AUTOMATED GRADING OF PROSTATE CANCER USING ARCHITECTURAL AND TEXTURAL IMAGE FEATURES

Scott Doyle, Mark Hwang,
Kinsuk Shah, Anant Madabhushi
Rutgers, State University of New Jersey,
Department of Biomedical Engineering,
New Brunswick, New Jersey, USA 08854.
Email: anantm@rci.rutgers.edu

Michael Feldman, John Tomaszeweski
University of Pennsylvania,
Department of Surgical Pathology,
Philadelphia, Pennsylvania, USA 19104.

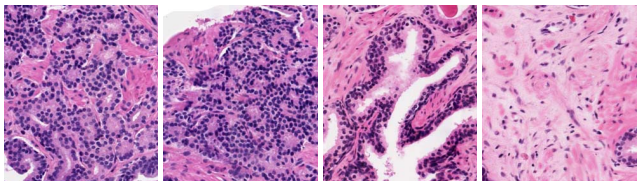
ABSTRACT

The current method of grading prostate cancer on histology uses the Gleason system, which describes five increasingly malignant stages of cancer according to qualitative analysis of tissue architecture. The Gleason grading system has been shown to suffer from inter- and intra-observer variability. In this paper we present a new method for automated and quantitative grading of prostate biopsy specimens. A total of 102 graph-based, morphological, and textural features are extracted from each tissue patch in order to quantify the arrangement of nuclei and glandular structures within digitized images of histological prostate tissue specimens. A support vector machine (SVM) is used to classify the digitized histology slides into one of four different tissue classes: benign epithelium, benign stroma, Gleason grade 3 adenocarcinoma, and Gleason grade 4 adenocarcinoma. The SVM classifier was able to distinguish between all four types of tissue patterns, achieving an accuracy of 92.8% when distinguishing between Gleason grade 3 and stroma, 92.4% between epithelium and stroma, and 76.9% between Gleason grades 3 and 4. Both textural and graph-based features were found to be important in discriminating between different tissue classes. This work suggests that the current Gleason grading scheme can be improved by utilizing quantitative image analysis to aid pathologists in producing an accurate and reproducible diagnosis.

1. INTRODUCTION

The current protocol for prostate cancer diagnosis involves manual analysis of prostate biopsy tissue by a pathologist to determine the presence or absence of cancer in a tissue sample, followed by Gleason grading to assign a number between 1 (early stage cancer) and 5 (highly infiltrative cancer) to grade the invasiveness of the cancer. Gleason grading is currently done by visual analysis of the arrangement, size, and shape of the gland structures and nuclei within the tissue sample. Recently however a number of problems have been identified with the Gleason system:

- *Variability:* Studies have shown significant levels of inter- and intra-observer variability in manual grading of prostate cancer [1] with the rates of undergrading tissue patterns (assigning a tissue a grade lower than its actual grade) as high as 47% [2], [3]. Figure 1 illustrates these clinical problems, as the difference between intermediate cancer grades (grade 3 (Figure 1 (a)) and grade 4 (Figure 1 (b))) is more difficult to distinguish than, say, between grades 1 and 5.
- *Lack of Standardization:* Changes in grading methodologies over the years and variations in the grading criteria at different institutions make consistent grading difficult [4].
- *Limited Feature Set:* The Gleason system distinguishes different cancer grades solely on tissue architecture while other attributes such as texture and cellular and glandular morphology are not accounted for.



(a) (b) (c) (d)

Fig. 1. Digitized histological tissue patches corresponding to (a) Gleason grade 3 adenocarcinoma, (b) grade 4 adenocarcinoma, (c) benign epithelium, and (d) benign stroma.

Computer-aided diagnosis (CAD) techniques for prostate cancer have been proposed to supplement manual visual analysis of biopsy core specimens. At low magnification, systems have been proposed that employ low-level image features such as color and intensity to classify image pixels as "cancer" or "non-cancer," i.e. only object detection [5]. In order to distinguish cancer grades, more detailed information at high magnification such as the arrangement of nuclei and glands is required. Benign epithelial tissue contains large, irregularly-shaped glands, while Gleason grades 2 and 3 have a fewer number of organized, circular glands. Cancer of Gleason grades 4 and 5 have a large number of nuclei in disorganized, infiltrative sheets, with gland lumen that is almost completely occluded and small or nonexistent stromal regions between the glands [4].

Attempts at automated prostate cancer grading have typically (i) only addressed classification between "high-grade" and "low-grade" cancer, which is not the most difficult clinical problem, and (ii) do not utilize the full range of attributes that could possibly be used. Most of the variability and error in pathologist scoring of Gleason grades is in discriminating between intermediate grades (3 and 4) [2].

In [6], we presented a CAD system to detect potentially cancerous areas on digitized prostate histology. In this work, we describe a system to classify prostate histology into one of four categories: Gleason grade 3 adenocarcinoma, grade 4 adenocarcinoma, benign epithelium, or benign stroma. Over 100 features are automatically

This work was made possible due to grants from the Coulter foundation (WHCF 4-29349, WHCF 4-29368), the Busch Biomedical Award and the Technology Commercialization Fund at Rutgers University

Authorized licensed use limited to: Rutgers University. Downloaded on November 14, 2008 at 15:27 from IEEE Xplore. Restrictions apply.

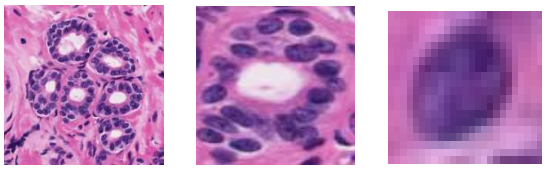
computed from within a tissue patch, including 13 features describing nuclear arrangement, 32 features describing gland size and arrangement, and 57 image texture features. Non-linear dimensionality reduction methods are applied to the feature set and a support vector machine (SVM) is used to classify the different digitized tissue patches into a lower-dimensional space. A Boosting algorithm [7] is applied to find the features that contribute the most discriminating information to the classifier. The novelty of our work is that our system focuses on distinguishing between intermediate Gleason grades (3 and 4), and we explore the efficacy of textural and morphological features in addition to tissue architecture to distinguish between cancer classes. In addition to automated prostate cancer grading, a second aim of this work is to integrate use of additional image features (texture and gland morphology) into the Gleason grading paradigm.

In Section 2, we describe our methodology to distinguish between prostate cancer grades, and in Section 3 we present the results of our computational model. Concluding remarks are presented in Section 4.

2. METHODS

2.1. Materials

Hematoxylin and eosin stained prostate biopsy cores are scanned into a computer using a high resolution whole slide scanner at 40x optical magnification at the Hospital at the University of Pennsylvania, Department of Surgical Pathology. An expert pathologist labels regions of tissue within each image as Gleason grades 3 adenocarcinoma, grade 4 adenocarcinoma, benign stroma, or benign epithelial tissue. A total of 54 labeled tissue patches were considered in this study, comprising 11 Gleason grade 3, 7 Gleason grade 4, 17 benign epithelial, and 12 benign stromal regions.



(a) (b) (c)

Fig. 2. (a) A region of benign epithelium comprising 6 glands, (b) a gland from (a) magnified, and (c) one of the nuclei surrounding the gland in (b), also magnified.

2.2. Feature Extraction

The feature set includes graph-based, morphological, and global textural features to capture the architectural tissue patterns and arrangement of glands and nuclei in the sample. A listing of the features considered in this work are given in Table 1.

2.2.1. Architectural Features

Figure 2 (a) illustrates a region of benign epithelial tissue made up of six glands, and a single gland is shown in Figure 2 (b). Each region also contains several nuclei, one of which has been magnified in Figure 2 (c). An image region R is made up of pixels c, containing m nuclei with centroids $c_n^1, c_n^2, \cdots, c_n^m$. In addition R may contain k glands with corresponding centroids $c_g^1, c_g^2, \cdots, c_g^k$.

The shape and arrangement of glandular and nuclear structures within a histological image region is related to tissue type [4]. Quantifying these structural arrangements into phenotypic signatures allows for efficient classification between Gleason grades. In this study 4 separate graphs are constructed and features based on these graphs are extracted. Graphs (A)-(C) describe the spatial arrangement of the nuclei, while graph (D) describes the spatial arrangement of glands. Graphs (A)-(C) were not constructed for gland structures because the low number of glands per region does not yield informative graphs. Figure 3 illustrates these graphs for Gleason grade 3 (Figure 3 (a)-(f)), grade 4 (Figure 3 (g)-(l)), benign epithelium (Figure 3 (m)-(r)), and benign stroma (Figure 3 (s)-(x)) tissues.

A. Voronoi Diagram - The Voronoi diagram $\mathbf{V}$ [8] comprises a set of polygons $\mathbf{P} = \{P_1, P_2, \cdots, P_m\}$. Any pixel $c \in R$ is included in polygon P_a if $\mathbf{d}(c, c_n^a) = \min_j\{||c - c_n^j||\}$ where $a, j \in \{1, 2, \cdots, m\}$ and $\mathbf{d}(c,d)$ is the Euclidean distance between any two pixels $c, d \in R$. The following features are computed from $\mathbf{V}$. The area $\mathcal{N}^A(P_j)$ of a polygon P_j is given as $|P_j|$ which is the cardinality of set P_j. The average area is computed as $\widehat{\mathcal{N}^A} = \frac{1}{m}\sum_{j=1}^{m} \mathcal{N}^A(P_j)$. The disorder of the area is defined as $D^A = 1/(1 + \frac{\sigma^A}{\widehat{\mathcal{N}^A}})$ where σ^A is the standard deviation over $\mathcal{N}^A(P_j)$ for $j \in \{1, 2, \cdots, m\}$ [9]. The roundness factor $\mathcal{N}^r(P_j)$ of polygon P_j is calculated as 4π divided by the sum of the lengths of each side of P_j. Average roundness factor $\widehat{\mathcal{N}^r}$ and disorder D^r are computed similar to $\widehat{\mathcal{N}^A}$ and D^A.

B. Delaunay Triangulation - The Delaunay graph $\mathbf{D}$ is constructed such that any two unique nuclear centroids c_n^a and c_n^b, where $a, b \in \{1, 2, \cdots, m\}$, are connected by an edge $E^{a,b}$ if P_a and P_b share a side in $\mathbf{V}$. The following features are computed from $\mathbf{D}$. The average edge length $\widehat{E}$ and the maximum edge length $E^{\max}$ are computed over all edges in $\mathbf{D}$. Each c_n^a is connected to B other nuclear centroids $c_n^{a,\alpha_1}, c_n^{a,\alpha_2}, \cdots, c_n^{a,\alpha_B}$ by corresponding edges $E^{a,\alpha_1}, E^{a,\alpha_2}, \cdots, E^{a,\alpha_B}$ in $\mathbf{D}$, where $\alpha_1, \alpha_2, \cdots, \alpha_B \in \{1, 2, \cdots, m\}$. The average length of the edges $E^{a,\alpha_1}, E^{a,\alpha_2}, \cdots, E^{a,\alpha_B}$ is calculated for each $a \in \{1, 2, \cdots, m\}$.

C. Minimum Spanning Tree - Given a connected, undirected graph, a spanning tree $\mathbf{S}$ of that graph is a subgraph which is a tree and connects all the vertices together. A single graph can have many different $\mathbf{S}$. Weights $\omega_{\mathbf{S}}^E$ are assigned to each edge E in each $\mathbf{S}$ based on the length of E in $\mathbf{S}$. The sum of all weights $\omega_{\mathbf{S}}^E$ in each $\mathbf{S}$ is determined to give the weight $\widehat{\omega}_{\mathbf{S}}$ assigned to each $\mathbf{S}$. The minimum spanning tree (MST) denoted by $\mathbf{S}^M$ has a weight $\widehat{\omega}_{\mathbf{S}}^M$ less than or equal to $\widehat{\omega}_{\mathbf{S}}$ for every other spanning tree $\mathbf{S}$. The features computed from $\mathbf{S}^M$ are (i) average length of all edges in $\mathbf{S}^M$ and (ii) edge length disorder computed as for D^A and D^r.

D. Co-Adjacency Matrix - To describe the arrangement of glands in the image, a co-adjacency matrix W is constructed where for the value of row u, column v, $W(u,v) = ||c_g^u - c_g^v||$, $u, v \in \{1, 2, \cdots, k\}$, and $W \in \mathbb{R}^{k \times k}$. This diagonally symmetric matrix describes the spatial relationships between each pair of glands in R analogous to the manner in which Haralick's co-occurrence matrix describes the relationship between pixel intensity values. We calculate 13 of Haralick's second-order features from W: angular second moment, contrast, correlation, variance, entropy, sum average, sum variance, sum entropy, difference variance, difference entropy, difference moment, and two measurements of correlation [10].

2.2.2. Nuclear Density and Gland Morphology

Nuclear density $\Pi^{\mathcal{D}}$ is computed as $\Pi^{\mathcal{D}} = \frac{m}{|R|}$, where $|R|$ is the cardinality of R. For each nuclear centroid c_n^a, $N(\mu, c_n^a)$ is the set of pixels $c \in R$ contained within a circle with its center at c_n^a and radius μ. We compute the number of c_n^j, $j \neq a$, $j, a \in \{1, 2, \cdots, m\}$ which are in set $N(\mu, c_n^a)$. The disorder of this feature D^N is calculated

Authorized licensed use limited to: Rutgers University. Downloaded on November 14, 2008 at 15:27 from IEEE Xplore. Restrictions apply.

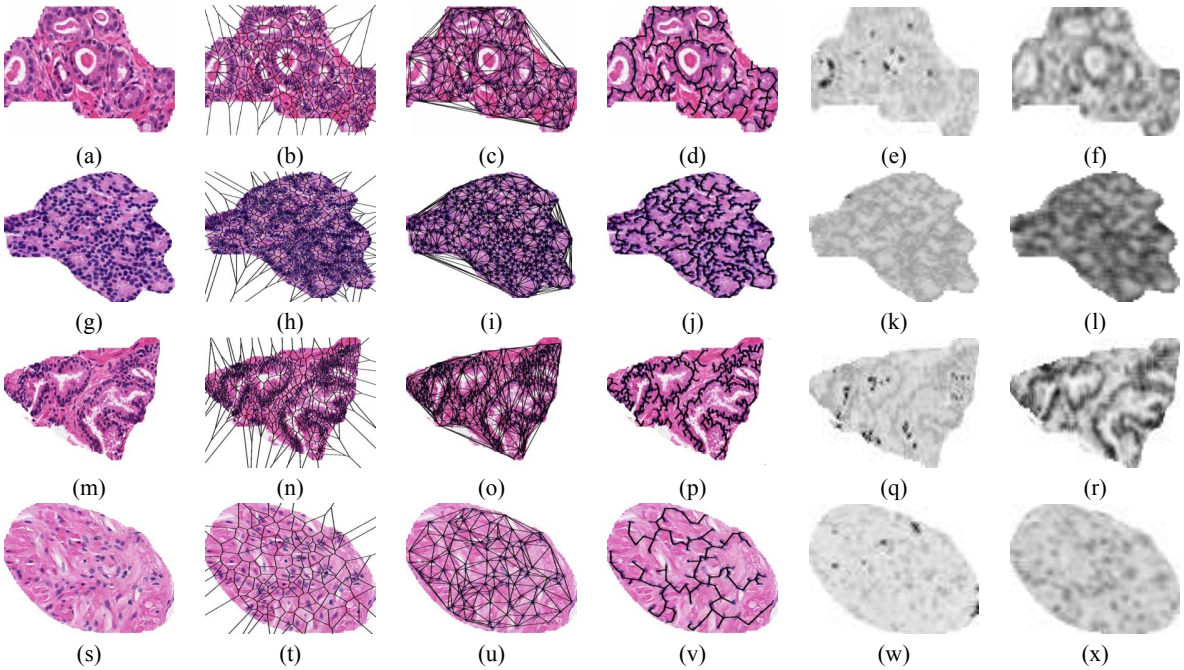


Fig. 3. Comparison of ((a)-(f)) Gleason grade 3 tissue, ((g)-(l)) grade 4 tissue, ((m)-(r)) benign epithelium, and ((s)-(x)) benign stroma. Superimposed on ((a), (g), (m), (s)) the original images are ((b), (h), (n), (t)) the Voronoi diagram, ((c), (i), (o), (u)) the Delaunay triangulation, ((d), (j), (p), (v)) the minimum spanning trees, ((e), (k), (q), (w)) pixel entropy texture feature, and ((f), (l), (r), (x)) Gabor filter ($s = 3$, $\theta = \frac{5\pi}{8}$).

as described in section 2.2.1 A. The size and shape of the structures within a set of glands are also important in discriminating tissue type [4]. These characteristics are quantified by calculating the average and standard deviation of the following 8 features within an image: total gland area, gland perimeter length, lumen area, ratio of lumen area to gland area, ratio of gland area to perimeter, number of nuclei surrounding the lumen, ratio of nuclei to gland perimeter, and the number of visible nuclei layers around the gland. In addition, we calculated the total area of non-stromal tissue in the image, the total lumen area in the image, and the ratio of these two values to measure the amount of the total region area that was taken up by gland structures. These bring the total number of tissue-based architectural features to 45.

2.2.3. Global Texture Features

The proliferation of nuclei and changes in gland morphology in cancerous tissue suggest that textural characteristics can distinguish between low- and high-grade cancer regions [11]. The following features are extracted from each region [6]:

A. First-Order Statistical Features - The average, median, standard deviation, and range of the pixel values $f(c)$, for $c \in R$, are calculated to quantify first-order statistics.

B. Second-Order Statistical Features - Second-order co-occurrence texture features are described by the 13 Haralick features described in section 2.2.1 D, using a co-occurrence matrix $Z \in \mathbb{R}^{\mathcal{M} \times \mathcal{M}}$, where $\mathcal{M}$ is the maximum pixel intensity of all $c \in R$, in place of the co-adjacency matrix. The value in $Z(f(c), f(d))$ is given by the number of times intensities $f(c), f(d)$ appear within a fixed displacement of each other and at a specified orientation. The 13 Haralick features are calculated from Z [10].

C. Steerable Filters - Gabor filters provide varying responses to textural differences in an image. The filter kernel G is constructed based on a scale and an orientation parameters as

$$G(x,y,\theta,s) = e^{-\frac{1}{2}((\frac{x'}{\sigma_x})^2+(\frac{y'}{\sigma_y})^2)} cos(2\pi s x'),$$

where (x,y) are the 2D Cartesian coordinates for pixel $c \in R$, $x' = x\cos\theta + t\sin\theta$, $y' = y\cos\theta + x\sin\theta$, $\theta \in \{0, \frac{\pi}{8}, \cdots, \frac{7\pi}{8}\}$ is an orientation parameter, $s \in \{0,1,\cdots,4\}$ is a scale parameter, and σ_x and σ_y are the standard deviations along the X and Y axes, respectively. For this study, 40 Gabor filters are constructed by varying θ and s. Thus a total of 57 texture-based features are obtained for each R.

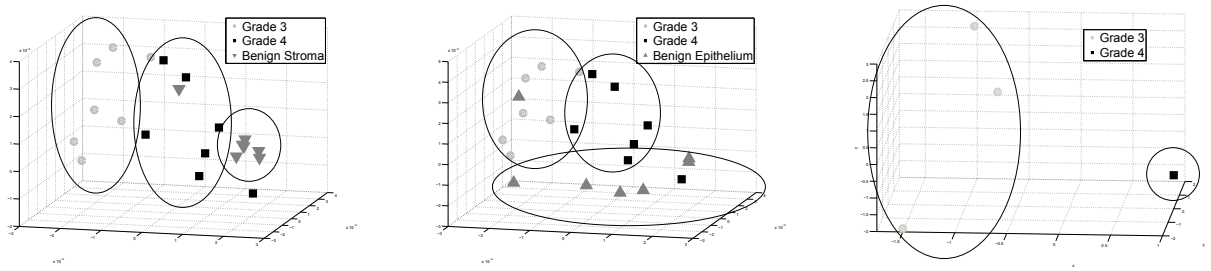
Feature Type	Number of Features	Representative Features
Nuclear Architecture	13	Voronoi, Delaunay, MST
Glandular Architecture	32	Gland area, lumen area
Global Textural	57	Haralick, greylevel, Gabor filter

Table 1. Summary of feature types, the number of each feature type, and representative features from each type.

2.3. Classification

An SVM uses the above features to discriminate between each pair of tissue types used in this study: Gleason grade 3 vs. grade 4, grade 3 vs. benign epithelium, grade 3 vs. benign stroma, grade 4 vs. benign epithelium, grade 4 vs. benign stroma, and benign epithelium vs. benign stroma. A third of the data from each group was used for training the classifier and the rest was used for testing. Since SVMs perform classification in a high-dimensional space, local linear embedding (LLE) [12] was applied to reduce the dimensionality of the feature space prior to classification. Cross-validation was used to obtain optimal parameters for the classifier.

Authorized licensed use limited to: Rutgers University. Downloaded on November 14, 2008 at 15:27 from IEEE Xplore. Restrictions apply.



(a) (b) (c)

Fig. 4. Scatter plots of tissue regions represented in reduced three dimensional space. Tissue regions belong to (a) Gleason grade 3 (circles), Gleason grade 4 (squares) and benign stromal tissue (downward-pointing triangles) are distinguished by textural and graph-based features, and (b) Gleason grades 3 and 4 with benign epithelium (upward-pointing triangles) are distinguished by textural and graph-based features. Also shown are the subset of tissue regions containing glandular features distinguished by graph-based, textural, and glandular morphological features (c). Superimposed ellipses indicate the clusters of similar tissue types.

3. RESULTS AND DISCUSSION

We generate phenotypic signatures for 54 regions consisting of 11 of Gleason grade 3 tissue, 7 of Gleason grade 4, 17 of benign epithelium, and 12 of benign stroma. Table 2 lists classification results for each paired comparison of the four classes. The accuracy and standard deviation σ over six trials are given for each classification task, where the training data is randomized for each trial. Automated classification yields approximately 77% accuracy when comparing between Gleason grade 3 and grade 4 adenocarcinoma, which is clinically a very difficult problem. Studies on agreement and reproducibility in clinical practice show that consensus grading among pathologists rarely approaches 70%, and agreement on a Gleason score of 4 or 3 is particularly low [2], [3]. In this context, classification accuracy greater than 75% is an improvement on existing grading methods. The standard deviation for all paired comparisons is less than or equal to 0.05, indicating the reproducibility of the classification results with respect to training data.

Figure 4 plots each tissue region in a low-dimensional space obtained by LLE. In Figure 4 (a) Gleason grade 3 (circles), grade 4 (squares), and benign stroma (downward triangles) are shown, and Figure 4 (b) shows grade 3 and grade 4 regions along with benign epithelium (upward triangles). The graphs illustrate the inter-class relationships between the data using this feature set. Because not all tissue regions contain glands, only graph-based and textural features were used to generate the graphs in Figure 4 (a) and (b). Figure 4 (c) plots the tissue regions for which morphological gland features could be calculated. The separation between grade 3 and grade 4 tissues improves with the inclusion of these features, indicating that their addition to the classifier improves accuracy.

Implicit feature selection via the AdaBoost algorithm [7] revealed that texture features are weighted heavily, particularly in discriminating the adenocarcinoma tissues from benign epithelium and in discriminating epithelium from stroma.

4. CONCLUDING REMARKS

In this paper, we have demonstrated a system that can quantify graph-based, textural, and morphological features in tissue patterns to accurately classify histological images according to Gleason grade. By developing quantitative phenotypic signatures for tissue patterns, we can begin to resolve some of the issues currently seen in practice, such as inter- and intra-observer variability, under- and over-grading, and subjective analysis of qualitative image features. Our analysis in determining feature importance revealed that texture features are capable of discriminating between tissue patterns and can extend the Gleason criteria for grading tissue patterns.

Class Distinction	Accuracy	σ	Highest Weighted Feature
Grade 3 vs. Grade 4	76.9%	0.049	Delaunay Nearest Neighbor
Grade 3 vs. Epithelium	85.4%	0.051	Haralick Difference Variance
Grade 3 vs. Stroma	92.8%	0.026	MST Average Edge Length
Grade 4 vs. Epithelium	88.9%	0.025	Haralick Variance
Grade 4 vs. Stroma	89.7%	0.040	Voronoi Nearest Neighbor
Epithelium vs. Stroma	92.4%	0.017	Haralick Difference Moment

Table 2. SVM classification accuracy, standard deviation, and the feature assigned the highest weight by the AdaBoost algorithm.

5. REFERENCES

[1] R. Montironi, R. Mazzuccheli, et al., "Gleason grading of prostate cancer in needle biopsies or radical prostatectomy specimens: contemporary approach, current clinical significance and sources of pathology discrepencies," *BJU Int.*, vol. 95, no. 8, pp. 1146–1152, 2005.

[2] W. Allsbrook, K.A. Mangold, et al., "Interobserver reproducibility of gleason grading of prostatic carcinoma: General pathologist," *Hum. Path.*, vol. 32, no. 1, pp. 81–88, 2001.

[3] C.R. King, "Patterns of prostate cancer biopsy grading: Trends and clinical implications," *Int. J. Cancer (Radiat. Oncol. Invest)*, vol. 90, pp. 305–311, 2000.

[4] J.I. Epstein, W.C. Allsbrook, et al., "The 2005 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma," *Am. J. of Surg. Path.*, vol. 29, no. 9, pp. 1228–1242, 2005.

[5] A.N. Esgiar, R.N.G. Naguib, et al., "Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa," *IEEE Trans. on Info. Tech. in Biomed.*, vol. 2, no. 3, pp. 197–203, 1998.

[6] S. Doyle, A. Madabhushi, et al., "A boosting cascade for automated detection of prostate cancer from digitized histology," *MICCAI*, vol. 4191, pp. 504–511, 2006.

[7] Y. Freund and R. Schapire, "Experiments with a new boosting algorithm," *Proc. of the Natural Conf. on Mach. Learn.*, pp. 148–156, 1996.

[8] G.T. Toussaint, "The relative neighbourhood graph of a finite planar set," *Patt. Recog.*, vol. 12, pp. 261–268, 1980.

[9] J. Sudbø, R. Marcelpoil, et al., "New algorithms based on the voronoi diagram applied in a pilot study on normal mucosa and carcinomas," *Anal. Cell. Path.*, vol. 21, no. 2, pp. 71–87, 2000.

[10] R.M. Haralick, I. Dinstein, et al., "Textural features for image classification," *IEEE Trans. on Sys., Man, and Cyber.*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[11] K. Jafari-Khouzani and H. Soltanian-Zadeh, "Multiwavelet grading of pathological images of prostate," *IEEE Trans. on Biomed. Eng.*, vol. 50, no. 6, pp. 697–04, 2003.

[12] S.T. Roweis and L.K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 22, pp. 2323–2326, 2000.

Authorized licensed use limited to: Rutgers University. Downloaded on November 14, 2008 at 15:27 from IEEE Xplore. Restrictions apply.