HHS Public Access

Author manuscript

Comput Methods Programs Biomed. Author manuscript; available in PMC 2021 June 02.

Published in final edited form as:

Comput Methods Programs Biomed. 2017 April ; 142: 91–99. doi:10.1016/j.cmpb.2017.02.023.

Multiview boosting digital pathology analysis of prostate cancer

Jin Tae Kwak[a,*], **Stephen M. Hewitt**[b]

[a]Department of Computer Science and Engineering, Sejong University, Seoul 05006, Korea

[b]Tissue Array Research Program, Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, MD 20852, USA

Abstract

Background and Objective: Various digital pathology tools have been developed to aid in analyzing tissues and improving cancer pathology. The multi-resolution nature of cancer pathology, however, has not been fully analyzed and utilized. Here, we develop an automated, cooperative, and multi-resolution method for improving prostate cancer diagnosis.

Methods: Digitized tissue specimen images are obtained from 5 tissue microarrays (TMAs). The TMAs include 70 benign and 135 cancer samples (TMA1), 74 benign and 89 cancer samples (TMA2), 70 benign and 115 cancer samples (TMA3), 79 benign and 82 cancer samples (TMA4), and 72 benign and 86 cancer samples (TMA5). The tissue specimen images are segmented using intensity- and texture-based features. Using the segmentation results, a number of morphological features from lumens and epithelial nuclei are computed to characterize tissues at different resolutions. Applying a multiview boosting algorithm, tissue characteristics, obtained from differing resolutions, are cooperatively combined to achieve accurate cancer detection.

Results: In segmenting prostate tissues, the multiview boosting method achieved ≥ 0.97 AUC using TMA1. For detecting cancers, the multiview boosting method achieved an AUC of 0.98 (95% CI: 0.97–0.99) as trained on TMA2 and tested on TMA3, TMA4, and TMA5. The proposed method was superior to single- view approaches, utilizing features from a single resolution or merging features from all the resolutions. Moreover, the performance of the proposed method was insensitive to the choice of the training dataset. Trained on TMA3, TMA4, and TMA5, the proposed method obtained an AUC of 0.97 (95% CI: 0.96–0.98), 0.98 (95% CI: 0.96–0.99), and 0.97 (95% CI: 0.96–0.98), respectively.

Conclusions: The multiview boosting method is capable of integrating information from multiple resolutions in an effective and efficient fashion and identifying cancers with high accuracy. The multiview boosting method holds a great potential for improving digital pathology tools and research.

Keywords

Prostate cancer; Digital pathology; Multiview boosting; Multi-resolution; Machine learning

[*]Corresponding author. jkwak@sejong.ac.kr, trline@gmail.com (J.T. Kwak).

1. Introduction

In 2016, it is expected that more than 180,000 men are diagnosed with prostate cancer and more than 26,000 men are died from the disease (second leading cause of deaths from cancer) [1]. For prostate cancer diagnosis, over one million biopsies are annually performed. Biopsied prostate specimens are sectioned and usually stained with hematoxylin and eosin (H&E) to enhance the contrast in specimens. The stained specimens are examined under an optical brightfield microscopy by pathologists. A majority of the specimens are negative for prostate cancer. Cancer specimens are assigned a histological grade (or disease status) according to the Gleason grading system [2]. Assessing the qualitative characteristics of tissues/cells, pathologists determine a Gleason grade between 1 (relatively benign) and 5 (highly aggressive). The Gleason grade offers the definitive diagnosis of prostate cancer and forms the basis for cancer management and treatment today. The manual and qualitative nature of prostate pathology, however, not only limits the speed and throughput but also impedes the accuracy and reliability due to the substantial inter- and intra-observer variations in grading [3–5]. Automated and quantitative tools for prostate pathology could, therefore, aid in improving prostate cancer diagnosis.

Digital pathology [6] is an emerging practice of computerized image processing, analysis, and interpretation of digitized tissue specimen images. Digitized tissue specimen images are often segmented into a number of histological objects/classes (e.g., epithelium, stroma, and nucleus) due to their biological, chemical, and functional differences [7]. The segmented histological objects are used to characterize and quantify tissue appearances and microstructures; for example, alteration of glandular shape and spatial distribution and arrangement of cells and nuclei. The quantitative measures are utilized, in a machine learning framework, to make a diagnostic decision. Due to the automated and quantitative nature, digital pathology holds a potential for advancing the current practice of cancer pathology [8]. Numerous digital pathology methods have been proposed for analyzing differing types of tissue specimen images [9,10]; for instance, prostate cancer [11], breast cancer [12], neuroblastoma [13], and colorectal cancer [14]. A variety of approaches have been employed to quantify tissue characteristics such as color [15], morphology [15–17], Fourier transform [18], wavelet transform [15,19], gray level co-occurrence matrix [20], fractal analysis [15,21], local binary pattern [22], probabilistic models [23], and graph theory [24]. Machine learning algorithms, including decision tree [25], k-nearest neighbor [19,15,21], Bayesian [26], support vector machine (SVM) [15,16], boosting [27], and deep learning [28,29] have been applied to determine the disease status.

Many of the previous works in digital pathology arbitrarily chose the resolution or scale at which a tissue image is processed and/or analyzed [30]. It is likely determined by the resolution at which the tissue specimen is digitized. This approach is contrary to a pathologist who examines a tissue specimen under a microscope at different resolutions; for example, the presence (or absence) of basal nuclei (indicative of cancer) is assessed at a higher resolution (e.g., ×20 or ×40 magnification) whereas a lower resolution is suitable to observe the glandular shape or formation. Accordingly, the current pathological review process adopts a multi-resolution (or scale) approach. There have been several multi-resolution digital pathology approaches, mainly based on a coarse-to-fine strategy

[13,31,26,32]. The coarse-to-fine strategy, in general, makes an initial decision (or diagnosis) at a coarser resolution, and then refines the decision at finer resolutions. In this strategy, the choice and order of coarse and fine resolutions are still, by and large, arbitrary. Further, it is not trivial to correct earlier mistakes (or misclassifications) as well as to determine the optimal resolutions for analysis. Alternatively, a multiview learning approach is able to combine information from multiple sources (multiple resolutions in our case) and to make a better decision. It was first developed in the context of the semi-supervised learning, so called co-training algorithm [33]. Later, several techniques were developed in the context of the supervised learning; for example, classifier fusion [34], local learning [35], and boosting [36].

In this study, we propose an automated, cooperative, and multi-resolution approach to improve prostate cancer diagnosis (Fig. 1). Tissue specimens are digitized and segmented using intensity- and texture-based image features [37]. Utilizing the segmented lumens and epithelial nuclei, a number of morphological features are computed to characterize tissues at varying resolutions. A multiview boosting classification method [38] is adopted to achieve a robust and accurate tissue segmentation and diagnosis through an effective and efficient cooperation between differing resolutions. We systematically evaluate the performance of the multiview boosting method using a large set of digitized tissue images from tissue microarrays (TMA).

The rest of this manuscript is organized as follows. In Materials and Methods section, we begin with describing the dataset and multiview boosting algorithm. In the following subsections, tissue segmentation, tissue morphological feature extraction, and classification method are described. In Results section, the experimental results, including tissue segmentation and cancer detection, are presented. In Discussions section, the implications and limitations of our study are discussed. Finally, we conclude in Conclusions section.

2. Materials and methods

2.1. Dataset

We employed five tissue microarrays (TMA) from Tissue microarray research program at the National Institutes of Health. Tissue specimen sample cores in TMA are stained with hematoxylin and eosin (H&E) and scanned under a standard brightfield microscope (Leica Biosystems) at 40× magnification, resulting in a spatial resolution of 0.228 μm × 0.228 μm. A tissue sample core may contain one or more glands. An experienced pathologist (S.M.H) reviewed each tissue specimen sample core and determined its disease status using the Gleason grading system. Five TMAs contain 205 (70 Benign, 135 Cancer), 163 (74 Benign, 89 Cancer), 185 (70 Benign, 115 Cancer), 161 (79 Benign, 82 Cancer), and 158 (72 Benign, 86 Cancer) tissue specimen samples, respectively (Table 1). The first TMA (“TMA1”) is used to train and test a tissue segmentation method; 101 (212,104 epithelium, 279,811 stroma, 72,827 nucleus, and 56,851 lm pixels) and 104 (219,235 epithelium, 275,893 stroma, 83,291 nucleus, and 63,144 lm pixels) tissue samples are used for the purpose of training and testing, respectively. For cancer diagnosis, the second TMA (“TMA2”; including 163 samples) is used as the training dataset, and third, fourth, and fifth TMAs

("TMA3", "TMA4", and "TMA5"; including 185, 161, and 158 samples, respectively) are used as the validation datasets.

Algorithm 1.

Multiview boosting.

Input

$S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ where $x_i \in X_1 \times \ldots \times X_m$,

$y_i \in \{1,..,k\}$, weak learning algorithm WL,

number of iteration T, baseline B

Initialize

$i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, m\}$, $l \in 1, \ldots, k\}$: $f_{0,j}(i, l) = 0$

Cost matrices: $C_{0,G}(i,l) = C_{0,j}(i,l) = \begin{cases} 1 & \text{if } y_i \neq l \\ -(k-1) & \text{if } y_i = l \end{cases}$

where $C_{0,G}$ is the global cost matrix

for $t = 1$ to T do

Train WL using $C_{t-1,1}, \ldots, C_{t-1,m}$

for $j = 1$ to m do

Get weak classifier $h_{t,j}$ and edge $\delta_{t,j}$ on $C_{t-1,j}$

Compute $\alpha_{t,j} = \frac{1}{2}\ln\frac{1+\delta_{t,j}}{1-\delta_{t,j}}$

for $j = 1$ to m do

Update cost matrices: $C_{t,j}(i,l) = \begin{cases} \exp\left(f_{t,j}(i,l) - f_{t,j}(i,y_i)\right) & \text{if } y_i \neq l \\ -\sum_{p=1, p\neq y_i}^{k} \exp\left(f_{t,j}(i,p) - f_{t,j}(i,y_i)\right) & \text{if } y_i = l \end{cases}$

where $f_{t,j}(i,l) = \sum_{z=1}^{t} 1\left[h_{z,j}(i) = l\right]\alpha_{z,j} d_{z,j}(i)$,

and $d_{z,j}(i) = \begin{cases} 1 & \text{if } h_{z,j}(i) = y_i \text{ or } \nexists q \in \{1, \ldots, m\}, h_{z,q}(i) = y_i \\ 0 & \text{else} \end{cases}$

Get $h_t = \underset{h_{t,j}}{\operatorname{argmax}}\left(\text{edge } h_{t,j} \text{ on } C_{t,G}\right)$ and δ_t on $C_{t,G}$

Compute $\alpha_{t,j} = \frac{1}{2}\ln\frac{1+\delta_t}{1-\delta_t}$

Update global cost matrix: $C_{t,G}(i,l) = \begin{cases} \exp\left(f_{t,G}(i,l) - f_{t,G}(i,y_i)\right) & \text{if } y_i \neq l \\ -\sum_{j\neq y_i}^{k} \exp\left(f_{t,G}(i,j) - f_{t,G}(i,y_i)\right) & \text{if } y_i = l \end{cases}$

where $f_{t,G}(i,l) = \sum_{z=1}^{t} 1[h_z(i) = l]\alpha_z$

Final hypothesis

$H(x) = \underset{l \in 1, \ldots, k}{\operatorname{argmax}} f_T(x,l)$ where $f_T(i,l) = \sum_{t=1}^{T} 1[h_t(i) = l]\alpha_t$

2.2. Multiview boosting

Consider a dataset $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ where $x_i \in X$ is an example and $y_i \in Y$ is a class label. X comprises a number of views $X = X_1 \times \ldots \times X_m$ and $Y = \{1, \ldots, k\}$. An

example x_i in a view m is represented as $x_{i,m}$. In a cost matrix C, each row and column denotes an example and a class label, respectively, i.e., $C(i, l)$ is the cost of classifying the example i as the class l. It maintains a cost matrix C_j for a view j and a global cost matrix C_G. A weak classifier h_j is learned on a view j, satisfying the edge condition if $C_j \cdot 1_{h_j} \leq C_j \cdot B$ where 1_h is the prediction matrix defined as $1_{h_j}(i,l)$ is 1 if $h_j(i) = l$ and 0 otherwise. A baseline B is a cost matrix $U\gamma$ defined as $U\gamma(i, l)$ is $\frac{(1-\gamma)}{k} + \gamma$ if $y_i = l$ and $\frac{(1-\gamma)}{k}$ if $y_i \neq l$. Edge of h_j is computed as $\delta_j = C_j \bullet U_j - C_j \bullet 1_j$. It maintains the cost matrices in a way that the hardest examples for a view are managed by the other views. To achieve this, the matrices are updated as follows: ith row in C_j is updated ($d_j(i) = 1$) only if the classifier from the view j correctly classifies the example i or if the rest of the classifiers misclassify it. Also, the global cost matrix C_G is utilized to select the best classifier from the classifiers built on different views at each iteration. Thereby, the final boosting classification model is a weighted sum of the classifiers from different views. Depending on the training procedure, a classifier from a particular view may be selected multiple times with different weights or completly missed from the model construction. The algorithm of multiview boosting is presented in Algorithm 1.

2.3. Tissue segmentation

To improve contrast and color representation, a tissue specimen image I (in RGB: red, green, and blue channels) is converted into 3 different color forms: 1) Histogram equalization 2) HSV (hue, saturation, and value) color space 3) La*b* (L: illumination, a* and b*: color-opponent dimensions) color space. In total, 9 color channels are generated. To quantify the image characteristics, we employ intensity- and texture-based features. Intensity-based features include average, standard deviation, kurtosis, and skewness. Texture-based features are computed using local binary pattern (LBP) [39], local directional derivative pattern (LDDP) [40], and variance measure (VAR) [39] with two neighborhood topologies (P,R) = {(16,2),(24,3)} where P and R are the number of neigh- boring pixels and the radius, respectively. These operators generate a (binary) pattern code per pixel. The pattern codes are summarized into a histogram.

For a pixel $x \in I$, we extract the intensity- and texture-based features at different resolutions (or views) by drawing a rectangular window of differing sizes w (w = 1, 3, 7, 15, and 31 pixels; pixel resolution = 0.228 μm). For w = 1, the intensity value of 9 color channels constitutes 9 features. For w = 3, 36 features (average, standard deviation, kurtosis, and skewness per color channel) are computed. For $w \geq 7$, we compute 36 intensity-based features and 108 texture-based features ((P,R) = (16,2): 18 features for LBP and LDDP and 10 features for VAR per color channel; (P,R) = (24,3): 26 features for LBP and LDDP and 10 features for VAR per color channel). A multiview boosting algorithm is adopted to cooperatively integrate the 5 different sets of features and to provide a histological class of a pixel.

A tissue specimen image is segmented into a lumen, nucleus, epithelium, and stroma in a cascaded manner (Fig. 2). The tissue specimen image is first segmented into lumen and non-lumen areas. Lumens are determined by using a threshold value (> 0.5) on the output of

the lumen vs. non-lumen multiview boosting classification (+ : lumen, −: non-lumen), followed by a size constraint (> 50 μm^2). Second, non-lumen areas are classified into nucleus and non-nucleus areas. Thresholding (> 0.5) the output of the nucleus vs. non-nucleus multiview boosting classification (+ : nucleus, −: non-nucleus), initial nuclei are identified. The size and shape of the initial nuclei are examined: If the size of a nucleus is smaller than 5 μm^2 or the ratio of the major and minor axis is greater than 5 when its size is smaller than 25 μm^2, then the nucleus is considered to be an artifact. Individual nucleus is identified by applying the Euclidean distance transform and a watershed algorithm (Fig. 3). Third, non-nucleus areas are grouped into epithelium and stroma. We identify epithelium by using a threshold value (> 0.5) on the output of the epithelium vs. stroma multiview boosting classification (+ : epithelium, −: stroma), followed by a size constraint (> 500 μm^2). Moreover, in order to identify and eliminate lumen artifacts, the perimeter of a lumen is examined. By definition, epithelial cells enclose a lumen in tissue. If < 40% of its perimeter is surrounded by epithelium, such lumen is excluded.

2.4. Tissue morphology: feature extraction

We quantify tissue characteristics using morphological features of lumens and epithelial nuclei and relational features between lumens and epithelial nuclei (Fig. 4). The morphological features of a lumen are:

1. Area L_{area}: A number of pixels within a lumen
2. Compactness: $L_{peri}/\sqrt{L_{area}}$ where L_{peri} is the perimeter of a lumen
3. Smoothness: L_{peri}/L_{bperi} where L_{bperi} is the perimeter of a bounding box of a lumen
4. Roundness: $(r \bullet L_{peri})/L_{area}$ where is r the radius of a circle of size L_{area}
5. Convex hull ratio: L_{conv}/L_{area} where L_{conv} is the size of a convex hull of a lumen
6. Major-minor axis ratio: Ratio of major and minor axes of a lumen
7. Extent: L_{peri}/L_{barea} where L_{barea} is the size of a bounding box of a lumen
8. Bounding circle ratio: L_{peri}/L_{bcirc} where L_{bcirc} is the size of a minimum bounding circle of a lumen
9. Distortion: $STD(L_d)/AVG(L_d)$ where L_d is the distance from the center of a lumen to the boundary of the lumen and AVG and STD represent the average and standard deviation
10. Epithelial perimeter ratio: Portion of the perimeter of a lumen that is connected to epithelium
11. Distance to the closest lumen
12. Symmetric index of a lumen boundary: Sum of vertical and horizontal symmetry. Vertical and horizontal symmetry are computed as $\Sigma|L_{Ti} - L_{Bi}|/\Sigma(L_{Ti} + L_{Bi})$ and $\Sigma|L_{Ri} - L_{Li}|/\Sigma(L_{Ri} + L_{Li})$, respectively, where L_{Ti} and L_{Bi} are vertical distances from a vertical axis to the boundary of a lumen and L_{Li} and L_{Ri} are horizontal

distances from a horizontal axis to the boundary of a lumen. The vertical axis runs along the longest diameter, and the horizontal axis runs perpendicularly to the horizontal axis passing the center of a lumen

13. Symmetric index of a lumen area: Sum of left-right area symmetry and top-bottom area symmetry. Left-right and top-bottom area symmetry are computed as $|L_{Larea} - L_{Rarea}|/(L_{Larea} + L_{Rarea})$ and $|L_{Tarea} - L_{Barea}|/(L_{Tarea} + L_{Barea})$, respectively, where L_{Larea}, L_{Rarea}, L_{Tarea}, and L_{Barea} denote the size of left, right, top, and bottom quadrants (determined by vertical and horizontal axes), respectively

14. Number of lumens

Similarly, epithelial nucleus features are computed:

1. Area
2. Compactness
3. Smoothness
4. Roundness
5. Convex hull ratio
6. Major-minor axis ratio
7. Extent
8. Bounding circle ratio
9. Distortion
10. Shape context [41]: For each nucleus, a log-polar histogram is constructed using 4 bins for r (bin_r) and 4 bins for θ ($bin\ \theta$). r is the distance between a nucleus and its neighboring nucleus. θ is the angle between the line of reference (x-axis) and the line through a nucleus and its neighboring nucleus. A histogram is defined to be the shape context of a nucleus

$$bin_r = \begin{cases} 1, & if\ r \le \log(20\mu\text{m}) \\ 2, & if\ \log(20\mu\text{m}) < r \le \log(40\mu\text{m}) \\ 3, & if\ \log(20\mu\text{m}) < r \le \log(60\mu\text{m}) \\ 4, & if\ \log(20\mu\text{m}) < r \le \log(80\mu\text{m}) \end{cases}$$

$$\text{bin}_\theta = \begin{cases} 1, & if\ 0° \le \theta < 45°\ or\ 180° \le \theta < 225° \\ 2, & if\ 45° \le \theta < 90°\ or\ 225° \le \theta < 270° \\ 3, & if\ 90° \le \theta < 135°\ or\ 270° \le \theta < 315° \\ 4, & if\ 135° \le \theta < 180°\ \text{or}\ 315° \le \theta < 360° \end{cases}$$

11. Distance fractal dimension: For each nucleus, the distances from a nucleus to its closest N neighboring nuclei are computed and sorted in an ascending order. The slope of the best linear least-squares fit is computed for the order versus distance plot. (N = 10, 20, 30)

12. Higuchi fractal dimension [42]: For each nucleus, the distances from a nucleus to its closest N (up to 30) neighboring nuclei are computed and sorted in an ascending order, forming a data series: $x(1)$, $x(2)$, …, $x(N)$. k new data series are constructed for $m = 1, \ldots, k : x_m^k = \left\{ x(m), x(m+k), \ldots, x\left(m + \frac{N-m}{k}k\right) \right\}$. The length of x_m^k is defined as $L_m(k) = \left(\sum_{i=1}^{\frac{N-m}{k}} |x(m+ik) - x(m+(i-1)k)|\right)\frac{N-1}{\frac{N-m}{k}k}$. The average length is computed as $L(k) = \frac{1}{k}\sum_{m=1}^{k} L_m(k)$. Repeating this for each k ranging from 1 to k_{max}, the best linear least-squares fit for the plot of ln $L(k)$ versus ln k gives the Higuchi fractal dimension.

13. Entropy of nuclei distribution: Dividing a tissue image into 50 μm X 50 μm disjoint partitions, the entropy is computed as $H(nuclei) = -\sum_{i=1}^{n}\sum_{j=1}^{n} p(x_{ij}) \log p(x_{ij})$ where $p()$ is the probability mass function of the number of nuclei in a partition and x_{ij} is the number of nuclei in (i, j) partition.

14. Distance to the closest epithelial nucleus

15. Number of nuclei

16. Number of neighboring epithelial nuclei: Number of epithelial nuclei that are within 20 μm, 40 μm, 60 μm, and 80 μm from a nucleus.

Also, relational features utilizing both lumens and epithelial nuclei are computed:

1. Distance from a lumen to its closest epithelial nucleus

2. Number of epithelial nuclei that are assigned to a lumen: A nucleus is assigned to its closest lumen

3. Distance from an epithelial nucleus to its closest lumen

4. Number of distal epithelial nuclei: A nucleus is designated as a distal nucleus if it is 40 μm away from its closest lumen.

For each histological class (a lumen and epithelial nucleus), the morphological features are computed at different resolutions (or views) by sliding a rectangular window of differing sizes w (w = 100, 300, 500, 700 μm; view1, 2, 3, and 4) and from a whole tissue sample (view5), i.e., 5 different resolutions are employed (Fig. 5). The quantities describing lumens and epithelial nuclei within a window are summarized by computing up to 5 statistics (AVG, STD, minimum, maximum, and total sum). AVG, STD, mini- mum, and maximum of the summarized statistics form the morphological features (from 1st to 4th resolution). In total, 670 features are computed (217 lm, 333 epithelial nucleus, and 120 relational features). We also calculate the 5 statistics of the morphological quantities from a whole tissue sample. These become the morphological features for the 5th view, comprising 175 features (54 lm, 91 epithelial nucleus, and 30 relational features).

2.5. Classification: Cancer vs. Benign

We employ a multiview boosting algorithm to cooperatively combine the 5 different sets of morphological features and to determine the disease status of tissue specimen samples. A linear support vector machine (SVM) [43] is adopted as a weak classifier. 10% of the training dataset is sampled based on the global cost matrix and used to build a weak classifier. For each view (or feature set), the discriminative morphological features are selected using Wilcoxon-rank sum test (p-value < 0.01; 17 features are selected on average) and used to learn the weak classifier. The boosting procedure terminates if it reaches 20 boosting steps.

In order to evaluate the performance of the proposed method, four TMAs (TMA2, TMA3, TMA4, and TMA5) are employed. We train a multiview boosting classifier on TMA2 (74 Benign, 89 Cancer) and test its performance on the rest of the TMAs (TMA3, TMA4, TMA5; 221 Benign, 283 Cancer). The correct and incorrect predictions by the multiview boosting classifier are summarized into a receiver operating characteristic (ROC) plot, and the area under the ROC curve (AUC) is computed. Boot-strap resampling with 20 0 0 repetitions is adopted to assess 95% confidential interval (CI) of AUCs and statistical significance of the differences between AUCs of the two ROC curves [44].

3. Results

We adopted three multiview boosting classifiers to segment tissue sample images: 1) lumen versus non-lumen 2) nucleus versus non-nucleus 3) epithelium versus stroma. Training and testing on the manually selected ROIs from TMA1, we obtained an AUC of 0.99 (95% CI: 0.99–1.00), 0.99 (95% CI: 0.99–0.99), 0.97 (95% CI: 0.97–0.98) for lumen versus non-lumen, nucleus versus non-nucleus, and epithelium versus stroma, respectively. Applying the three multiview boosting classifiers in a cascaded fashion, lumens and epithelial nuclei were identified from the tissue specimen samples in TMA2, TMA3, TMA4, and TMA5.

Extracting and utilizing morphological features from lumens and epithelial nuclei at 5 different views (w = 100, 300, 500, 700 μm, whole tissue), a multiview boosting classifier was trained on TMA2 and tested on the validation datasets, including TMA3, TMA4, and TMA5 (Table 2). The classifier achieved an AUC of 0.98 (95% CI: 0.97–0.99). Moreover, we compared the performance of the multiview boosting classification to that of single-view boosting classifications. The same boosting scheme was adopted to train and validate the single-view boosting classifiers. An individual boosting classifier was constructed using the morphological features from each view. The AUCs of the single-view classifications were: 0.95 (95% CI: 0.93–0.97) for view1, 0.96 (95% CI: 0.94–0.97) for view2, 0.95 (95% CI: 0.93–0.97) for view3, 0.95 (95% CI: 0.94–0.970) for view4, and 0.96 (95% CI: 0.94–0.97 for view5 (Table 1). A single-view boosting classification on the entire feature set (concatenating 5 views) was also constructed and examined, and an AUC of 0.96 (95% CI: 0.95–0.98) was obtained. The AUC of the multiview boosting classification was significantly larger than that of the single-view classifications (p-$value$ < 0.01) (Table 2).

To ensure the robustness of the multiview boosting cancer detection, we performed the following experiments. Using each of the validation datasets (TMA3, TMA4, and TMA5),

we trained a separate multiview boosting classifier and tested on the rest of the datasets including TMA2. For examples, a multiview boosting classifier was trained on TMA3 and tested on TMA2, TMA4, and TMA5. Trained on TMA3, TMA4, and TMA5, these experiments resulted in an AUC of 0.97 (95% CI: 0.96–0.98), 0.98 (95% CI: 0.96–0.99), and 0.97 (95% CI: 0.96–0.98), respectively. No significant difference was found in comparison to the above experiment where TMA2 was used as the training dataset (0.98 AUC; 95% CI: 0.97–0.99). Moreover, for each experiment, single-view boosting classifiers were constructed and compared to the multiview boosting classifier. For all the three experiments, statistically significant differences between the multiview boosting classification and single-view boosting classifications were found ($p\text{-}value < 0.05$).

Further, the classification result of the multiview boosting method was compared to standard machine learning algorithms – SVM and random forests. Similar to the above single-view boosting classifications, SVM with a radial basis kernel using all features and random forests with 20 trees and random selection of $\sqrt{n}$ features (n: number of features) were trained and tested using the morphological features per view and the combined feature set of 5 views. The multiview boosting classification outperformed SVM (≤ 0.97 AUC) and random forests (≤ 0.96 AUC) classifications (Table 3). The performance of the multiview boosting classification significantly differed from that of both SVM and random forests classifications ($p\text{-}value < 0.05$).

The importance of the morphological features was examined by measuring the fraction of boosting iterations that used each of the features (Fig. 6). Lumen features were, in general, more frequently selected than epithelial nucleus features. This is consistent with a previous report [16]. The best-5 features included two lumen features (convex hull ratio and symmetric index of a lumen boundary), one epithelial nucleus feature (shape context), and two relational features (distance from a lumen to its closest epithelial nucleus and number of epithelial nuclei assigned to a lumen). That is, the shape of a lumen, nuclei distribution, and relation between lumens and epithelial nuclei are most informative in distinguishing cancers from benign tissue samples.

The computational performance of our method was evaluated. The method was implemented in MATLAB and performed on a PC with 3.4 GHz processor and 8GB of RAM. The spatial dimension of a tissue specimen sample is ~5000 × 5000 pixels on average. It requires 3143 s to process and classify a single tissue specimen image. Notably, the tissue segmentation step takes > 90% of the entire running time. This is because tissue segmentation is performed per pixel basis. A substantial reduction in computation cost would be possible via a parallelization of the process. The cancer detection by the multiview boosting takes a minimal amount of time (< 1 s), i.e., no extra computation burden in comparison to other machine learning algorithms. The summary of the computational performance of our method is available in Table 4.

4. Discussions

This study has focused on examining the ability of the multiview boosting algorithm to improve the automated cancer detection. The experimental results suggest that the multiview

boosting approach could utilize tissue morphology from multiple views to identify prostate cancers and outperform the standard machine learning algorithms.

Several multi-resolution approaches have been proposed to detect and characterize cancers. The previous methods typically analyze tissues at multiple resolutions, i.e., learn a classification model per resolution, and combine the classification models/results via a coarse-to-fine strategy [13,31,26,32] or aggregation/averaging scheme [45,46]. The procedures are mainly pre-determined by heuristics whereas pathologists implicitly integrate the information from multiple resolutions (or views) and make diagnostic decisions. The multiview boosting approach does not explicitly determine how to combine the tissue information or classification results from differing views, but rather automatically and interactively integrates them.

The multiview boosting approach outperformed the single-view boosting approaches, in particular, the single-view approach of concatenating the entire 5 views. This demonstrates that the multiview boosting approach does not simply merge information from differing views but is capable of optimizing the use of the different information in a synergistic manner. Moreover, no significant difference was found in choosing different training and test dataset, i.e., the performance of the multiview boosting approach was not dataset dependent.

Multimodal imaging/data approaches may take advantage of the multiview boosting approach; for instance, combining microscopy imaging with chemical imaging [16,11], radiology imaging [47], or genomic data [48–50]. Multimodal imaging/data approaches often extract a set of features from each modality, simply combine the feature sets together, and construct a classification model. The multiview boosting algorithm could facilitate the effective and efficient fusion of the differing imaging/data modalities as well as the reliable and accurate cancer diagnosis.

There are several limitations to our study. First, the presented multiview boosting approach only identifies cancer tissue specimen samples. The extended study is desirable to assess the capability of the multiview boosting approach to provide Gleason grade of tissue specimens. Second, the morphological features were computed from a $w \times w$ rectangular window of 4 different sizes (w = 100, 300, 500, 700 μm) and a whole tissue sample. The choice of the size and number of the window is still arbitrary. Optimizing these factors could provide an improved cancer detection. Third, we only utilized morphological features of lumens and epithelial nuclei in this study. Morphological features of glands and stroma have been proposed to characterize cancers [51,12,52]. Also, other types of image features, e.g., texture features, can be incorporated to our approach. Last, our method was developed and evaluated on TMAs that were prepared at one institute. The efficacy of our approach on whole slide prostate specimen images has not been examined. A validation study, including multi-institute datasets and whole slide specimen images, could further ensure the reliability and validity of our approach.

5. Conclusion

In this study, we have demonstrated a multiview boosting approach for prostate cancer diagnosis. The experimental results are sufficiently promising to warrant prospective validation in the clinics and to be extended to more complex pathological models. The proposed method obtains tissue characteristics from multiple resolutions and cooperatively combines them to automatically identify cancers with high accuracy. We anticipate that the application of the multiview boosting approach will augment the accuracy, robustness, and utility of digital pathology tools by facilitating the multi-resolution process or analysis of tissues as well as the integration of information from various imaging technologies, leading to improved cancer pathology.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2016R1C1B2012433).

References

[1]. Siegel RL, Miller KD, Jemal A, Cancer statistics, 2016, CA Cancer J. Clin. 66 (2016) 7–30. [PubMed: 26742998]

[2]. Gleason DF, Classification of prostatic carcinomas, Cancer Chemother Rep. 50 (1966) 125–128. [PubMed: 5948714]

[3]. Cintra ML, Billis A, Histologic grading of prostatic adenocarcinoma: intraobserver reproducibility of the Mostofi, Gleason and Bocking grading systems, Int. Urol. Nephrol. 23 (1991) 449–454. [PubMed: 1938244]

[4]. Ozdamar SO, Sarikaya S, Yildiz L, Atilla MK, Kandemir B, Yildiz S, Intraob- server and interobserver reproducibility of WHO and Gleason histologic grading systems in prostatic adenocarcinomas, Int. Urol. Nephrol. 28 (1996) 73–77. [PubMed: 8738623]

[5]. Egevad L, Ahmad AS, Algaba F, Berney DM, Boccon-Gibod L, Comperat E, Evans AJ, Griffiths D, Grobholz R, Kristiansen G, Langner C, Lopez-Beltran A, Montironi R, Moss S, Oliveira P, Vainer B, Varma M, Camparo P, Standardization of Gleason grading among 337 European pathologists, Histopathology 62 (2013) 247–256. [PubMed: 23240715]

[6]. Madabhushi A, Lee G, Image analysis and machine learning in digital pathology: challenges and opportunities, Med. Image Anal. 33 (2016) 170–175. [PubMed: 27423409]

[7]. Kwak JT, Reddy R, Sinha S, Bhargava R, Analysis of variance in spectroscopic imaging data from human tissues, Anal. Chem. 84 (2012) 1063–1069. [PubMed: 22148458]

[8]. Pantanowitz L, Digital images and the future of digital pathology, J. Pathol. Inform. 1 (2010).

[9]. Mulrane L, Rexhepaj E, Penney S, Callanan JJ, Gallagher WM, Automated image analysis in histopathology: a valuable tool in medical diagnostics, Expert Rev. Mol. Diagn. 8 (2008) 707–725. [PubMed: 18999923]

[10]. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B, Histopathological image analysis: a review, IEEE Rev. Biomed. Eng. 2 (2009) 147–171. [PubMed: 20671804]

[11]. Kwak JT, Hewitt SM, Kajdacsy-Balla AA, Sinha S, Bhargava R, Automated prostate tissue referencing for cancer detection and diagnosis, BMC Bioinform. 17 (2016) 227.

[12]. Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ, West RB, van de Rijn M, Koller D, Systematic analysis of breast cancer morphology uncovers stromal features associated with survival, Sci. Transl. Med. 3 (2011) 108ra113.

[13]. Kong J, Sertel O, Shimada H, Boyer KL, Saltz JH, Gurcan MN, Computer-aided evaluation of neuroblastoma on whole-slide histology images: classifying grade of neuroblastic differentiation, Pattern Recogn. 42 (2009) 1080–1092.

[14]. Kalkan H, Nap M, Duin RP, Loog M, Automated colorectal cancer diagnosis for whole-slice histopathology, Med. Image Comput. Comput. Assist. Interv. 15 (2012) 550–557. [PubMed: 23286174]

[15]. Tabesh A, Teverovskiy M, Pang HY, Kumar VP, Verbel D, Kotsianti A, Saidi O, Multifeature prostate cancer diagnosis and Gleason grading of histological images, IEEE Trans. Med. Imaging 26 (2007) 1366–1378. [PubMed: 17948727]

[16]. Kwak JT, Hewitt SM, Sinha S, Bhargava R, Multimodal microscopy for auto- mated histologic analysis of prostate cancer, BMC Cancer 11 (2011) 62. [PubMed: 21303560]

[17]. Gorelick L, Veksler O, Gaed M, Gomez JA, Moussa M, Bauman G, Fenster A, Ward AD, Prostate histopathology: learning tissue component histograms for cancer detection and classification, IEEE Trans. Med. Imaging 32 (2013) 1804–1818. [PubMed: 23739794]

[18]. Smith Y, Zajicek G, Werman M, Pizov G, Sherman Y, Similarity measurement method for the classification of architecturally differentiated images, Comput. Biomed. Res. 32 (1999) 1–12. [PubMed: 10066352]

[19]. Jafari-Khouzani K, Soltanian-Zadeh H, Multiwavelet grading of pathological images of prostate, IEEE Trans. Biomed. Eng. 50 (2003) 697–704. [PubMed: 12814236]

[20]. DiFranco M, O'Hurley G, Kay E, Watson W, Cunningham P, Automated gleason scoring of prostatic histopathology slides using multi-channel co-occurrence texture features, in: Proceedings of International Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB'08), 2008.

[21]. Huang PW, Lee CH, Automatic classification for pathological prostate images based on fractal analysis, IEEE Trans. Med. Imaging 28 (2009) 1037–1050. [PubMed: 19164082]

[22]. Gertych A, Ing N, Ma ZX, Fuchs TJ, Salman S, Mohanty S, Bhele S, Velasquez-Vacca A, Amin MB, Knudsen BS, Machine learning approaches to analyze histological images of tissues from radical prostatectomies, Comput. Med. Imag. Grap. 46 (2015) 197–208.

[23]. Monaco JP, Tomaszewski JE, Feldman MD, Hagemann I, Moradi M, Mousavi P, Boag A, Davidson C, Abolmaesumi P, Madabhushi A, High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models, Med. Image Anal. 14 (2010) 617–629. [PubMed: 20493759]

[24]. Nguyen K, Sarkar A, Jain AK, Prostate cancer grading: use of graph cut and spatial arrangement of nuclei, IEEE Trans. Med. Imaging 33 (2014) 2254–2270. [PubMed: 25029379]

[25]. Doyle S, Feldman MD, Shih N, Tomaszewski J, Madabhushi A, Cascaded dis- crimination of normal, abnormal, and confounder classes in histopathology: gleason grading of prostate cancer, BMC Bioinform. 13 (2012) 282.

[26]. Doyle S, Feldman M, Tomaszewski J, Madabhushi A, A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies, IEEE Trans. Biomed. Eng. 59 (2012) 1205–1218. [PubMed: 20570758]

[27]. Doyle S, Madabhushi A, Feldman M, Tomaszeweski J, A boosting cascade for automated detection of prostate cancer from digitized histology, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 4191, Springer Berlin Heidelberg, 2006, pp. 504–511.

[28]. Xu J, Luo XF, Wang GH, Gilmore H, Madabhushi A, A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images, Neurocomputing 191 (2016) 214–223. [PubMed: 28154470]

[29]. Litjens G, Sanchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, Hulsbergen-van de Kaa C, Bult P, van Ginneken B, van der Laak J, Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis, Sci Rep. 6 (2016). [PubMed: 28442741]

[30]. Doyle S, Rodriguez C, Madabhushi A, Tomaszeweski J, Feldman M, Detecting prostatic adenocarcinoma from digitized histology using a multi-scale hierarchical classification approach, in: 2006 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1–15, 2006, pp. 5558–5561.

[31]. Roullier V, Lezoray O, Ta VT, Elmoataz A, Multi-resolution graph-based analysis of histopathological whole slide images: application to mitotic cell extraction and visualization, Comput. Med. Imag. Grap. 35 (2011) 603–615.

[32]. Barker J, Hoogi A, Depeursinge A, Rubin DL, Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles, Med. Image Anal. 30 (2016) 60–71. [PubMed: 26854941]

[33]. Blum A, Mitchell T, Combining labeled and unlabeled data with co-training, in: Proceedings of the Eleventh Annual Conference on Computational Learning Theory, ACM, 1998, pp. 92–100.

[34]. Wozniak M, Jackowski K, Some remarks on chosen methods of classifier fusion based on weighted voting, in: Corchado E, Wu X, Oja E, Herrero Á, Baruque B (Eds.), Hybrid Artificial Intelligence Systems: 4th International Conference, HAIS 2009, Salamanca, Spain, June 10–12, 2009. Proceedings, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 541–548.

[35]. Zhang D, Wang F, Zhang C, Li T, Multi-view local learning, AAAI (2008) 752–757.

[36]. Peng J, Barbu C, Seetharaman G, Fan W, Wu X, Palaniappan K, Shareboost: boosting for multi-view learning with performance guarantees, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2011, pp. 597–612.

[37]. Kwak JT, Xu S, Pinto PA, Turkbey B, Bernardo M, Choyke PL, Wood BJ, A multiview boosting approach to tissue segmentation, in: Proc SPIE, 2014, p. 9041.

[38]. Koco S, Capponi C, A boosting approach to multiview classification with cooperation, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 6912, Springer Berlin Heidelberg, 2011, pp. 209–228.

[39]. Ojala T, Pietikainen M, Maenpaa T, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. 24 (2002) 971–987.

[40]. Guo ZH, Li Q, You J, Zhang D, Liu WH, Local directional derivative pattern for rotation invariant texture classification, Neural Comput. Appl. 21 (2012) 1893–1904.

[41]. Belongie S, Malik J, Puzicha J, Shape context: a new descriptor for shape matching and object recognition, NIPS (2000) 3.

[42]. Higuchi T, Approach to an irregular time-series on the basis of the fractal theory, Physica D 31 (1988) 277–283.

[43]. Chang CC, Lin CJ, LIBSVM: a library for support vector machines, ACM Trans. Intel Syst. Tec 2 (2011).

[44]. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M, pROC: an open-source package for R and S plus to analyze and compare ROC curves, BMC Bioinform. 12 (2011).

[45]. Basavanhally A, Ganesan S, Feldman M, Shih N, Mies C, Tomaszewski J, Madabhushi A, Multi-field-of-view framework for distinguishing tumor grade in ER plus breast cancer from entire histopathology slides, IEEE Trans. Biomed Eng. 60 (2013) 2089–2099. [PubMed: 23392336]

[46]. Sharma H, Zerbe N, Heim D, Wienert S, Behrens H-M, Hellwich O, Hufnagl P, A multi-resolution approach for combining visual information using nuclei segmentation and classification in histopathological images, VISAPP 3 (2015) 37–46.

[47]. Kwak JT, Sankineni S, Xu S, Turkbey B, Choyke PL, Pinto PA, Merino M, Wood BJ, Correlation of magnetic resonance imaging with digital histopathology in prostate, Int. J. Comput. Ass. Rad. 11 (2016) 657–666.

[48]. Madabhushi A, Agner S, Basavanhally A, Doyle S, Lee G, Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data, Comput. Med. Imag. Grap. 35 (2011) 506–514.

[49]. Cooper LAD, Kong J, Gutman DA, Wang FS, Gao JJ, Appin C, Cholleti S, Pan T, Sharma A, Scarpace L, Mikkelsen T, Kurc T, Moreno CS, Brat DJ, Saltz JH, Integrated morphologic analysis for the identification and characterization of disease subtypes, J. Am. Med. Inform. Assn. 19 (2012) 317–323.

[50]. Kong J, Cooper LAD, Wang FS, Gao JJ, Teodoro G, Scarpace L, Mikkelsen T, Schniederjan MJ, Moreno CS, Saltz JH, Brat DJ, Machine-based morphologic analysis of glioblastoma using whole-slide pathology images uncovers clinically relevant molecular correlates, PLoS One 8 (2013).

[51]. Doyle S, Hwang M, Shah K, Madabhushi A, Feldman M, Tomaszeweski J, Automated grading of prostate cancer using architectural and textural image features, Biomedical imaging: from nano

to macro, 2007, in: ISBI 2007. 4th IEEE international symposium on, IEEE, 2007, pp. 1284–1287.

[52]. Chen JM, Qu AP, Wang LW, Yuan JP, Yang F, Xiang QM, Maskey N, Yang GF, Liu J, Li Y, New breast cancer prognostic factors identified by computer-aided image analysis of HE stained histopathology images, Sci Rep. 5 (2015).

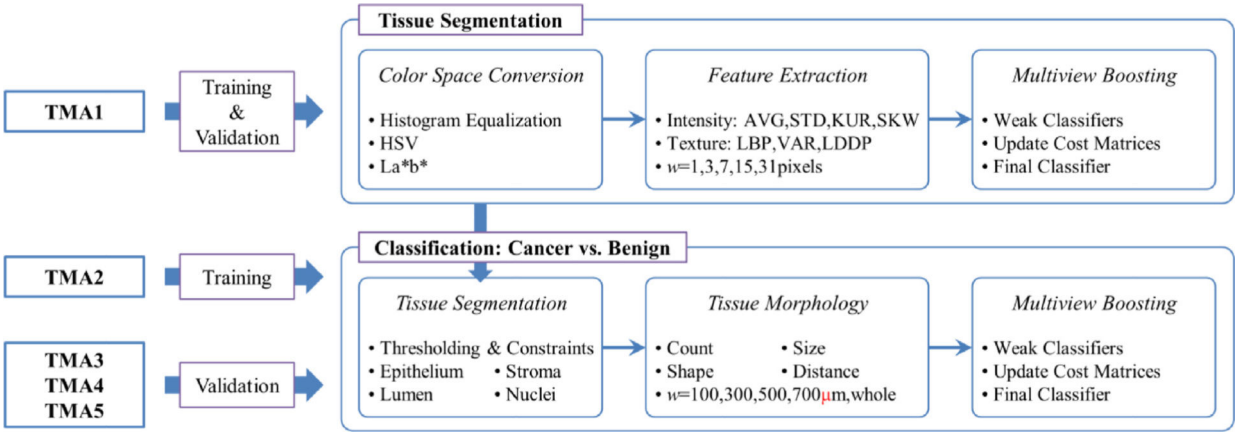

Fig. 1. A flowchart of multiview boosting cancer detection.

A cascaded multiview boosting tissue segmentation method is trained and validated using TMA1. Using the tissue segmentation, a multiview boosting cancer detection method, utilizing tissue morphology, is constructed based on TMA2 and validated using TMA3, TMA4, and TMA5.

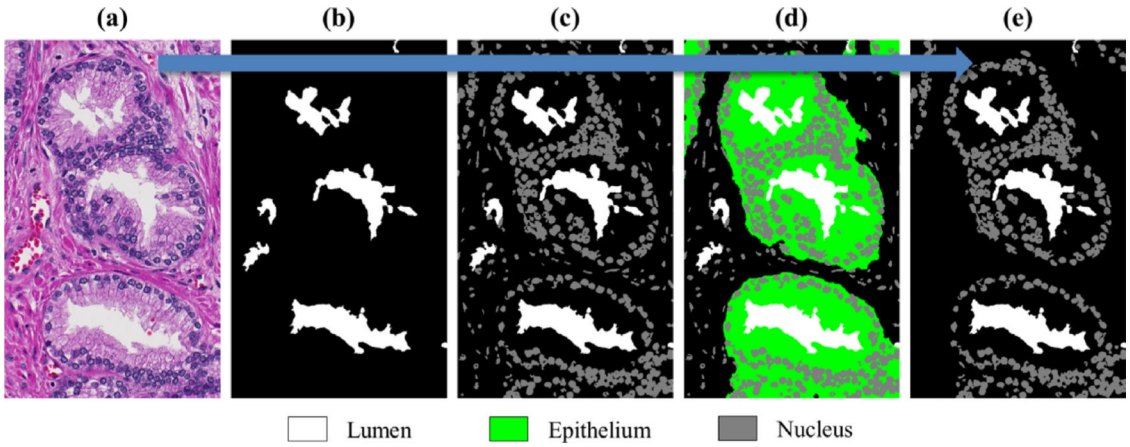

Fig. 2. Illustration of cascaded tissue segmentation.

(a) H&E tissue image is segmented into (b) lumen (white), (c) nucleus (grey), and (d) epithelium (green). (e) Epithelial nuclei are identified and luminal artifacts are eliminated.

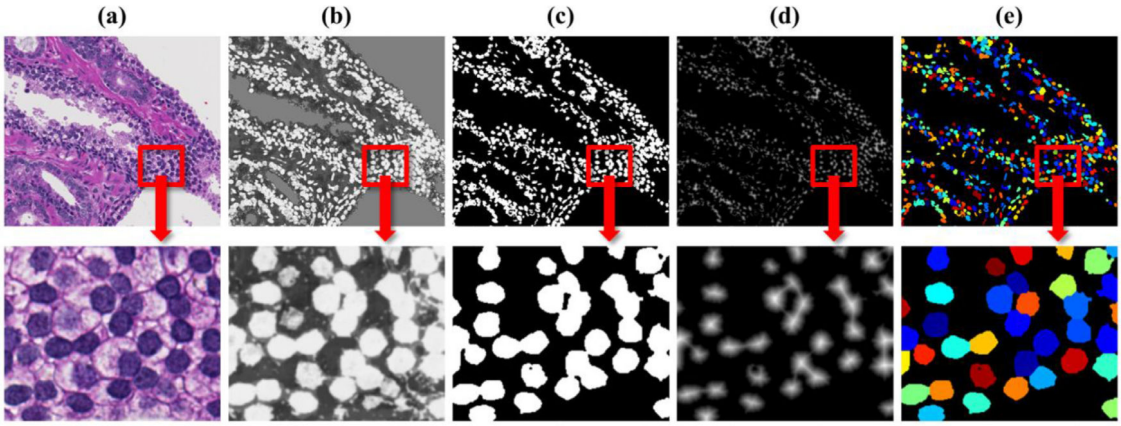

Fig. 3. Illustration of nuclus segmentation.

(a) H&E tissue image (b) the output of nucleus vs. non-nucleus multiview boosting classification (c) the result of thresholding and artifact removal (d) a distance map generated by the Euclidean distance transform (e) Individual nuclei are identified by a watershed algorithm and marked in color.

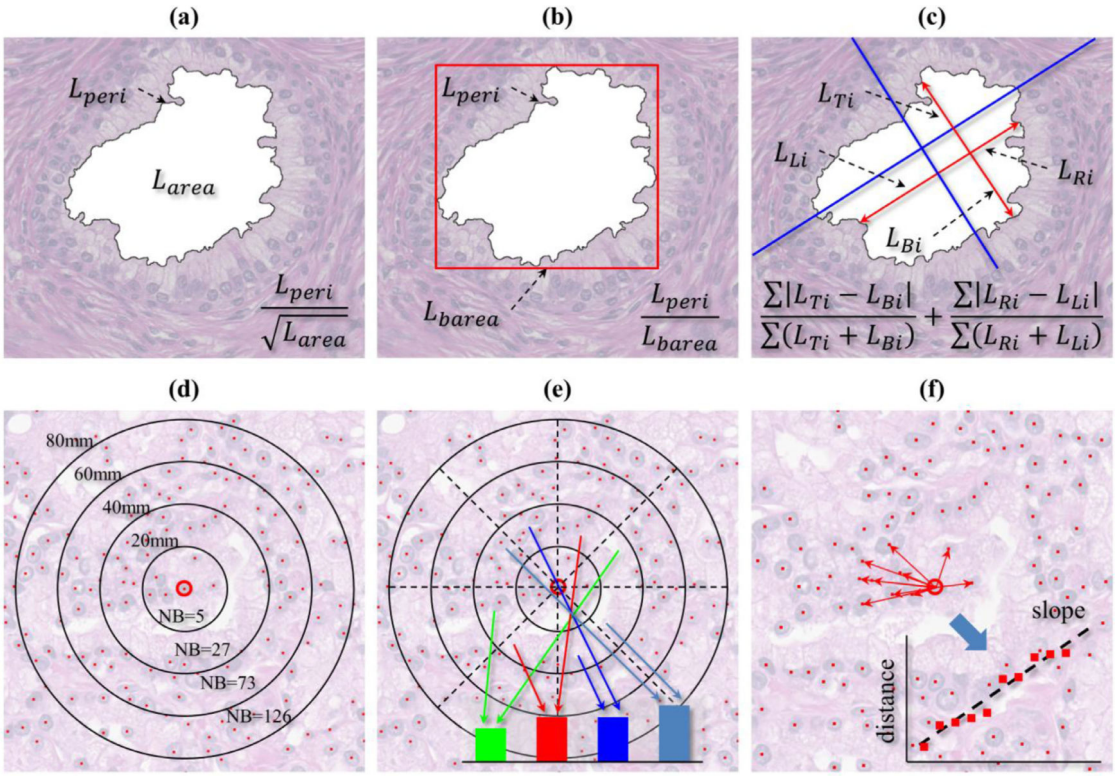

Fig. 4. Morphological feature extraction.

Examples of morphological feature extraction for (a)(b)(c) a lumen and (d)(e)(f) a nucleus. (a) Compactness (b) Extent (c) Symmetric index of a lumen boundary (d) Number of neighboring epithelial nuclei (e) Shape context (f) Distance fractal dimension.

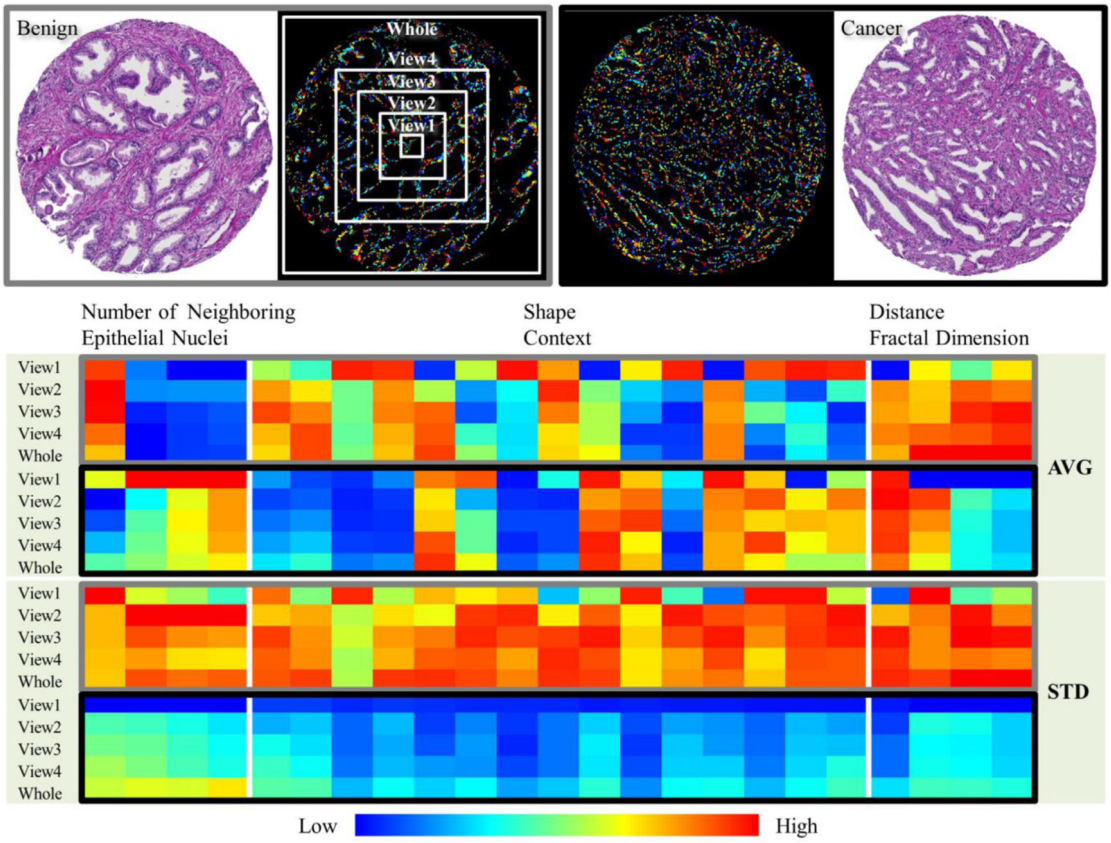

Fig. 5. Multiview feature extraction.

(Top) A set of morphological features are computed at multiple scales from a benign (grey rectangle) and cancer (black rectangle) tissue sample. Epithelial nuclei features are summarized within a window of differing sizes w (w = 100, 300, 500, 700 μm, and a whole tissue sample; View1, 2, 3, 4, and Whole) using average (AVG) and standard deviation (STD). (Bottom) The AVG and STD of the epithelial nuclei features are presented and compared between a benign (grey rectangle) and cancer (black rectangle) tissue sample. Color bar represents the intensity of feature values.

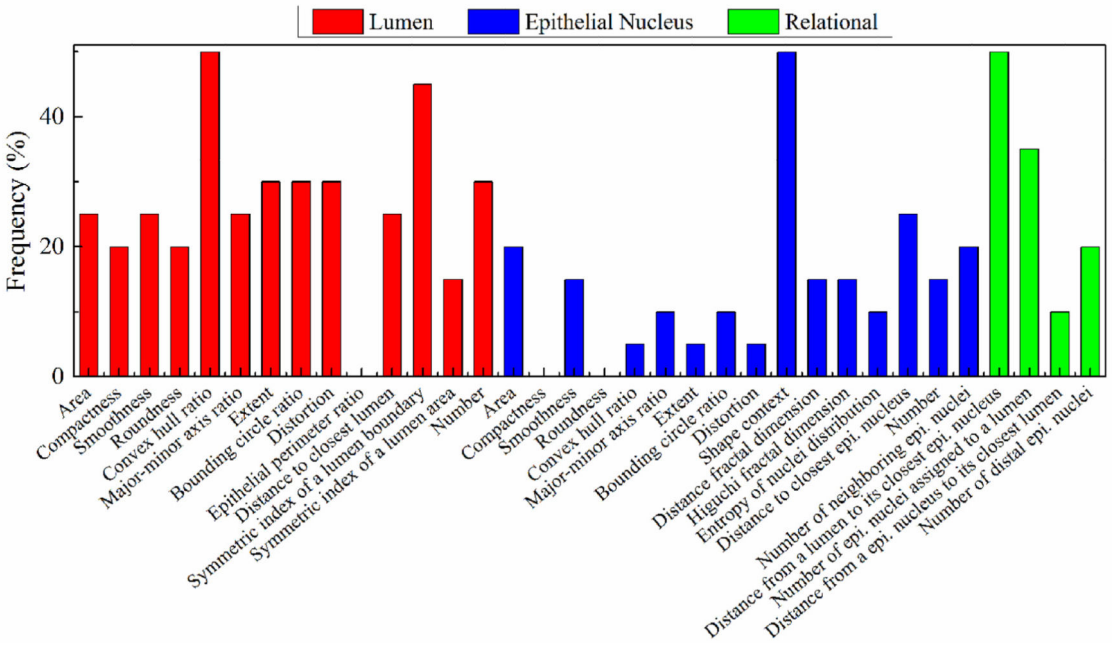

Fig. 6. Importance of morphological features.

The frequency of the morphological features of lumens and epithelial nuclei are shown. The frequency represents the fraction of boosting iterations that used each of the morphological features.

Table 1

Details of datasets.

TMA	TMA1	TMA2	TMA3	TMA4	TMA5
Benign, (n_s)	70	74	70	79	72
Cancer, (n_s)	135	89	115	82	86
Total, (n_s)	205	163	185	161	158
Gleason Grade	TMA1	TMA2	TMA3	TMA4	TMA5
3 + 3	18	0	0	0	0
3 + 4	48	49	36	33	39
4 + 3	20	20	30	18	26
4 + 4	35	10	35	16	13
4 + 5	14	8	14	10	5
5 + 4	0	2	0	5	3
Tissue Type	Lumen	Epithelium	Stroma	Nucleus	
Training, (n_p)	56,851	212,104	279,811	72,827	
Testing, (n_p)	63,144	219.235	275,893	83,291	

n_S and n_P denote the number of tissue sample cores and the number of pixels, respectively.

Table 2

Results of cancer detection.

	AUC	95% CI	p-value
Multiview	0.98	0.97–0.99	–
View1	0.95	0.93–0.97	<1e-4
View2	0.96	0.94–0.97	<1e-4
View3	0.95	0.93–0.97	<1e-3
View4	0.95	0.94–0.97	<1e-3
View5	0.96	0.94–0.97	<1e-3
View1+ 2 + 3 + 4 + 5	0.96	0.95–0.98	<1e-2

AUC and CI represent area under ROC curve and confidence interval, respectively.

Table 3

Comparison of cancer detection results.

	AUC	95% CI	p-value
Multiview	0.98	0.97–0.99	–
SVM-View1	0.94	0.92–0.96	<1e-6
SVM-View2	0.95	0.93–0.96	<1e-5
SVM-View3	0.95	0.93–0.97	<1e-5
SVM-View4	0.95	0.93–0.96	<1e-4
SVM-View5	0.97	0.96–0.98	<0.05
SVM-View1 + 2 + 3 + 4 + 5	0.96	0.95–0.97	<1e-3
RF-View1	0.96	0.94–0.98	<1e-2
RF-View2	0.95	0.93–0.97	<1e-3
RF-View3	0.95	0.94–0.97	<1e-4
RF-View4	0.96	0.95–0.97	<1e-2
RF-View5	0.96	0.95–0.98	<1e-2
RF-View1 + 2 + 3 + 4 + 5	0.96	0.94–0.98	<1e-3

SVM and RF denote support vector machine and random forest, respectively. AUC and CI represent area under ROC curve and confidence interval, respectively.

Table 4

Evaluation of computational performance.

Multiview boosting cancer detection procedure		Time (seconds)
Tissue	Feature Extraction: Intensity and Texture	1447
Segmentation	Segmentation	1400
Cancer	Feature Extraction: Morphology	296
Detection	Classification: Cancer vs. Benign	<1