IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 10, 2017, accepted August 20, 2017, date of publication August 31, 2017, date of current version September 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2747838

Nuclear Architecture Analysis of Prostate Cancer via Convolutional Neural Networks

JIN TAE KWAK[1], (Member, IEEE), AND STEPHEN M. HEWITT[2]

[1]Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea
[2]Experimental Pathology Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

Corresponding author: Jin Tae Kwak (jkwak@sejong.ac.kr)

This work was supported by the National Research Foundation of Korea through the Korea Government (MSIP) under Grant 2016R1C1B2012433.

ABSTRACT In this paper, we present an approach of convolutional neural networks (CNNs) to identify prostate cancers. Prostate tissue specimen samples were obtained from the tissue microarrays and digitized. For each sample, epithelial nuclear seeds were identified and used to generate a nuclear seed map, i.e., only the location information of epithelial nuclei was utilized. From the nuclear seed maps, CNNs sought to learn the high-level feature representation of nuclear architecture and to detect cancers. Applying data augmentation technique, CNNs were trained on the training data set including 73 benign and 89 cancer samples and validated on the testing data set comprising 217 benign and 274 cancer samples. In detecting cancers, CNNs achieved an AUC of 0.974 (95% CI: 0.961–0.985). In comparison with the approaches of utilizing hand-crafted nuclear architecture features and the state of the art deep learning networks with standard machine learning methods, CNNs were significantly superior to them (p-value $<$ 5e-2). Moreover, stromal nuclei were incapable of improving the cancer detection performance. The experimental results suggest that our approach offers the ability to aid in improving prostate cancer pathology.

INDEX TERMS Computer-aided diagnosis, microscopy, artificial neural network, pattern recognition, cancer detection.

I. INTRODUCTION

Mannually-conducted histologic assessment of a biopsied tissue specimen by a pathologist forms the definitive diagnosis of prostate cancer today. The identified cancer is assigned a histological grade between 1 (relatively benign) and 5 (highly aggressive) according to the Gleason grading system [1]. The Gleason grade serves as the basis for cancer management and treatment. However, the current practice of prostate pathology is limited in several ways. Every year $>$1 million biopsies are expected in the US [2], and each biopsy produces 10$\sim$14 tissue specimen samples. This places a large demand for high-throughput pathology services, but the manual process impedes the speed. Such heavy workload increases pathologists' fatigue and restricts the time allocated per case, likely decreasing the quality of the service. The Gleason grading also suffers from inter- and intra-observer variations [3]–[5], questioning the accuracy and reliability of prostate pathology. Moreover, the majority of the biopsied tissue specimens are, in fact, negative for cancer, demonstrating the inefficiency of the current prostate pathology. Therefore, automated, fast, and objective tools for accurate, robust, and reproducible cancer detection could aid in resolving the current issues with prostate pathology and improving the diagnostic performance.

Digital pathology systems have been proposed to improve the current practice of cancer pathology [6]. Tissue specimens are digitized at a high resolution and processed to analyze and interpret tissue characteristics. It often involves computerized image processing, pattern recognition, and machine learning techniques. The typical procedure of digital pathology systems is as follows: 1) Obtain digitized tissue specimen images 2) Segment several cellular/sub-cellular components (e.g., epithelium, stroma, nucleus, and etc.) due to their biological, chemical, and functional differences [7] 3) Quantify tissue characteristics (appearance and microstructures) using the segmented cellular/sub-cellular components; for example, disruption of glandular shape and spatial distribution and arrangement of cellular/sub-cellular components 4) Decision making via a machine learning framework. Several methods have been developed for analyzing differing types of tissues and disease [8], [9] such as prostate cancer [10], breast cancer [11], [12], pancreatic cancer [13], colorectal cancer [14], [15], and brain tumor [16], [17]. These methods have often adopted, so called, hand-crafted image features

to quantify tissue characteristics; for instance, color [18], morphology [19], [20], and texture, including wavelet transform [21], gray level co-occurrence matrix [22], fractal analysis [23], local binary pattern [24], and graph theory [25]. Utilizing such features, the tissue or disease status is determined by machine learning algorithms such as decision tree [26], k-nearest neighbor [21], [23], Bayesian [27], and support vector machine [18], [19]. Multi-scale (or -resolution) approaches have been also applied to integrate the information from differing scales [17], [27], [28].

Recently, deep learning has gained much attention for its success in image recognition [29]. Deep learning can be described as a structural/hierarchical learning algorithm, based upon multiple layers of (nonlinear) processing units (e.g., artificial neural networks), that is able to learn high-level features/representations of the observational (raw) data. Several types of deep learning approaches have been proposed, including convolutional neural networks (CNNs), stacked denoising autoencoders, and deep belief networks. Contrary to the previous approaches of utilizing hand-crafted image features (e.g., color, morphology, and texture), deep learning approaches sought to learn image features directly from a vast amount of (raw) datasets. With the advent of a high-quality and large-quantity of digitized image datasets, deep learning has also been applied to medical domains such as prostate cancer [30], breast cancer, [30]–[33] lung cancer [34], and brain cancer [34], [35]. Specifically, CNNs were used for detecting prostate cancer, mitosis [31] and invasive ductal carcinoma in breast cancer [32], and metastases in lymph nodes [30] as well as segmenting neuronal membranes [36] and epithelium and stroma of breast [33]. CNNs-driven image features were also combined with hand-crafted features for detecting mitosis [37]. Stacked denoising autoencoders were utilized for detecting and segmenting cells in lung and brain cancer [34].

Herein, we propose an approach of CNNs for detecting prostate cancers. Prostate tissue specimens are digitized and processed to segment nuclei within tissue. Utilizing the segmentation results, CNNs learn the high-level representation of the complex nuclear architecture and determines the disease status. Employing a large set of digitized tissue specimen images from tissue microarrays (TMAs), the performance of the proposed approach is systematically evaluated. The rest of this paper is organized as follows. In Methodology section, the details of the nuclear detection and CNNs are described. In Experiments section, the dataset and experimental setup are presented. In Results section, the experimental results are demonstrated. In Discussions section, the implications and limitations of this study are discussed. Finally, we conclude in Conclusions section.

II. METHODOLOGY

A. NUCLEAR SEED DETECTION

For each tissue specimen image (in RGB; red, green, and blue channels), color deconvolution is performed to separate hematoxylin and eosin stains. By Lambert-Beer's law, the optical density (OD) of a sample can be written as $-\log \frac{I_1}{I_0} = \alpha \cdot c$ where I_0 is the intensity of light entering the sample, I_1 is the intensity of light through the sample, α is the absorption coefficient, and c is the concentration of a stain. Let D be the OD of red, green, and blue channels and C be the amount of stains (hematoxylin, eosin, and background). Define a 3×3 unmixing matrix M where rows represent red, green, and blue channels and columns represent hematoxylin, eosin, and background, then $D = MC$. The amount of each stain in a sample, thus, can be attained by $C = M^{-1}D$. The details of this process are described in [38].

Applying Otsu's thresholding to the hematoxylin stain image, any connected components (or a group of pixels) are designated as an initial nucleus, and its size and shape are examined: If the size of a nucleus is smaller than $5\mu m^2$ or the ratio of the major and minor axis is greater than 5 when its size is smaller than $25\mu m^2$, then the nucleus is considered to be an artifact. Applying the Euclidean distance transformation and a watershed algorithm, each individual nucleus is identified. The surrounding pixels of nuclei are examined to discriminate epithelial nuclei from stromal nuclei (>50% of the perimeter). We adopt a multiview boosting approach to identify epithelial cells. The detailed description of the approach is available in [39]. Considering only the epithelial nuclei, the local maxima from the Euclidean distance transformation are denoted as seeds (or centroids) of individual nuclei. The procedure is described in Figure 1.

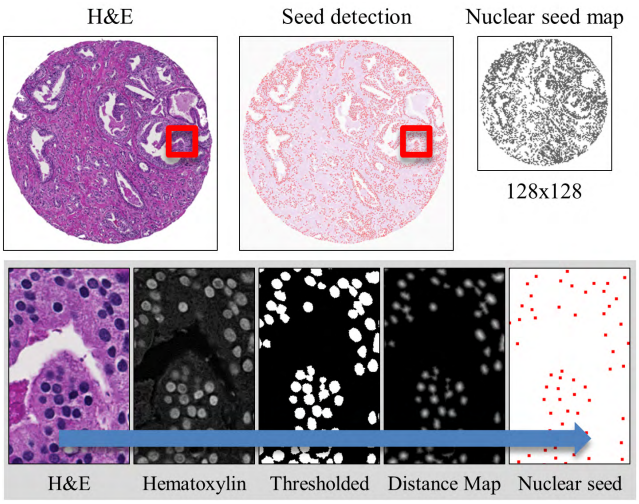

FIGURE 1. Nuclear seed map generation. (top) For an H&E tissue image, nuclear seed are detected and a nuclear seed map is generated. (bottom) Hematoxylin stains are extracted from an H&E image and thresholded to compute a distance map. Local maxima are designated as nuclear seeds.

B. DATASET CONSTRUCTION AND AUGMENTATION

For each tissue specimen image, a 2-dimensional spatial transformation, which rescales the original image to a fixed size of 128×128, is calculated. Using the spatial transformation, the corresponding locations of the identified nuclear seeds (in Section 2.2) are computed and marked in

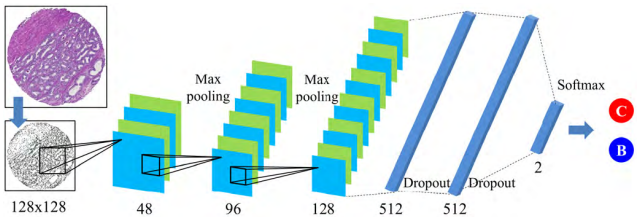

FIGURE 2. Structure of CNNs. Nuclear seed maps are fed into 6 layers of artificial neurons: three convolutional layers, two full-connected layers, and one softmax layer.

a 128 × 128 blank square patch, yielding a nuclear seed map (Figure 1).

We enlarge the dataset by performing label-preserving transformations, mimicking plausible physical variations of tissue specimen images, i.e., nuclear seed maps. A nuclear seed map is randomly transformed using a scale factor $s \in [0.5, 1.2]$, a rotation angle $\theta \in [0°, 180°]$, and horizontal flipping with probability 0.5. If up-scaled, the central 128 × 128 patch from the transformed image is cropped. If down-scaled, the edges of the transformed image are zero-padded to yield a 128 × 128 patch. Repeating this procedure, a hundred variants of each nuclear seed map is produced. This data augmentation is known to be helpful in reducing overfitting [29].

C. CONVOLUTIONAL NEURAL NETWORKS (CNNs)

Convolutional Neural Networks (CNNs) employ 6 layers of artificial neurons: three convolutional layers, two fully-connected layers, and a softmax layer for the classification (Figure 2). The rectified linear unit (ReLU) is adopted as a neuron activation function. The convolutional layer convolves the input image with a filter kernel of size $w \times w$. Each pixel in the resultant image contains the sum of products of the filter kernel and the local corresponding region ($w \times w$) of the input image. The input image from the preceding layer is convoluted with the filter kernels with a stride of 2 pixels. Local response-normalization is performed for improving generalizability [29]:

$$y_{ijk} = \frac{x_{ijk}}{\left(k + \alpha \sum_{l \in U} x_{ijl}^2\right)^{\beta}} \quad (1)$$

where x_{ijk} is the filter response at position (i, j), y_{ijk} is the normalized filter response, and U is the neighboring kernels. We set the parameters $\alpha = 10^{-4}$, $\beta = 0.75$, $k = 2$. Max-pooling is also used for summarizing the filter response of neighboring pixels (i.e., a max operation). It attempts to attain spatial-invariant features. It also reduces the spatial size of the image, leading to the reduction of the amount of parameters and computational complexity.

In the fully-connected layer, the neurons are connected to all neurons in the preceding layer. In order to avoid neuron co-adaptations or overfitting, the "dropout" technique [40] which drops (hidden) neurons with probability 0.5, is applied to the resultant image of the fully-connected layers.

In the softmax layer, the probability that the input image belongs to a class c is computed as

$$p(y = c|x) = \frac{exp(y_c(x))}{\sum_{k \in \sin\{cancer, benign\}} exp(y_k(x))} \quad (2)$$

where x is the input, y is the predicted class label and $y_c(\cdot)$ is the predicted score that the input belongs to a class c.

The first convolutional layer contains 48 kernels of size 13 × 13, the second convolutional layer comprises 96 kernels of size 5 × 5 × 48, and the third convolutional layer consists of 128 kernels of size 3 × 3 × 96. For the first two convolutional layers, local response-normalization is applied. Max-pooling with a window of size 3 × 3 and a stride of 2 pixels is applied for all three convolutional layers. The two fully-connected layers contain 512 neurons and are followed by dropout. The last 2-way softmax layer identifies cancer tissue specimen images.

III. EXPERIMENTS

A. DATASET

Four tissue microarrays (TMAs) were obtained from Tissue microarray research program at the National Institutes of Health. Tissue specimen sample cores were stained with hematoxylin and eosin (H&E) and digitized using a whole slide scanner (Leica Biosystems) at 40x magnification (a spatial resolution of 0.228 μm × 0.228 μm). Each tissue specimen sample core was evaluated by an experienced pathologist (S.M.H) to determine its disease status according to the Gleason grading system. These TMAs contain 162 (73 Benign and 89 Cancer; TMA A), 185 (70 Benign and 115 Cancer; TMA B), 149 (76 Benign and 73 Cancer; TMA C), and 157 (71 Benign and 86 Cancer; TMA D) tissue specimen sample cores, respectively. CNNs were trained on TMA A ("training dataset") and validated on the remaining three TMAs (TMA B, TMA C, and TMA D; "testing dataset").

B. EVALUATION OF CNNs APPROACH

We assess the ability of CNNs to learn the high-level nuclear architectural features and to detect prostate cancers. Nuclear seed maps are generated and fed into CNNs (Figure 2). CNNs are trained on the training dataset (TMA A) and tested on the validation dataset (comprising TMA B, TMA C, and TMA D). The correct and incorrect predictions in the test dataset are summarized into a receiver operating characteristic (ROC) plot. The area under ROC curve (AUC) and a 95% confidence interval (CI) are computed with the trapezoidal rule. Boot-strap resampling with 2000 repetitions is adopted to assess 95% CI of AUCs and statistical significance of the differences between AUCs of the two ROC curves [41].

C. COMPARISON WITH CONVENTIONAL APPROACHES

We compare the performance of CNNs with the conventional approaches of utilizing hand-crafted features. To quantify the nuclear architectural characteristics, several hand-crafted

features are adopted from the previous literature [19], [26]. The nuclear architectural features are:

1) Area: A number of pixels within a nucleus. The total sum, average, standard deviation, minimum, and maximum are computed.
2) Nuclear Entropy: Dividing a tissue image I into 50μm $\times$ 50μm disjoint partitions, the nuclear entropy is computed as

$$\mathrm{H}(X) = -\sum_{i=1}^{n}\sum_{j=1}^{n} p\left(x_{ij}\right)\log p\left(x_{ij}\right) \quad (3)$$

where $p(\cdot)$ is the probability mass function of the number of nuclei in a partition and $x_{ij} \in X$ is the number of nuclei in (i, j) partition.
3) Voronoi Diagram: Given nuclear centroids Q, Voronoi diagram divides a tissue image I into a set of disjoint Voronoi cells. A non-centroid pixel $p \in I$ lies in the cell corresponding to a centroid $q_i \in Q$ if and only if $\|p - q_i\| = \|p - q_j\|$ where $\|\cdot\|$ is the Euclidean distance between two points. For each Voronoi cell, the area and perimeter are computed. Then, their average (μ), standard deviation (σ), disorder $(1 - 1/(1 + \mu/\sigma))$, and minimum to maximum ratio are computed.
4) Delaunay Triangulation: Delaunay triangulation is a triangulation of nuclear centroids Q such that no centroid is inside the circumcircle of any triangle. The area and perimeter of each triangle is computed, and the average, standard deviation, disorder, and minimum to maximum ratio of them are computed.
5) Minimum Spanning Tree: Minimum spanning tree of nuclear centroids Q is a spanning tree that connects all centroids with the minimal total weighting for the edges. Edge weights are computed as the distance between two centroids. The average, standard deviation, disorder, and minimum to maximum ratio of the edge weights are computed.
6) Nuclear Density: Drawing a circle of a radius r (r=10, 20, 30, 40 and 50 pixels) around each nuclear centroid, the number of neighboring nuclei is counted, and the average, standard deviation, and disorder of them are computed. Also, for each nuclear centroid, the distance to the nearest 3, 5, and 7 neighboring nuclei is computed. Their average, standard deviation, and disorder are computed.

In order to assess the performance of the above hand-crafted features for detecting cancers, we employ four standard machine learning algorithms: 1) a support vector machine (SVM) with a polynomial (SVM-POLY) and a radial basis (SVM-RBF) kernel function 2) random forests with random selection of $\sqrt{n}$ features (n: number of entire features) using 100 (RF-100) and 500 (RF-500) trees 3) k-nearest neighbor algorithm with k=5 (kNN-5) and k=10 (kNN-10) 4) Naïve Bayes. These algorithms are separately trained on the training dataset and validated on the testing dataset.

D. COMPARISON WITH DEEP LEARNING APPROACHES

We compare the performance of CNNs with the state of the art deep learning networks. Five previous (pre-trained) networks, shown to be effective in image recognition, are adopted: 1) Krizhevsky [29] 2) VGG-S [42] 3) VGG-VD [43] 4) GoogLeNet [44] 5) Residual Networks (ResNets) [45]. For each of these networks, the output of the preceding layer of the softmax layer is fed to the standard machine learning algorithms (used in the previous section). In other words, these networks are used as a feature extractor. For each tissue image, the original RGB image and nucleus map are reduced to a fixed size of 224 $\times$ 224 and separately provided with the networks, generating two sets of image features. Based on the extracted features, the above experiments are repeated.

E. EVALUATION OF EPITHELIAL AND STROMAL NUCLEI

We investigate the efficacy of distinguishing epithelial nuclei from stromal nuclei. Using the identical architecture of CNNs as well as training and testing dataset, we perform two experiments: 1) CNNs using stromal nuclei alone 2) CNNs using both epithelial and stromal nuclei. The overall procedure is identical, but the nuclear seed map generation is different; using stromal nuclei alone and both epithelial and stromal nuclei.

F. IMPLEMENTATION AND EXPERIMENTAL ENVIRONMENT

Nuclear seed detection and data augmentation were coded and executed on MATLAB R2016a. An open-source implementation of CNNs "MatConvNET" [46] was adopted and used for training and validation experiments. The entire experiment was performed on a PC with Intel(R) Core(TM) i7 3.4 GHz processors, 24GB of RAM, and NVIDIA GeForce GTX 750 Ti 2GB Graphical Processor Unit (GPU).

TABLE 1. Comparison with the conventional approaches.

Classification		AUC (95% CI)	p-value
CNNs		**0.974 (0.961-0.985)**	-
Hand-crafted	SVM-POLY	0.880 (0.850-0.909)	<1e-7
	SVM-RBF	0.884 (0.855-0.913)	<1e-6
	RF-100	0.894 (0.864-0.921)	<1e-6
	RF-500	**0.898 (0.869-0.924)**	<1e-5
	KNN-5	0.850 (0.816-0.882)	<1e-11
	KNN-10	0.860 (0.827-0.891)	<1e-10
	Naïve Bayes	0.841 (0.806-0.875)	<1e-11

IV. RESULTS

A. RESULTS FOR CANCER DETECTION

In detecting prostate cancers, our CNNs achieved an AUC of 0.974 (95% CI: 0.961-0.985) (Table 1). Figure 3 shows the process of our CNNs to determine the disease status of tissue sample cores. CNNs first revealed local and low-level features. In deeper layers, coarser but high-level features were obtained and utilized to identify cancers.

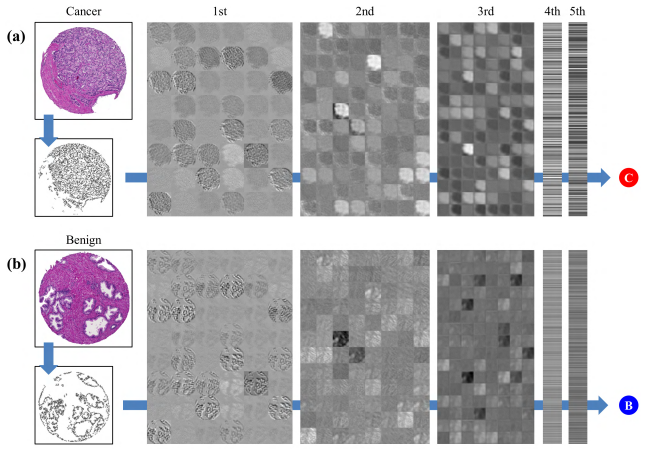

FIGURE 3. Illustration of cancer detection via CNNs. A nuclear seed map is generated for (a) cancer and (b) benign tissue sample cores and is fed into CNNs. The output of each convolutional layer is demonstrated.

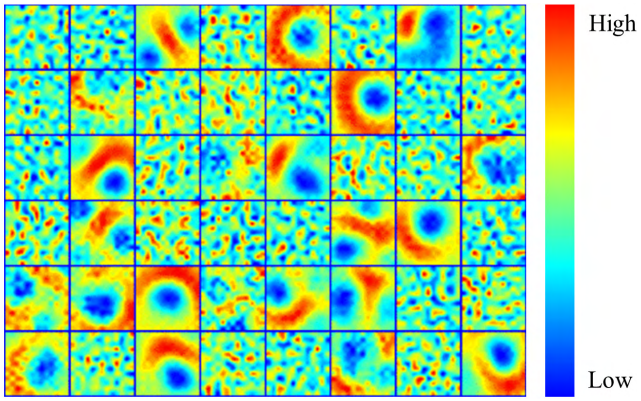

FIGURE 4. Illustration of the filter kernels of the first convolutional layer.

We examined the filter kernels of the first convolutional layer (Figure 4). This shows that CNNs have learned the low-level architectural characteristics of the prostate. In benign tissues, epithelial cells surround empty circular or elliptical regions (lumens), i.e., epithelial nuclei line lumens. Some of the filter kernels contain a straight or curved band of high weight elements. This could recognize the alignment of nuclei in benign prostate glands. In cancerous tissues, cells grow in and out of the gland and disrupt the overall shape and arrangement of prostate glands. This could be recognized by the filter kernels that contain small and scattered high-weight elements. Hence, the first convolutional layer utilizes the local arrangement of nuclei with or without lumens.

We also assessed mispredicted tissue sample cores (false positives and false negatives) by CNNs (Figure 5). The filter responses by false negatives (missed cancer cores) and false positives (missed benign cores) were similar to true negatives (benign cores) and true positives (cancer cores), respectively. In the output of the 2nd and 3rd convolutional layers, several filter kernels that have relatively lower responses for benign (Figure 3b and Figure 5a) and higher responses for cancer (Figure 3a and Figure 5b) were observed. These are likely to be responsive to high cellular regions in a tissue, i.e., recognizing the overall distribution of nuclei.

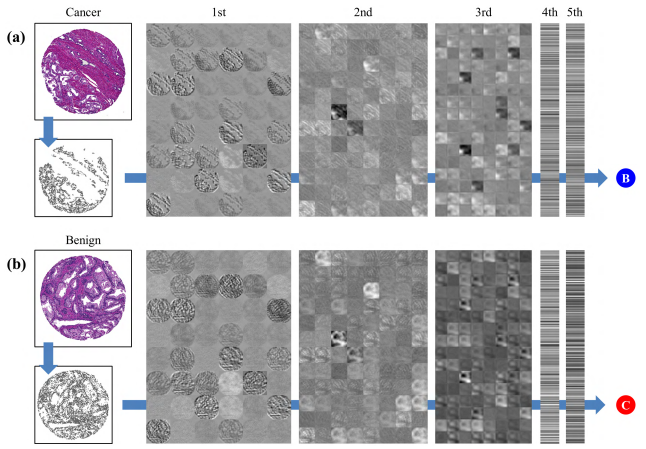

FIGURE 5. False prediction by CNNs. (a) Cancer and (b) benign tissue sample cores are classified into benign and cancer class, respectively.

B. RESULTS FOR COMPARISON EXPERIMENTS

Our CNNs outperformed the conventional cancer detection approaches (Table 1). Using hand-crafted nuclear architectural features, SVM, random forests, kNN, and Naïve bayes achieved $\leq$0.898 AUC. The performance of our CNNs (0.974 AUC) was significantly better than that of the standard machine learning algorithms (p-$value$<1e-5). In addition, stromal nuclei alone could not distinguish cancers from benign samples, obtaining an AUC of 0.555 (95% CI: 0.504-0.604). Using both epithelial and stromal nuclei, an AUC of 0.947 (95% IC: 0.927-0.964) was achieved; however, the performance was significantly lower than that of our CNNs using epithelial nuclei alone (p-$value$<1e-3).

In comparison to the state of the art deep learning approaches, our CNNs outperformed other approaches. Table 2 shows the performance of the five networks (Krizhevsky, VGG-S, VGG-VD, GooLeNet, and ResNets), combined with the standard machine learning algorithms. ResNets with RF-500 using RGB images performed the best among the previous deep learning approaches, obtaining an AUC of 0.956 (95% CI: 0.936-0.973). However, our CNNs significantly outperformed all the previous deep learning approaches (p-$value$<5e-2).

C. COMPUTATIONAL COMPLEXITY

The computational performance of our approach was assessed (Table 3). For a tissue sample image of approximately 5000 $\times$ 5000 pixels, the nuclear seed detection step takes 547s on average due to the pixel-by-pixel process of the epithelium classification (>90% of the nuclear seed detection step). Training CNNs with 100 epochs requires a substantial amount of time (9800s). In testing phase, however, it only requires a minimal amount of time (<1s), i.e., no computational overload compared to the standard machine learning methods.

V. DISCUSSIONS

We empirically chose the architecture of CNNs presented in this study. The architecture of CNNs plays a pivotal role in improving the performance. Finding the optimal

TABLE 2. Comparison with the other deep learning approaches.

Input Type	Deep Learning	Classification	AUC (95% CI)	p-value
Nucleus Map	Krizhevsky	SVM-POLY	0.883 (0.854-0.909)	<1e-8
		SVM-RBF	**0.937 (0.915-0.956)**	<1e-2
		RF-100	0.936 (0.915-0.955)	<1e-2
		RF-500	0.935 (0.914-0.955)	<1e-3
		KNN-5	0.897 (0.868-0.925)	<1e-6
		KNN-10	0.913 (0.886-0.937)	<1e-4
		Naïve Bayes	0.776 (0.742-0.814)	<1e-23
	VGG-S	SVM-POLY	0.877 (0.847-0.906)	<1e-8
		SVM-RBF	0.948 (0.927-0.966)	<3e-2
		RF-100	0.946 (0.926-0.964)	<2e-2
		RF-500	**0.949 (0.928-0.966)**	<5e-2
		KNN-5	0.936 (0.911-0.956)	<1e-2
		KNN-10	0.937 (0.916-0.958)	<1e-2
		Naïve Bayes	0.804 (0.768-0.838)	<1e-18
	VGG-VD	SVM-POLY	0.945 (0.926-0.962)	<1e-2
		SVM-RBF	**0.947 (0.925-0.965)**	<2e-2
		RF-100	0.945 (0.926-0.964)	<2e-2
		RF-500	0.941 (0.920-0.961)	<1e-2
		KNN-5	0.917 (0.892-0.940)	<1e-4
		KNN-10	0.929 (0.903-0.951)	<1e-3
		Naïve Bayes	0.777 (0.739-0.811)	<1e-23
	GooLeNet	SVM-POLY	0.923 (0.898-0.947)	<1e-3
		SVM-RBF	**0.944 (0.923-0.961)**	<1e-2
		RF-100	0.943 (0.921-0.960)	<1e-2
		RF-500	0.938 (0.917-0.957)	<1e-2
		KNN-5	0.885 (0.855-0.913)	<1e-7
		KNN-10	0.927 (0.902-0.949)	<1e-3
		Naïve Bayes	0.761 (0.726-0.798)	<1e-26
	ResNets	SVM-POLY	0.936 (0.912-0.957)	<1e-2
		SVM-RBF	0.945 (0.925-0.963)	<2e-2
		RF-100	0.948 (0.929-0.965)	<2e-2
		RF-500	**0.953 (0.934-0.968)**	<5e-2
		KNN-5	0.908 (0.880-0.933)	<1e-5
		KNN-10	0.931 (0.907-0.953)	<1e-2
		Naïve Bayes	0.886 (0.857-0.913)	<1e-7
RGB	Krizhevsky	SVM-POLY	0.891 (0.859-0.918)	<1e-6
		SVM-RBF	0.909 (0.881-0.933)	<1e-5
		RF-100	0.910 (0.883-0.934)	<1e-5
		RF-500	**0.914 (0.889-0.937)**	<1e-4
		KNN-5	0.871 (0.837-0.901)	<1e-9
		KNN-10	0.881 (0.850-0.910)	<1e-7
		Naïve Bayes	0.776 (0.740-0.812)	<1e-24
	VGG-S	SVM-POLY	0.923 (0.896-0.944)	<1e-3
		SVM-RBF	**0.939 (0.917-0.958)**	<1e-2
		RF-100	0.929 (0.903-0.950)	<1e-3
		RF-500	0.933 (0.911-0.954)	<1e-3
		KNN-5	0.900 (0.871-0.926)	<1e-5
		KNN-10	0.923 (0.898-0.946)	<1e-3
		Naïve Bayes	0.817 (0.782-0.850)	<1e-18
	VGG-VD	SVM-POLY	0.933 (0.912-0.953)	<1e-3
		SVM-RBF	**0.947 (0.927-0.962)**	<2e-2
		RF-100	0.938 (0.917-0.957)	<1e-2
		RF-500	0.939 (0.918-0.958)	<1e-2
		KNN-5	0.902 (0.871-0.927)	<1e-5
		KNN-10	0.913 (0.887-0.937)	<1e-4
		Naïve Bayes	0.798 (0.764-0.830)	<1e-22
	GooLeNet	SVM-POLY	0.886 (0.851-0.914)	<1e-6
		SVM-RBF	**0.928 (0.905-0.949)**	<1e-3
		RF-100	0.917 (0.892-0.941)	<1e-4
		RF-500	0.917 (0.892-0.941)	<1e-4
		KNN-5	0.887 (0.856-0.915)	<1e-6
		KNN-10	0.907 (0.880-0.932)	<1e-5
		Naïve Bayes	0.725 (0.685-0.764)	<1e-32
	ResNets	SVM-POLY	0.936 (0.914-0.957)	<1e-2
		SVM-RBF	0.954 (0.935-0.971)	<5e-2
		RF-100	0.948 (0.927-0.964)	<2e-2
		RF-500	**0.956 (0.936-0.973)**	<5e-2
		KNN-5	0.921 (0.898-0.941)	<1e-4
		KNN-10	0.936 (0.913-0.954)	<1e-2
		Naïve Bayes	0.895 (0.867-0.919)	<1e-6

TABLE 3. Computational performance.

CNNs cancer detection procedure		Time (seconds)
Nuclear seed map generation	Nuclear seed detection	43
	Epithelium detection	504
CNNs cancer detection	Training CNNs	9800
	Classification	<1

architecture (or parameters) and the theoretical justification, however, is still an ongoing research topic. Several strategies have been proposed; for example, hidden layer supervision [47], [48], deformable pooling layer [49], residual functions [45], stretching and symmetrical [50]. Differing training methodologies have been proposed [51]. It is infeasible to investigate all the strategies. Nevertheless, the addition of the proper strategies could help to improve the performance of CNNs in detecting and characterizing prostate cancers.

CNNs were only provided with the location information of epithelial nuclear centroids and outperformed the conventional hand-crafted features and standard machine learning methods. This demonstrates that CNNs could offer the ability to better characterize the nuclear architecture. However, the location information of stromal nuclear centroids alone was incompetent to detect cancers. The arrangement and distribution of stromal nuclei may not provide sufficient information that CNNs could utilize to determine the disease status. Including both epithelial nuclei and stromal nuclei, we were not able to improve the performance of CNNs. This suggests that the epithelium detection step that distinguishes epithelial nuclei from stromal nuclei is essential to reveal the high-level feature representation of prostate cancers as well as to attain the improved cancer detection performance.

Whereas our approach is based on the location information of nuclei, many other deep learning networks as well as hand-crafted feature extraction methods, in general, exploit the whole histology objects (e.g., nuclei) or images. The five state of the art deep learning networks were adopted and used to detect cancers. Although their performance was inferior to our CNNs, they consistently outperformed the conventional hand-crafted features as combined with SVM-RBF, random forests, and kNN. This supports the idea of utilizing deep learning approach for characterizing tissues and identifying cancers. Fine tuning the previous networks on tissue images may improve the overall performance in detecting cancers; however, this is beyond the scope of our study. It is noticeable that the best performing combination mostly uses nucleus maps, not RGB images. Provided with nucleus maps, the networks may have been directed to focus on nuclear characteristics. The effect of the color space of input data (or images) on deep learning has been noted and discussed [52]. Our results suggest that the selection of the suitable form of input data is another crucial factor to the performance of CNNs.

Conventional methods of utilizing color decomposition, thresholding, watershed algorithm, and boosting were adopted to identify nuclei and segment epithelium. CNNs have also shown to be effective in detecting nuclei and cells [33], [53]–[56]. This may offer the potential for

improving nuclei detection; in particular, resolving densely clustered nuclei or inhomogeneous stains. The improvement in detecting nuclei and cells will be beneficial to our CNNs. Further study will be conducted to develop the full CNNs approach for identifying nuclei and cells as well as detecting cancers.

We employed data augmentation to enlarge the dataset and to avoid overfitting. The strength of deep learning lies in the ability to explore a huge amount of data and to extract relevant features. Despite of the relatively small size of the dataset, our CNNs were capable of detecting cancers with high accuracy, i.e., the size of the dataset was not a limiting factor in this study. Nonetheless, an extended validation study could further ensure the utility and reliability of our approach.

VI. CONCLUSIONS

We have presented an approach of CNNs for identifying prostate cancers. CNNs were only fed with the location information of nuclear centroids and able to learn complex and complicated nuclear architecture characteristics. Our approach was significantly superior to the conventional hand-crafted features and standard machine learning methods. The experimental results are promising to justify further investigation of our approach for other applications. We will apply our approach to other types of cancers and disease.

REFERENCES

[1] D. F. Gleason, "Classification of prostatic carcinomas," *Cancer Chemotherapy Rep.*, vol. 1, no. 50, pp. 125–128, Mar. 1966.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA, A Cancer J. Clin.*, vol. 66, no. 1, pp. 30–37, 2016.

[3] M. L. Cintra and A. Billis, "Histologic grading of prostatic adenocarcinoma: Intraobserver reproducibility of the Mostofi, Gleason and Böcking grading systems," *Int. Urol. Nephrol.*, vol. 23, no. 5, pp. 449–454, 1991.

[4] S. O. Özdamar, S. Sarikaya, L. Yildiz, M. K. Atilla, B. Kandemir, and S. Yildiz, "Intraobserver and interobserver reproducibility of WHO and Gleason histologic grading systems in prostatic adenocarcinomas," *Int. Urol. Nephrol.*, vol. 28, no. 1, pp. 73–77, 1996.

[5] L. Egevad *et al.*, "Standardization of Gleason grading among 337 European pathologists," *Histopathology*, vol. 62, no. 2, pp. 247–256, Jan. 2013.

[6] L. Pantanowitz, "Digital images and the future of digital pathology," *J. Pathol. Inform.*, vol. 1, p. 15, Jun. 2010.

[7] J. T. Kwak, R. Reddy, S. Sinha, and R. Bhargava, "Analysis of variance in spectroscopic imaging data from human tissues," *Anal. Chem.*, vol. 84, pp. 1063–1069, Jan. 2012.

[8] L. Mulrane, E. Rexhepaj, S. Penney, J. J. Callanan, and W. M. Gallagher, "Automated image analysis in histopathology: A valuable tool in medical diagnostics," *Expert Rev. Mol. Diagnostics*, vol. 8, pp. 707–725, Nov. 2008.

[9] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, no. 7, pp. 71–147, Jul. 2009.

[10] J. T. Kwak, S. M. Hewitt, A. A. Kajdacsy-Balla, S. Sinha, and R. Bhargava, "Automated prostate tissue referencing for cancer detection and diagnosis," *BMC Bioinform.*, vol. 17, p. 227, May 2016.

[11] A. H. Beck *et al.*, "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival," *Sci. Translational Med.*, vol. 3, pp. 108–113, Nov. 2011.

[12] J. Xu *et al.*, "Sparse non-negative matrix factorization (SNMF) based color unmixing for breast histopathological image analysis," *Comput. Med. Imag. Graph.*, vol. 46, pp. 20–29, Dec. 2015.

[13] L. Langer, Y. Binenbaum, L. Gugel, M. Amit, Z. Gil, and S. Dekel, "Computer-aided diagnostics in digital pathology: Automated evaluation of early-phase pancreatic cancer in mice," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, no. 7, pp. 1043–1054, 2015.

[14] H. Kalkan, M. Nap, R. P. Duin, and M. Loog, "Automated colorectal cancer diagnosis for whole-slice histopathology," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 15. 2012, pp. 550–557.

[15] A. I. Wright, D. Magee, P. Quirke, and D. Treanor, "Incorporating local and global context for better automated analysis of colorectal cancer on digital pathology slides," *Procedia Comput. Sci.*, vol. 90, pp. 125–131, Dec. 2016.

[16] J. Kong, O. Sertel, H. Shimada, K. L. Boyer, J. H. Saltz, and M. N. Gurcan, "Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation," *Pattern Recognit.*, vol. 42, no. 6, pp. 1080–1092, Jun. 2009.

[17] J. Barker, A. Hoogi, A. Depeursinge, and D. L. Rubin, "Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles," *Med. Image Anal.*, vol. 30, pp. 60–71, May 2016.

[18] A. Tabesh *et al.*, "Multifeature prostate cancer diagnosis and Gleason grading of histological images," *IEEE Trans. Med. Imag.*, vol. 26, no. 10, pp. 1366–1378, Oct. 2007.

[19] J. T. Kwak, S. M. Hewitt, S. Sinha, and R. Bhargava, "Multimodal microscopy for automated histologic analysis of prostate cancer," *BMC Cancer*, vol. 11, p. 62, Feb. 2011.

[20] L. Gorelick *et al.*, "Prostate histopathology: Learning tissue component histograms for cancer detection and classification," *IEEE Trans. Med. Imag.*, vol. 32, no. 10, pp. 1804–1818, Oct. 2013.

[21] K. Jafari-Khouzani and H. Soltanian-Zadeh, "Multiwavelet grading of pathological images of prostate," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 697–704, Jun. 2003.

[22] M. DiFranco, G. O'Hurley, E. Kay, W. Watson, and P. Cunningham, "Automated gleason scoring of prostatic histopathology slides using multi-channel co-occurrence texture features," in *Proc. Int. Workshop Microscopic Image Anal. Appl. Biol. (MIAAB)*, 2008.

[23] P.-W. Huang and C.-H. Lee, "Automatic classification for pathological prostate images based on fractal analysis," *IEEE Trans. Med. Imag.*, vol. 28, no. 7, pp. 1037–1050, Jul. 2009.

[24] A. Gertych *et al.*, "Machine learning approaches to analyze histological images of tissues from radical prostatectomies," *Comput. Med. Imag. Graph.*, vol. 46, pp. 197–208, Dec. 2015.

[25] K. Nguyen, A. Sarkar, and A. K. Jain, "Prostate cancer grading: Use of graph cut and spatial arrangement of nuclei," *IEEE Trans. Med. Imag.*, vol. 33, no. 12, pp. 2254–2270, Dec. 2014.

[26] S. Doyle, M. D. Feldman, N. Shih, J. Tomaszewski, and A. Madabhushi, "Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer," *BMC Bioinform.*, vol. 13, p. 282, Sep. 2012.

[27] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, "A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1205–1218, May 2012.

[28] J. T. Kwak and S. M. Hewitt, "Multiview boosting digital pathology analysis of prostate cancer," *Comput. Methods Programs Biomed.*, vol. 142, pp. 91–99, Apr. 2017.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[30] G. Litjens *et al.*, "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," *Sci. Rep.*, vol. 23, no. 6, p. 26286, May 2016.

[31] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 16. 2013, pp. 411–418.

[32] A. Cruz-Roa *et al.*, "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," *Proc. SPIE*, vol. 9041, p. 904103, Mar. 2014.

[33] J. Xu, X. F. Luo, G. H. Wang, H. Gilmore, and A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214–223, May 2016.

[34] H. Su, F. Xing, X. Kong, Y. Xie, S. Zhang, and L. Yang, "Robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 383–390.
[35] M. G. Ertosun and D. L. Rubin, "Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks," in *Proc. AMIA Annu. Symp.*, 2015, pp. 1899–1908.
[36] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 2843–2851.
[37] H. Wang *et al.*, "Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features," *J. Med. Imag.*, vol. 1, no. 3, p. 034003, Oct. 2014.
[38] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol.*, vol. 23, pp. 291–299, Aug. 2001.
[39] J. T. Kwak *et al.*, "Correlation of magnetic resonance imaging with digital histopathology in prostate," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, pp. 657–666, Apr. 2016.
[40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[41] X. Robin *et al.*, "pROC: An open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinform.*, vol. 12, p. 77, Mar. 2011.
[42] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. (May 2014). "Return of the devil in the details: Delving deep into convolutional nets." [Online]. Available: https://arxiv.org/abs/1405.3531
[43] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[44] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
[45] K. He, X. Zhang, S. Ren, and J. Sun. (Dec. 2015). "Deep residual learning for image recognition." [Online]. Available: https://arxiv.org/abs/1512.03385
[46] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 689–692.
[47] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Proc. AISTATS*, 2015, p. 6.
[48] L. Wang, C.-Y. Lee, Z. Tu, and S. Lazebnik. (May 2015). "Training deeper convolutional networks with deep supervision." [Online]. Available: https://arxiv.org/abs/1505.02496
[49] W. Ouyang *et al.* (Sep. 2014). "DeepID-Net: Multi-stage and deformable deep convolutional neural networks for object detection." [Online]. Available: https://arxiv.org/abs/1409.3505
[50] S. Shankar, D. Robertson, Y. Ioannou, A. Criminisi, and R. Cipolla. (Apr. 2016). "Refining architectures of deep convolutional neural networks." [Online]. Available: https://arxiv.org/abs/1604.06832
[51] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks," *J. Mach. Learn. Res.*, vol. 10, no. 1, pp. 1–40, Jan. 2009.
[52] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.
[53] F. Liu and L. Yang, "A novel cell detection method using deep convolutional neural network and maximum-weight independent set," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 349–357.
[54] Y. Xie, X. Kong, F. Xing, F. Liu, H. Su, and L. Yang, "Deep voting: A robust approach toward nucleus localization in microscopy images," *Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, pp. 374–382, Oct. 2015.
[55] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 550–566, Feb. 2016.
[56] X. Pan *et al.*, "Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks," *Neurocomputing*, vol. 229, pp. 88–99, Mar. 2017.

JIN TAE KWAK received the B.E. degree in electrical engineering from Korea University, Korea, in 2005, the M.Sc. degree in electrical and computer engineering from Purdue University, USA, in 2007, and the Ph.D. degree in computer science from the University of Illinois at Urbana–Champaign, USA, in 2012. From 2012 to 2016, he was a Research Fellow with the Center for Interventional Oncology, Clinical Center, National Institutes of Health, USA. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Sejong University, South Korea. His research interests include multimodal medical imaging, digital pathology, computer-aided diagnosis, radiology-pathology correlation, data mining, and machine learning. His research has been funded by the National Research Foundation of Korea.

STEPHEN M. HEWITT received the bachelor's degree in philosophy from Johns Hopkins University, the Ph.D. degree in genetics from the Graduate School of Biomedical Sciences, The University of Texas, having completed his thesis with the Laboratory of Grady Saunders, University of Texas M. D. Anderson Cancer Center, the M.D. degree from the University of Texas Medical School, Houston, and the Residents in anatomic pathology from the Laboratory of Pathology, National Cancer Institute. He has served as the Head of the Experimental Pathology Laboratory within the National Cancer Institute, National Institutes of Health. He led the Tissue Array Research Program from 2000 to 2014, and the Applied Molecular Pathology Laboratory from 2008 to 2014. He has served as a Consultant to the Hematology and Pathology Devices Panel, Center for Devices and Radiological Health, Food, and Drug Administration. He also has co-authored over 220 articles and served on the editorial board of several peer-reviewed journals. His research interests are in the development of tissue-based biomarkers for cancer diagnosis, prognosis, and prediction of response to therapy. Specific areas of concentration include tissue microarrays, biobanking, tissue proteomics, whole-slide imaging and image analysis, and cancers of the aerodigestive and urogenitial tracts. He is an active member of the Histochemical Society, the Association for Pathology Informatics, and the American Society for Investigative Pathology. He is a board-certified Anatomic Pathologist and has served as a member of the Clinical Laboratory Standards Institute Immunology and Ligand Assay Consensus Committee, and also serves as the Co-Chair of the Subcommittee on Immunohistochemical Assays.