sample variance of  $Y_{ij}$ . Scaling the features to the same standard deviation prevents the features with larger variances from dominating the Euclidean distance.

Parameter  $k$  is selected such that the LOO classification accuracy on the training set is maximized. In our experiments,  $k$  was chosen from the set  $\{1, 3, 5, 11, 21, 51\}$ . The combination of SFS and the  $k$ NN classifier was realized using Tooldiag [12], which in turn was wrapped in a MATLAB [13] script to implement CV.

3) *SVM classifier*: The SVM classifier first proposed by Vapnik [14] is given by

$$f(\mathbf{X}) = \mathbf{w}^T \boldsymbol{\Phi}(\mathbf{X}) + b. \quad (8)$$

where  $\boldsymbol{\Phi}(\cdot)$  is a nonlinear function, and parameters  $\mathbf{w}$  and  $b$  are determined from the training set. This is accomplished via minimization of the *structural risk* function  $J(\mathbf{w}, \boldsymbol{\xi})$  given by

$$J(\mathbf{w}, \boldsymbol{\xi}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^L \xi_i, \quad (9)$$

subject to  $d_i f(\mathbf{X}) \geq 1 - \xi_i$  and  $\xi_i \geq 0$ ,  $i = 1, \dots, L$ ,

where  $C$  is a user-specified, positive parameter,  $L$  is the number of training samples,  $\xi_i$  are slack variables, and  $d_i = -1$  for class 0 and  $+1$  for class 1, respectively.

The risk function in (9) constitutes a balance between model complexity (first term) and classification error (second term). Parameter  $C$  controls this trade-off. The use of model complexity to constrain the minimization of empirical risk aims to avoid overfitting to the training data.

The dual representation of the minimization in (9) yields an equivalent form of (8 as

$$f(\mathbf{X}) = \sum_{i=1}^{N_s} d_i \alpha_i K(\mathbf{X}, \mathbf{S}_i) + b. \quad (10)$$

where  $\alpha_i$  denote the Lagrange multipliers associated with the constraints  $d_i f(\mathbf{X}) \geq 1 - \xi_i$ ,  $\mathbf{S}_i$  are support vectors,  $N_s$  is the number of support vectors, and  $K(\mathbf{X}, \mathbf{S}_i) = \boldsymbol{\Phi}(\mathbf{X})^T \boldsymbol{\Phi}(\mathbf{S}_i)$  is the kernel function. In this paper, we used the linear and Gaussian kernels. The linear kernel is given by

$$K(\mathbf{X}, \mathbf{S}_i) = \mathbf{X}^T \mathbf{S}_i, \quad (11)$$

and the Gaussian kernel is given by

$$K(\mathbf{X}, \mathbf{S}_i) = \exp(-\gamma \|\mathbf{X} - \mathbf{S}_i\|^2). \quad (12)$$

where  $\gamma$  denotes the kernel width. We also considered combining the above kernels with whitening the features. The whitened kernel  $K_w(\mathbf{X}, \mathbf{S}_i)$  is given by

$$K_w(\mathbf{X}, \mathbf{S}_i) = K(\mathbf{A}\mathbf{X}, \mathbf{S}_i), \quad (13)$$

where  $\mathbf{A} = \Sigma^{-1/2}$ .

Parameters  $C$  and  $\gamma$  are jointly selected such that the classification accuracy on the training set is maximized. Values considered for  $C$  and  $\gamma$  were  $\{1, 10, 10^2, 10^3, 10^4\}$  and  $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$ , respectively. We used the LIBSVM [15] implementation of SVM classifiers and wrapped it in a MATLAB [13] script to perform SFS and CV.

### C. Error Estimation

We have used the  $K$ -fold CV method for error estimation in our studies [16]. CV consists of splitting the dataset into  $K$  roughly equal sets. Classifier training and feature selection is done  $K$  times, each time using the samples in  $K-1$  sets. Each of these training sets is referred to as a CV training set. The samples in the remaining set, referred to as a CV test set, are then used for error estimation. The overall performance of the  $K$  trained models is obtained as the mean of the accuracies over the  $K$  left-out sets. We set  $K=5$  in our experiments.

On each of the  $K$  CV training sets, classifier design consists of feature selection and classifier training. Feature subsets were evaluated via the LOO method on the CV training set [11], where LOO is a special case of CV with  $K=n$ . For SVM classifiers, the 5-fold CV method was used due to the computational complexity of training these classifiers.

## IV. EXPERIMENTAL RESULTS

### A. Image Data

Tissue samples used for tumor classification were obtained from tissue microarray cores stained using H&E. The specimens consisted of 218 tumor and 149 non-tumor samples, totaling 367 cores. The images were captured at  $20\times$  magnification using a light microscope equipped with a SPOT Insight QE Color Digital Camera (KAI2000) producing  $1600\times 1200$ -pixel color images and were stored as 24-bit-per-pixel TIFF images. All images were captured under similar illumination conditions and labeled by an expert pathologist. Images labeled as tumor had between 5% and 100% of their area covered by tumor, and contained Gleason grades 2 to 5. Images labeled as non-tumor included samples of PIN, prostatitis, and normal tissue. Examples of the images are shown in Fig. 1.

Prior to feature extraction, the images were subjected to histogram matching against a reference image [10]. This was accomplished via applying the transformation  $F_r^{-1}[F_l(x)]$  to the pixel values  $x$  in each of the red, green, and blue channels, where  $F_l$  and  $F_r$  denote the cumulative distribution function

(cdf) of pixel values for the input and reference images, respectively. Histogram matching aims to alleviate color variations due to staining and illumination conditions.

### B. Results

1) *Univariate analysis*: To gain insight into the meaning and value of features, univariate analysis was performed to select the individually best features. Features resulting in the highest LOO accuracy on the entire dataset of 367 images were selected from each feature set using the quadratic Gaussian classifier. Table I(a) shows the results of univariate analysis. The following observations can be made about the results. First, incorporation of color information and rotational invariance consistently improves classification accuracy regardless of the choice of color space.

Second, for both invariant and non-invariant feature sets, the highest accuracy is achieved using a feature obtained from a chrominance component in a perceptually uniform or approximately uniform color space. Specifically, for the non-invariant and invariant features, the B component of the CIELAB space and the saturation component of the HSV space yielded the highest accuracies, respectively. This is perhaps due to the fact that the chrominance components are less sensitive to different illumination and staining conditions across the image set.

Third, the highest accuracy is always achieved for a feature in the third or fourth decomposition level. This observation indicates that a larger proportion of the tumor classification information lies in those subbands than others.

Table I(b) shows the means and standard deviations of the features in Table I(a). For color images, the statistics of the feature that performed best across all color spaces is reported. Note that the mean variance is always smaller for the non-tumor class than for the tumor class. This is true for all features except for the mean variance of the approximation (LL) subband, where the trend is in the opposite direction. These patterns are consistent with our intuition, mentioned in Section

TABLE I  
(A) TOP FEATURES VIA UNIVARIATE ANALYSIS AND (B) THEIR CORRESPONDING MEANS AND STANDARD DEVIATIONS

(a)

Color Space	Non-invariant		Invariant	
	C, L, F <sup>a</sup>	Accuracy (%)	C, L, F	Accuracy (%)
RGB	2, 3, HH	89.6	1, 3, min	91.6
HSV	2, 4, HH	90.2	2, 3, min	92.9
CIELAB	3, 4, HH	91.0	3, 3, min	91.8
KLT	1, 4, HH	89.6	1, 3, min	91.3
Intensity	1, 4, HH	89.9	1, 3, min	91.0

<sup>a</sup> Triplet (C, L, F) denote color component, decomposition level, and feature, respectively. Decomposition levels 1 and 4 correspond to the highest and lowest resolution levels, respectively. For invariant features, feature may be one of minimum (min), maximum (max), or HH. For non-invariant features, feature may be one of HL, LH, or HH.

(b)

Features	Color Space	mean $\pm$ std dev	
		Non-Tumor	Tumor
Non-invariant	Intensity	7770 $\pm$ 1720	12400 $\pm$ 1930
	CIELAB	270.0 $\pm$ 64.5	462.0 $\pm$ 72.7
Invariant	Intensity	2830 $\pm$ 716	5100 $\pm$ 922
	HSV	0.0303 $\pm$ 0.0072	0.0503 $\pm$ 0.0084

II, that non-tumor images have coarser texture than tumor images.

2) *Multivariate analysis*: Table II shows the results of multivariate analysis. For color images, the 5-fold CV accuracies are reported for the rotationally invariant features. As baselines for comparison, the accuracies for non-invariant and invariant intensity features are shown as well. For each feature set, feature selection and classification was conducted using the linear and quadratic Gaussian,  $k$ NN, and SVM (with RBF, whitened RBF, and whitened linear kernels) classifiers. The results in the table correspond to the classifier achieving the highest accuracy for a given color space.

As in univariate analysis, the classification accuracies for color wavelet features are consistently higher than for intensity features. The highest accuracy of 95.6% was achieved using the CIELAB color space and the SVM classifier with the whitened linear kernel (SVM-Lin-W). Non-invariant intensity features achieved an accuracy of 91.3% using the same classifier. The two-sample  $t$  test [17] shows that the difference between these two accuracies is statistically significant at a level of 0.0093.

The table also suggests that simpler classifiers (e.g., linear Gaussian, and SVM-Lin-W) are often preferred over more complex ones (e.g., SVM-RBF).

## V. CONCLUSIONS

In this paper, we presented a study of wavelet-based color texture features for tumor classification in histological images of prostate. An automated approach brings objectivity and reproducibility to cancer diagnosis and is suited to high-throughput applications. We studied the impact of the choice of color space, rotational invariance, and classifiers.

Our results indicate that incorporation of color and rotational invariance has a consistently positive impact on the classification accuracy. We obtained a 5-fold CV error of 8.7% for intensity images and no rotational invariance. Incorporation of color, via the perceptually uniform CIELAB space, and rotational invariance lowered the error to 4.4%. Both results were obtained using SVM classifiers along with the linear kernel. The improvement achieved in classification accuracy corresponds to a significance level of 0.0093.

The results of our univariate analysis show that most of the classification information lies in the coarser resolution levels. This implies that low-power images of the cores may be used for rapid spotting of candidates for more elaborate analysis at higher magnifications.

TABLE II  
RESULTS OF MULTIVARIATE ANALYSIS

Color Space	Classifier	Accuracy (%)	CI* (%)	Sens/Spec <sup>b</sup> (%)
RGB	Linear	94.3	2.4	94.5 / 94.0
HSV	Quadratic	94.0	2.4	95.4 / 91.9
CIELAB	SVM-Lin-W	95.6	2.1	95.0 / 96.6
KLT	SVM-RBF	94.3	2.4	95.4 / 92.6
Intensity <sup>c</sup>	SVM-Lin-W	91.3	2.9	91.7 / 90.6
Intensity <sup>d</sup>	Linear	92.4	2.7	92.2 / 92.6

\* The 95% confidence interval for the accuracy estimate [17].

<sup>b</sup> Sensitivity and specificity. Sensitivity is measured with respect to the tumor class.

<sup>c</sup> Rotationally non-invariant.

<sup>d</sup> Rotational invariant.

## ACKNOWLEDGMENTS

We would like to express our gratitude to Aureon Laboratories' CEO, Dr. Vijay Aggarwal, and scientific founders, Professors Carlos Cordon-Cardo, Jose Costa, and Robert Singer, for their continuous support of this research. We also thank Drs. Ricardo Mesa-Tejada and Gerardo Fernandez for helpful discussions of the pathology of prostate cancer.

## REFERENCES

- [1] American Cancer Society, *Cancer Facts & Figures 2005*. Atlanta, GA: American Cancer Society, 2005.
- [2] D. F. Gleason, "Histologic grading of prostate cancer: a perspective," *Human Pathol.*, vol. 23, pp. 273-279, 1992.
- [3] J. Diamond, N. Anderson, P. Bartels, R. Montironi, and P. Hamilton, "The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia," *Human Pathol.*, vol. 35, pp. 1121-1131, 2004.
- [4] M. A. Roula, J. Diamond, A. Bouridane, P. Miller, and A. Amira, "A multispectral computer vision system for automatic grading of prostatic neoplasia," in *Proc. Proc. IEEE Int. Symp. Biomed. Imaging*, Washington, DC, 2002, pp. 193- 196.
- [5] A. Tabesh, V. P. Kumar, D. Verbel, A. Kotsianti, M. Teverovskiy, and O. Saidi, "Automated prostate cancer diagnosis and Gleason grading of tissue microarrays," in *Proc. SPIE Int. Symp. on Medical Imaging*, vol. 5747, San Diego, CA, Feb. 2005, pp. 58-70.
- [6] A. Tabesh, M. Teverovskiy, H. Pang, V. P. Kumar, D. Verbel, A. Kotsianti, and O. Saidi "Multi-feature prostate cancer diagnosis and Gleason grading of histological images," to appear in *IEEE Trans. Medical Imag.*, 2007.
- [7] A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 1186-1191, 1993.
- [8] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
- [9] G. Paschos, "Perceptually uniform color spaces for color texture analysis: An empirical evaluation," *IEEE Trans. Image Processing*, vol. 10, pp. 932-937, 2001.
- [10] R. C. Gonzales and R. E. Woods, *Digital Image Processing*. New York: Addison-Wesley, 1992.
- [11] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.
- [12] <http://www.inf.ufes.br/~thomas/home/tooldiag.html>.
- [13] <http://www.mathworks.com>.
- [14] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
- [15] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," 2001. Available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [16] R. O. Duda, R. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
- [17] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 5th ed. Pacific Grove, CA: Duxbury, 2000.