Prostate Cancer Grading: Use of Graph Cut and Spatial Arrangement of Nuclei

Kien Nguyen, Anindya Sarkar, Anil K. Jain, *Fellow, IEEE*

***Abstract*—Tissue image grading is one of the most important steps in prostate cancer diagnosis, where the pathologist relies on the gland structure to assign a Gleason grade to the tissue image. In this grading scheme, the discrimination between grade 3 and grade 4 is the most difficult, and receives the most attention from researchers. In this study, we propose a novel method (called nuclei-based method) that (i) utilizes graph theory techniques to segment glands and (ii) computes a *gland-score* (based on the spatial arrangement of nuclei) to estimate how similar a segmented region is to a gland. Next, we create a fusion method by combining this nuclei-based method with the lumen-based method presented in our previous work to improve the performance of grade 3 vs grade 4 classification problem (the accuracy is now improved to 87.3% compared to 81.1% of the lumen-based method alone). To segment glands, we build a graph of nuclei and lumina in the image, and use the normalized cut method to partition the graph into different components, each corresponding to a gland. Unlike most state-of-the-art lumen-based gland segmentation method, the nuclei-based method is able to segment glands without lumen or glands with multiple lumina. Moreover, another important contribution in this research is the development of a set of measures to exploit the difference in nuclei spatial arrangement between grade 3 images (where nuclei form closed chain structure on the gland boundary) and grade 4 image (where nuclei distribute more randomly in the gland). These measures are combined to generate a single gland-score value, which estimates how similar a segmented region (which is a set of nuclei and lumina) is to a gland.**

***Index Terms*—Prostate cancer, Gleason grading, gland segmentation, nuclei, lumen, stroma, normalized cut**

I. Introduction

According to the American Cancer Society [1], prostate cancer is the most prevalent type of cancer in men. In the year 2013, the estimated new cases of prostate cancer is 238,590, accounting for 28% of all the new cancer cases. Furthermore, the estimated number of deaths from this type of cancer is 29,720, which is the second highest (after lung and bronchus).

In prostate cancer diagnosis, one of the most important stages is the examination and grading of the tissue slide using Gleason grading method [2]. The tissue slide is obtained from the prostate biopsy and is digitized for convenience in examination. The Gleason grading method defines five numerical grades, from least aggressive (1) to most aggressive (5), based on the gland structures in the image (Fig. 2). Due to the developments in digital pathology (e.g., high resolution tissue slide scanners), the volume of image data being generated has increased immensely. Moreover, the high cost and the large interobserver variability [3] of the grading procedure are also challenging problems in digital pathology. As a result, an automatic image analysis based grading method is desired.

According to a recent study [4], grades 1 and 2 together are considered as normal tissue. Moreover, grade 5 is not commonly present in prostate tissue databases. Hence, for the tissue image grading problem in this paper, we focus on the three-class classification problem: normal, grade 3 and grade 4 prostate tissue types. Examples of these three grades are shown in Fig. 1. The differences in glandular structures among these three classes are as follows:

- In normal tissue images, glands are typically large with arbitrary shape. These glands usually contain abundant nuclei on the boundary. In contrast, grade 3 images usually contain small glands, with circular shape and a thin nuclei layer on the boundary. Finally, glands in grade 4 images are generally smaller, with fewer nuclei on the boundary than those in grade 3 images.
- Compared to glands in normal images, glands in grade 3 and grade 4 images stay closer together, and neighboring glands have similar shape and size.
- Glands in grade 4 tissue images tend to be fused together rather than being isolated as those in grade 3 images. Hence, the gland structures are poorly-defined in grade 4 images.

We first present a lumen-based method based on our previous work to address different classification problems (normal vs grade 3, normal vs grade 4, grade 3 vs grade 4, and complete three-class classification). Among these problems, the grade 3 vs grade 4 classification is the most difficult problem because the gland structures in grade 3 and 4 tissue images are not as well-defined as those in normal tissue images (where glands have clear, large lumen, clear boundary and are well separated from each other). As a result, we propose the nuclei-based method and its combination with the lumen-based method to improve the grade 3 vs grade 4 classification result.

II. Related work

The published studies addressing the automatic prostate cancer grading problem can be classified into three different approaches: texture-based, nuclei-architecture-based, and gland-segmentation-based approaches. *It is important to*

K. Nguyen and A. Sarkar are with Ventana Medical Systems, Inc., Moutain View, CA 94043 USA. The work was done when Kien Nguyen was with Michigan State University. E-mail: kien.nguyen@contractors.roche.com, anindya.sarkar@ventana.roche.com.

A.K. Jain is with the Dept. of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824 USA. He is also with the Department of Brain and Cognitive Engineering, Korea University, Anamdong, Seongbukgu, Seoul 136-713, Republic of Korea. E-mail: jain@cse.msu.edu.

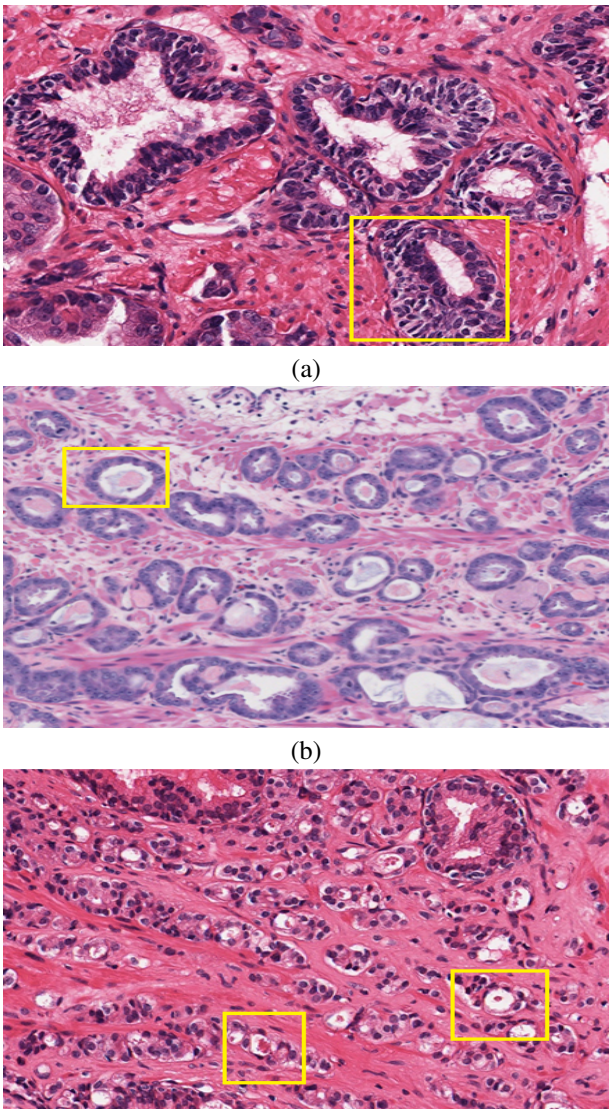


Fig. 1. Three types of tissue classes of interest for the Gleason grading problem: (a) Normal tissue, (b) Gleason grade 3 tissue and (c) Gleason grade 4 tissue. Yellow rectangles denote typical glands in these images.

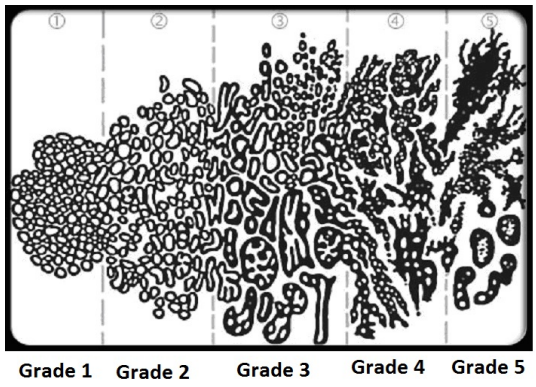


Fig. 2. The five different grades in the Gleason grading method.

note that all these studies have used different tissue image databases because of the lack of public databases. However, in this paper, we compare different methods using the same database.

1. Texture-based approach: This is the most popular approach in the literature. Different types of texture features in the image were used for tissue classification. In [5], features computed from the gray level co-occurrence matrix (GLCM) were combined with lumen area to classify an image region into normal, cancer, and stroma. An accuracy of 79% was achieved for the classification of the image regions in 8 tissue images. In [6], the authors applied a multiwavelet transform on the images and computed the entropy and energy of the multiwavelet coefficients. They reported 97% accuracy when classifying 100 images into grades 2, 3, 4 and 5. A similar technique, cardinal multiridgelet transform, was proposed in [7] to classify an image into grade 3 vs grade 4. By evaluating the method on 42 images, the authors reported 93.75% accuracy. In [8], Khurd et al. applied the popular bag-of-words (BoW) model to the tissue image grading problem to classify an image into grade 3 vs grade 4. In this method, the filter responses obtained by applying a bank of invariant filters on the images were clustered into 16 clusters (called textons). The image pixels were assigned to different textons and the image was represented by a histogram of texton. Using this method, they reported 94% accuracy on a database of 75 images. Tai et al. [9] computed the fractal dimension features for the images and the sub-bands of the images to classify them into normal, grade 3, 4 and 5. They performed the evaluation on a database of 1,000 images and reported 86.3% accuracy.

The texture features used in the published studies are generic features computed from the "low level" image information (local information derived from neighboring image pixels). No domain knowledge about prostate tissue image and glandular structures is utilized in this approach. As a result, it is unclear how these features are able to capture the gland structure differences of different Gleason grades (which pathologists use for the grading). Moreover, these texture features are computed from all pixels in the image, while only the glandular regions in the image are necessary in determining the Gleason grade. Hence, the information from the non-glandular regions may introduce noise in the feature extraction stage.

2. Nuclei-architecture-based approach: The features used in this approach are computed from the distribution of all nuclei in the image. In [10], Khurd et al. first built a network of all nuclei in the image. Next, network cycle features were computed from the statistics about the cycles in the network. These features resulted in 91.5% accuracy when evaluated for grade 3 vs grade 4 classification on 75 images. In [11], Doyle et al. computed nuclei architecture features from the nuclei density, Voronoi diagram, Delaunay triangulation and minimum spanning tree formed by all the nuclei. Further, they combined these features with texture features (first-order statistics, co-occurrence features, and

steerable filters) to classify prostate tissue images into seven classes, including: PIN, epithelium, stroma, atrophy, Gleason grades 3, 4 and 5. They reported accuracies in identifying grade 3 and grade 4 images among all image types of 77% and 76%, respectively.

Since these features were computed from all nuclei in the image but not only from the glands, the intuition on how these features describe the structural differences between glands of different grades is not apparent. Further, tissue images always contain stromal nuclei (nuclei in the stroma background) that are not correlated with the gland structures. Hence, using these nuclei to extract features may generate noisy information.

3. Gland-segmentation-based approach: In this approach, glands are first segmented followed by gland feature extraction. Naik et al. [12] used the level set method [13] to segment glands and then extracted different shape features of the glands and lumen. Using a database of 44 images, they reported the following results for different two-class classification problems: 86.35% for normal vs grade 3, 92.9% for normal vs grade 4, and 95.19% for grade 3 vs grade 4. In [14], the authors used the region growing method to perform gland segmentation. Gland area was the only feature used for the normal vs cancer tissue image classification problem. The method was evaluated on a database of 62 images and the authors reported an AUC (area under the ROC curve) of 0.92. Gland segmentation using level set and region growing methods was also used in many other studies [15]–[17], although the objectives of these studies are not to classify tissue images based on Gleason grading. In our previous work [18], we also introduced a method for gland segmentation called nuclei-lumen association. The features we extracted for the glands include structural and contextual features. However, in this problem, we performed gland classification instead of tissue image grading.

The shortcoming of state-of-the-art gland segmentation methods is that they rely on the presence of lumen, i.e., they fail to detect glands without lumen or glands with multiple lumina.

The method proposed in this paper can be considered as the combination of the approaches 2 and 3 mentioned above, i.e., both the nuclei spatial arrangement (a type of nuclei architecture information) and gland structure information are used. Gland structure information describes properties of the gland components (lumen, cytoplasm, nuclei) such as color intensity, size, and shape, which are salient for glands with clear lumen. On the other hand, nuclei spatial arrangement describes how the nuclei distribute in the glands, which is very useful for glands with no lumen (these glands are not found based on lumen, so gland structure information cannot be computed).

Contributions of the paper:

1) We propose a lumen-based method for tissue image grading that is built on top of our previous work on gland segmentation and gland feature extraction [18]. This method leads to comparable classification accuracies to the state-of-the-art grading method [11].
2) We propose a novel nuclei-based gland segmentation method by formalizing the gland segmentation problem in a graph cut framework. This method is able to detect glands without lumen and glands with multiple lumina, which overcomes the limitation of state-of-the-art gland segmentation methods.
3) We propose different shape measures to extract the nuclei spatial arrangement from the segments; this information is useful to discriminate between grade 3 and grade 4 tissue images. Unlike nuclei-architecture-based methods [10], [11], we extract nuclei arrangement information from the glands instead of from all nuclei in the tissue. This way, we are able to (i) discard the stromal nuclei and (ii) design features that intuitively describe the differences in nuclei spatial arrangement between glands of these grades.
4) Since the lumen-based and nuclei-based methods complement each other, we combine them to improve the grade 3 vs grade 4 classification results (an improvement of 5% accuracy over the state-of-the-art grading method [11] is obtained). Grade 3 vs grade 4 classification is an important and a very challenging task even for pathologists.

The remainder of the paper is organized as follows. In Sec. III, we describe the lumen-based method and show its performance for the three-class Gleason grading problem. The nuclei-based gland segmentation method is presented in Sec. IV, where we discuss: tissue component detection (Sec. IV-A), nuclei-lumina graph construction (Sec. IV-B), normalized cut for gland segmentation and gland-score definition (Sec. IV-C and IV-D), and the use of gland-score for grade 3 vs grade 4 image classification (Sec. IV-E and IV-F). In Sec. V, we present the fusion method and its classification result. Several discussions are presented in Sec. VI before we conclude the paper in Sec. VII.

III. Lumen-based Grading Method

Since lumen is the central component of the gland, we aim to rely on lumen to perform gland segmentation. To achieve this, we utilize the nuclei-lumen association (NLA) method[1] and the structural-contextual features proposed in our previous work [18] (see Fig. 3):

i. **Extract discriminative features**: We segment glands from the tissue image and extract 22 structural-contextual features from each gland. Among these features, there are 19 structural features (color information of the cytoplasm and nuclei, nuclei density, lumen size and shape, and gland morphology), and 3 contextual features (neighborhood crowdedness, size similarity, and shape similarity) [18].
ii. **Remove artifacts:** We identify artifacts (noisy regions) from these extracted glands[2]. In order to identify artifacts,

[1]Since the NLA method segments glands by first detecting the lumen followed by finding nuclei associated with the lumen, we call it a lumen-based gland segmentation method.

[2]Artifact removal was also mentioned in [12] and [19]

TABLE I
GLOSSARY OF MAIN NOTATIONS USED IN THE PAPER.

Notation	Description
NLA method	Nuclei-lumen association gland segmentation method [18]
CV	Cross-validation
$G = (V, E)$	A generic graph with vertex set V and edge set E
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	The nuclei-lumina graph for the entire image ($\mathcal{V}$: vertices, $\mathcal{E}$: edges)
$\mathcal{N} = \{n_1, \ldots, n_N\}$	The set of nuclei detected in the image
$\mathcal{L} = \{l_1, \ldots, l_L\}$	The set of lumina detected in the image
$\mathbf{S}$	The mask of detected stroma region
NNLink, NLLink	Nucleus-nucleus-link, nucleus-lumen-link
Ω_j	Conical search region (used to search for nuclei in NNLink and NLLink creation procedures)
$(r_n, \theta_n), (r_l, \theta_l)$	Radius and angle of Ω_j used for NNLink creation (Alg. 1) and NLLink creation (Alg. 2), respectively
p_j^l	Lumen-points (the points sampled on the lumen boundary) used to perform NLLink creation
Ncut	The normalized cut value: the fraction of the total connection between two components (resulted from the cut) to the total connection between the components and the entire graph (Eq. 1)
δ_c	The cut threshold, used to determine when to stop the normalized cut process (Sec. IV-C)
MST	Minimum spanning tree of a graph (Sec. IV-D)
P^*	Backbone of the MST, i.e., the path with most vertices in the MST (Sec. IV-D)
$M^1_{MD}, M^1_V, M^1_L, M^1_{CI}$	Closed chain structure measures: mean degree, fraction of backbone in the MST, closeness index (Sec. IV-D)
M^2_E, M^2_I, M^2_{CI}	Ellipse measures: fitting error, percentage of inliers, coverage index (Sec. IV-D)
$\{p_i\}$	The set of points (nuclei) whose closed chain structure and ellipse measures are computed
$g(C)$	The gland-measure vector (containing all close chain structure and ellipse measures) of a graph component C (Eq. 4)
ψ	The gland-score function (or gland-score value of a component) (Sec. IV-D)
ψ^I	The image-gland-score (Sec. IV-F)

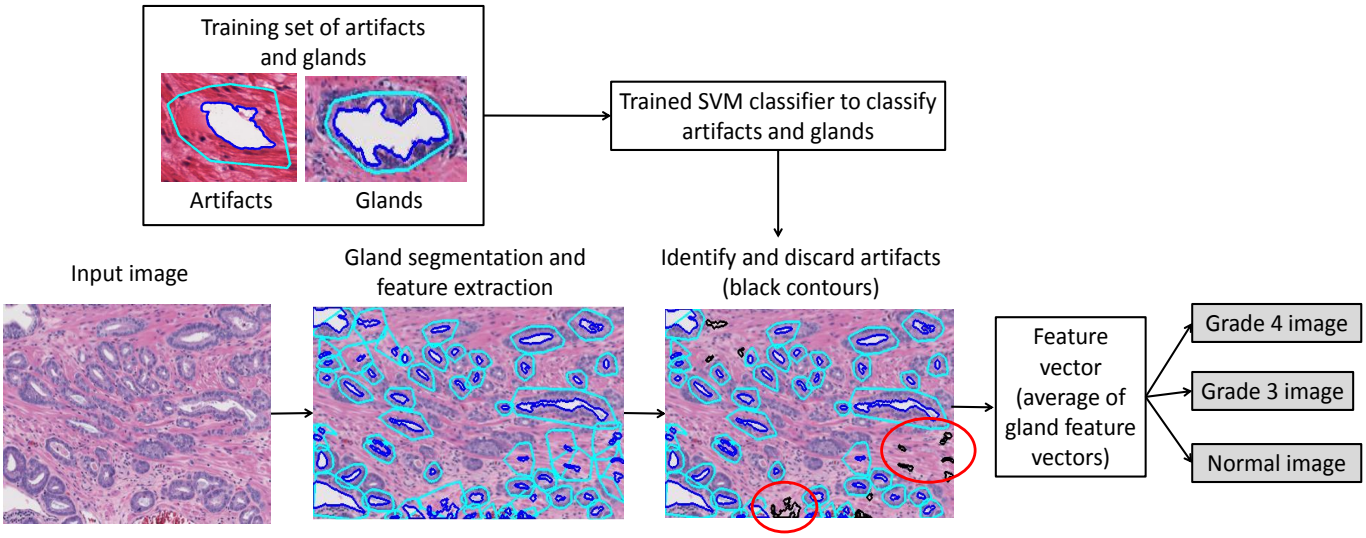


Fig. 3. Flowchart of the lumen-based grading method, which utilizes the gland segmentation and gland feature extraction methods proposed in [18].

we obtain an independent gland dataset, each of which is associated with a label (artifact or true gland), and learn a SVM classifier to discriminate them. The artifacts are discarded from subsequent processing.

iii. **Perform the grading**: We compute the averages of the gland features to create a 22-dim feature vector to represent the image, and use the SVM classifier (RBF kernel) to classify the image into normal, grade 3 or grade 4.

A flowchart of the lumen-based method is shown in Fig. 3.

Evaluation of the Lumen-based Grading Method

A. Database and Evaluation Strategy

The database used in our experiment includes 317 tissue images at 20× magnification with an average size of 1,400 × 1,380 pixels. These tissue images are taken from 29 patients. Among these 317 images, there are 113 normal, 134 grade 3 and 70 grade 4 images. The grade of each image was determined by a pathologist. Examples of the images are shown in Fig. 1. Using SVM classifier (RBF kernel), we perform 10-fold cross-validation (CV) and compute the average classification accuracy and the standard deviation to evaluate the proposed method. In each iteration of the 10-fold CV, the database is divided into training data $\mathbf{T}_i$ and test data $\mathbf{t}_i$. We use a different 5-fold CV on $\mathbf{T}_i$ to find the best parameters for SVM (c and *gamma*). These parameters are applied on $\mathbf{t}_i$ for prediction. The same evaluation strategy is applied to all other methods evaluated in this paper. Note that this strategy is adopted as an effort to avoid the overfitting problem due to the small size of the dataset. A large independent test dataset (which is not available), however, is a better solution

TABLE II
AVERAGE CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION OBTAINED BY DIFFERENT METHODS FOR FOUR DIFFERENT CLASSIFICATION PROBLEMS. THE ACCURACIES OF THE TWO LEADING METHODS ARE SHOWN IN BOLD.

Grading Method	All three classes	Normal vs Grade 3	Normal vs Grade 4	Grade 3 vs Grade 4
Proposed lumen-based method	**84.2** (5.5)	**96.3** (4.0)	**97.2** (3.9)	**81.1** (9.3)
Texture-Nuclei architecture [11]	**84.2** (6.2)	**96.7** (3.2)	**98.4** (4.2)	**82.3** (10.1)
Method in [20]	79.0 (7.2)	92.0 (7.5)	93.2 (5.2)	77.5 (8.5)
GLCM [21]	75.4 (8.1)	90.3 (4.7)	90.8 (5.6)	74.9 (9.1)
Multiwavelet [6]	74.2 (7.0)	86.4 (9.4)	89.1 (4.3)	74.9 (12.2)
BoW-filter [8]	78.6 (6.7)	88.9 (8.2)	94.0 (6.1)	75.2 (11.6)
BoW-SIFT [22]	78.1 (6.6)	91.0 (7.4)	92.9 (6.4)	78.9 (8.7)
Spatial pyramid matching [23]	77.6 (6.6)	89.8 (6.6)	88.7 (5.6)	74.4 (7.2)

to evaluate all the methods mentioned in this paper.

B. Tissue Image Grading Evaluation

To show the robustness of the proposed lumen-based method, we compare it with seven different methods:

i. Method in [20]: Our previous method for tissue image classification.
ii. GLCM: features computed from the gray-level co-occurrence matrix [21].
iii. Multiwavelet: features computed using multiwavelet transform (similar to [6]). See Sec. II.
iv. BoW-filter: bag-of-words (BoW) model and the maximum response filters (similar to [8]). See Sec. II.
v. BoW-SIFT: bag-of-words model and the scale-invariant feature transform (SIFT) algorithm [22].
vi. The spatial pyramid matching method [23] (which also uses SIFT to compute the features).
vii. Texture-Nuclei architecture: the combination of texture features and nuclei architecture features (similar to [11]). See Sec. II.

Methods (ii) and (iii) are popular texture analysis methods for histological images. Methods (iv-vi) are popular methods for image classification (or recognition) in computer vision. Finally, method (vii) is the most recent published study on automatic Gleason grading. We have already shown in [18] that the proposed lumen-based method outperformed other gland-segmentation-based methods.

Besides the three-class classification problem, we also report results of the following three two-class classification problems: normal vs grade 3, normal vs grade 4, and grade 3 vs grade 4 (Table II). Note that for the grade 3 vs grade 4 classification problem, we only use the 19 structural features and exclude the 3 contextual features because in these images, glands have similar contextual properties (multiple glands staying close together). Based on the results in Table II, we make the following observations:

- The proposed lumen-based method obtains comparable results to the state-of-the-art method (texture-nuclei architecture [11]). These two methods obtain the best results among all the methods.
- The normal vs cancer (grade 3 and grade 4) classification is easier than the grade 3 vs grade 4 classification because the differences in gland structures between normal and cancer images are larger than those between grade 3 and grade 4 images.

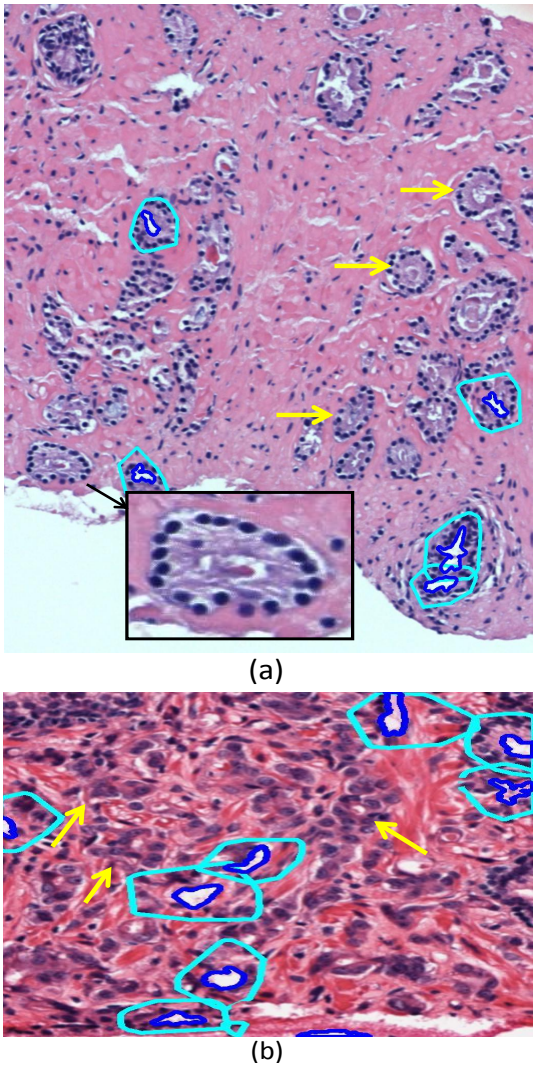


Fig. 4. Limitation of the lumen-based method. Many glands are not detected by the lumen-based method in both grade 3 image (a) and grade 4 image (b) (yellow arrows). The structural features do not capture the nuclei arrangement difference between grade 3 and grade 4 images. In grade 3 image, nuclei arrange as closed chain structure on the gland boundary (yellow and black arrows in (a)). In grade 4 image, nuclei are more randomly distributed since glands fuse together and are poorly defined (yellow arrows in (b)).

In the next section, we propose a *nuclei-based method* to improve the grade 3 vs grade 4 classification result.

IV. NUCLEI-BASED GRADING METHOD

The lumen-based method presented in Sec. III has the following limitations (Fig. 4):

i. In some of the grade 3 and grade 4 images, lumen is not present inside the glands (yellow arrows). As a result, the

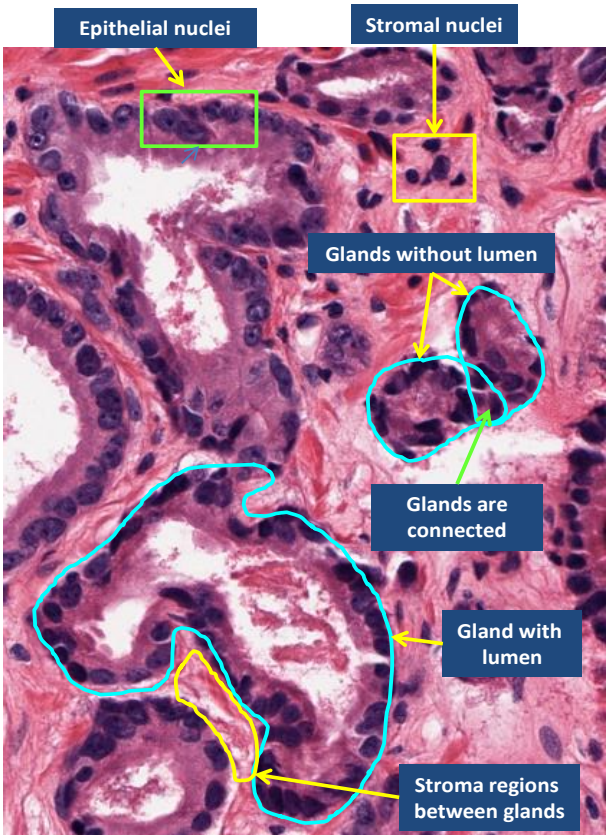


Fig. 5. A different perspective for gland segmentation.

lumen-based method cannot detect these glands[3].

ii. The structural features [18] do not capture the difference in nuclei arrangement between grade 3 and grade 4 images.

To overcome these limitations, we propose a nuclei-based gland segmentation method, and a method to extract the nuclei arrangement information in the image.

The nuclei-based method relies on a different perspective for gland segmentation (Fig. 5):

i. Each gland is considered as a group of epithelial nuclei [4] that are close together, which may or may not contain lumen in the center. Nuclei usually form a closed chain structure, or an ellipse on the gland boundary.

ii. Stroma rarely appears inside the gland area but mostly in between different glands. However, *glands are not perfectly separated by stroma*, i.e., we may see non-stroma area in between neighboring glands, or glands can be connected.

In the nuclei-based method, we aim to seek a strategy to group the nuclei and lumina (if lumina are available) belonging to the same gland together. We achieve this goal by using graph theoretic techniques: we model the relationship between nuclei and lumina in the image by a nuclei-lumina graph $\mathcal{G}$, where each nucleus or lumen is considered as a vertex in $\mathcal{G}$. Each edge in $\mathcal{G}$ is created as a "link" between a nucleus and a nucleus or between a nucleus and a lumen. The links indicate which nuclei and lumina are likely to belong to the same gland. Finally, the normalized cut method [24] is applied on this graph to find and remove the weakest sets of links (the links that are likely to connect different glands) to partition them into different connected components, each of which corresponds to a gland. Once glands are segmented, we compute a set of measures to evaluate how much the segment resembles a gland (nuclei forming a closed chain structure), and derive a gland-score as a discriminative feature for grade 3 vs grade 4 classification.

A flowchart of the nuclei-based method is illustrated in Fig. 6. Note that the gland-score is used for two purposes: (i) to find the best gland segment obtained during the recursive normalized cut process, and (ii) to use as an image feature for classification (after the segmentation is complete). A segmentation example of the nuclei-based method and a qualitative comparison with the lumen-based segmentation method are shown in Fig. 7. Here, the nuclei-based method is able to find glands without detected lumen, and it does not generate multiple segments for glands with multiple detected lumina[5] as is the case with the lumen-based method.

A. Tissue Component Detection

We first detect the tissue components in the image, namely, lumen, nuclei, and stroma.

1. Detect Nuclei:
Since most nuclei have circular shapes and are of similar size, we use the radial-symmetry-based method [25] to detect the nuclei. The goal of the method is to detect the centers of the circular regions by using a voting scheme. To perform the voting, pixels with strong gradient magnitude in the image are selected (referred to as voting pixels) to cast the votes for its neighborhood region (voting region). Pixels with strong gradient magnitude are chosen because they are likely to be the pixels on the nuclei boundary, which can effectively vote for the nuclei centers. The advantage of this method is that it can detect the clumped nuclei [25].

2. Identify Epithelial Nuclei:
We classify the nuclei into epithelial nuclei (e-nuclei) and stromal nuclei (s-nuclei). To perform the classification, we compute textural features in the nuclei neighborhood of size s $\times$ s pixels (s = 40 is used here)[6]. These textural features capture the information about the neighborhood of the nuclei, which can be either stroma regions (for s-nuclei) or cytoplasm regions (for e-nuclei). Hence, they can be used in discriminating the two nuclei types. We use an SVM classifier (RBF kernel) to perform the classification[7]. The e-nuclei

[3]This is also the limitation of state-of-the-art gland segmentation methods [12], [14]–[17] since they also rely on lumen.

[4]Epithelial nuclei are nuclei on the gland boundary, while stromal nuclei are nuclei scattered in the stroma region.

[5]The multiple lumina observed in a gland are the result of the 2D projection of a branched lumen within a glandular epithelial mass in 3D.

[6]The features we used include color histogram and those computed from the GLCM [21].

[7]We apply the cross-validation technique on a set of training nuclei to find the best SVM parameters (c and gamma). The best CV accuracy obtained in this learning process is 99%.

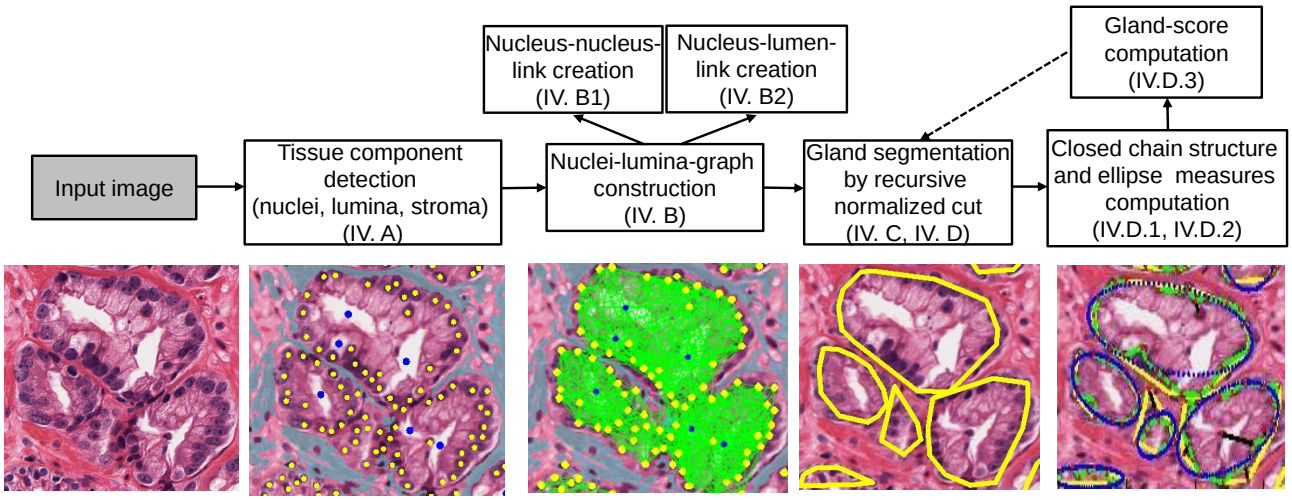
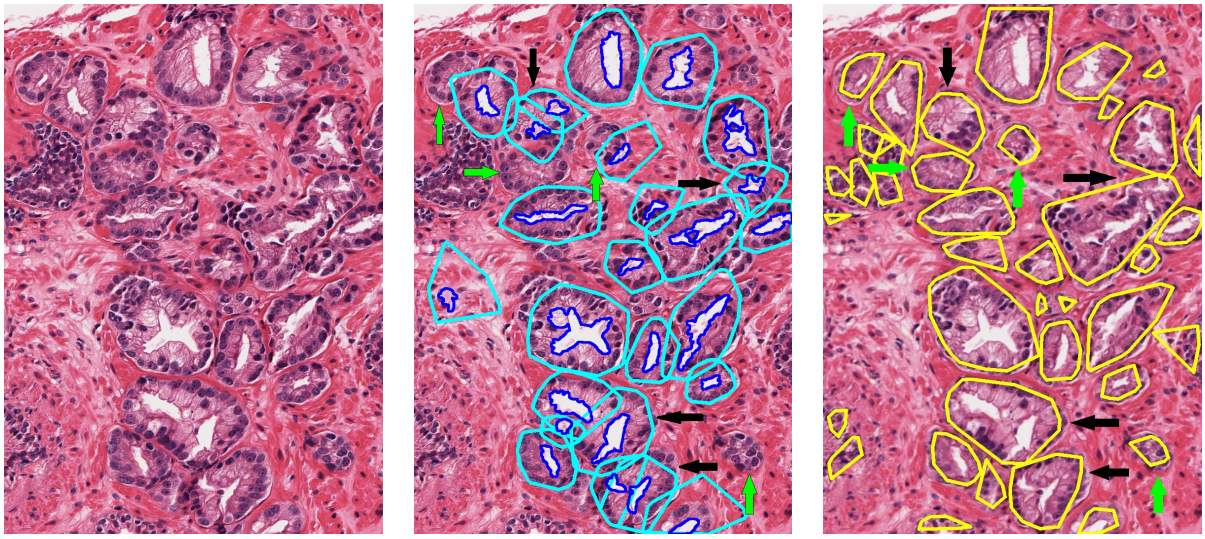
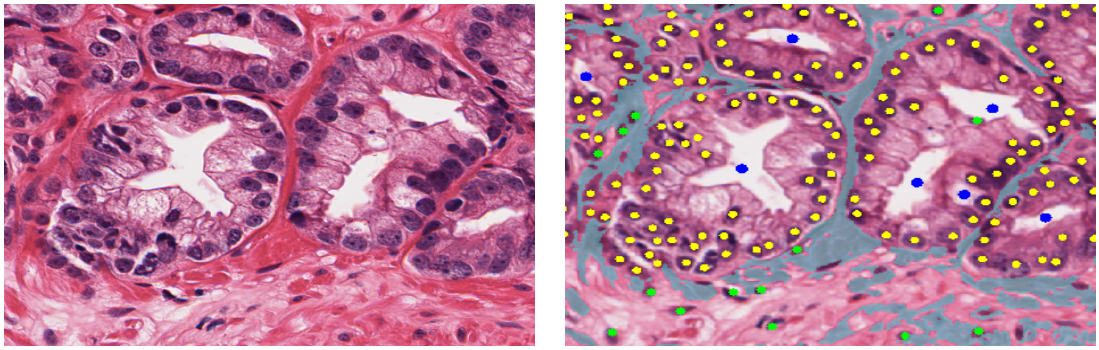


Fig. 6. Flowchart of the nuclei-based method (Sec. IV) to segment glands and compute their gland-scores, which are useful for grade 3 vs grade 4 classification. The subsection describing each module is indicated.



(a) (b) (c)

Fig. 7. Qualitative comparison between the proposed nuclei-based and lumen-based gland segmentation methods. (a) Input tissue image. (b) Segmentation result of the lumen-based method, where detected lumina are shown as blue contours. (c) Segmentation result of the nuclei-based method. The black arrows indicate glands with multiple lumina, while the green arrows indicate glands with no detected lumen. The nuclei-based method segments these glands successfully while the lumen-based method does not.



(a) (b)

Fig. 8. Results of tissue component detection. (a) Input image. (b) Detection result where yellow and green dots denote epithelial and stromal nuclei, respectively. Blue dots denote the lumen centers. Cyan regions denote stroma.

detected from the image are used for further processing while the s-nuclei are discarded. For simplicity, we will use the term "nuclei" to indicate the e-nuclei in the remainder of the paper.

Finally, stroma and lumina are detected using the k-means clustering procedure described in [18]. An example of tissue component detection is shown in Fig. 8.

B. Nuclei-Lumina Graph Construction

The nuclei-lumina graph demonstrates the relationship between (i) nuclei and nuclei and (ii) lumina and nuclei in the image, i.e., which nuclei and lumina should belong to the same gland. We formalize the problem as following.

i. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denotes the nuclei-lumina graph, where $\mathcal{V}$ denotes the vertex set and $\mathcal{E}$ denotes the edge set of the graph.

ii. Let $\mathcal{N} = \{n_1, n_2, \ldots, n_N\}$ and $\mathcal{L} = \{l_1, l_2, \ldots, l_L\}$ denote the sets of N nuclei and L lumina detected from the image, respectively.

iii. Define $\mathcal{V} = \mathcal{N} \cup \mathcal{L}$. $\mathcal{V}$ can also be described as $\mathcal{V} = \{v_1^n, v_2^n, \ldots, v_N^n, v_1^l, v_2^l, \ldots, v_L^l\}$, where $\{v_1^n, v_2^n, \ldots, v_N^n\}$ are the N vertices corresponding to the nuclei set $\mathcal{N}$, and $\{v_1^l, v_2^l, \ldots, v_L^l\}$ are the L vertices corresponding to the lumina set $\mathcal{L}$.

iv. Construct the edge set $\mathcal{E}$ ($\mathcal{E} \leftarrow \emptyset$ in the beginning) so that if $\exists v_i v_j \in \mathcal{E}$, the two vertices v_i, v_j (v_i, v_j can be either a nucleus or a lumen) have potential to belong to the same gland[8]. To construct $\mathcal{E}$, we use two procedures: nucleus-nucleus-link creation and nucleus-lumen-link creation.

1. Nucleus-Nucleus-Link Creation

Because nuclei belonging to the same glands stay close together, we aim to create links between the neighboring nuclei to segment the glands.

We define a *link* between two nuclei n_i and n_j if they are likely to belong to the same gland. We develop an algorithm, termed as **NNLink** (Algorithm 1), to find all the nuclei that link to a nucleus of interest n_i. See Fig. 9. Since the objective of the nucleus-nucleus-link creation is to find glands without lumen (mostly small-sized and average-sized glands), we choose the radius $r_n = 100$ pixels (at 20×), which corresponds to the size of an average-sized gland. For the conical angle θ_n, we select $\theta_n = \pi/12$ by estimating the density of nuclei on the boundary of the glands.

The reason for choosing the closest nucleus in each conical search region (line 6 in Algorithm 1) is that this nucleus is most likely to belong to the same gland as n_i. For example, there are some nuclei in Fig. 9c that fall within the conical search regions, yet do not belong to the same gland as n_i (e.g., the nuclei indicated by red arrows).

Although there are some *bad links* (links between n_i and the nuclei not belonging to the same gland as n_i) created (e.g., the links indicated by black arrows in Fig. 9d), these links are generally outnumbered by the *good links* (links between n_i and the nuclei belonging to the same gland as n_i), due to the assumption about stroma (stroma mostly appears in between glands rather than inside the glands). Hence, by applying a global method like normalized cut (Sec. IV-C), the bad links are likely to be removed.

[8]We use the term "potential" since the final decision that the two vertices belong to the same gland or not will be decided by the normalized cut procedure applied on the graph (Sec. IV-C).

Algorithm 1 Nucleus-Nucleus-Link Creation

Input: Nuclei set $\mathcal{N} = \{n_1, n_2, \ldots, n_N\}$, with corresponding coordinates $\{(x_1^n, y_1^n), (x_2^n, y_2^n), \ldots, (x_N^n, y_N^n)\}$. Nucleus of interest $n_i = (x_i^n, y_i^n)$. Stroma mask $\mathbf{S}$. Parameters r_n, θ_n.

Output: The set of nuclei that link to n_i, denoted by Γ_i^n

1: $\Gamma_i^n \leftarrow \emptyset$
2: $\Theta \leftarrow [0, \theta_n, 2\theta_n, \ldots, 2\pi]$ (angles corresponding to the conical search regions)
3: Generate $|\Theta| - 1$ conical search regions, Ω_j:
4: $\forall j \in [1, |\Theta| - 1], \Omega_j = \{(x, y) | \|(x - x_i^n, y - y_i^n)\|_2 < r_n$ and $\angle(x - x_i^n, y - y_i^n) > \Theta_j$ and $\angle(x - x_i^n, y - y_i^n) < \Theta_{j+1}\}$
5: **for each** Ω_j **do**
6: Find the closest nucleus to n_i in Ω_j, n^*
7: $n^* = (x^{n*}, y^{n*}) \in \Omega_j \cap \mathcal{N}$ such that $\|(x^{n*} - x_i^n, y^{n*} - y_i^n)\|_2 \leq \|(x^n - x_i^n, y^n - y_i^n)\|_2$, $\forall n = (x^n, y^n) \in \Omega_j \cap \mathcal{N}$
8: Let $l(n^*, n_i)$ denote the line connecting n^* and n_i
9: **if** $l(n^*, n_i) \cap \mathbf{S} = \emptyset$ **then**
10: (there is no stroma in between the two nuclei)
11: $\Gamma_i^n \leftarrow \Gamma_i^n \cup \{n^*\}$
12: **end if**
13: **end for**

2. Nucleus-Lumen-Link Creation

Algorithm 2 Nucleus-Lumen-Link Creation

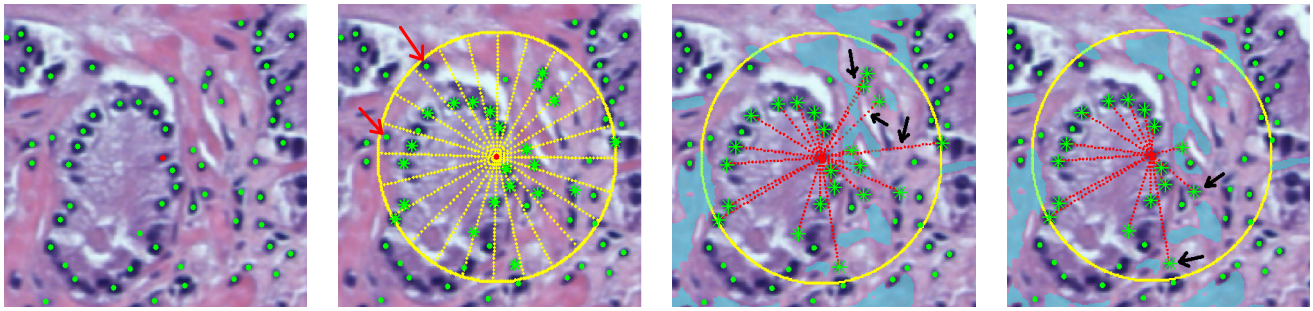
Input: Nuclei set $\mathcal{N} = \{n_1, n_2, \ldots, n_N\}$. Stroma mask $\mathbf{S}$. Lumen of interest l_i. Parameters r_l, θ_l.

Output: The set of nuclei that link to l_i, denoted by Γ_i^l

1: $\Gamma_i^l \leftarrow \emptyset$
2: Select the lumen-points $\{p_j^l\}$ by sampling at interval r_l from the boundary of l_i
3: **for each** p_j^l **do**
4: Apply Algorithm 1 on p_j^l, with parameters r_l and θ_l (which means $(x_i^n, y_i^n) \leftarrow p_j^l$ and $(r_n, \theta_n) \leftarrow (r_l, \theta_l)$) to find the set of nuclei that link to p_j^l, denoted by Γ_p
5: $\Gamma_i^l \leftarrow \Gamma_i^l \cup \Gamma_p$
6: **end for**

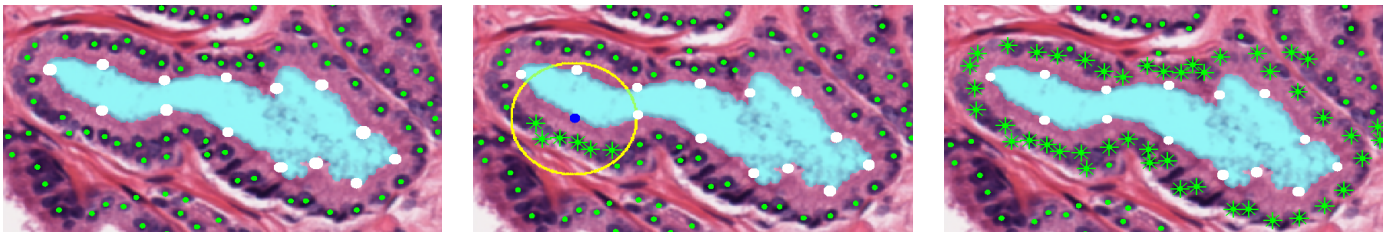
As has been mentioned, the goal of the nucleus-nucleus-links is to deal with small-sized and average-sized glands without lumen. For glands with large size, lumen is commonly present. Hence, we utilize lumen to enhance the connection (the density of links) between nuclei within the gland.

Given a lumen of interest l_i, we develop an algorithm, termed as **NLLink** (Algorithm 2), to find the nuclei that link to l_i (Fig. 10). The radius r_l of the neighborhood region that we apply on the lumen-point p_j^l (the yellow circle in Fig. 10b) is chosen as 50 pixels (at 20×), which is the estimated



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

Fig. 9. The nucleus-nucleus-link creation procedure. (a) A tissue image where the nucleus of interest n_i is indicated in red and the remaining nuclei are indicated in green. (b) The conical search regions, shown in yellow. The closest nucleus to n_i in each conical search region is shown as a green star in the region. The red arrows denote nuclei falling in the search regions but not belonging to the same gland as n_i. (c) The detected stroma regions, shown in cyan and the lines connecting n_i to the closest nuclei, shown in red. The lines intersecting stroma are indicated by black arrows. (d) The final nuclei connected to n_i. The black arrows indicate some of the bad links.



(a) (b) (c)

Fig. 10. The nucleus-lumen-link creation procedure. (a) A tissue image where the cyan region denotes the lumen of interest, the white dots denote the lumen-points sampled on the lumen boundary, and the green dots denote the detected nuclei. (b) A selected lumen-point (blue dot) and nuclei that link to this lumen-point (green stars). (c) All the detected nuclei (green stars) that link to the lumen.

maximum distance between lumen and nuclei on the gland boundary. The interval to sample the lumen-points on the lumen boundary is also chosen as the same value of r_l so that we can efficiently cover the region surrounding the lumen when finding nuclei.

3. Constructing the Edge Set $\mathcal{E}$ of the Nuclei-Lumina Graph

By applying the two algorithms NNLink and NLLink to all the nuclei and lumina in the image, we are able to create all the links. Note that, if two nuclei n_i and n_j have links to a lumen l_i, we also create a link for n_i and n_j; this strengthens the connection between the distant nuclei of the same gland. Each link in the image corresponds to an edge in $\mathcal{E}$.

C. Normalized Cut for Gland Segmentation

Recall that the nucleus-nucleus-links and nucleus-lumen-links created are based on the local information at the nuclei and lumina, without considering the global structure of the glands in the image. As a result, besides the good links, we may get some bad links that connect nuclei of different glands together. Therefore, the nuclei-lumina graph created for an image is likely to contain different connected components, each of which may correspond to a single gland or a group of multiple connected glands. See Fig. 11 for an example. To segment individual glands, we need to find a way to partition each connected component into glands. The normalized cut method [24] is a suitable solution for this task. Normalized cut is a global method, which takes into account all the links in the image and finds the weakest set of links for the partitioning. Intuitively, the weakest set of links mostly contains the bad links since the bad links are less dense than the good ones. Formally, the normalized cut method aims to partition the graph $G = (V, E)$ into two components A and B such that

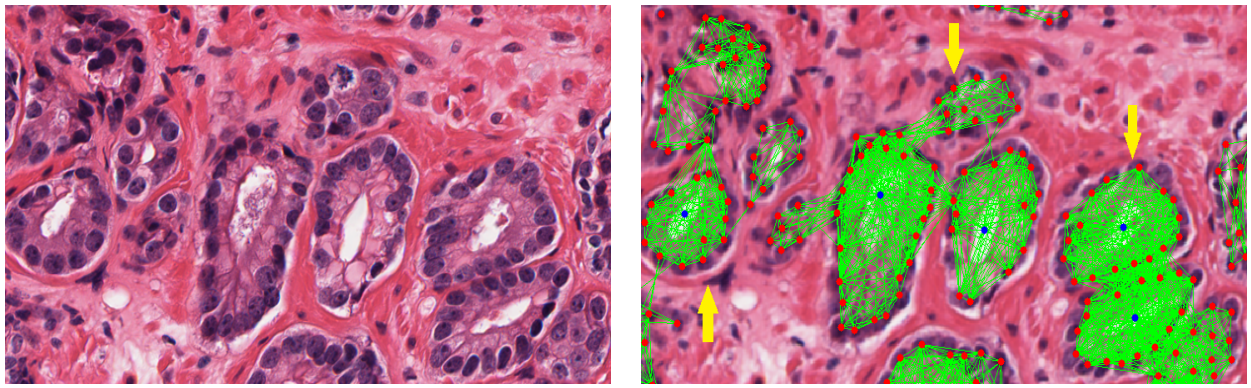
$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(B, A)}{assoc(B, V)} \tag{1}$$

is minimized. In this equation,

$$cut(A, B) = \sum_{u \in A, v \in B} w(u, v); \quad assoc(A,V) = \sum_{u \in A, v \in V} w(u, v). \tag{2}$$

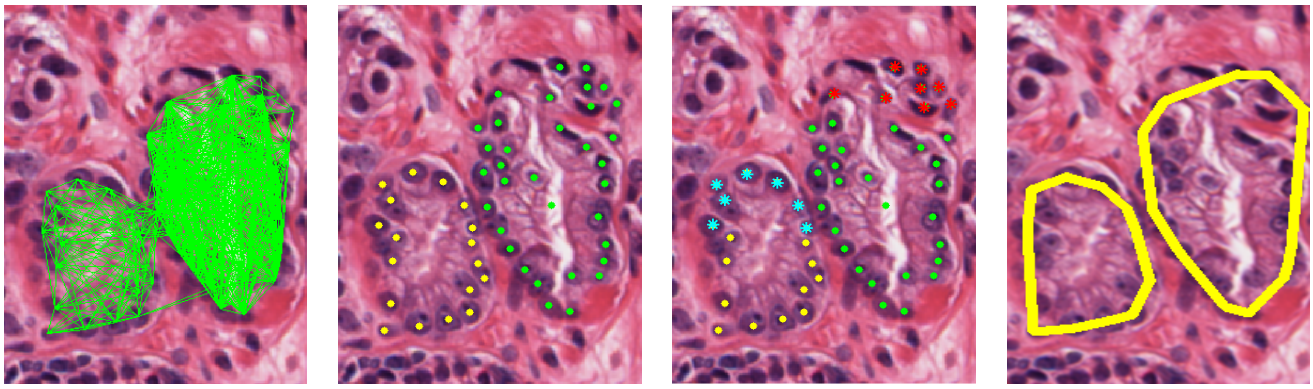
We assign all edges in the nuclei-lumina graph the same weight. More precisely, $\forall (v_i, v_j) \in \mathcal{V}$, $w_{ij} = 1$ if $v_i v_j \in \mathcal{E}$, otherwise $w_{ij} = 0$. Moreover, we perform the normalized cut in a recursive manner, i.e., we partition each connected component in the graph into two sub-components, and recursively partition the sub-components. One possible method to stop the process is to examine the *Ncut* value (Equation 1) and stop the process if this value is higher than a predefined threshold δ_c. The final components obtained are considered the segmentation results.

An example of the recursive cut process is shown in Fig. 12. Given the graph in Fig. 12a, denoted as C_0, we apply normalized cut on C_0 and obtain two components: C_1 (yellow dots) and C_2 (green dots) in Fig. 12b. The *Ncut* value for



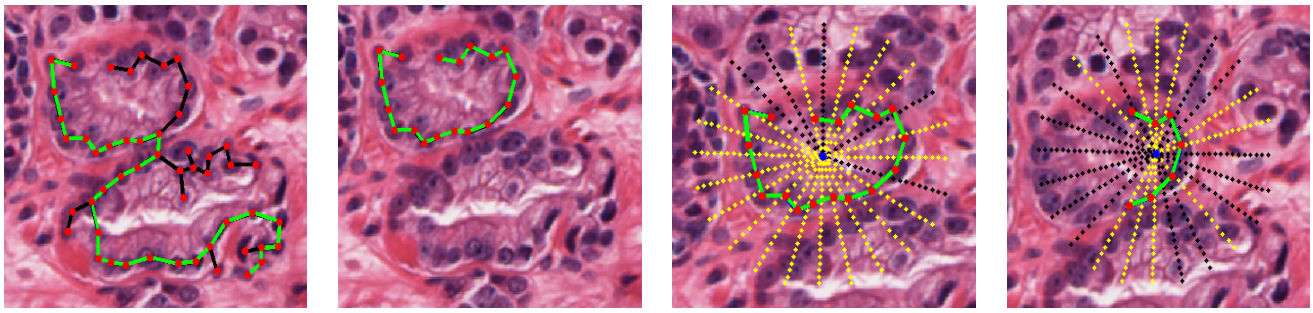
(a) (b)

Fig. 11. The nuclei-lumina graph constructed for an image. (a) Input tissue image. (b) The nuclei-lumina graph in which nuclei are denoted by red dots, lumina are denoted by blue dots, and the edges (links) are denoted by green lines. Three different connected components in the graph, on which the normalized cut is applied are indicated by yellow arrows.



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

Fig. 12. The recursive normalized cut process. (a) A connected component in the nuclei-lumina graph (C_0), whose edges are shown as green lines. (b) The two components, C_1 (yellow dots) and C_2 (green dots) obtained by applying normalized cut on C_0 (with *Ncut* = 0.15). (c) Results of the normalized cut applied on C_1 and C_2: C_1 is partitioned into C_{11} (yellow dots) and C_{12} (cyan stars) with *Ncut* = 0.68, while C_2 is partitioned into C_{21} (green dots) and C_{22} (red stars) with *Ncut* = 0.72. (d) The final result showing the two gland segments (C_1 and C_2), when the threshold $\delta_c = 0.5$ is used.



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

Fig. 13. Computing closed chain structure measures for a set of points (nuclei). (a) and (b) The MSTs and the MST backbones computed for two sets of points, one not forming the chain structure (a) and one forming the chain structure (b). The MST is shown as the green and black lines connecting the points (red dots). The green path is the MST backbone, while the black edges denote the branches in the MST. (c) and (d) Computing the closeness index. A path (green line) with a large value of closeness index (c) and a path with a small value of closeness index (d). The center of the point set is shown as the blue dot. The dotted lines separate the angular bins, where yellow lines denote the presence of points in the bin on its right (clockwise order), while the black lines denote the absence of points in the bin on its right (clockwise order).

this cut is 0.15. If we further partition C_1 and C_2, we will obtain the components C_{11}, C_{12}, C_{21}, and C_{22}, denoted by yellow dots, cyan stars, green dots and red stars in Fig. 12c, respectively. The *Ncut* value for C_1 and C_2 are 0.68 and 0.72, respectively. For this example, if we choose the cut threshold $\delta_c = 0.5$, no further splitting of C_1 and C_2 is done and the two components C_1 and C_2 are chosen as the final segmentation results. In Fig. 12d, we show the convex hull of these two components. This appears to be the best solution for this example.

In general, we do not know the suitable value of δ_c to stop the recursive cut to obtain a good segmentation result for an arbitrary graph. Hence, we aim to use a large value of δ_c (e.g., $\delta_c = 0.9$) to obtain multiple levels of recursive partitioning so that the good components (components corresponding to the complete glands) are likely to be created during the process. To determine good components, we rely on the nuclei arrangement information, i.e., we compare the nuclei arrangement between these components and the gland. Intuitively, the nuclei arrangement in a gland is similar to a *closed chain structure*, or in some cases, also similar to an *ellipse*. Fig. 12b illustrates this intuition. Hence, we aim to find different measures to estimate the similarity between the components[9] to these structures. Next, we combine these measures into a single number, called the *gland-score*, to estimate the likelihood that the component being a gland.

D. Normalized Cut Using Gland-Scores

As have been discussed, we need to compute a gland-score for the graph component by first computing the closed chain structure and ellipse measures. For generalization purposes, given a set of points $\{p_i\}$, we are interested in developing different measures to estimate how similar the arrangement of $\{p_i\}$ is to (i) a closed chain structure and (ii) an ellipse.

1. Closed Chain Structure Measures

To compute the closed chain structure measures for a point set $\{p_i = (x_i, y_i)\}$, we:

i. Construct a graph $G = \{V, E\}$, where $V = \{p_i\}$ and $E = \{e = v_i v_j\}\ \forall v_i, v_j \in V$.
ii. Compute the weight (or length) of the edge as $w_{ij} = \|(x_i - x_j, y_i - y_j)\|_2$.
iii. Compute the minimum spanning tree (MST) of G [26].
iv. By denoting the path between two vertices v_i, v_j in the MST as $P_{ij} = (v_i - v_j)$, the path length (total edge length) as $|P_{ij}|_l$, and the number of vertices in the path as $|P_{ij}|_v$, we find the path with most vertices, P^*, i.e., $|P^*|_v \geq |P|_v$, $\forall P \in$ MST[10], and refer to this path as the MST *backbone*.

In Figs. 13a and 13b, we show the MSTs computed for the components C_0 and C_1 mentioned earlier. The MST backbones and the branches (the edges not belonging to the backbone) are shown as green lines and black lines, respectively.

We compute the following measures to estimate how similar the MST is to a closed chain structure (refer to Figs. 13a and for 13b).

i. Mean degree (M^1_{MD}): the average degree of non-leaf vertices (vertices with degree greater than 1) in the tree.
ii. The ratio of the number of vertices in P^* to the total number of vertices, $M^1_V = \frac{|P^*|_v}{|V|}$.
iii. The ratio of the length of P^* to the total length of all the edges in the MST, $M^1_L = \frac{|P^*|_l}{\sum_{e \in \text{MST}} |e|}$.
iv. The closeness index, M^1_{CI}, which determines the closeness of P^* (see Figs. 13c, 13d). To compute this measure, we first compute the center of P^*, $C_0 = (x_0, y_0)$, with $x_0 = mean(x_j)$ and $y_0 = mean(y_j)$, where $v_j = (x_j, y_j) \in P^*$. Next, we partition the region surrounding C_0 into b angular bins, $\{\Omega_i\}_{i=1}^b$, each with an angle of $\pi/12$ (which means $b = 24$). We compute M^1_{CI} as the ratio of the number of bins that contain vertices in P^* (the bins with yellow lines on the left border (clockwise order)) to the total number of bins, i.e.,

$$M^1_{CI} = \frac{|\{\Omega_i | \exists v \in P^* \cap \Omega_i\}|}{b}$$

For a nuclei-group similar to a closed chain structure, we expect that M^1_{MD} is close to 2, while M^1_V, M^1_L, M^1_{CI} are close to 1.

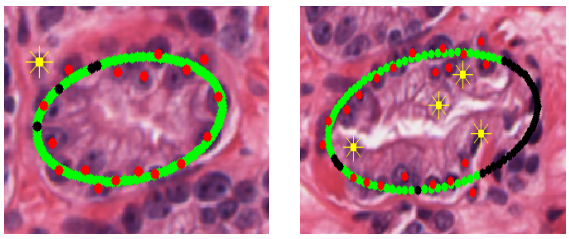
2. Ellipse Measures

To estimate how similar the arrangement of a point set $\{p_i\}$ is to an ellipse, we first fit an ellipse to $\{p_i\}$. The conic equation of an ellipse is

$$E = ax^2 + bxy + cy^2 + dx + ey + f = 0. \quad (3)$$

We use the least square method to fit this ellipse model [27], i.e., estimating the parameters a, b, c, d, e using $\{p_i\}$. To make the fitting more robust to noise, the random sample consensus (RANSAC) algorithm [28] is employed.

According to the RANSAC algorithm, when applying the ellipse fitting procedure to $\{p_i\}$, we may or may not find an ellipse. If an ellipse $\mathcal{M}^*$ is found, we compute the following measures[11] (see Fig. 14 for an illustration).



(a) (b)

Fig. 14. The ellipse fitting results on different point sets. The fitted ellipses are shown in green, while the fractions not covered by the points are shown in black. Inlier points are shown as red dots while outlier points are shown as yellow stars. The point set in (a) are arranged more similar to an ellipse than the one in (b).

[9]We also denote the component as the nuclei-group since we only use nuclei, but not lumen in the component for subsequent computations.

[10]There may be several such paths, but we select only one of them randomly.

[11]If an ellipse is not found, we assign zero values to all these measures.

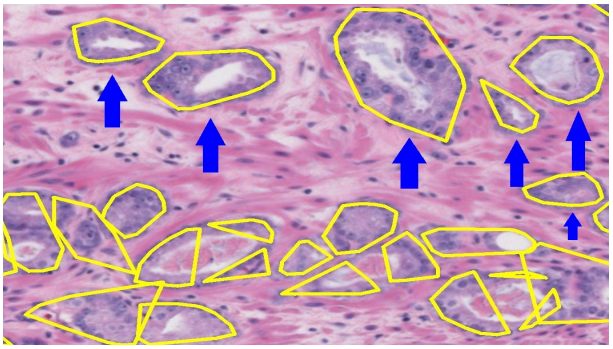


Fig. 15. Examples of the training data used for learning the gland-score function ψ. The segments indicated by blue arrows are used as gland segments, while the remaining segments are used as non-gland segments. The cut threshold $\delta_c = 0.5$ is used for the recursive cut segmentation.

i. The average fitting error M_E^2. To compute this measure, we sample m points on the ellipse $\mathcal{M}^* = \{q_1, q_2, \ldots, q_m\}$. The fitting error for a point p_i is computed as $\epsilon_i = \min_{q_j} \|p_i - q_j\|_2$, $\forall q_j \in \mathcal{M}^*$. We average the errors for all the points in $\{p_i\}$ to generate M_E^2.

ii. The percentage of inliers M_I^2, i.e., ratio of the number of inliers to the total number of points. A point $p_i \in \{p_i\}$ is considered an inlier if $\exists q^* \in \mathcal{M}^*$ such that $\|p_i - q^*\| < \delta_e$. We use $\delta_e = 13$ pixels (which is based on the estimated deviation of nuclei from the nuclei chain in the gland boundary).

iii. The coverage index M_{CI}^2, i.e., the fraction of the ellipse that is covered by $\{p_i\}$. A point $q_i \in \mathcal{M}^*$ is considered covered by $\{p_i\}$ if $\exists p^* \in \{p_i\}$ so that $\|p^* - q_i\|_2 < \delta_e$. We compute the coverage index as the ratio of the number of points in $\mathcal{M}^*$ that are covered to the total number of points in $\mathcal{M}^*$. In Fig. 14, the fraction of the ellipse that is not covered by $\{p_i\}$ is shown in black.

For a nuclei-group that is similar to an ellipse, we expect that the M_E^2 is small, M_I^2 is high and M_{CI}^2 is high.

3. Computing the Gland-Score for a Nuclei-Group

We aim at combining all the closed chain structure and ellipse measures computed for a nuclei-group[12] (a component generated during the normalized cut process) to generate a single *gland-score* to estimate how similar the arrangement of the nuclei-group is to the arrangement of nuclei in a gland. For convenience, we define the *gland-measure* vector of a nuclei-group C, denoted by $\mathbf{g}(C)$, as the seven-dimensional vector constructed by concatenating all the closed chain structure and ellipse measures, i.e.,

$$\mathbf{g}(C) = \{M_{MD}^1, M_V^1, M_L^1, M_{CI}^1, M_E^2, M_I^2, M_{CI}^2\}. \quad (4)$$

To compute the gland-score s of a nuclei-group C with gland-measure $\mathbf{g}(C)$, we need a function $\psi(\mathbf{g}(C)) = s$. We build this function using a learning framework as follows: (i) obtain a training set of *gland nuclei-groups* (nuclei-groups corresponding to the complete glands) and *non-gland nuclei-groups* (nuclei-groups corresponding to parts of the gland, or nuclei-groups with random nuclei arrangement), (ii) compute the gland-measure vectors from them, and (iii) learn an SVM classifier (RBF kernel) to separate the two types of groups. This SVM classifier serves as the function ψ. To compute gland-score of a test sample, we compute the distance from the sample to the decision boundary, and convert this distance to a probability output (with value in [0, 1]), i.e., the closer the output to 1, the higher the confidence that the sample is a gland nuclei-group[13].

To obtain the training set in step (i), we first perform segmentation on the training images by using a fixed value of δ_c (we use $\delta_c = 0.5$) as the stopping criterion. The final segments (nuclei-groups) obtained are manually labeled as gland and non-gland segments (we use 130 non-gland segments and 100 gland segments with a large variation in their shape and size, obtained from 30 training images). Examples of the training data are shown in Fig. 15.

4. Using Gland-Score for Segmentation

Algorithm 3 Normalized cut gland segmentation using gland-score

Input: Nuclei-lumina graph $\mathcal{G}$. Gland-score function ψ. Cut threshold δ_c.

Output: The set of components (the sub-graphs generated during normalized cut) with the highest gland-scores, Λ^*

$\Lambda_1 \leftarrow \mathcal{G}$ (components to be partitioned)
$\Lambda_2 \leftarrow \emptyset$ (all components being generated)
while $\Lambda_1 \neq \emptyset$ **do**
  Select an arbitrary element $C \in \Lambda_1$
  $\Lambda_2 \leftarrow \Lambda_2 \cup \{C\}$ (save C in Λ_2)
  $\Lambda_1 \leftarrow \Lambda_1 \backslash \{C\}$ (remove C from Λ_1 before partitioning it)
  Perform normalized cut on C to obtain C_1, C_2, and *Ncut* value
  if *Ncut* $< \delta_c$ **then**
    $\Lambda_1 \leftarrow \Lambda_1 \cup \{C_1, C_2\}$ (include new components for later partitioning)
  end if
end while
$\Lambda^* \leftarrow \emptyset$
while $\Lambda_2 \neq \emptyset$ **do**
  Select $C^* = argmax_{C \in \Lambda_2} \psi(\mathbf{g}(C))$ (component with highest gland-score)
  $\Lambda_2 \leftarrow \Lambda_2 \backslash \{C^*\}$ (C^* will not be considered again)
  if $\forall C \in \Lambda^*$, $C^* \cap C = \emptyset$ **then**
    $\Lambda^* \leftarrow \Lambda^* \cup \{C^*\}$ (save C^* in the result list)
  end if
end while

We now present the use of gland-score function ψ in the recursive normalized cut process for gland segmentation (Algorithm 3). We first compute gland-scores for all of the components created in this process. Using these gland-scores, we determine the final segmentation results as the

[12]The point set $\{p_i\}$ is the nuclei-group in this case.

[13]We have also tried to use the kernel logistic regression method to estimate the probability output of the classification. However, the performance of this method is worse than that of the SVM classifier.

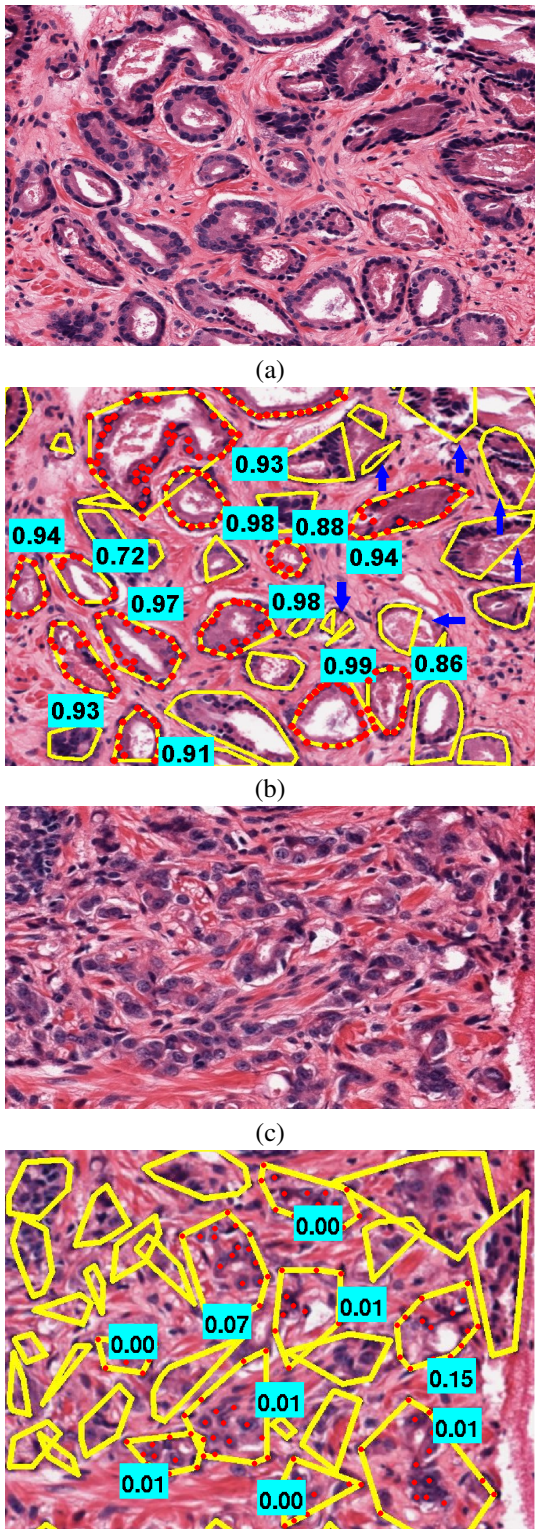


Fig. 16. The nuclei-based gland segmentation results for a grade 3 and a grade 4 image. (a) A grade 3 image. (c) A grade 4 image. (b) and (d) Segmentation results of the images in (a) and (c), respectively. Representative segments in each image are indicated by red dots (nuclei in the segments), with gland-scores shown in cyan boxes. The gland-scores for the segments in the grade 3 image are higher than those in the grade 4 image.

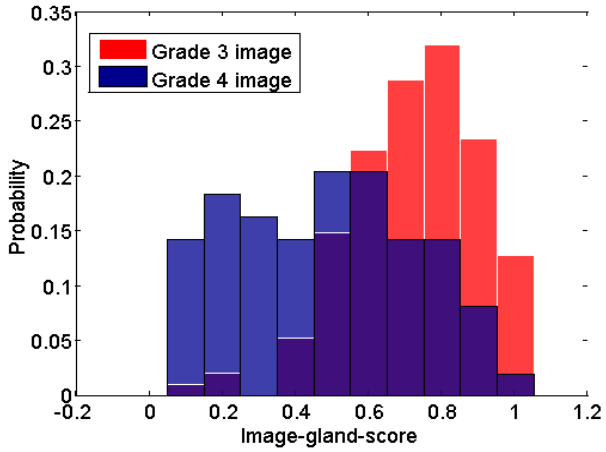


Fig. 17. The distribution of image-gland-score (ψ^I value) of grade 3 and grade 4 images in the database.

components that (i) have the highest gland-scores, (ii) are mutually exclusive, and (iii) their union is equal to the original component. Recall that in this algorithm, we use $\delta_c = 0.9$, which is a sufficiently high value to ensure that we will find the complete glands in the resulting components.

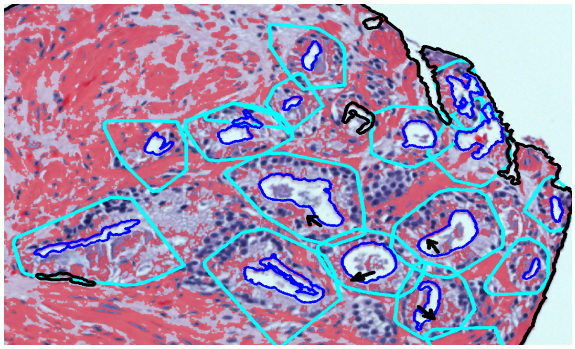
To explain Algorithm 3, we reuse the example in Fig. 12[14] with $\delta_c = 0.9$. Since the *Ncut* values for the partitioning of C_{11}, C_{12}, C_{21}, C_{22} are all greater than 0.9, these are the final components being obtained. We denote $\Lambda = \{C_0, C_1, C_2, C_{11}, C_{12}, C_{21}, C_{22}\}$ as the set of all resulting components (note that these components are not mutually exclusive). Next, we compute the gland-scores for these components and sort the components based on the gland-scores (from high to low), which results in the ordered sequence $C_1, C_2, C_{21}, C_{11}, C_{12}, C_0, C_{22}$, with corresponding gland-scores 0.96, 0.87, 0.77, 0.55, 0.17, 0.04, and 0.01, respectively. We iteratively choose the component with the highest score in Λ to include it into the final result Λ^* (which is initialized as $\emptyset$), such that it does not overlap the current components in Λ^*. We stop the iteration when all nuclei in the original component C_0 are included in Λ^*. As a result, the components chosen in this example are first C_1 followed by C_2.

An example of segmentation output by the nuclei-based method was shown previously in Fig. 7.

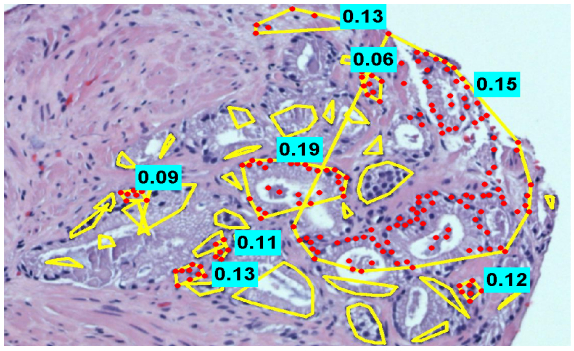
E. Using Gland-Score for Grade 3 vs Grade 4 Tissue Classification

In grade 3 images, nuclei in the segments are more likely to form a closed chain structure than those in grade 4 images (nuclei in the segments of grade 4 images are more randomly distributed because glands tend to fuse together). Hence, we are interested in analyzing the gland-scores of the segments obtained for grade 3 and grade 4 images. Fig. 16 shows a comparison between the gland segmentation results for a grade

[14]When denoting the gland-score value s of a component C, instead of writing $\psi(\mathbf{g}(C)) = s$, we can simply write $\psi = s$.



(a)



(b)

Fig. 18. Limitation of the nuclei-based method. (a) A grade 3 image with detected stroma regions, shown in red. Stroma is incorrectly detected within the gland regions (indicated by black arrows). The lumen-based method does not rely on stroma, thus, can still obtain good segmentation results (cyan contours). (b) Nuclei-based segmentation result, which is not good due to its dependence on stroma. The segments with ψ values (gland-score) in the top 20% are indicated by red dots. The ψ values of these segments are shown in cyan boxes, yielding the image-gland-score $\psi^I = 0.12$.

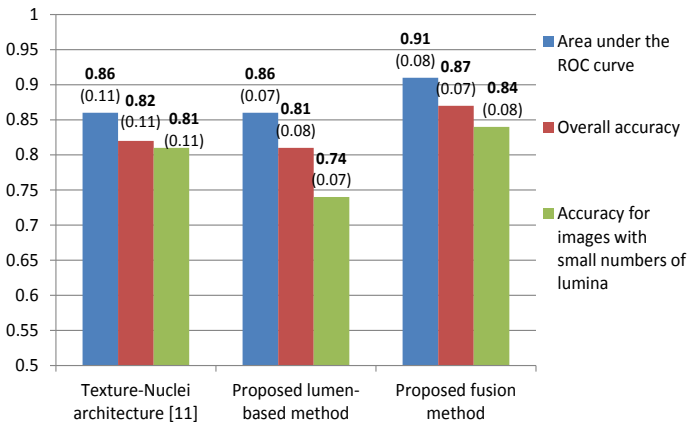


Fig. 19. Summary of grade 3 vs grade 4 classification results (cross-validation accuracies and standard deviations) for the best grading methods (from Table II).

3 and a grade 4 image. It is reasonable to expect that the gland-scores for the segments in the grade 3 image are higher than those in the grade 4 image.

F. Image-gland-score

We aim to average the gland-scores of the individual segments in the image to derive a single score to represent the image, referred to as the *image-gland-score* and denoted as ψ^I. Due to inevitable noise in the segmentation results (e.g., the blue arrows in Fig. 16b indicate some noisy segments in the image), we only use 20% of the segments with the highest ψ values to compute ψ^I, i.e., given m segments sorted by gland-scores, $\psi_1 > \psi_2 > \cdots > \psi_m$,

$$\psi^I = \frac{\sum_{i=1}^{k} \psi_i}{k}, \text{where } k = \lceil 0.2m \rceil. \quad (5)$$

In Fig. 17, we plot the histogram of the image-gland-score (ψ^I) values for all grade 3 and grade 4 images. Although there is a certain overlap between the two distributions of the ψ^I values of the grade 3 and grade 4 images, we can still see some separation between the two distributions. This suggests that the ψ^I values could be helpful for the grade 3 vs grade 4 tissue image classification.

Limitation of the nuclei-based method: The limitation of the nuclei-based method is its dependence on stroma detection result. In Fig. 18, due to the incorrectly detected stroma within the gland regions (black arrows in Fig. 18a), the connection between nuclei of the same glands is weak (sparse links). Hence, the segments obtained do not capture the complete gland area but only some of the nuclei in the gland, resulting in low ψ values (Fig. 18b). The lumen-based method, in contrast, does not depend on stroma, thus, it still provides good segmentation result for this image (cyan contours in Fig. 18a). Hence, we combine the nuclei-based and lumen-based methods to enhance the grade 3 vs grade 4 classification result.

V. A Fusion Method for Grade 3 vs Grade 4 Tissue Image Classification

To improve the grade 3 vs grade 4 tissue classification, we combine:

i. the lumen-based method which involves segmenting glands using the NLA method and computing the 19 structural features for the image (Sec. III), denoted by F_{19}, and
ii. the nuclei-based method which involves segmenting glands using the nuclei-based gland segmentation method and computing the image-gland-score ψ^I.

A 20-dimensional feature vector $F_{20} = \{F_{19}, \psi^I\}$ is now used for the grade 3 vs grade 4 tissue image classification problem.

Evaluation of the Fusion Method

We use the same database of all grade 3 and grade 4 images and the same evaluation strategy as discussed in Sec. III-A to evaluate this fusion method. We compare it with the lumen-based and the texture-nuclei architecture [11] methods (the best methods in Table II) in Fig. 19. Apart from the classification accuracies, we also compute the "area under the receptive operation characteristics (ROC) curves" to better evaluate the methods.

To further demonstrate the advantage of the fusion method over the lumen-based method, we re-evaluate these methods on images with small numbers of lumina. We select 81 grade 3 images and 40 grade 4 images which contain less than 38

lumina (this is the average number of lumina contained in each image in the original grade 3 and grade 4 image dataset). The classification accuracies for this case are also included in Fig. 19. The large difference in accuracies between the lumen-based and fusion methods in this case demonstrates the effectiveness of the fusion method on images with only a few lumina detected.

It is important to note that our final goal is to classify a given tissue image, not necessarily to obtain perfect gland segmentation results. In fact, the nuclei-based method does not require a perfect segmentation result for every gland since it only uses the glands with the top ψ^I values for feature computation. As a result, we only quantitatively evaluated the classification performance but not the segmentation performance of the proposed method.

In Fig. 20, we show a grade 3 image that is misclassified as grade 4 when only the lumen-based method is used. The reason is that glands are not well-detected by the lumen-based method since only a few lumina are detected (Fig. 20b). Moreover, the lumina in this image are very small, similar to what is observed in grade 4 images. However, when using the fusion method, we obtain the correct classification result because the nuclei-based method is able to detect the glands without lumen, and obtain high ψ^I value for this image (Fig. 20c).

Finally, the fusion method is tested for the full three-class classification problem, obtaining an average CV accuracy of 0.88 (with a standard deviation of 0.06), which is also an improvement over both the lumen-based method (with an accuracy of 0.84 (0.05)) and the texture-nuclei architecture method (with an accuracy of 0.84 (0.06)) (Table II).

VI. Discussion

1. Closed chain structure vs ellipse measures: To evaluate the contributions of these measures, we compute the gland-scores and then perform grade 3 vs grade 4 classification using only closed chain structure or ellipse measures alone. The classification accuracies obtained for close chain structure and ellipse measures are 85.1% and 86%, respectively. Since we do not see significance difference between their performance, a combination of these measures is a reasonable solution, which also leads to a better result.

2. Computation cost: The computational complexity of each individual step of the proposed nuclei-based method is presented in Table III. Both the proposed nuclei-based method and the methods in [10] and [11] need to detect nuclei and build a graph of nuclei (or graph of nuclei and lumina). However, the proposed nuclei-based method spends additional computation on texture feature extraction for nuclei classification (to discard irrelevant stroma nuclei), and computation on the gland segmentation by normalized cut (from which more meaningful features can be extracted). By empirical experiments in Matlab programming language, we observe that most of the computation time is spent on the nuclei classification (step 2 in Table III) due to the heavy computation of the GLCM features, followed by the gland segmentation by normalized cut (step 6 in Table III). Possible solutions to speed up the proposed method are: (i) explore simpler alternative texture features (to replace the GLCM features) for nuclei classification, (ii) use parallel computing for nuclei classification (the feature extraction for each nucleus is independent, thus, can be parallelized) and gland segmentation (the recursive cut on each connected component in $\mathcal{G}$ is independent and can be parallelized), and (iii) find better stopping criteria for early termination of the recursive cut.

3. Classification of Gleason scores 3+4 vs 4+3: Although some of our images contain both Gleason grade 3 and grade 4 glands, we do not have the ground truth for 3+4 and 4+3 scoring. Hence, we are unable to evaluate the performance of the proposed method on this problem. However, if we apply the nuclei-based method on this problem, we expect that the image-gland-scores of the images of score 3+4 are higher than those of images of score 4+3, since there are more grade 3 glands in the images of score 3+4 compared to those of score 4+3. Moreover, to the best of our knowledge, there is no previous work that performs the 3+4 vs 4+3 discrimination automatically.

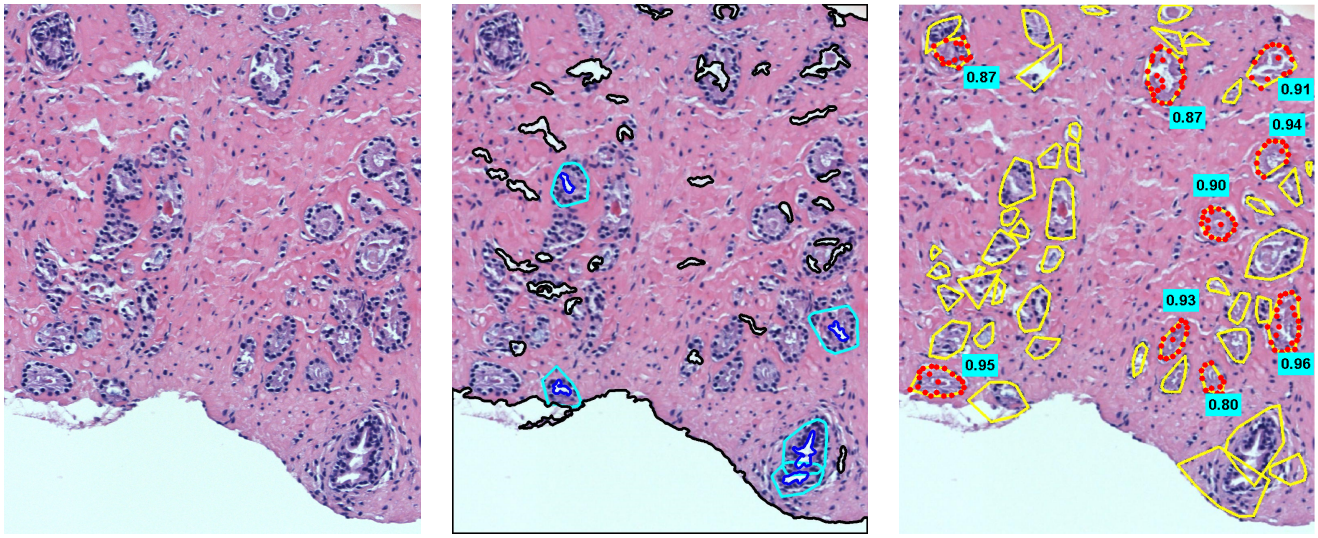
4. Cribriform Pattern: Since cribriform pattern is a common pattern in grade 4 images (according to a recent study [4], all cribriform regions should be assigned grade 4), we discuss the effect of the nuclei-based method when applying to the cribriform pattern. A cribriform pattern resembles a group of multiple small glands connected together without stroma lying between them (e.g., Figs. 21a, 21c, 21e). In case the cribriform contains clear lumina surrounded by well-defined nuclei-chain structure, the nuclei-based method may generate segments similar to glands and assign high gland-scores to these segments (the segments indicated by the cyan boxes Fig. 21b). Hence, the image can be misclassified as grade 3 due to the high ψ^I value. However, in case the cribriform does not contain clear lumina (Fig. 21c), or nuclei do not form solid chain structure around each lumen (Fig. 21e), the segmented regions commonly do not receive high gland-scores (red boxes in Figs. 21d, 21f). In such cases, the images are correctly classified as grade 4. Moreover, we further observe that among 41 grade 4 images containing cribriform pattern in the database, only nine are mis-classified using the proposed fusion method. These observations suggest that the nuclei-based method is useful in classifying several images with cribriform pattern, although additional techniques (or features) maybe necessary to improve the identification of this pattern. One potential strategy to address this pattern in the future work is to consider the "fusion zone", i.e., the area in between the neighboring lumina. See Fig. 22. Since the cribriform pattern resembles a group of glands fused together, the fusion zone created inside this pattern does not contain stroma, and contains randomly distributed nuclei. This is not the case for the fusion zone in a grade 3 image, where glands are separated. We anticipate that useful features can be computed from this zone to identify cribriform pattern.

5. Summary of the proposed methods: As discussed earlier, the limitation of the proposed lumen-based method is that it is not able to detect glands without lumen, while the proposed nuclei-based method is able to do so. On the other hand, the nuclei-based method has its own limitation that it relies on the

TABLE III

COMPUTATIONAL COMPLEXITY OF DIFFERENT STEPS IN THE NUCLEI-BASED METHOD. REFER TO TABLE I FOR THE DESCRIPTION OF THE NOTATIONS USED HERE. A MORE DETAILED DISCUSSION (E.G., USE PARALLEL COMPUTING TO IMPROVE COMPUTATION TIME) IS INCLUDED IN SEC. VI.2.

Step	Computational complexity
1. Nuclei detection by radial-symmetry-based method	$O(m)$ (compute gradient at all m pixels in the image) + $O(m_v)$ (perform the radial symmetry voting at m_v voting pixels)
2. Nuclei classification	$O(\lvert\mathcal{N}\rvert)$ (extract texture features, i.e., GLCM and color histogram, from a fixed-size neighborhood area of each nucleus)
3. Stroma and lumen segmentation using k-means clustering	$O(m \times i)$, where m and i denote the number of pixels in the image and number of iterations, respectively
4. Nucleus-nucleus-link creation	$O(\lvert\mathcal{N}\rvert)$ (search in a fixed-size neighborhood area of each nucleus)
5. Nucleus-lumen-link creation	$O(\lvert\mathcal{L}\rvert \times \lvert\{p_j^l\}\rvert)$ (search in a fixed-size neighborhood area of each lumen-point p_j^l)
6. Segmentation by normalized cut	$O(\lvert CC\rvert \times \lvert\Lambda\rvert \times c_N \times (x_1 + x_2))$, where $\lvert CC\rvert$ denotes the number of connected components (CC) in $\mathcal{G}$ (Sec. IV-C and Fig. 11), $\lvert\Lambda\rvert$ denotes the number of sub-components generated during the recursive cut process applied on each CC, c_N denotes the cost of a normalized cut operation (described in [24]), and x_1, x_2 are mentioned in the following.
6.1. Compute close-chain structure measures of a graph component $G = (V, E)(x_1)$	$O(\lvert E\rvert log\lvert E\rvert)$ (compute the MST by Kruskal algorithm) + $O(\lvert V\rvert \times \lvert E\rvert)$ (find the MST backbone by applying the depth-first search algorithm from each node)
6.2. Compute ellipse measures of a graph component $G = (V, E)(x_2)$	$O(r \times \lvert V\rvert^2)$ (fit the ellipse to the vertices), where r denotes the number of iterations in RANSAC algorithm
7. Compute image-gland-score	$O(\lvert\Lambda^*\rvert \times (x_1 + x_2))$, where $\lvert\Lambda^*\rvert$ denotes the number of segmented glands in the image (the best components being selected)



(a) (b) (c)

Fig. 20. Contribution of the nuclei-based method in grade 3 vs grade 4 image classification. (a) A grade 3 image. (b) Gland segmentation result of the lumen-based method (cyan contours), where detected lumina are shown as blue contours and detected artifacts are shown as black contours. (c) Gland segmentation result of the nuclei-based method, in which the segments with ψ values in the top 20% are indicated by red dots (nuclei in the segments), yielding $\psi^I = 0.9$.

segmentation result of stroma, while the lumen-based method does not. As a result, we combine these two methods into a fusion method to solve the tissue image grading problem, which leads to an improvement of the classification accuracies.

VII. CONCLUSION

In this paper, we first presented a lumen-based method to address the prostate tissue image classification problem, and obtained comparable accuracies with state-of-the-art methods. Next we focus on developing a nuclei-based method as an attempt to improve the grade 3 vs grade 4 classification result. The nuclei-based method segments glands by utilizing both the local information at the nuclei and lumina level (to construct the nuclei-lumina graph), and the global information in the image (to find the weakest set of links in the graph using the normalized cut method). This method is able to segment glands without lumen and glands with multiple lumina, which are the limitations of the current gland segmentation methods. We further develop a novel method to exploit the difference in nuclei spatial arrangement between grade 3 and grade 4 images. This is done by computing a set of closed chain structure and ellipse measures, leading to a gland-score to estimate how similar the segment is to a gland. Finally, we combine the lumen-based and nuclei-based methods to improve the grade 3 vs grade 4 classification accuracy, compared to the state-of-the-art method. For future work, we plan to further improve the classification accuracy by focusing on cribriform pattern. Moreover, we also plan to reduce the computation time of the method.

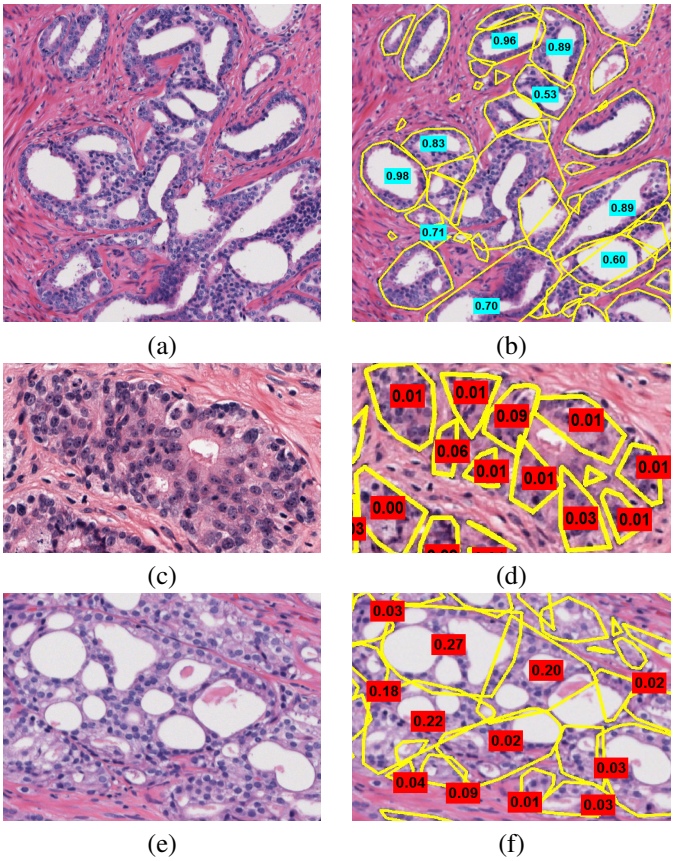


Fig. 21. Effect of the nuclei-based method on grade 4 images that contain cribriform pattern. (a), (b) The case where gland-scores are high and (c)-(f) the case where gland-scores are low.

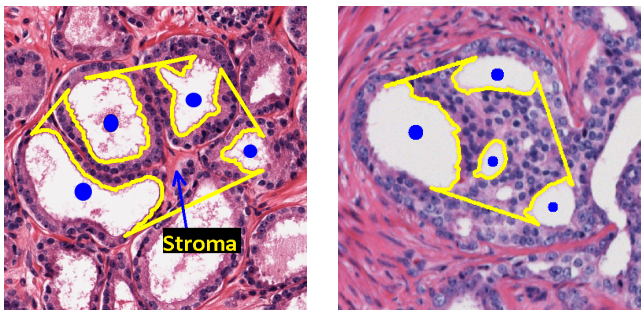


Fig. 22. Extracting information from fusion zones, a potential method to identify cribriform in grade 4 images. Fusion zones (regions inside yellow contours), created as regions in between the neighboring lumina (blue dots), in a grade 3 image (a) and in a cribriform pattern in a grade 4 image (b).

REFERENCES

[1] R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2013," *CA: A Cancer Journal for Clinicians*, vol. 63, pp. 11–30, 2013.

[2] D. Gleason, "Histologic grading of prostate cancer: A perspective," *Human Pathology*, vol. 23, pp. 273–279, 1992.

[3] R. Singh, et al., "Interobserver reproducibility of Gleason grading of prostatic adenocarcinoma among general pathologists," *Indian Journal of Cancer*, vol. 48, pp. 448–495, 2011.

[4] J. I. Epstein, "An update of the Gleason grading system," *J. Urology*, vol. 183, no. 2, pp. 433–440, 2010.

[5] J. Diamond, et al., "The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia," *Human Pathology*, vol. 35, pp. 1121–1131, 2004.

[6] K. J. Khouzani and H. S. Zadeh, "Multiwavelet grading of pathological images of prostate," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 6, pp. 697–704, June 2003.

[7] H. J. Yoon, C. C. Li, C. Christudass, R. Veltri, J. I. Epstein, and Z. Zhang, "Cardinal multiridgelet-based prostate cancer histological image classification for Gleason grading," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2011, pp. 315–320.

[8] P. Khurd, et al., "Computer-aided Gleason grading of prostate cancer histopathological images using texton forests," in *International Symposium on Biomedical Imaging: From Nano to Macro*, 2010, pp. 636–639.

[9] S. K. Tai, et al., "Classification of prostatic biopsy," in *International Conference on Digital Content, Multimedia Technology and its Applications (IDC)*, August 2010, pp. 354–358.

[10] P. Khurd, et al., "Network cycle features: Application to computer-aided Gleason grading of prostate cancer histopathological images," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2011, pp. 1632–1636.

[11] S. Doyle, et al., "Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer." *BMC Bioinformatics*, vol. 13, p. 282, 2012.

[12] S. Naik, et al., "Automated gland segmentation and Gleason grading of prostate histology by integrating low-, high-level and domain specific information," *MIAAB workshop, Piscataway, NJ*, 2007.

[13] C. Li, et al., "Level set evolution without re-initialization: A new variational formulation," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 430–436.

[14] Y. Peng, et al., "Segmentation of prostatic glands in histology images," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2011, pp. 2091–2094.

[15] J. Vidal, et al., "A fully automated approach to prostate biopsy segmentation based on level-set and mean filtering," *Journal of Pathology Informatics*, vol. 2, no. 2, p. 5, 2011.

[16] J. Xu, et al., "High-throughput prostate cancer gland detection, segmentation, and classification from digitized needle core biopsies," in *Prostate Cancer Imaging. Computer-Aided Diagnosis, Prognosis, and Intervention*, 2010, vol. 6367, pp. 77–88.

[17] J. Monaco, et al., "High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models," *Medical Image Analysis*, vol. 14, pp. 617–629, 2010.

[18] K. Nguyen, A. Sarkar, and A. Jain, "Structure and context in prostatic gland segmentation and classification," in *In: N. Ayache et al. (eds.) MICCAI 2012, LNCS, vol. 7510, pp. 115-123, Springer, Heidelberg*, 2012, pp. 115–123.

[19] A. Tabesh, M. Teverovskiy, and H. Pang, "Multifeature prostate cancer diagnosis and Gleason grading of histological images," *IEEE Transactions on Medical Imaging*, vol. 26, no. 10, pp. 1366–1378, 2007.

[20] K. Nguyen, R. Allen, and A. Jain, "Automated gland segmentation and classification for gleason grading of prostate tissue images," in *Proc. International Conf. Pattern Recognition*, 2010, pp. 1497–1500.

[21] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern. SMC*, vol. 3, pp. 610–621, 1973.

[22] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 1150–1157.

[23] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, Washington, DC, USA, 2006, pp. 2169–2178.

[24] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[25] B. Parvin, Q. Yang, J. Han, H. Chang, B. Rydberg, and M. Barcellos-hoff, "Iterative voting for inference of structural saliency and characterization of subcellular events," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 615–623, 2007.

[26] C. Zahn, "Graph-theoretical methods for detecting and describing gestalt clusters," *IEEE Transactions on Computers*, vol. C-20, no. 1, pp. 68–86, 1971.

[27] "Ellipse fitting implementation," www.mathworks.com/matlabcentral/fileexchange/3215-fitellipse.

[28] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.