

# Convolutional Neural Networks for Prostate Cancer Recurrence Prediction

Neeraj Kumar<sup>a</sup>, Ruchika Verma<sup>a</sup>, Ashish Arora<sup>a</sup>, Abhay Kumar<sup>a</sup>, Sanchit Gupta<sup>a</sup>, Amit Sethi<sup>a</sup>, and Peter H. Gann<sup>b</sup>

<sup>a</sup>Indian Institute of Technology Guwahati, Guwahati-781039, Assam, India

<sup>b</sup>University of Illinois at Chicago, Chicago-60607, Illinois, USA

## ABSTRACT

Accurate prediction of the treatment outcome is important for cancer treatment planning. We present an approach to predict prostate cancer (PCa) recurrence after radical prostatectomy using tissue images. We used a cohort whose case vs. control (recurrent vs. non-recurrent) status had been determined using post-treatment follow up. Further, to aid the development of novel biomarkers of PCa recurrence, cases and controls were paired based on matching of other predictive clinical variables such as Gleason grade, stage, age, and race. For this cohort, tissue resection microarray with up to four cores per patient was available. The proposed approach is based on deep learning, and its novelty lies in the use of two separate convolutional neural networks (CNNs) – one to detect individual nuclei even in the crowded areas, and the other to classify them. To detect nuclear centers in an image, the first CNN predicts distance transform of the underlying (but unknown) multi-nuclear map from the input H&E image. The second CNN classifies the patches centered at nuclear centers into those belonging to cases or controls. Voting across patches extracted from image(s) of a patient yields the probability of recurrence for the patient. The proposed approach gave 0.81 AUC for a sample of 30 recurrent cases and 30 non-recurrent controls, after being trained on an independent set of 80 case-controls pairs. If validated further, such an approach might help in choosing between a combination of treatment options such as active surveillance, radical prostatectomy, radiation, and hormone therapy. It can also generalize to the prediction of treatment outcomes in other cancers.

**Keywords:** Prostate cancer recurrence, deep learning, convolutional neural networks, distance transform, annotations

## 1. INTRODUCTION

Prostate cancer (PCa) is one of the leading causes of death among American men. In 2017, 161,360 new cases and 26,730 deaths are projected to occur due to PCa.<sup>1</sup> Prostate specific antigen (PSA) screening and the use of pathological and demographic variables such as the clinical stage, Gleason score, age, race, etc. remains the current gold standard in PCa treatment planning and decision making. PCa recurrence is a prominent cause of concern among the men undergoing curative treatment, especially radical prostatectomy (RP). The need for precise PCa recurrence prediction stems from the fact that about 16% of the PCa patients who undergo radical prostatectomy experience biochemical recurrence, signified by a rising serum PSA level, within 5 years of the surgery.<sup>2</sup> This uncertainty poses a challenge for the treating physicians who want to minimize unnecessary treatments, and yet save lives. Guidelines developed to accurately predict PCa recurrence using demographic and clinical variables such as PSA levels, positive surgical margin, clinical grade, and Gleason score include risk-grouping,<sup>3</sup> look-up tables,<sup>4</sup> nomograms,<sup>5,6</sup> and risk scoring.<sup>7,8</sup> In current clinical research and patient care, the Kattan nomogram<sup>5,6</sup> and CAPRA-S score<sup>8</sup> are the most popular tools for predicting PCa recurrence after RP. Although the accuracy of these prediction models surpass manual decision making capabilities, their performance remains suboptimal on the most common prostate tumors that have intermediate grade and moderate PSA levels.<sup>8,9</sup> Even the use of additional clinical, genetic, and multi-modal imaging data has failed to boost recurrence

---

Author contact information: (Send correspondence to Neeraj Kumar)

Email:(neeraj.kumar.iitg,verma.ece.ruchika,ashish.arora.iitg,kabhay518,sanchit257)@gmail.com,amitsethi@iitg.ernet.in, pgann@uic.edu

prediction accuracy on such cases.<sup>10–16</sup> Therefore, high precision PCa recurrence tools are required to provide better patient care through effective treatment decision making.

Towards achieving this goal, we present an algorithm to predict PCa biochemical recurrence with high accuracy using H&E stained tissue images. Our algorithm automatically identifies morphometric cues characterizing PCa recurrence to estimate probability of recurrence of a tissue image. Although, our algorithm was trained and tested on tissue resections from prostatectomy samples, it can also be used on biopsies at the time of diagnosis. The main contributions of this paper are as follows:

1. Unlike previous attempts to predict PCa recurrence using only H&E stained tissue images, our approach doesn't rely on hand-crafted features. Instead, we use deep learning to automatically learn a hierarchy of features to differentiate recurrent from non-recurrent morphological patterns.
2. We propose a two-stage deep learning-based approach to classify H&E stained tissue images to predict PCa recurrence probability. This approach can be extended to predicting other fine-grained tissue classes such as cancer grades, types, and molecular sub-types of cancer of various organs for precision treatment planning. In the first stage, we use a convolutional neural network (CNN) to detect the nuclear centers with high accuracy in a given tissue image. The proposed nucleus detection algorithm was trained to detect both epithelial and stromal nuclei in tumor as well as non-tumor regions of tissue images.
3. The second stage uses another CNN which takes patches centered around the detected nuclear centers as input to estimate patch-wise cancer recurrence probability as its output. This stage can easily be modified and retrained to predict the probability of different sub-types of cancers in different types of tissue images. The final sub-type probability a given patient (probability of recurrence) is determined by the aggregating its patch-wise probabilities determined by the CNN.
4. We use color normalization as a pre-processing step to undo the effect of unwanted variations in tissue images due to staining and scanning differences across hospitals.

To train, validate, and test the proposed algorithm we used a cohort from the Cooperative Prostate Cancer Tissue Resource (CPCTR)<sup>17</sup> of PCa patients who underwent radical prostatectomy such that the cases had cancer recurrence, while the controls did not. All selected patients for this study had mid-grade and moderate PSA level prostate cancer. Further, to aid the discovery of novel PCa recurrence biomarkers, the cases and controls were matched on several variables including age, race, Gleason score, and pathological stage. We obtained the H&E stained tissue microarray (TMA) images of the selected case-control pairs.

To establish the background for the proposed algorithm we first review some of the previous approaches to predict PCa biochemical recurrence using only H&E stained tissue images in Section 2. Then, we describe the details of the proposed algorithm in Section 3, followed by some experimental results in Section 4. We conclude and identify future directions in Section 5.

## 2. RELATED WORK

There have been numerous attempts in the recent past to develop approaches based on quantitative image processing and machine learning to accurately predict PCa recurrence. Accurate prediction of recurrence at the time of initial diagnosis can allow medical practitioners to decide between treatment options such as radical prostatectomy, radiation therapy, chemotherapy, hormone therapy, and active surveillance for more indolent tumors. Accurate prediction of tumor response to treatment can improve long-term treatment outcomes and reduce risks due to unnecessary treatment. There is an acute need for developing image processing based recurrence prediction models because previous attempts to do so with clinical, pathological and demographic variables have not been able to accurately stratify the bulk of patients who fall in intermediate cancer grade such as Gleason scores of 3+4 or 4+3, and generally overestimate the recurrence probability for the low-risk patients.<sup>8,13</sup> It should be noted that in most image analysis based approaches the available image data from several patients is often split into case-control pairs such that the cases include those patients who experienced biochemical recurrence after specified follow-up time (usually five years) and controls are the patients without

recurrence. The case-control pairs are usually matched on the clinical and demographic variables such as age, Gleason score, pathological stage and race to aid the development of novel biomarkers. Now we will describe some of the previous approaches for predicting PCa recurrence using only H&E stained images of the prostate tissue.

Jafari-Khouzani et al. presented one of the earliest efforts in PCa recurrence prediction using second order image intensity texture features obtained from the co-occurrence matrices of H&E stained tissue images.<sup>18</sup> Comparative analysis of the predictive power of image-based features such as nuclear shape, texture etc. and clinical features such as Gleason score for PCa recurrence prediction is presented by Teverovskiy et al.<sup>19</sup> Lee et al. presented a data fusion scheme leveraging the advantages of both imaging and non-imaging data (e.g. proteomics).<sup>20</sup> An extension of this data fusion scheme using histological images along with proteomics data using a computationally efficient supervised regularized canonical correlation analysis (SRCCA) algorithm was presented by Golugula et al.<sup>21</sup> Graph-based features for differentiating between recurrent and non-recurrent cases using co-occurring gland tensors,<sup>22</sup> co-occurring gland angularity,<sup>2</sup> and cell cluster graphs<sup>23</sup> have also been proposed. Image texture-based features such as first order statistical intensity and steerable gradient filters (e.g. Gabor filters) have also been explored to predict PCa recurrence.<sup>24</sup> Recently, Lee et al. have presented a new data integration methodology termed as supervised multi-view canonical correlation analysis (sMVCCA) to predict PCa recurrence by integrating data from histology images and proteomic assays.<sup>25</sup>

Although, most of the aforementioned methods use supervised learning to estimate the likelihood of recurrence for a given patient using H&E stained tissue images, they rely mostly on the hand-crafted features to do so. These hand-crafted features often require enormous human effort in feature selection and usually rely on expert knowledge. Moreover, these features normally do not generalize well across a wide range of patients and some of them also require manual tuning of their parameters (e.g. thresholds for the gradients in texture-based features) before applying them to a new image. Thus, it is important to develop techniques for automatic feature extraction to improve overall generalization capabilities of the PCa recurrence prediction algorithms. Motivated by this, we present a deep learning-based approach to differentiate between recurrent and non-recurrent patients who have undergone radical prostatectomy, using only their tissue images and no other source of information. We hope that with sufficient training, the proposed model will generalize well on independent external validation datasets. Before presenting the proposed PCa recurrence approach we give a brief overview of the exciting field of deep learning.

When a large number of annotated examples are available, deep learning techniques, especially convolutional neural networks, have shown state-of-the-art performance for image classification.<sup>26</sup> Traditional machine learning techniques were limited in their ability to automatically learn meaningful features from the given data in their raw form (e.g. images). Hence, conventional pattern recognition techniques required enormous effort in terms of feature engineering and selection to transform the raw data (e.g. pixels of an image) into suitable feature vectors (for example higher order moments of a gray level co-occurrence matrix) to train an accurate classification model. Deep learning-based architectures, on the other hand, automatically learn the representations of the data with multiple levels of abstraction embedded in multiple layers of artificial neurons arranged in a hierarchical order. Due to their architectural innovations, training algorithms, and advancement in computer hardware, these methods have dramatically improved the state-of-the-art accuracy in problems with large amounts of data, such as speech recognition, visual object recognition, and many other domains such as drug discovery and genomics.<sup>26</sup> Leveraging the power of deep convolutional neural networks (CNNs) to automatically extract meaningful representations from the raw image data, we present an approach to automatically predict PCa recurrence from tissue images of patients who underwent radical prostatectomy. Now we describe the details of the proposed algorithm.

### 3. PROPOSED ALGORITHM

CNNs are the most popular deep learning models for processing multi-dimensional array data such as color images. A typical CNN consists of multiple convolutional and pooling layers followed by a few fully-connected layers to simultaneously learn a feature hierarchy and classify images. It uses error backpropagation – an efficient form of gradient descent – to update the weights connecting its inputs to the outputs through its multi-layered architecture.<sup>27</sup> In this paper, we present a two-stage approach using two separate CNNs. The first CNN detects

nuclei in a given tissue image while the second CNN takes patches centered at the detected nuclear centers as input to predict the probability that the patch belongs to a case of PCa recurrence. Before describing our CNN models, we present the details of the data we used to develop the proposed PCa recurrence model.

### 3.1 Dataset Description

We used H&E stained tissue microarray (TMA) core sections obtained from the Cooperative Prostate Cancer Tissue Resource (CPCTR).<sup>17</sup> Multiple core samples - in most cases four - were available in the TMA from each patient. For this study, we randomly selected 110 cases with PCa recurrence and 110 controls without recurrence from the CPCTR data, as determined based on biochemical recurrence (elevated PSA) in a five-year followup after radical prostatectomy. Additionally, the selected case-control pairs were matched on Gleason grade, pathological stage, age, and race. Digital images of the TMA cores were obtained using 20 $\times$  objective magnification (200 $\times$  overall) and cut into images of size 2000  $\times$  2000 pixels. It is worth noting that the case-control pairs were not matched on PSA levels. We labeled the tissue images of the patients with PCa recurrence (cases) as 1 and without recurrence (controls) as 0. We randomly divided this dataset into 160 training and 60 testing images, with an equal number of cases and controls in both subsets.

#### 3.1.1 Nuclear Annotations

The first stage of our PCa recurrence model is a nucleus detection module which predicts the nuclear centers in a given test tissue image using CNN as the learning machine. To train the nucleus detection CNN we annotated around 20,000 nuclear boundaries in Aperio ImageScope® from the training images. We annotated approximately an equal number of nuclei from the TMAs of recurrent and non-recurrent patients. Both epithelial and stromal nuclei were annotated in any given tissue image. For accurate nuclear boundary annotations the images were enlarged to 200 $\times$  on a 25" monitor such that each image pixel occupied 5  $\times$  5 screen pixels for clear visibility, and the annotations were drawn with a laser mouse. The annotators were engineering undergraduate and graduate students. The generated XML files containing pixel coordinates of the annotated nuclear boundaries were saved for data preparation for CNN training. It is worth noting that for overlapping nuclei, each multi-nuclear pixel was assigned to the largest nucleus containing that pixel. An example of our nuclear annotations is shown in Figure 1a

#### 3.1.2 Tumor Annotations

Since our goal was to predict PCa recurrence with high accuracy, we annotated tumor regions (including both tumor cells and tumor-associated stroma) in the TMA images of the patients of both classes in training and test sets. Nuclei detected *within* the tumor regions were used to train a second stage CNN for highly accurate PCa recurrence prediction. Tumor annotations were done by an expert pathologist by visual assessment of the architectural patterns and nuclear appearances in a given TMA image viewed on a 25" monitor. The pathologist outlined the tumor regions by drawing an approximate tumor boundaries in Aperio ImageScope® as illustrated in Figure 1b.

### 3.2 Pre-processing: Color Normalization

The performance of the computational approaches for H&E stained tissue image analysis often suffers because of a large variation in image colors due to H&E reagent concentration, staining process, and stain absorption pattern caused by differences in tissue fixation and processing.<sup>28,29</sup> To neutralize these unwanted color variations various color normalization (CN) approaches have been proposed in the literature. A recent CN approach that preserves biological structure by basing color mixture modeling on sparse nonnegative matrix factorization (SNMF) has demonstrated an improvement in the tissue segmentation quality.<sup>28,29</sup> This technique first assesses a stain density map for a given image as follows. For a pixel vector  $\vec{x}$  with  $(R, G, B)$  components in the 8-bit pixel range [0, 255], Beer-Lambert transform converts it into an optical density vector  $\vec{y}$  with components  $(r, g, b)$  such that,  $r = -\log(R/255)$ ,  $g = -\log(G/255)$   $b = -\log(B/255)$ . Then, stain density is obtained using SNMF over the data matrix  $\mathbf{Y}$  corresponding to all the pixels such that one matrix contains the  $(r, g, b)$  optical density components of each stain prototype (two prototypes for H&E), and the other contains the stain mixing components of each pixel, or stain density maps. The tradeoff between sparseness and reconstruction accuracy is controlled by a hyper-parameter  $\lambda$  that is multiplied with the L1 norm of the weight (stain density) matrix,

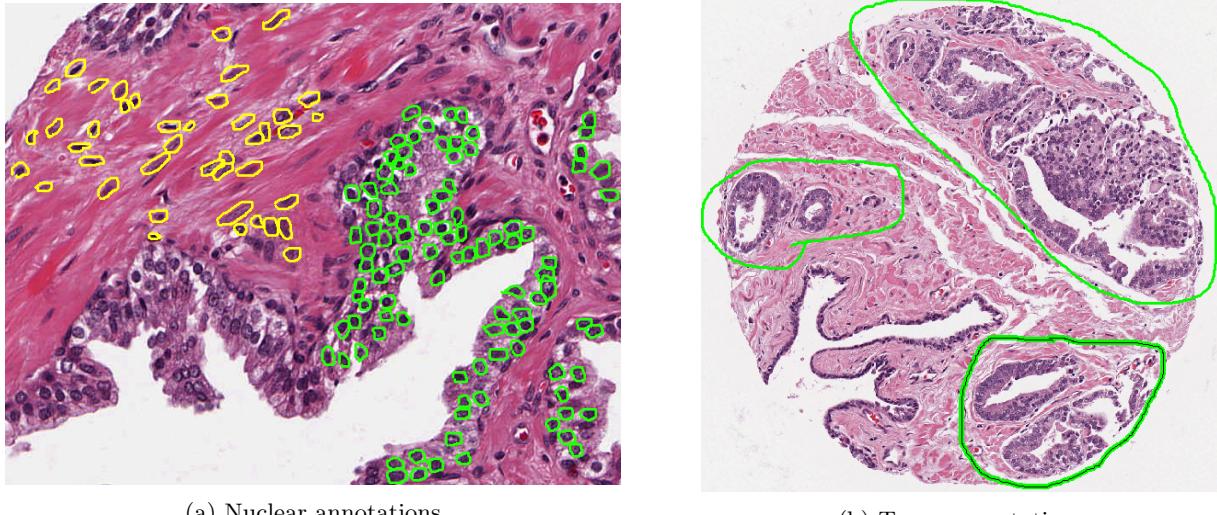


Figure 1: We annotated nuclear boundaries and tumor regions in the available TMA images to train and evaluate the proposed PCa recurrence algorithm. Zoomed in portion of an example TMA image is shown in (a) to illustrate our annotations for both epithelial (in green) and stromal (in yellow) nuclei. Tumor regions marked by the pathologist are shown in (b).

which was recommended to be set to 0.1.<sup>28</sup> Introduction of sparseness emulates the specificity of target material for each stain in the staining process, and thus preserves the biological structure. A color normalized image is computed by multiplying its stain density map with a color prototype basis matrix of a standard image, and taking the inverse Beer-Lambert transform. The target image was selected by a pathologist based on what was subjectively determined to be a well-stained image. Sample results of the CN procedure are shown in Figure 3.

### 3.3 Stage 1: Nucleus Detection

After pre-processing, we first detect nuclei using a CNN because nuclear morphometric features are often the key predictors of tumor grade and stage, and are associated with adverse events such as recurrence. Thus, we hypothesized that once we accurately detect the nuclei present in a given tumor image we can use the information around each of them in the second stage to predict the recurrence probability. A second reason for applying the second CNN to the patches centered at the nuclear centers detected by the first CNN was that it reduces the amount of information that the second CNN has to learn to ignore. Thus, instead of learning to ignore the location of a nucleus in its input patch, the second stage always gets a nucleus at the center of its input patch, and has to concentrate only on learning differences between centered nuclei for recurrent vs. non-recurrent decision. Our nucleus detection scheme is based on the distance transform of the annotated nuclear regions and is described as follows.

- 1. Data preparation:** We used nuclear boundary annotations (Section 3.1.1) to create binary masks where one (white pixel) represented nuclear pixels and zero (black pixel) represented the background pixels. Then we computed the distance transform<sup>30</sup> of each annotated nucleus from its binary mask using Euclidean distance as the distance metric. Distance transform assigns each non-zero pixel in the given binary map a value equal to the distance of that pixel from the closest nuclear boundary. It assigns a zero value to pixels on or outside nuclear boundaries. Thus, the distance transform has high values at the center of a nucleus and its value decreases as we move closer to the nuclear boundary, and remains zero outside all nuclei. An illustration of the distance transform representation of the annotated nuclei is shown in Figure 2. The distance transform is zero on nuclear boundaries, including those between two touching or overlapping nuclei. Being able to predict low values of distance transform on boundaries of two touching nuclei is what will allow our technique to disambiguate between multiple nuclei in the crowded regions.



Figure 2: (Left) A portion of the original H&E image with overlaid nuclear annotations in green; (Middle) Binary mask representing annotated nuclear regions with 1 (white pixels) and background as 0 (black pixels); (Right) Distance transform of the binary nuclear mask in which the peak values at the nuclear centers (in red) and low values near the nuclear boundaries can be noticed.

Table 1: CNN Architecture for predicting distance transform

Layer	Filter size	Activation	Output size	Dropout rate
Input	—	—	$51 \times 51 \times 3$	—
Conv 1	$5 \times 5$	ReLU	$47 \times 47 \times 20$	0.1
Pool 1	$2 \times 2$	Max	$23 \times 23 \times 20$	—
Conv 2	$5 \times 5$	ReLU	$19 \times 19 \times 50$	0.2
Pool 2	$2 \times 2$	Max	$9 \times 9 \times 50$	—
FC 1	—	ReLU	1024	0.5
FC 2	—	ReLU	1024	0.5
Output	—	Linear	49	—

2. **Learning to predict distance transform:** We trained a CNN that took a patch of size  $51 \times 51$  from H&E stained tissue images as input and predicted the corresponding distance transform of a smaller patch of size  $7 \times 7$  as output centered at the same pixel as that of the input patch. To create the training set we randomly sampled 120,000 input-output patch pairs of sizes  $51 \times 51$  and  $7 \times 7$ , respectively, of which 85% were centered around the nuclear pixels within the annotated nuclear boundaries and the rest were centered around the pixels outside all nuclei. To ensure correctness in patch sampling for the outside nuclei class, we used a standard protocol of applying a threshold on the average intensity of the color-deconvolved image as proposed in.<sup>31</sup> The rationale is based on the fact that the outside class pixels should not strongly absorb hematoxylin. Similarly, a validation set of 50,000 input-output patch pairs randomly sampled from the training images was also prepared for the performance evaluation of the trained model. The output patch size was kept smaller than the input to preserve smoothness among the neighboring patches. It should be noted that the patch-pairs used in training and validation sets were obtained from the TMA images of mutually exclusive patients. Further, roughly equal number of patch-pairs from the TMA images of the recurrent cases and non-recurrent nuclear controls were used. We used cross-validation on a partial training set to fix CNN hyper-parameters such as the required number of convolutional and fully-connected layers, number and size of filters in each layer, dropout value, etc. The selected CNN architecture for distance transform prediction is shown Table 1. To train the CNN we minimized the mean squared error between the ground truth and the predicted distance transform by updating the CNN weights through error backpropagation with gradient descent algorithm<sup>27</sup> at a learning rate of 0.001. Note that this CNN was used for nonlinear regression (predicting a continuous range of distance transform), and not for classification, which is more commonly done.

3. **Post-processing:** The distance transform maps predicted by the first CNN contained some noise (see Figure 4, column 4) which we suppressed using H-minima suppression<sup>32</sup> with the depth parameter of 2. Local maxima were detected from the distance transform maps thus obtained to get the desired nuclear centers. A few examples of the detected nuclear centers overlaid on the original H&E images are shown in the last column of Figure 4.

For any new prostate TMA image, overlapping patches of size  $51 \times 51$  with one pixel shift were extracted and fed as input to the trained CNN of stage 1 to get the corresponding  $7 \times 7$  predicted distance transform maps. Patch-wise distance transform maps were then rearranged to obtain the final predicted distance transform map. Prediction from 49 overlapping patches was averaged to smoothen and reduce noise in the predicted distance transform map. Using step 3 in Section 3.3 the desired nuclear centers were obtained using these maps.

### 3.4 Stage 2: Recurrence score prediction

Now we describe how we used a second CNN for PCa recurrence probability estimation using the nuclei detected in stage 1.

1. **Data preparation:** We extracted patches of size  $101 \times 101$  around the detected nuclear centers from H&E stained TMA and assigned them the label 1 if the patches we sampled from the images of recurrent cases, and 0 otherwise. It should be noted that the stage 1 CNN was trained to detect nuclear centers of both epithelial and stromal nuclei in tumor as well as non-tumor regions of the recurrent and non-recurrent input images. However, we used tumor annotations (Section 3.1.2) to extract patches around the detected nuclear centers from *only* the tumorous regions in stage 2. Further, we used 80 recurrent cases and 80 non-recurrent controls to prepare the training data for the second stage CNN. The remaining 30 case-control pairs were used to evaluate the performance of the proposed PCa recurrence model. We randomly sampled around 90,000 patches, each of size  $101 \times 101$ , from the 80 training case-control pairs with equal number of patches extracted from the images of the recurrent and non-recurrent patients.
2. **Learning to predict patch-wise recurrence probability:** The training data prepared in the previous step was used for training the second stage CNN that predicted the label (1 or 0) at its output for a given  $101 \times 101$  H&E patch as input. The CNN was trained by minimizing binary cross-entropy loss between the ground truth and the predicted labels by updating the CNN weights through error backpropagation using gradient descent algorithm<sup>27</sup> at a learning rate of 0.01. We used a part of the training data to optimize the CNN hyper-parameters such as the number of convolutional layers, number and size of filters in each layer, etc. through cross-validation as done for the stage 1 CNN. We selected that CNN architecture for patch-wise recurrence probability prediction which gave saturation in performance on the validation data set. The proposed architecture is shown in Table 2. It is worth noting that the last layer in Table 2 is a softmax layer that gave the probability of recurrence for a given input patch. The class label for each patch was decided using a recurrence probability threshold of 0.5.

Table 2: CNN architecture for patch-wise recurrence prediction

Layer	Filter size	Activation	Output size	Dropout rate
Input	—	—	$101 \times 101 \times 3$	—
Conv 1	$11 \times 11$	ReLU	$91 \times 91 \times 30$	0.1
Pool 1	$2 \times 2$	Max	$45 \times 45 \times 30$	—
Conv 2	$7 \times 7$	ReLU	$39 \times 39 \times 48$	0.2
Pool 2	$2 \times 2$	Max	$19 \times 19 \times 48$	—
Conv 3	$5 \times 5$	ReLU	$15 \times 15 \times 70$	0.25
Pool 3	$2 \times 2$	Max	$7 \times 7 \times 70$	—
FC 1	—	ReLU	1024	0.5
FC 2	—	ReLU	1024	0.5
Output	—	Softmax	2	—

3. **Predicting PCa recurrence score:** We used majority voting among the predicted patch-wise labels of a given H&E image to categorize that image as recurrent or non-recurrent. The final PCa recurrence score was assigned to an H&E image on the basis of the average patch-wise recurrence probability of its patches that belonged to the majority class.

Once the proposed PCa recurrence model is satisfactorily trained, it can be applied to any new test image by first applying the proposed stage 1 nucleus detector (Section 3.3) to estimate the nuclear centers in the given image. Then, with the annotated tumor regions, the second stage CNN can be used to predict its patch-wise labels and recurrence probabilities using the patches sampled around the detected nuclear centers in the tumorous regions. These patch-wise labels and recurrence probabilities can then be aggregated through majority voting among labels to estimate the final class label and PCa recurrence score for that image in accordance with step 3 of Section 3.4. Now, we will describe the results of the experimental evaluation of the proposed PCa recurrence scheme.

## 4. EXPERIMENTAL RESULTS

We first demonstrate the importance of color normalization in our PCa recurrence scheme followed by an illustration of the results of our nucleus detection module and finally present the quantitative performance results of the proposed PCa recurrence score prediction algorithm.

### 4.1 Importance of color normalization

Color normalization is a useful pre-processing technique to neutralize the color variations in the H&E stained images as shown in Figure 3. It is clearly noticeable from Figure 3 that color normalization reduces the inter-image color variation while preserving the intra-image contrast between nuclei and the surrounding tissue. It has been previously demonstrated that color normalization can improve the performance of higher level computer vision tasks such as image segmentation, etc.<sup>28,29</sup> In our experiments as well, we saw an increase in the PCa recurrence prediction accuracy by 3-5% when color normalization was used. However, it should be noted that although color normalization improves the performance of higher level computer vision tasks such as ours, the choice of reference image doesn't have much impact on the segmentation quality. This is because color normalization only serves as a means to project all tissue slides onto the same color space to account for the variations in the staining patterns across multiple hospitals. For our experiments, an expert pathologist subjectively selected a well stained image as the reference for our color normalization.

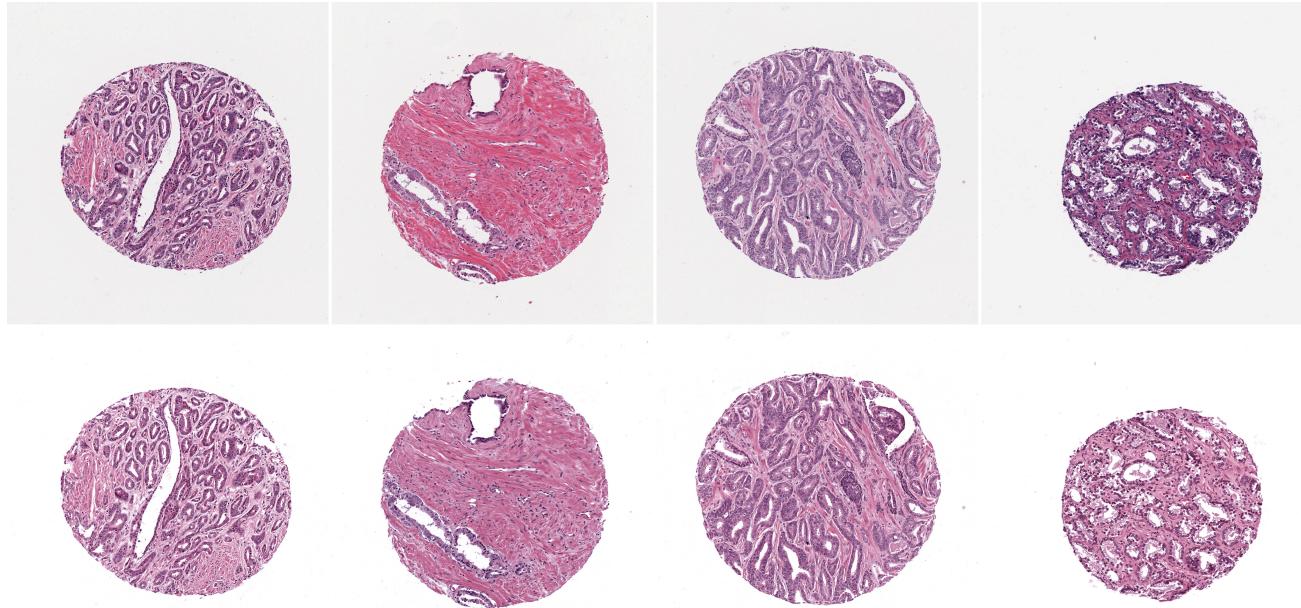


Figure 3: Color normalization (CN) reduces the color variations in H&E stained images as shown here for sample TMA images used in this study. Original TMA images are shown in the first row and the corresponding color normalized images are shown in the second row.

## 4.2 Nucleus detection performance

As mentioned in Section 3.3, we first detect nuclear centers from a given test image by identifying the local peaks of the distance transform map predicted by a CNN. We trained the stage 1 CNN (Table 1) for about 200 epochs after which the decrease in the mean squared error between the ground truth and the predicted distance transform almost saturated. The CNN thus trained gave an accuracy of 89.6% for nucleus detection on an independent set of validation images. Qualitative results of our nucleus detection module are shown in Figure 4. It can be seen that our nucleus detection scheme not only detects the annotated nuclei but also successfully detects unannotated ones even though it leaves a few crowded nuclei undetected due to merging of the two or more distance transform peaks (predicted). In the future we will work on rectifying this problem.

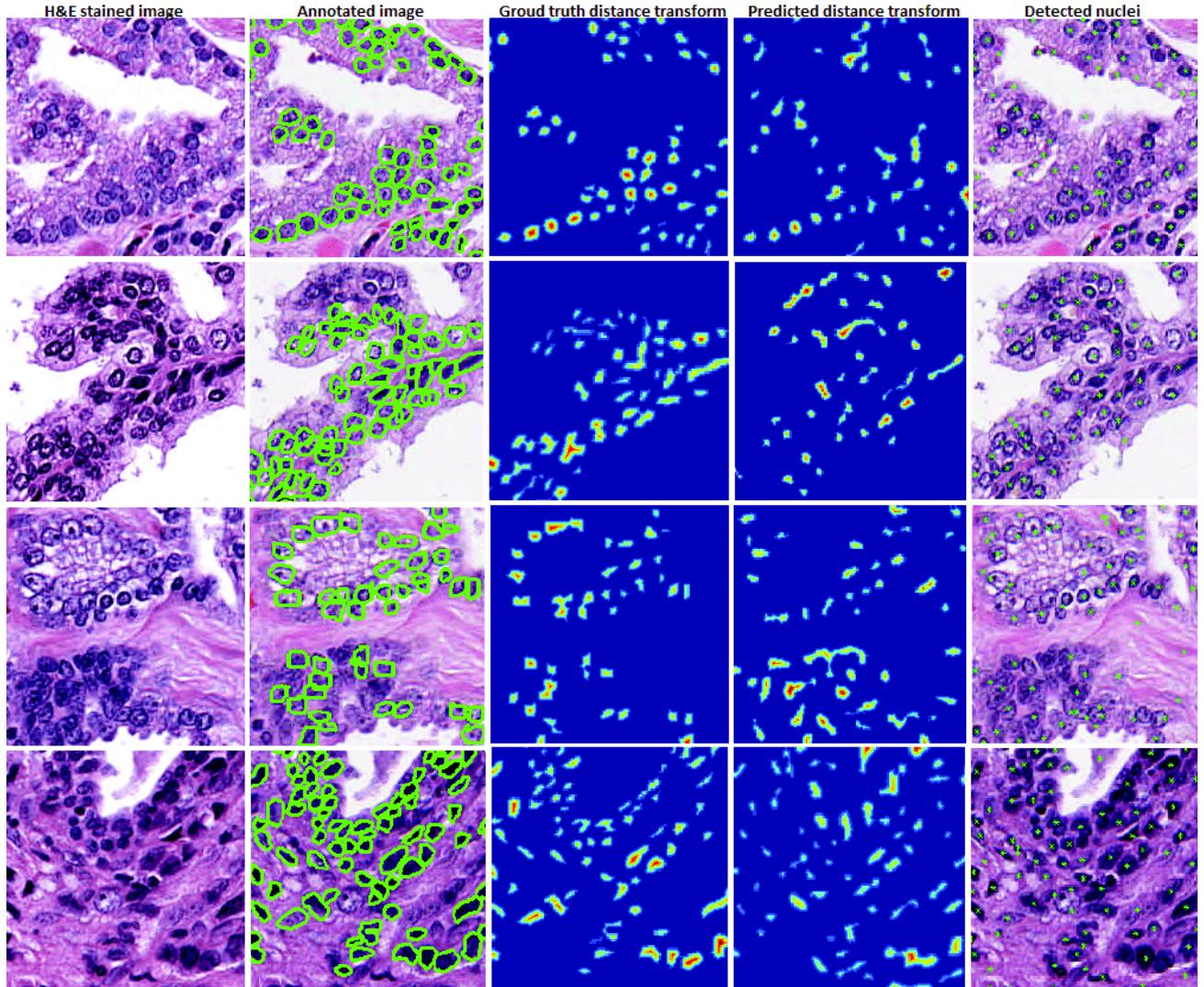


Figure 4: Results of the proposed nucleus detection module.

## 4.3 PCa recurrence prediction performance

After nucleus detection, we used the stage 2 CNN to predict patch-wise labels (0 or 1) and recurrence probabilities of the patches sampled from test cases and controls. Final class labels (0 or 1) and PCa recurrence scores were assigned to the test images by majority voting among their patch-wise labels and average patch-wise recurrence probability of their majority class, respectively. We used an independent set of 30 case-control pairs to evaluate the performance of our recurrence prediction module trained on a separate set of 80 case-control pairs. For

comparison, we also computed the class labels using clinical features such as Gleason primary grade, Gleason secondary grade, Gleason sum score, PSA level, and age at initial diagnosis in a logistic regression model developed using the clinical data from the same training (80) case-control pairs that were used to train the CNNs. The receiver-operating characteristic (ROC) thus obtained for the proposed approach had 0.81 area under curve (AUC) as compared to 0.59 AUC of the ROC computed using the usual clinical features. These results are shown in Figure 5. These results indicate that deep learning based image analysis of H&E stained tissue images can not only complement but even outperform clinical information based regression models for accurate PCa recurrence prediction.

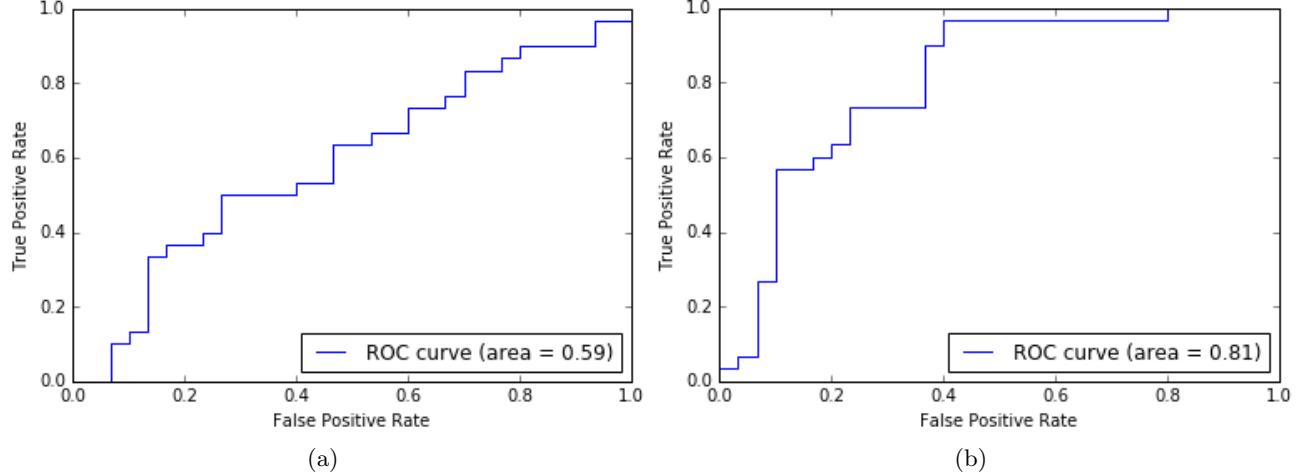


Figure 5: Receiver operating characteristics for (a) logistic regression model using clinical variables, and (b) the proposed model for PCa recurrence prediction.

## 5. CONCLUSION

In this paper, we presented a new machine learning based approach to predict the biochemical recurrence in prostate cancer from H&E stained prostate cancer images. The proposed method detects nuclei from the input digital tissue sample using a convolutional neural network (CNN) in its first stage. Then, using tumor region annotations, patches extracted around the detected nuclear centers from tumor regions are classified as recurrent or non-recurrent by the second stage CNN. Finally, the probability of recurrence for each image is computed based on majority voting on its predicted patch labels. For this paper, we used 80 recurrent cases and 80 non-recurrent controls for training our system, which gave 0.81 AUC for a separate test dataset containing 30 cases and 30 controls (all mid-grade, intermediate stage PCa patients). Our algorithm is particularly useful because the current approaches for predicting PCa outcome using Gleason grade and other clinical measures often fail to predict biochemical recurrence among mid-grade prostate cancer patients, who form the bulk of all PCa patients. If validated further, our approach might serve as an valuable addition to the pathologists toolbox for effective treatment planning and decision making.

## ACKNOWLEDGMENTS

All authors except Peter Gann were partly supported by Ministry of Human Resource Development, Government of India. Peter Gann and Neeraj Kumar were partly supported by US NIH/NCI grant RO1CA155301. Amit Sethi thanks Amazon Web Services for cloud computing credits.

## REFERENCES

- [1] Siegel, R. L., Miller, K. D., and Jemal, A., “Cancer statistics, 2017,” *CA: A Cancer Journal for Clinicians* **67**(1), 7–30 (2017).

- [2] Lee, G., Sparks, R., Ali, S., Shih, N. N., Feldman, M. D., Spangler, E., Rebbeck, T., Tomaszewski, J. E., and Madabhushi, A., “Co-occurring gland angularity in localized subgraphs: predicting biochemical recurrence in intermediate-risk prostate cancer patients,” *PLoS one* **9**(5), e97954 (2014).
- [3] AV, D., R, W., S, M., and et al, “Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer,” *JAMA* **280**(11), 969–974 (1998).
- [4] Partin, A., Yoo, J., Carter, H. B., Pearson, J., Chan, D., Epstein, J., and Walsh, P., “The use of prostate specific antigen, clinical stage and gleason score to predict pathological stage in men with localized prostate cancer.,” *The Journal of urology* **150**(1), 110–114 (1993).
- [5] Kattan, M. W., Wheeler, T. M., and Scardino, P. T., “Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer,” *Journal of Clinical Oncology* **17**(5), 1499–1499 (1999).
- [6] Stephenson, A. J., Scardino, P. T., Eastham, J. A., Bianco, F. J., Dotan, Z. A., Fearn, P. A., and Kattan, M. W., “Preoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy,” *Journal of the National Cancer Institute* **98**(10), 715–717 (2006).
- [7] Cooperberg, M. R., Pasta, D. J., Elkin, E. P., Litwin, M. S., Latini, D. M., Du Chane, J., and Carroll, P. R., “The university of california, san francisco cancer of the prostate risk assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy,” *The Journal of urology* **173**(6), 1938–1942 (2005).
- [8] Cooperberg, M. R., Hilton, J. F., and Carroll, P. R., “The capra-s score,” *Cancer* **117**(22), 5039–5046 (2011).
- [9] Greene, K. L., Meng, M. V., Elkin, E. P., Cooperberg, M. R., Pasta, D. J., Kattan, M. W., Wallace, K., and Carroll, P. R., “Validation of the kattan preoperative nomogram for prostate cancer recurrence using a community based cohort: results from cancer of the prostate strategic urological research endeavor (capsure),” *The Journal of urology* **171**(6), 2255–2259 (2004).
- [10] Cheng, G. C., Chen, M.-H., Whittington, R., Malkowicz, S. B., Schnall, M. D., Tomaszewski, J. E., and DAmico, A. V., “Clinical utility of endorectal mri in determining psa outcome for patients with biopsy gleason score 7, psa 10, and clinically localized prostate cancer,” *International Journal of Radiation Oncology\* Biology\* Physics* **55**(1), 64–70 (2003).
- [11] Glinsky, G. V., Glinskii, A. B., Stephenson, A. J., Hoffman, R. M., and Gerald, W. L., “Gene expression profiling predicts clinical outcome of prostate cancer,” *The Journal of clinical investigation* **113**(6), 913–923 (2004).
- [12] Stephenson, A. J., Smith, A., Kattan, M. W., Satagopan, J., Reuter, V. E., Scardino, P. T., and Gerald, W. L., “Integration of gene expression profiling and clinical variables to predict prostate carcinoma recurrence after radical prostatectomy,” *Cancer* **104**(2), 290–298 (2005).
- [13] Gallina, A., Karakiewicz, P. I., Hutterer, G. C., Chun, F. K.-H., Briganti, A., Walz, J., Antebi, E., Shariat, S. F., Suardi, N., Graefen, M., et al., “Obesity does not predispose to more aggressive prostate cancer either at biopsy or radical prostatectomy in european men,” *International journal of cancer* **121**(4), 791–795 (2007).
- [14] Shukla-Dave, A., Hricak, H., Ishill, N., Moskowitz, C. S., Drobnjak, M., Reuter, V. E., Zakian, K. L., Scardino, P. T., and Cordon-Cardo, C., “Prediction of prostate cancer recurrence using magnetic resonance imaging and molecular profiles,” *Clinical Cancer Research* **15**(11), 3842–3849 (2009).
- [15] Karnes, R. J., Cheville, J. C., Ida, C. M., Sebo, T. J., Nair, A. A., Tang, H., Munz, J.-M., Kosari, F., and Vasmatzis, G., “The ability of biomarkers to predict systemic progression in men with high-risk prostate cancer treated surgically is dependent on erg status,” *Cancer research* **70**(22), 8994–9002 (2010).
- [16] Wassberg, C., Akin, O., Vargas, H. A., Shukla-Dave, A., Zhang, J., and Hricak, H., “The incremental value of contrast-enhanced mri in the detection of biopsy-proven local recurrence of prostate cancer after radical prostatectomy: effect of reader experience,” *AJR. American journal of roentgenology* **199**(2), 360 (2012).
- [17] Berman, J. J., Datta, M., Kajdacsy-Balla, A., Melamed, J., Orenstein, J., Dobbin, K., Patel, A., Dhir, R., and Becich, M. J., “The tissue microarray data exchange specification: implementation by the cooperative prostate cancer tissue resource,” *BMC bioinformatics* **5**(1), 1 (2004).

- [18] Jafari-Khouzani, K. and Soltanian-Zadeh, H., "Multiwavelet grading of pathological images of prostate," *IEEE Transactions on Biomedical Engineering* **50**, 697–704 (June 2003).
- [19] Teverovskiy, M., Kumar, V., Ma, J., Kotsianti, A., Verbel, D., Tabesh, A., Pang, H.-Y., Vengrenyuk, Y., Fogarasi, S., and Saidi, O., "Improved prediction of prostate cancer recurrence based on an automated tissue image analysis system," in [*2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)*], 257–260 Vol. 1 (April 2004).
- [20] Lee, G., Doyle, S., Monaco, J., Madabhushi, A., Feldman, M. D., Master, S. R., and Tomaszewski, J. E., "A knowledge representation framework for integration, classification of multi-scale imaging and non-imaging data: Preliminary results in predicting prostate cancer recurrence by fusing mass spectrometry and histology," in [*2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*], 77–80 (June 2009).
- [21] Golugula, A., Lee, G., Master, S. R., Feldman, M. D., Tomaszewski, J. E., Speicher, D. W., and Madabhushi, A., "Supervised regularized canonical correlation analysis: integrating histologic and proteomic measurements for predicting biochemical recurrence following prostate surgery," *BMC Bioinformatics* **12**(1), 483 (2011).
- [22] Lee, G., Sparks, R., Ali, S., Madabhushi, A., Feldman, M. D., Master, S. R., Shih, N., and Tomaszewski, J. E., "Co-occurring gland tensors in localized cluster graphs: Quantitative histomorphometry for predicting biochemical recurrence for intermediate grade prostate cancer," in [*2013 IEEE 10th International Symposium on Biomedical Imaging*], 113–116 (April 2013).
- [23] Ali, S., Veltri, R., Epstein, J. A., Christudass, C., and Madabhushi, A., "Cell cluster graph for prediction of biochemical recurrence in prostate cancer patients from tissue microarrays," (2013).
- [24] Doyle, S., Feldman, M. D., Shih, N., Tomaszewski, J., and Madabhushi, A., "Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer," *BMC Bioinformatics* **13**(1), 282 (2012).
- [25] Lee, G., Singanamalli, A., Wang, H., Feldman, M. D., Master, S. R., Shih, N. N. C., Spangler, E., Rebbeck, T., Tomaszewski, J. E., and Madabhushi, A., "Supervised multi-view canonical correlation analysis (smvcca): Integrating histologic and proteomic features for predicting recurrent prostate cancer," *IEEE Transactions on Medical Imaging* **34**, 284–297 (Jan 2015).
- [26] Lecun, Y., Bengio, Y., and Hinton, G., "Deep learning," *Nature* **521**, 436–444 (5 2015).
- [27] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in neural information processing systems*], 1097–1105 (2012).
- [28] Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A. M., Esposito, I., and Navab, N., "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Transactions on Medical Imaging* **35**, 1962–1971 (Aug 2016).
- [29] et. al., A. S., "Empirical comparison of color normalization methods for epithelial-stromal classification in h and e images," *Journal of Pathology Informatics* (2016).
- [30] Kimmel, R., Kiryati, N., and Bruckstein, A. M., "Sub-pixel distance maps and weighted distance transforms," *Journal of Mathematical Imaging and Vision* **6**(2), 223–233 (1996).
- [31] Ruifrok, A. and Johnston, D., "Quantification of histochemical staining by color deconvolution," *Analytical and quantitative cytology and histology* **23**, 291–299 (aug 2001).
- [32] Soille, P., [*Morphological Image Analysis: Principles and Applications*], Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2 ed. (2003).