



# Prostate cancer grading: Gland segmentation and structural features <sup>☆</sup>

Kien Nguyen <sup>a</sup>, Bikash Sabata <sup>b</sup>, Anil K. Jain <sup>a,\*,1</sup>

<sup>a</sup> Michigan State University, East Lansing, MI 48824, USA

<sup>b</sup> Ventana Medical Systems, Inc., 919 Hermosa Ct, Sunnyvale, CA 94085, USA

## ARTICLE INFO

### Article history:

Available online 10 October 2011

### Keywords:

Prostate cancer  
Benign  
Carcinoma  
Gleason grading system  
Gland segmentation  
Nuclei

## ABSTRACT

In this paper, we introduce a novel approach to grade prostate malignancy using digitized histopathological specimens of the prostate tissue. Most of the approaches proposed in the literature to address this problem utilize various textural features computed from the prostate tissue image. Our approach differs in that we only focus on the tissue structure and the well-known Gleason grading system specification. The color space representing the tissue image is investigated and basic components of the prostate tissue are detected. The components and their structural relationship constitute a complete gland region. Tissue structural features extracted from gland morphology are used to classify a tissue pattern into three major categories: benign, grade 3 carcinoma and grade 4 carcinoma. Our experiments show that the proposed method outperforms a texture-based method in the three-class classification problem and most of the two-class classification problems except for the grade 3 vs grade 4 classification. Based on these results, we propose a hierarchical (binary) classification scheme which utilizes the two methods and obtains 85.6% accuracy in classifying an input tissue pattern into one of the three classes.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

Prostate cancer is a type of cancer that occurs in men's reproductive system. In the United States, it is the second most prevalent cancer in men and it is also one of the leading causes of death by cancer (in 2006, prostate cancer developed in 203,415 men and killed 28,372 men) ([US Cancer Statistics Working Group, 2010](#)). Prostate cancer grows slowly with very few symptoms; it develops mostly in men over the age of fifty. Prostate cancer is considered serious because of the threat of its invasion (metastasis) into other organs such as bones, bladder and rectum. The prognosis involves a screening (such as digital rectal examination or prostate-specific antigen (PSA) test ([Catalona et al., 1991](#))) and, if necessary, a follow-up prostate biopsy. After an unsuspected cancer is revealed via the screening, a biopsy is used to confirm it. A CT scan or a bone scan can be employed additionally to determine the spread of the cancer.

The biopsy is conducted by a radiologist or a urologist. First, a prostate tissue sample is removed from the patient for inspection under a microscope. A grade is then reported for the tumor derived from the tissue. The most widely used grading method is Gleason

grading ([Gleason, 1977, 1992](#)), which assigns a numerical grade from 2 to 10 to the tumor. The grade is based solely on structural features of the tissue and excludes cytological features ([Mason, 1964](#)). In this grading method, a pathologist finds the most predominant and the second most predominant histological carcinoma patterns in the tissue, assigns each of them a score (from 1 to 5) and adds the two scores together to obtain the final Gleason grade (2 to 10) for the tissue. The grade of each carcinoma pattern is based on its differentiation (how much of its structure resembles a normal pattern structure). A grade 1 carcinoma pattern is very well differentiated and a grade 5 carcinoma pattern is very poorly differentiated. The change in tissue structure is good evidence for this differentiation. More specifically, in Gleason grades 1 and 2, most of the glands appear as single units, separated from each other, densely packed, and there is no infiltration of these glands into benign tissue areas (this is very close to the structure of a normal tissue). Gleason Grade 3, the most common case of carcinoma, is characterized by the invasion of small glands into the muscle (stroma). In Gleason grade 4, glands are fused with each other and poorly defined; glands are not well-separated by stroma as in lower grades. Finally, in Gleason grade 5, there is no evidence of the formation of gland units in the pattern. A visual summarization of these five grades can be found in [Fig. 1](#). Pathologists face a number of difficulties in manually diagnosing prostate cancer, i.e. to look at the prostate tissue under a microscope is tedious and time-consuming. Moreover, the diagnostic accuracy depends on the personal skill and experience of a pathologist. These problems motivate the research and development for automating the diagnosis and prognosis processes.

<sup>☆</sup> A preliminary version of this paper appeared in the Proceedings of the 20th ICPR, Istanbul, August 2010.

\* Corresponding author. Tel.: +1 517 355 9282.

E-mail addresses: [nguye231@msu.edu](mailto:nguye231@msu.edu) (K. Nguyen), [bikash.sabata@bioimagene.com](mailto:bikash.sabata@bioimagene.com) (B. Sabata), [jain@cse.msu.edu](mailto:jain@cse.msu.edu) (A.K. Jain).

<sup>1</sup> Also affiliated to Department of Brain & Cognitive Engineering, Korea University, Seoul.

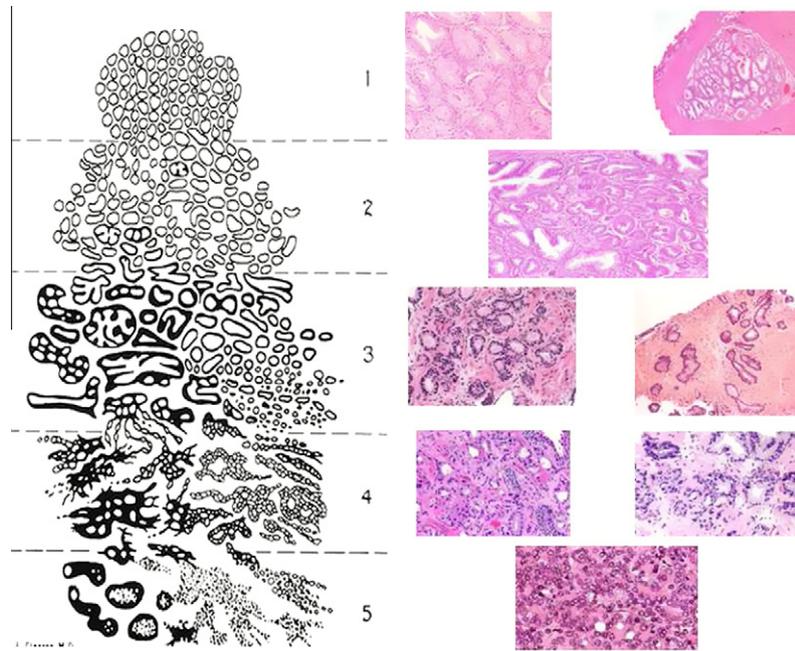


Fig. 1. Five grades of the Gleason grading applied to histological patterns of the prostate tissue.

In most digital pathology studies on computer-aided prognosis for prostate cancer, textural features of the image and structural features of the tissue have been widely used. [Diamond et al. \(2004\)](#) used co-occurrence texture features ([Haralick et al., 1973](#)) to classify each  $100 \times 100$  sub-region in a tissue image into either stroma or prostatic carcinoma. In addition, lumen area was used to discriminate benign tissue from the other two classes. They reported 79.3% accuracy when evaluating the algorithm on sub-regions of 8 tissue images ( $40\times$  magnification). A cancer vs non-cancer classification problem which used 594 features including first-order statistics (average, median, standard deviation), co-occurrence and wavelet features was addressed in [Doyle et al. \(2006\)](#). The algorithm was implemented at three different scales of the image. At each scale, a Bayes classifier was designed for each feature individually, resulting in 594 base learners for AdaBoost. The reported accuracy was 88% on a dataset of 22 images ( $40\times$  magnification). In [Tai et al. \(2010\)](#), fractal dimension features were calculated for the tissue image and the low frequency sub-bands of the image to discriminate the textural discrepancy between low grade and high grade carcinoma. By using an SVM classifier with leave-one-out technique, the method achieved 86.3% accuracy for the classification of 1,000 prostatic biopsy images into normal, grade 3, grade 4 and grade 5 classes. A multiwavelet transform was used as the main texture analysis tool in [Khouzani and Zadeh \(2003\)](#). The features used for classification included entropy and energy derived from the multiwavelet coefficients of the image. Ten different types of multiwavelet were evaluated on a dataset of 100 prostate sample images ( $100\times$  magnification) of grades 2, 3, 4 and 5, resulting in the best accuracy of 97%. In another study, [Tabesh et al. \(2007\)](#) employed both global features of the entire image and local features of every object in the image. Global features included color histogram, fractal features, texture and morphometry of the image. Local features were computed for histological objects such as nuclei, stroma and lumen, which were extracted by the MAGIC system ([Tevevskiy et al., 2004](#)). They achieved 96.7% accuracy for tumor-nontumor classification (fivefold cross validation with 367 images) and 81% accuracy for low grade-high grade classification (fivefold cross validation with 268 images). All images were at  $20\times$  magnification. A segmentation-based method was presented in [Naik et al.](#)

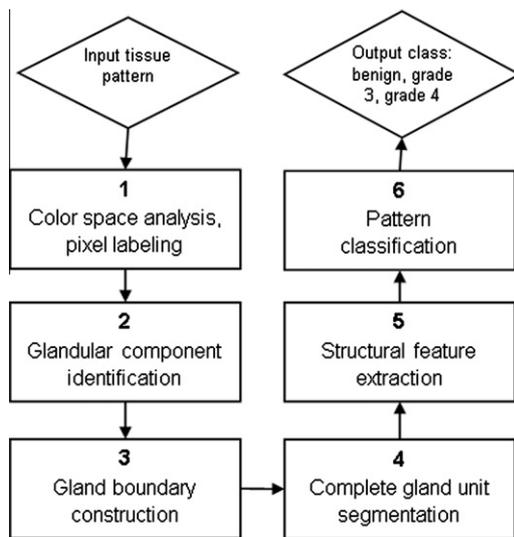
(2008). They first used a Bayesian classifier to place every pixel in the image into one of the three classes: lumen, nuclei and cytoplasm based on its color. Lumen pixels were first grouped together and lumen objects were then identified as the groups satisfying the gland size constraint. The inner boundary of the glands, which is the border of the nuclei and the cytoplasm surrounding the lumen, was detected using a level set formulation. Eight shape features for each of the lumen and the gland inner boundary were calculated. A tissue was classified into benign, a grade 3 carcinoma or a grade 4 carcinoma via an SVM classifier. However, by using a dataset that included 44 images at  $40\times$  magnification, they only reported results of two-class classifications: 86.35% accuracy when classifying grade 3 carcinoma and benign, 92.9% accuracy when classifying grade 4 carcinoma and benign, and 95.19% accuracy when classifying grade 3 carcinoma and grade 4 carcinoma. Three-class classification result was not reported. [Table 1](#) summarizes the related studies discussed in this section.

In this study, we present a segmentation-based method to classify a tissue pattern into three common cases based on Gleason grading: benign, grade 3 and grade 4 carcinoma. However, unlike [Naik et al. \(2008\)](#), we incorporate nucleus and blue mucin information into the glandular structures which are used for the classification. It is apparent from the tissue pattern image that nucleus distribution changes remarkably among various cancer stages (in benign tissue, nuclei form a ring on the gland boundary and scatter in other areas ([Fig. 8\(a\)](#)) while in grade 4 carcinoma, nuclei distribute more uniformly over the glandular regions ([Fig. 8\(c\)](#))) and mucin appears commonly in cancerous glands ([Fig. 8\(b\)](#))). While a gland region in [Naik et al. \(2008\)](#) solely consists of lumen and internal cytoplasm region, our segmentation procedure leads to complete glands which include their nucleus boundaries. Moreover, the structural features extracted in our method do not require a very high magnification (like  $40\times$  in [Naik et al., 2008](#)) to achieve state of the art classification results. The proposed algorithm is designed to work for images created from the Hematoxylin and Eosin (H&E) staining method ([Kiernan, 2001](#)). The outline of the methodology is delineated in the flowchart of [Fig. 2](#). Given an input tissue pattern, we first segment glands from the stroma area (this comprises steps 1, 2, 3, 4 in the flowchart). Once gland regions

**Table 1**  
Summary of major classification studies on prostate cancer.

Authors	Features used	Dataset size (magnification)	Classes	Accuracy
Diamond et al. (2004)	Co-occurrence texture features and lumen area	100 × 100 sub-regions of 8 tissue images (40× magnification)	Stroma, benign tissue <sup>a</sup> and prostatic carcinoma <sup>a</sup>	79.3%
Doyle et al. (2006)	First-order statistics (average, median, standard deviation), co-occurrence and wavelet features	22 images (40× magnification)	Cancer <sup>a</sup> vs noncancer <sup>a</sup>	88%
Tai et al. (2010)	Fractal dimension features	1000 images	Normal <sup>a</sup> , grade 3, grade 4 and grade 5	86.3%
Khouzani and Zadeh (2003)	Entropy and energy of the multiwavelet coefficients	100 images (100× magnification)	Grades 2, 3, 4 and 5 carcinoma	97%
Tabesh et al. (2007)	Global features of the image and local features of histological objects	268 images (20× magnification)	Low grade vs high grade	81%
Naik et al. (2008)	Shape features of the lumen and the gland inner boundary	367 images (20 × magnification) 44 images (40 × magnification)	Tumor <sup>a</sup> vs non-tumor <sup>a</sup> Grade 3 vs benign Grade 4 vs benign Grade 3 vs grade 4	96.7% 86.35% 92.9% 95.19%
Proposed hierarchical scheme	Glandular structural features and co-occurrence texture features	82 ROIs (10× magnification)	Benign, grade 3 and grade 4 carcinoma	85.6%

<sup>a</sup> Different terminologies which have the same meaning were used in the related work. Cancer, tumor and carcinoma refer to the tissues which are detected to have malignant properties of a cancer (cells grow aggressively, invade the surrounding tissues and spread to the non-adjacent tissues). Noncancer, normal, benign and nontumor refer to the tissues which do not have the properties of cancer.



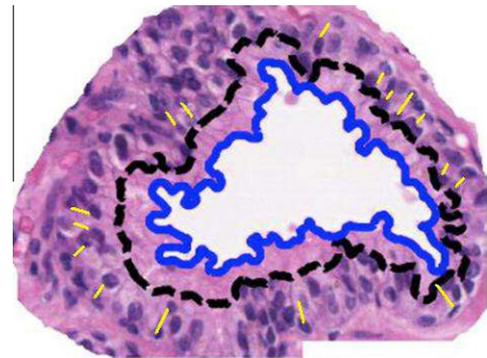
**Fig. 2.** Methodology outline.

are identified, we extract fifteen features from the pattern based on glandular structures (step 5 of the flowchart). Finally, these features become inputs to different classifiers to determine the carcinoma grade of the input pattern (step 6 in the flowchart).

The rest of the paper is organized as follows. Section 2 describes the gland segmentation algorithm, Section 3 explains the features extracted from the gland regions, Section 4 presents experimental results using different classifiers and Section 5 concludes the paper.

## 2. Segmentation of gland units

In a normal prostate tissue (Fig. 4(a)), stroma (pink<sup>2</sup> regions) serves as background and gland units are foreground objects. A gland unit does not have any fixed shape or size; it can be oval, round or branchy and it can be either small or very large. Hence, we cannot utilize Active Shape Model (Cootes et al., 1995) or Active Appearance

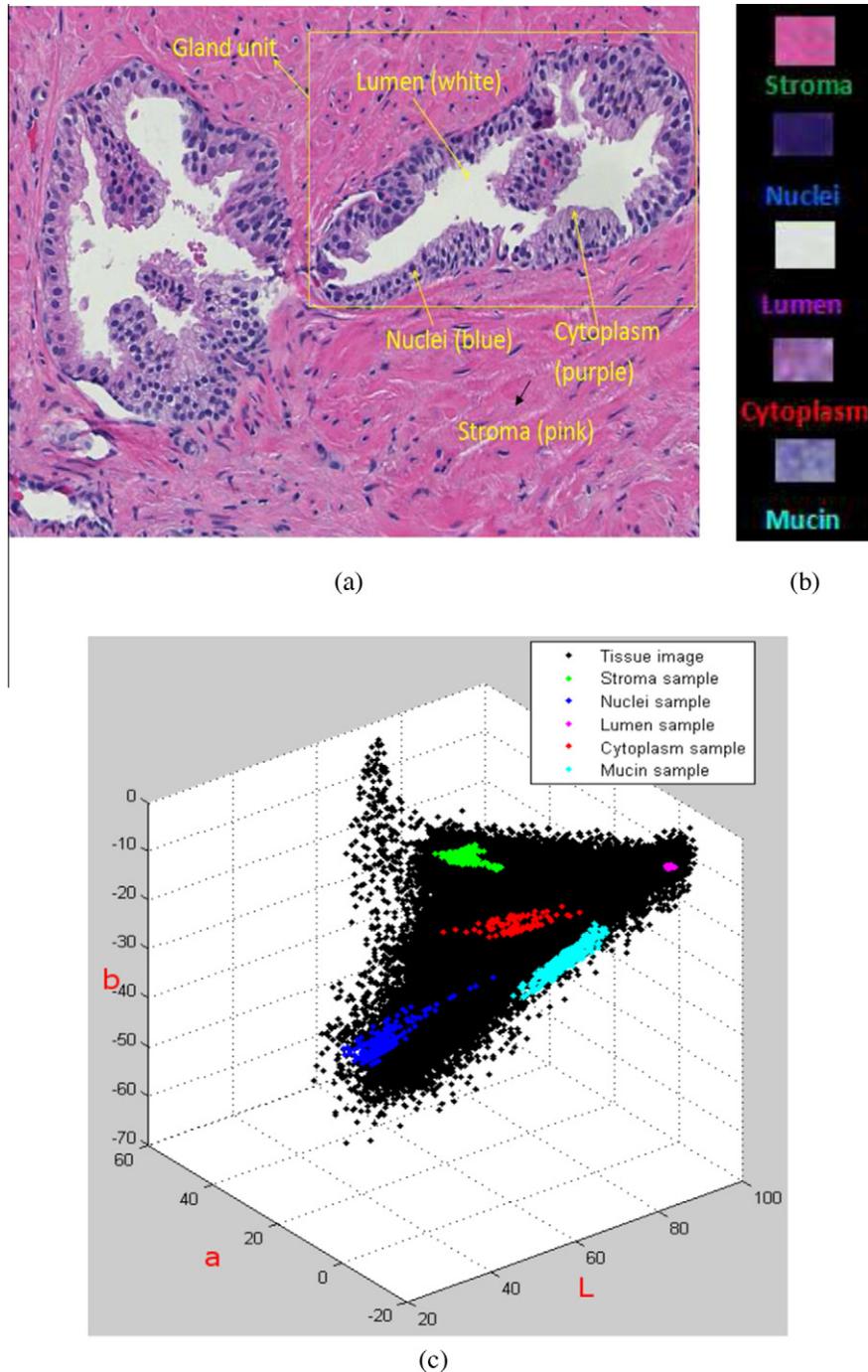


**Fig. 3.** A gland segmentation result obtained by a level set method in Naik et al. (2008). The blue and black curves show the segmented gland which only includes lumen and cytoplasm regions. The yellow lines delineate the gaps between the nuclei layers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Model (Cootes et al., 2001) to detect gland objects. In Naik et al. (2008), the authors used a level set approach to segment glands in the tissue. However, the gland segments obtained in their approach only include lumen and cytoplasm regions (Fig. 3) while we want to capture the entire gland area which also includes nuclei on the gland boundary. Since there are usually multiple layers of nuclei on the gland boundary and there are gaps between these layers (the yellow lines in Fig. 3 delineate these gaps), it is difficult to force the level set curve to capture all of the nuclei on the gland boundary. In a standard energy minimization formulation of the level set, the curve is attracted toward high gradient magnitude regions (Li et al., 2005). These high gradient magnitude regions can be the edges of the nuclei or can be the nuclei likelihood image as defined in Naik et al. (2008). Since the gaps between nuclei layers must have low gradient magnitude, it is difficult for a level set curve to pass these gaps and embrace all the nuclei.

Consequently, we rely on the structure of glands to segment them from the tissue background. Each gland unit has a boundary of epithelial cells which include epithelial nuclei (blue dots) mixed with epithelial cytoplasm (purple) and lumina (white) in the center. In some cancer tissues, blue mucin may be found to invade the lumina (Fig. 8(b)). In short, nuclei, cytoplasm, stroma, lumen and mucin are five basic components which appear in different colors in a prostate tissue image stained by the H&E method. As a

<sup>2</sup> For interpretation of color in Figs. 4 and 8, the reader is referred to the web version of this article.



**Fig. 4.** *Lab* color space. (a) Prostate tissue image. (b) Sampled regions of the five tissue components whose pixels are used as training pixels. (c) *Lab* color space: black points are *Lab* color points of tissue image pixels and other colored points are *Lab* color points of training pixels of the five tissue components.

result, an analysis of the color space of the prostate tissue image should help to identify different components of the tissue and pave the way for the segmentation of glands.

### 2.1. Classification in the color space

We represent all pixels in the tissue image in the *Lab* (also known as  $L^*a^*b^*$  or CIELAB) color space (Jain, 1989). The *Lab* color space is preferred over *RGB* because it was designed to approximate the color perception in human visual system. *Lab* is specified by the CIE (International Commission on Illumination) to separate

the lightness of the color (*L* channel) from the spectral properties of color (negative values of *a* channel indicate green and positive values of *a* indicate red, while negative values of *b* channel indicate blue and positive values of *b* indicate yellow). To convert the conventional *RGB* space to the corresponding *Lab* space, we first convert the *RGB* space to *XYZ* space by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4900 & 0.3100 & 0.2000 \\ 0.1769 & 0.8124 & 0.0107 \\ 0.0000 & 0.0099 & 0.9901 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (1)$$

The elements of the transformation matrix were derived in Fairman et al. (1997). Next, the  $XYZ$  space is converted to  $Lab$  space by

$$L = 116f(Y/Y_n) - 16 \quad (2)$$

$$a = 500[f(X/X_n) - f(Y/Y_n)] \quad (3)$$

$$b = 200[f(Y/Y_n) - f(Z/Z_n)] \quad (4)$$

where

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > (\frac{6}{29})^3 \\ \frac{1}{3}(\frac{29}{6})^2 t + \frac{4}{29} & \text{otherwise,} \end{cases} \quad (5)$$

and  $X_n, Y_n, Z_n$  are the white point tristimulus values in  $XYZ$  (Wikipedia contributors, 2010). Each pixel  $p$  in the image domain is mapped to a point  $c(p)$  in the  $Lab$  space, where  $c(p)$  is the three-dimensional  $(L, a, b)$  color vector of pixel  $p$ . Notice that this is a many-to-one mapping because several pixels can have the same  $(L, a, b)$  color vector. In Fig. 4(c), we demonstrate the  $Lab$  color space of the image in Fig. 4(a). Our goal is to determine to which component of the tissue each image pixel  $p$  belongs. Based on the fact that the number of points in the  $Lab$  space is smaller than the number of pixels in the image, we can (i) classify points in the  $Lab$  color space into five classes (representing the five components) and (ii) use the classification results to assign each image pixel a label of the corresponding component. This can be done due to the known mapping between pixels in the image and points in the  $Lab$  space. To facilitate the classification, we sample pixels in local regions (approximately  $18 \times 22$  pixels) for each component (Fig. 4(b)) from a training tissue image. Let  $C_i$  denote the class corresponding to the  $i$ th tissue component (we refer to stroma, nuclei, cytoplasm, lumen, mucin as the 1st, 2nd, 3rd, 4th and 5th components, respectively) and  $D_i = \{x_j^i\}_{j=1}^{n_i}$  denote the set of  $n_i$  training pixels of class  $C_i$ . The  $Lab$  color points of these training pixels (we call training points for short), denoted by  $\{c(D_i) = \{c(x_j^i)\}_{j=1}^{n_i}\}_{i=1}^5$ , are shown in color in Fig. 4(c) while the  $Lab$  color points of the tissue image pixels are shown in black in the same figure.

The training points create a Voronoi tessellation (Franz, 1991) of the  $Lab$  space (each training point  $c(x_j^i)$  is associated with one convex polygon which includes all points closer to it than any other training point). Each unclassified point  $c(p)$  (black points in Fig. 4(c)) is assigned to the same class associated with the training

point of the polygon to which it belongs. Once the points in the  $Lab$  space are classified (Fig. 5(a)), each pixel in the tissue image is assigned one of the five labels in  $L = \{L_S, L_N, L_C, L_L, L_M\}$  corresponding to the five components (stroma, nuclei, cytoplasm, lumen, mucin) (Fig. 5(b)) via the mapping of points in the  $Lab$  color space and pixels in the image. Formally, a pixel  $p$  is assigned a label  $L_i \in L, l(p) = L_i$  such that  $l(p) = \arg \min_{L_i} \left( \min_{j \in [1, n_i]} d(c(p), c(x_j^i)) \right)$ , where  $d(c(p), c(x_j^i))$  is the Euclidean distance between  $c(p)$  and  $c(x_j^i)$  in the  $Lab$  space.

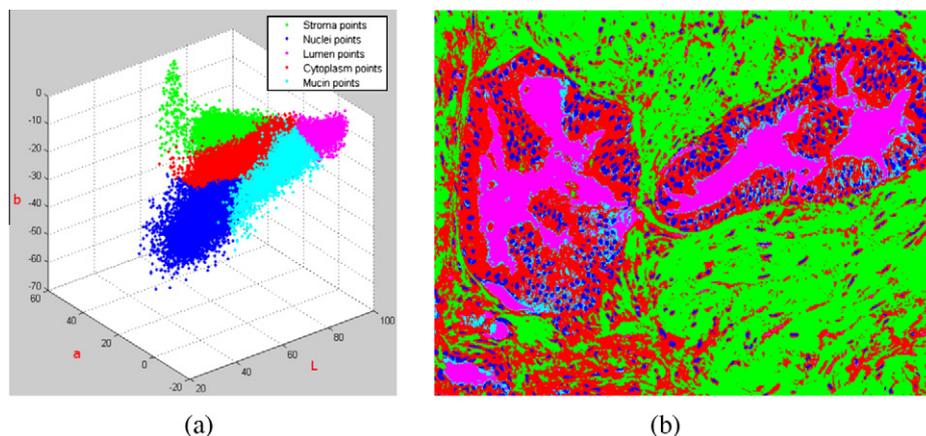
## 2.2. Identify glandular components

Once all the pixels have been classified, we identify nuclei and lumina, the two most important components of the gland. A binary image indicating nucleus pixels is derived from pixel labels (Fig. 6(a)) and a connected component algorithm is applied to this binary image to form the nucleus objects. The four-connectivity property, which only considers the top, left, bottom, and right neighbors of each pixel, is employed. Lumen objects (commonly located in the center of glands) are also created in the same manner (Fig. 7(a)). Nucleus objects and lumen objects are sequentially utilized in the following two procedures.

## 2.3. Construct gland boundary

In the anatomical structure of a prostate tissue, epithelial cell layers comprising nuclei and cytoplasm constitute the gland boundary. Moreover, as we can see in Fig. 4(a) both nuclei and cytoplasm gather densely around the gland but scatter sparsely in other areas (for example, stroma). This motivates us to develop a two-step algorithm for constructing gland boundary.

In the first step, nucleus objects (obtained from Section 2.2) are enlarged by combining them with cytoplasm pixels. An enlarged instance  $N'_i$  of a nucleus object  $N_i$  is defined as:  $N'_i = \{p | p \in N_i \text{ or } (l(p) = L_C \text{ and } \|p, \text{centroid}(N_i)\|_2 \leq d_n)\}$ , where  $p$  is a pixel in the image,  $L_C$  is the label of the cytoplasm pixel (defined in the previous part) and parameter  $d_n$  is empirically estimated by half the average distance between neighboring nuclei in the gland boundary and  $\text{centroid}(N_i)$  is the centroid of the object  $N_i$ . The goal of this step is to facilitate the grouping of neighboring nuclei in the second step.



**Fig. 5.** Classification in the  $Lab$  space (a) and tissue pixel labeling (b). The same color is used for a classified point in the  $Lab$  space and the associated labeled pixel in the image; cytoplasm, nuclei, lumen, stroma and blue mucin are denoted by green, blue, pink, red, cyan, respectively in both (a) and (b). Note that (i) several points of each class are invisible because they are occluded by points of other classes and (ii) multiple pixels in the image can be mapped to one point in the  $Lab$  color space. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

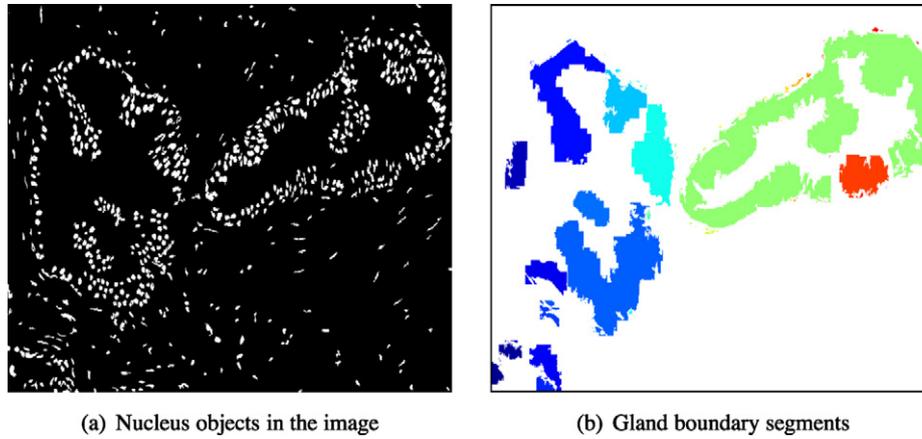


Fig. 6. Gland boundary segments are generated from nucleus objects.

In the second step, we group enlarged nuclei which intersect each other to construct gland boundary segments. Each segment may contain only one isolated enlarged nucleus or several enlarged nuclei, depending on the nucleus density in the area. More concretely, a gland boundary segment  $GB$  can be (i) a single enlarged nucleus  $N'_i$  if  $\forall N'_j \neq N'_i, N'_i \cap N'_j = \emptyset$  or (ii) a group of enlarged nuclei  $(N'_k, \dots, N'_{k+m})$ , where  $\forall N'_i (k \leq i \leq k+m), \exists N'_j (k \leq j \leq k+m)$  such that  $i \neq j$  and  $N'_i \cap N'_j \neq \emptyset$ .

As nuclei are not uniformly dense everywhere on the boundary, we may not obtain a complete boundary segment for each gland. Thus, this stage can generate multiple segments for each gland boundary. Fig. 6(b) demonstrates the output of gland boundary

construction applied to the image of Fig. 4(a). The next procedure combines these segments into the final gland regions.

#### 2.4. Segment complete gland units

Again, we rely on the glandular structure (lumina lie in the center of glands and are embraced by gland boundary) to implement an algorithm for unifying lumen objects (obtained from Section 2.2) with gland boundary segments (obtained from Section 2.3) to form complete gland units. The non-tissue areas (which appear white near the image boundary, e.g. in Fig. 8(a) and (c)) are discriminated from lumen objects by their contacts with the image boundary. The

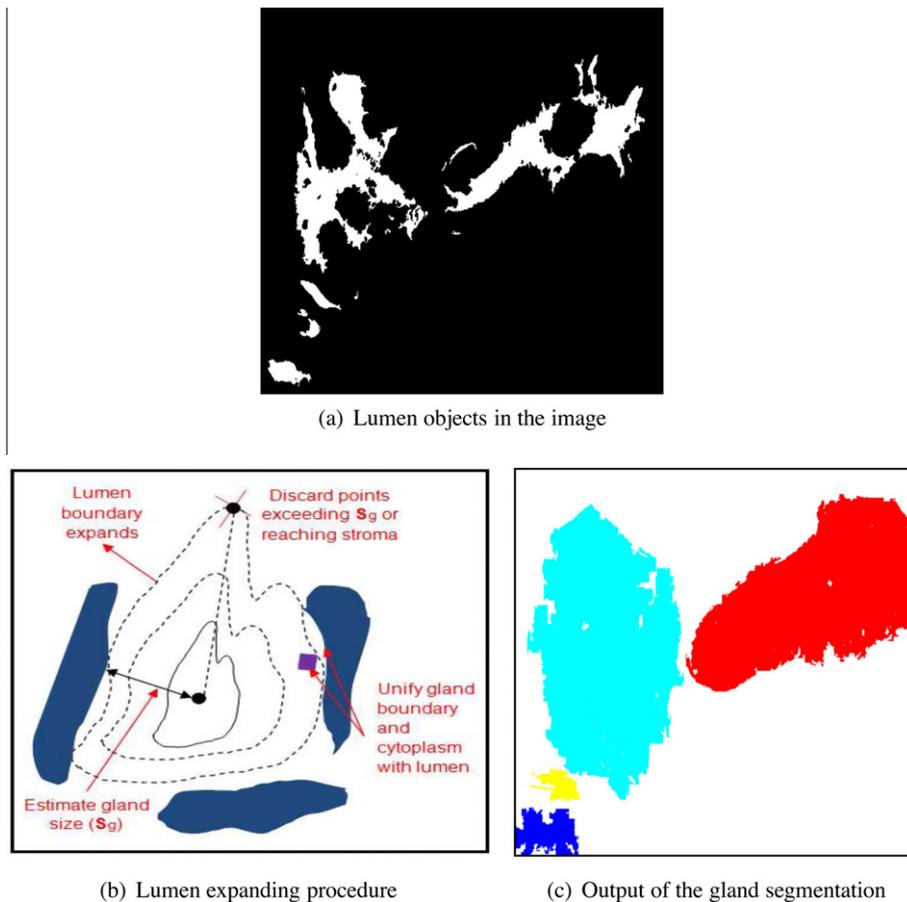


Fig. 7. Expansion procedure of the lumen to unify glandular components and form complete gland units.

algorithm is based on an expansion procedure of lumen objects (this is a different procedure from conventional morphological operations), which is described in Algorithm 1.

---

**Algorithm 1:** Glandular component unification by lumen expanding

---

**Input:** A set of  $N$  gland boundary segments  $\{GB_i\}_{i=1}^N$ . A set of  $M$  lumen objects  $\{LU_i\}_{i=1}^M$ . Label for pixel  $(x, y)$  in the tissue image is  $l(x, y) \in (L_S, L_N, L_C, L_L, L_M)$

**Output:**  $M$  complete gland units  $\{GU_i\}_{i=1}^M$

**for**  $i = 1$  **to**  $M$  **do**  
 2:  $GU_i \leftarrow LU_i$  (initialize gland unit by the lumen object)  
    Compute the centroid  $(x_0, y_0)$  of lumen object  $LU_i$   
 4: Find the set of boundary points of  $LU_i$ ; denote this set by  
     $B \leftarrow \{(x_i, y_i)\}_{i=1}^b$   
    Initialize the estimated gland size  
     $S_g \leftarrow \alpha \max_{(x_i, y_i) \in B} \|(x_i, y_i), (x_0, y_0)\|_2$   
 6: ( $\alpha > 1$  indicates the estimated ratio between gland size and lumen size)  
    Initialize the scale factor for the expanding process  $sf \leftarrow 1$   
 8: **while**  $B \neq \emptyset$  **do**  
    Initialize the remove set  $R$  (containing the points to be removed):  $R \leftarrow \emptyset$   
 10:  $sf \leftarrow sf + t_0$  (increase the scale factor by a step-size  $t_0 < 1$ )  
    **for** each  $(x_i, y_i)$  in  $B$  **do**  
 12: Calculate the expansion version  $(x'_i, y'_i)$  of  $(x_i, y_i)$ :  
     $x'_i \leftarrow (x_i - x_0) \times sf, y'_i \leftarrow (y_i - y_0) \times sf$   
 14: Create a window  $W$  of size  $S_W \times S_W$  ( $S_W$  is proportional to the step-size  $t_0$ ) centered at  $(x'_i, y'_i)$   
    **if**  $\exists(x, y) \in W$  such that  $l(x, y) = L_S$  **then**  
 16:  $R \leftarrow R \cup (x_i, y_i)$  (remove the point reaching stroma background)  
    **end if**  
 18: **if**  $\exists(x, y) \in W$  such that  $l(x, y) = L_C$  **then**  
     $GU_i \leftarrow GU_i \cup (x, y)$  (merge internal cytoplasm with gland)  
    **end if**  
 20: **if**  $\exists GB_j$  such that  $GB_j \cap W \neq \emptyset$  **then**  
 22:  $GU_i \leftarrow GU_i \cup GB_j$  (merge gland boundary segment with gland)  
    **if**  $S_g < \|(x'_i, y'_i), (x_0, y_0)\|_2$  **then**  
 24:  $S_g \leftarrow \|(x'_i, y'_i), (x_0, y_0)\|_2$  (update the gland size estimation)  
    **end if**  
 26:  $R \leftarrow R \cup (x_i, y_i)$  (remove the point reaching gland boundary)  
    **end if**  
 28: **if**  $\|(x'_i, y'_i), (x_0, y_0)\|_2 > S_g$  **then**  
     $R \leftarrow R \cup (x_i, y_i)$  (remove the point exceeding estimated gland size)  
    **end if**  
 30: **end for**  
 32:  $B \leftarrow B \setminus R$   
    **end while**  
 34: **end for**

---

Fig. 7(b) illustrates the algorithm. Since the tissue may contain artifacts (non-lumen white regions which can be formed due to small broken areas in the tissue), the algorithm creates some small regions as well. A minimum size constraint is applied on the segmented gland regions to eliminate such artifacts. The final gland segmentation output of Fig. 4(a) is shown in Fig. 7(c).

### 3. Glandular structural features

We observe the following characteristics of glandular structure of benign, Gleason grade 3 and Gleason grade 4 patterns.

- i. A benign pattern has large, densely-packed and separated gland units. Each gland has large lumen regions and a thick boundary with prominent nuclei (due to multi-layer epithelial cells). Nuclei form a chain on the gland boundary. There is a lot of variation in lumen shape, from circular to oval or branchy. Fig. 8(a) describes these benign structures.
- ii. A Gleason grade 3 pattern has small, circular lumina and thin gland boundaries (because there is only one layer of epithelial cells circumscribing the lumen). Glands are also smaller and more circular than in a benign pattern. Blue mucin is commonly found to mix with lumina. Fig. 8(b) illustrates these properties.
- iii. In a Gleason grade 4 pattern, glandular structures are observed to be altered dramatically due to the fusion of glands. Glands are poorly-defined (we cannot see individual gland units separated by stroma, with their own lumina and well defined epithelial cell layers on the boundary). Multiple glands are mixed together and nuclei distribute almost uniformly in the gland region instead of forming well-formed rings as in benign patterns. These properties can be seen in Fig. 8(c).

Given these observed properties, we extracted fifteen features from each gland region.

- i. Ten lumen features consisting of area statistics (average, variance, maximum), perimeter statistics (average, variance, maximum), circularity statistics (average and variance); percentage of the entire gland area to be lumen area and number of lumina in the gland.
- ii. Two nucleus features consisting of nucleus density and percentage of total gland area to be nucleus area.
- iii. Two gland morphology features consisting of average and variance of gland radius (gland radius is estimated by the distance between lumen centroids and gland boundary)
- iv. One feature related to blue mucin, which is the percentage of gland area covered by blue mucin.

To calculate the nucleus density ( $ND$ ) of a gland region, we first divide the region into a grid of patches (the size of each patch is  $S_c \times S_c$  pixels) and calculate the ratio of nucleus pixels to total number of pixels in each patch. A patch is called an  $N$ -patch if this ratio exceeds an empirically pre-defined threshold  $t_N$ . Then the nucleus density is defined as  $(\text{Number of } N\text{-patches}) / (\text{Total number of patches})$ . The choice of parameters  $S_c$  and  $t_N$  will have influence on the  $ND$  feature but the final classification result is not dramatically affected because the classification also depends on the other 14 features. In regions where nuclei are uniformly distributed,  $ND$  is large. On the other hand, in a region where nuclei focus on a specific area (gland boundary) and are very sparse in other areas (internal area of glands),  $ND$  is small.

The feature vector of a tissue pattern  $I$  is denoted by  $F(I) = \{F_i(I)\}_{i=1}^{15}$ . This is obtained by averaging features of all the gland regions in  $I$ , i.e.

$$F_i(I) = \frac{1}{N} \sum_{j=1}^N f_i(j),$$

where  $f_i(j)$  is the feature  $i$  of gland region  $j$  in  $I$  and  $N$  is the number of gland regions in  $I$ ,  $i = 1, 2, \dots, 15$ .

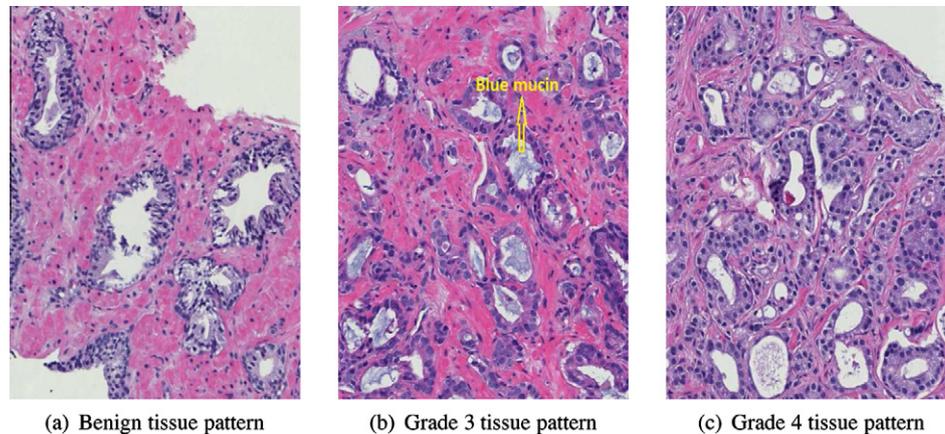


Fig. 8. Three classes of tissue patterns of interest.

An over-segmentation of glands does not have much influence on the feature extraction accuracy because there are only two features that depend on the gland morphology. As a consequence, in grade 4 carcinoma, where gland units are ill-defined and each gland region is not necessarily a gland unit but can be a combination of multiple glands fused with each other, the extracted features still provide good discriminating power.

## 4. Experimental results

### 4.1. Dataset of tissue patterns

From 26 digitized tissue specimens at  $20\times$  magnification (approximate size is  $90,000 \times 45,000$  pixels) of 17 patients obtained from Lakewood Pathology and Ventana Medical Systems, Inc., we manually selected 82 regions of interest (ROI) at  $10\times$  magnification to form the dataset. A  $10\times$  magnification is sufficient for the proposed structural-based method (higher magnifications ( $20\times$  or  $40\times$ ) may be required when computing cytological features). The average size of each ROI is approximately  $620 \times 550$  pixels. Each ROI represents one tissue pattern belonging to one of the three classes of interest. Among 82 ROIs, there are 34 benign regions, 28 grade 3 regions and 20 grade 4 regions. The grade of each ROI is determined by an expert pathologist. Examples of these ROIs with corresponding classes are shown in Fig. 8(a)–(c).

### 4.2. Classification results

In the first experiment, we address the three-class classification problem. We use the classifiers which were widely utilized in the literature, i.e. SVM (Naik et al., 2008; Tabesh et al., 2007; Tai et al., 2010), Nearest Neighbor (Tabesh et al., 2007; Khouzani and Zadeh, 2003), Adaboost (Doyle et al., 2006), Bayes (Tabesh et al., 2007; Naik et al., 2008) and two other popular classifiers which are feedforward Neural Network (FNN) and Decision Tree. We used their implementations provided in Weka (Hall et al., 2009), a popular data mining software. To evaluate the performance of all the classifiers, we use stratified  $K$ -fold cross-validation technique (stratified folds indicate folds which have similar class distribution with the dataset). Two cross-validation (CV) techniques, which are ROI-based CV and specimen-based CV, are employed. In ROI-based CV, assuming that all ROIs in the dataset are independent, we randomly divide the dataset into 10 stratified folds. However, in specimen-based CV, we want all ROIs of the same specimen to lie within one fold. By doing this, no specimen can be present in both the training and the testing data. Due to the constraint, the dataset

can only be divided into 4 stratified folds. The classification results using both CV techniques are shown in Table 2. Because there is a high variation in image intensity (due to the variation in staining procedure) and high variation in tissue structures of different specimens, the accuracy of the specimen-based CV is lower than the ROI-based CV.

In both CV techniques, FNN and SVM give the best results. The one-vs-one approach is implemented in SVM to deal with the three-class problem. The FNN is trained by the backpropagation method using sigmoid activation functions for all nodes. It has three output units corresponding to three class labels. An input  $\mathbf{x}$  is assigned the class label of the output unit which has the largest output value. Based on the number of features and classes, we test the FNN with a wide-range of number of hidden nodes (from 8 to 20) and obtained the best classification result when this number is 16. Since the dataset is small, one misclassified image can make a large difference in the final accuracy. This explains the large difference in the performance of the best and the worst classifiers as well as the large standard deviation of the CV accuracies.

Fig. 9 presents examples of the classification results obtained by FNN classifier. The grade 3 pattern in Fig. 9(a) is misclassified as benign due to the presence of a benign gland (the gland whose structure resembles gland structure in benign patterns) and the large lumen of the gland in the center. The grade 3 pattern in Fig. 9(b) is misclassified as grade 4 because some glands at the top are starting to merge with each other and change their regular structure. In this region, the cancer is evolving from grade 3 to grade 4. Finally, the reasons for the misclassification of the tissue pattern in Fig. 9(c) are as follows: although glands are mixed up, (i) lumen objects are skinny and elongated (lumen circularity is lower than in other grade 4 patterns) (ii) nuclei form a thick ring structure which resembles gland boundaries of benign glands and (iii) the presence of an individual gland on the right side of the pattern.

We also compare the performance of the proposed method with a texture-based method. The co-occurrence features which were widely used in the literature (Diamond et al., 2004; Doyle et al., 2006) are adopted. We compute 13 co-occurrence features for each of the  $L$ ,  $a$ ,  $b$  channel of the image and concatenate them to create a 39-dimensional feature vector for each image. We use the same classifiers and the same CV techniques to evaluate the performance of this method, which is also shown in Table 2. As can be seen, the proposed method outperforms the texture-based method in all the tests.

Similar to Naik et al. (2008), two-class classification is also considered. By using two CV techniques, a total of 8 classification problems need to be solved. In these problems, we use the two

**Table 2**

Accuracy (%) and standard deviation of three-class classification for the proposed method and texture-based method evaluated by ROI-based CV and specimen-based CV. Bold values are the best accuracies in each column.

Classifier	ROI-based CV		Specimen-based CV	
	Proposed method	Texture-based method	Proposed method	Texture-based method
Adaboost (with decision stump as weak classifier)	75.3 (12.7)	56.1 (10.9)	68.3 (06.7)	52.2 (04.3)
Nearest Neighbor	80.3 (14.0)	72.9 (13.7)	69.5 (11.4)	57.5 (19.5)
Decision Tree (C4.5)	79.0 (11.2)	72.1 (16.3)	68.4 (11.4)	62.4 (18.9)
Naive Bayes	81.5 (08.6)	70.4 (18.1)	71.9 (09.3)	59.9 (21.7)
SVM (squared exponential kernel)	<b>87.8</b> (09.6)	81.7 (10.9)	<b>75.1</b> (10.6)	69.9 (15.7)
Feedforward Neural Network with one hidden layer	<b>87.8</b> (13.7)	<b>83.0</b> (13.8)	74.5 (07.8)	<b>71.4</b> (15.5)

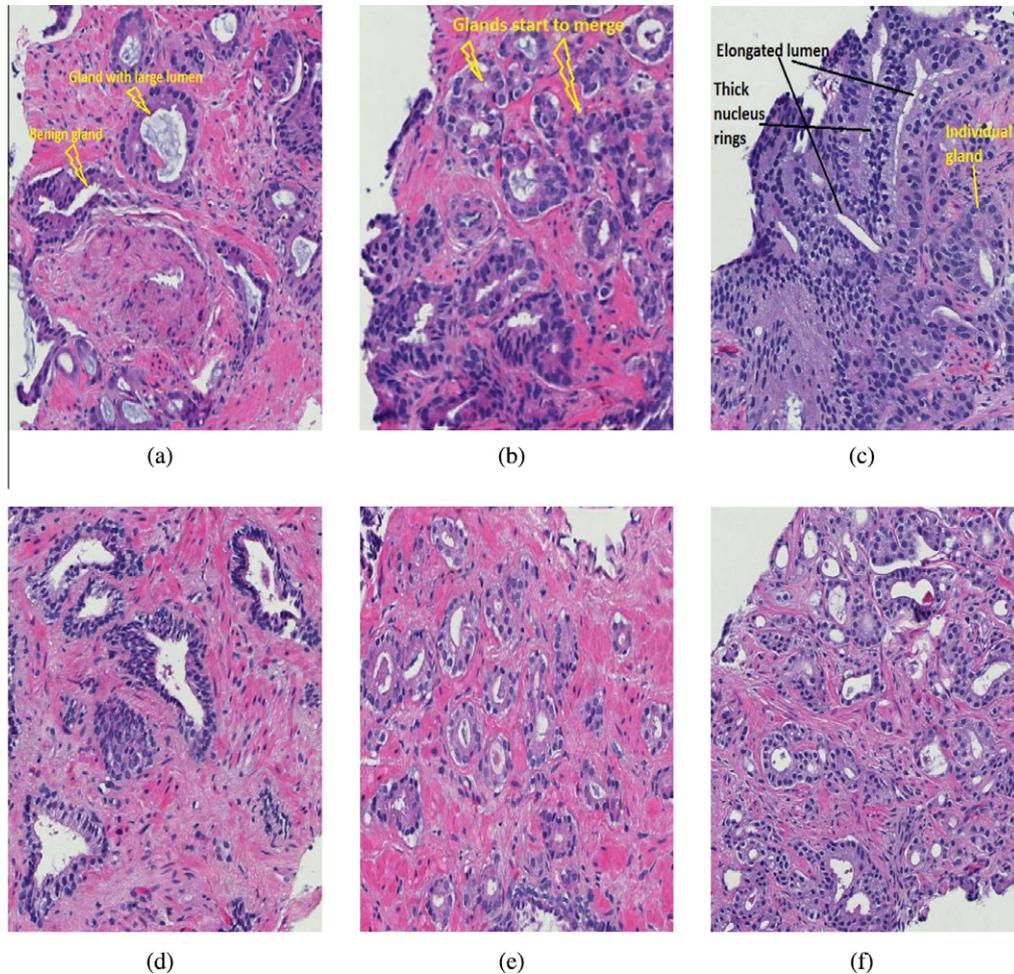
classifiers which performed the best in the three-class problem (SVM and FNN). We also compare the proposed method and the texture-based method on these problems. For the proposed method, SVM gives better results than FNN in all two-class classification

**Table 3**

Best accuracies achieved for 8 different two-class classification problems of the two methods (proposed and texture-based) evaluated by two cross-validation (CV) techniques. For the proposed method, SVM is better than FNN for all the 8 problems. For the texture-based method, the best classifier and its accuracy are shown for each problem. Bold values indicate the method with higher accuracy for each classification problem.

Classification problem	ROI-based CV		Specimen-based CV	
	Proposed method	Texture-based method	Proposed method	Texture-based method
Benign vs grade 3	<b>98.3</b> (05.0)	86.9 (14.5) (by FNN)	<b>93.2</b> (05.5)	74.9 (08.3) (by FNN)
Benign vs grade 4	<b>96.0</b> (12.0)	93.0 (11.4) (by SVM)	<b>89.5</b> (15.0)	87.8 (13.7) (by FNN)
Grade 3 vs grade 4	85.5 (13.1)	<b>86.0</b> (12.8) (by SVM)	70.7 (22.5)	<b>84.4</b> (12.3) (by SVM)
Benign vs carcinoma	<b>97.5</b> (05.0)	86.5 (12.8) (by SVM)	<b>94.2</b> (07.4)	80.3 (08.3) (by FNN)

problems. The accuracy of FNN is approximately 4% lower than SVM, on average. However, in the texture-based method, SVM outperforms FNN in some classification problems and FNN outperforms SVM in other problems. The best classification results of the two methods in 8 different classification problems are reported in Table 3. As can be seen, the proposed method is better than the texture-based method in all problems except for the grade 3 vs grade 4 classification problem. Since the dataset contains several



**Fig. 9.** Examples of classification results. Misclassifications: (a) Grade 3 pattern misclassified as benign, (b) Grade 3 pattern misclassified as grade 4 and (c) Grade 4 pattern misclassified as benign; Correct classifications: (d) Benign pattern, (e) grade 3 pattern and (f) grade 4 pattern.

images in which cancer is evolving from grade 3 to grade 4, i.e. some glands still appear as single units while other glands are merging with each other, the structural features of these ROIs are ambiguous which can degrade the performance of the proposed method. Moreover, due to the appearance of crystallized protein or due to the cutting direction when a tissue is sampled, the lumina of some glands in a tissue may be partially or totally occluded. These occluded lumina also affect the performance of our system because both the segmentation and the classification stages utilize lumen information. In our dataset, there are three cases in which grade 3 ROIs are misclassified as grade 4 because some of the lumina are occluded and also because the cancer in those ROIs are in the transition stage from grade 3 to grade 4. We do not see the affect of occluded lumina in other benign or grade 4 ROIs in our dataset.

Based on this observation, we propose a hierarchical (binary) classification scheme (Fig. 10) to classify a ROI  $I$  using both methods. In the first stage, we classify  $I$  as benign or carcinoma by using the proposed method because it performs better for this classification problem. If the result is benign, we stop. Otherwise, we continue to determine whether it is grade 3 or grade 4 carcinoma in the second stage by applying the texture-based method which is dominant for this problem.

We utilize the two-class classification results obtained in Table 3 to calculate the accuracy of the hierarchical scheme (HS), which gives the same results if performing separate experiments on the HS because the two classification stages are independent. Let  $p(\omega_i)$  denote the prior probability of class  $\omega_i$ , which is estimated from the dataset. Let  $p(\text{correct}|\omega_i)$  denote the probability that we correctly classify  $I$  when  $I$  belongs to class  $\omega_i$ . Since the probability that  $I$  is correctly classified,  $p(\text{correct})$ , depends on whether  $I$  is a benign, grade 3 or grade 4 region, we can calculate the overall accuracy as follows:

$$p(\text{correct}) = \sum_{i=1}^3 p(\text{correct}|\omega_i)p(\omega_i), \quad (6)$$

where  $\omega_1$ ,  $\omega_2$  and  $\omega_3$  denote the benign, grade 3, and grade 4 classes, respectively. While  $p(\text{correct}|\omega_1)$  is the accuracy of the first classification stage only,  $p(\text{correct}|\omega_2)$  and  $p(\text{correct}|\omega_3)$  involve the accuracies of both the stages since they require both to be correct. Let  $a_{BC}$  and  $a_{34}$  denote the accuracies of benign vs. carcinoma and grade 3 vs. grade 4 classification problems obtained from Table 3, respectively. We have  $p(\text{correct}|\omega_1) = a_{BC}$  and  $p(\text{correct}|\omega_2) = p(\text{correct}|\omega_3) = a_{BC}a_{34}$ .

Based on the distribution of three classes in the dataset, we have  $p(\omega_1) = 34/82$ ,  $p(\omega_2) = 28/82$  and  $p(\omega_3) = 20/82$ . Using the ROI-based CV technique ( $a_{BC} = 0.975$  and  $a_{34} = 0.86$ ), we have  $p(\text{correct}|\omega_1) = 0.975$  and  $p(\text{correct}|\omega_2) = p(\text{correct}|\omega_3) = 0.975 \cdot 0.86 = 0.838$ . So  $p(\text{correct}) = 89.5\%$  which is higher than the best accuracy

obtained by direct three-class classification, 87.8%. Using the specimen-based CV technique ( $a_{BC} = 0.942$  and  $a_{34} = 0.844$ ), we have  $p(\text{correct}|\omega_1) = 0.942$  and  $p(\text{correct}|\omega_2) = p(\text{correct}|\omega_3) = 0.942 \cdot 0.844 = 0.795$ . So  $p(\text{correct}) = 85.6\%$  which is higher than the best accuracy obtained by direct three-class classification, 75.1%. Hence, to classify a ROI into one of the three classes, it is better to use a HS which employs both the proposed method and the texture-based method.

Now, we show that using the two-class classification accuracies  $a_{BC}$  and  $a_{34}$  to compute the accuracy of the HS by Eq. (6) is valid as following. First, the input to the HS which is also the input to stage 1 of the HS can be benign or carcinoma. Hence, using  $a_{BC}$  as the accuracy of this stage is valid. Second, the input to stage 2 of the HS can also be benign or carcinoma since in stage 1, we may misclassify a benign sample as carcinoma. However, in this case, this benign sample is already counted as a misclassification of the HS from stage 1 and this misclassification is already included in the accuracy  $a_{BC}$ . So, all benign samples which are misclassified in stage 1 should be disregarded when considering the accuracy of stage 2. As a result, the accuracy of stage 2 is still  $a_{34}$ .

## 5. Summary and conclusions

We have proposed a novel method to analyze the glandular structure of a prostate tissue pattern in order to grade it as benign, grade 3 or grade 4 carcinoma. Basic underlying components of the tissue (nuclei, lumina, cytoplasm, blue mucin, stroma) and eventually gland units are segmented. Fifteen structural features are extracted to classify the tissue pattern, achieving state of the art classification results. Our algorithm utilizes nucleus and blue mucin components which were not used in the previous segmentation-based studies. Nucleus information not only facilitates the pattern grading but can also be helpful in detecting other structures of the prostate tissue such as seminal vesicles, paraganglia, eosinophilic crystalloids, perineural indentation or PIN (Prostatic Intraepithelial Neoplasia). Gland regions, which are not available in non-segmentation based approaches, can be used as landmarks for registering images of tissue slides in the same prostate region that are generated by different staining methods (H&E and IHC) to enhance the grading results. Gland regions can also be used to retrieve glands from a tissue image dataset, which may be of interest to pathologists. We plan to further improve the accuracy of grade 3 vs grade 4 classification by analyzing distinctive cytological features at a higher magnification scale. Finally, we are addressing the problem of detecting adenocarcinoma from a digitized tissue specimen, which requires processing a very large image (with an approximate size of  $90,000 \times 45,000$  pixels).

## Acknowledgement

Anil K. Jain's research was partially supported by WCU (World Class University) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R31-10008) to Korea university.

## References

- Catalona, W., Smith, D., Ratliff, T., 1991. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *New Engl. J. Med.* 324 (17), 1156–1161.
- Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape models—their training and application. *Comput. Vision Image Understand.* 61 (1), 38–59.
- Cootes, T., Edwards, G., Taylor, C., 2001. Active appearance models. *IEEE Transactions on Pattern Anal. Machine Intell.* 23, 681–685.
- Diamond, J., Anderson, N., Bartels, P., Montironi, R., Hamilton, P., 2004. The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia. *Hum. Pathol.* 35, 1121–1131.

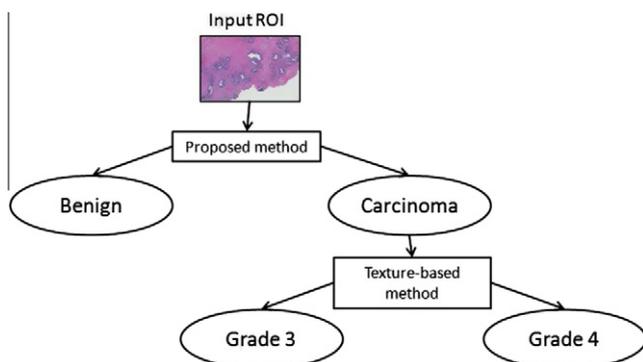


Fig. 10. The proposed hierarchical classification scheme based on binary classifications.

- Doyle, S., Madabhushi, A., Feldman, M., Tomaszewski, J., 2006. A boosting cascade for automated detection of prostate cancer from digitized histology. *MICCAI*, 504–511.
- Fairman, H.S., Brill, M.H., Hemmendinger, H., 1997. How the CIE 1931 color-matching functions were derived from Wright–Guild data. *Color Res. Appl.* 22, 11–23.
- Franz, A., 1991. Voronoi diagrams – A survey of a fundamental geometric data structure. *ACM Comput. Surv.* 23 (3), 345–405.
- Gleason, D., 1977. Histologic grading and clinical staging of prostatic carcinoma. In: Tannenbaum, M. (Ed.), *Urologic Pathology: The Prostate*. Lea and Febiger, Philadelphia, PA, pp. 171–198.
- Gleason, D., 1992. Histologic grading of prostate cancer: A perspective. *Hum. Pathol.* 23, 273–279.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I., 2009. The Weka data mining software: An update, *SIGKDD Explorations*, issue 1.
- Haralick, R., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. *IEEE Trans. Systems Man Cybernet.* SMC 3, 610–621.
- Jain, A., 1989. *Fundamentals of Digital Image Processing*. Prentice Hall.
- Khouzani, K.J., Zadeh, H.S., 2003. Multiwavelet grading of pathological images of prostate. *IEEE Trans. Biomed. Eng.* 50 (6), 697–704.
- Kiernan, J.A., 2001. *Histological and Histochemical Methods: Theory and Practice*. A Hodder Arnold Publication.
- Li, C., Xu, C., Gui, C., Fox, M.D., 2005. Level set evolution without re-initialization: A new variational formulation. *IEEE Internat. Conf. on Computer Vision and Pattern Recognition (CVPR)* 1, 430–436.
- Mason, M., 1964. Cytology of the prostate. *J. Clin. Pathol.*, 581–590.
- Naik, S., Doyle, S., Madabhushi, A., Tomaszewski, J., Feldman, M., 2008. Automated nuclear and gland segmentation and gleason grading of prostate histology by integrating low-, high-level and domain specific information. In: *Special Workshop on Computational Histopathology (CHIP)*, in conjunction with 5th IEEE Internat. Symp. on Biomedical Imaging (Invited Paper), pp. 284–287.
- Tabesh, A., Teverovskiy, M., Pang, H., 2007. Multifeature prostate cancer diagnosis and gleason grading of histological images. *IEEE Trans. Med. Imaging* 26 (10), 1366–1378.
- Tai, S.K., Li, C.Y., Wu, Y.C., Jan, Y.J., Lin, S.C., August 2010. Classification of prostatic biopsy. *Digital Content, Multimedia Technology and its Applications (IDC)*. In: 2010 6th Internat. Conf. on, pp. 354–358.
- Teverovskiy, M., Kumar, V., Ma, J., Kotsiantis, A., Verbel, D., Tabesh, A., Pang, H., Vengrenyuk, Y., Fogarasi, S., Saidi, O., 2004. Improved prediction of prostate cancer recurrence based on an automated tissue image analysis system. *Proc. IEEE Internat. Symp. Biomed. Imag.*, 257–260.
- US Cancer Statistics Working Group, 2010. *United states cancer statistics: 1999–2006 incidence and mortality web-based report*. Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute. <<http://www.cdc.gov/uscs>>.
- Wikipedia contributors, 2010. *Lab color space*. <[http://en.wikipedia.org/wiki/Lab\\_color\\_space](http://en.wikipedia.org/wiki/Lab_color_space)>.