PROCEEDINGS OF SPIE

Automated end-to-end deep learning framework for classification and tumor localization from native non-stained pathology images

Bayat, Akram, Anderson, Connor, Shah, Pratik

SPIE.

Automated end-to-end deep learning framework for classification and tumor localization from native non-stained pathology images

Akram Bayat[a], Connor Anderson[a], Pratik Shah*[a]

[a]Program in Media Arts and Sciences, Media Laboratory, Massachusetts Institute of Technology, 77 Mass. Ave., Cambridge, MA 02139, USA

ABSTRACT

Deep learning based quantitative assessment of digital pathology images and understanding the underlying reasons for a specific clinical decision is challenging, and automatic histology pattern classification and tumor localization in whole-slide pathology images are critical for interpretable learning systems. In this study, we propose an end-to-end deep learning framework for automatic detection and localization of tumors directly from non-stained whole slide prostate core biopsy images (WSI). We use a previously described Generative Adversarial Network (GAN)-based model from our laboratory for computational Hematoxylin and Eosin (H&E) staining of native non-stained pathology images. A convolutional neural network to detect and classify tumor regions in 1024×1024 virtually stained H&E pixel patches, and a concurrent deep weakly supervised (WSL) model that provides localization of predominant histologic patterns used for tumor classification without the need for pixel-level annotations are reported in this study for the first time. The end-to-end system was evaluated on a 17K hold out set of 1024×1024 non-stained patches extracted from 13 whole slide prostate biopsy images. Experimental results yielded 86.37% patch-level classification accuracy with 85.05% precision, and achieved a Dice index of 65.07±1.99 (compared to 70.24±1.86 Dice index in the U-Net reference model for pixel-level segmentation). The end-to-end deep learning framework thus automates digital pathology image workflow from tissue staining to interpretable prostate tumor classification and can be valuable for accurate grading of prostate cancer and generalized to other whole-slide image classification tasks.

Keywords: Pathology image analysis, prostate cancer, automated classification, tumor localization, deep learning

1. INTRODUCTION

Microscopic examination of histopathology slides is the criterion standard in diagnosing most types of cancers including those in prostate[1]. Automated histopathology image analysis can provide a cost effective, fast, and supportive method to complement expert pathologist opinion for diagnosis and prognosis of cancer predictions[2]. The widespread adaptation of WSI for digital pathology workflows has created opportunities for developing computer-aided deep learning algorithms for image analysis and diagnosis[3]. Pathologists use dye-based H&E staining to visualize the tissue structures to evaluate and identify important features for cancer[4]. However, limited tissue availability due to low sampling of tissue volumes[5,6], and irreversible H&E dye staining of tissues staining[7] are major challenges in dye-based histopathology image analysis. Deep learning-based approaches for computational virtual H&E staining of images of non-stained core biopsy samples can help address some limitations of traditional H&E dye staining by providing rapid and automated readouts and early assessments of tissue samples. We have previously reported generative deep learning algorithms and models (GAN-CS) that convert native non-stained whole slide images of prostate core biopsy to their computationally H&E dye-stained versions with high precision[8,9]. Detailed computer vision and physician ratification for detection and Gleason scoring of prostate tumors showed that the computational H&E-stained images were able to provide accurate diagnosis matching dye-stained images[9].

*pratiks@mit.edu; phone 1 617 324-3859; media.mit.edu/people/pratiks/

In this work, an end-to-end deep learning system that can perform virtual H&E staining, classification, and localization of tumors from native and unlabeled non-stained prostate core biopsy images is described for the first time. We report optimization of patch-level classification of WSIs including patch size and ratio of no-tumor and tumor data distribution. We develop deep learning models for classification of H&E dye-stained images (Classifier I) and computational H&E-stained images (Classifier II) and a deep weakly localization algorithm for localization of tumors on WSI. The localization maps compared to the pixel-level segmentation results for validation of the end-to-end system described in this study.

2. BACKGROUND AND RELATED WORKS

Clinical interpretation of pathology images remains the criterion standard for diagnosing several diseases including most types of cancer[10,11,12]. However, widespread adoption of WSI for tumor detection and cancer diagnosis by pathologists has increased the demand for effective and efficient gigapixel image analysis tools and methods. Recently, a plethora of reports advocating for integration of deep learning algorithms for automated tumor detection from H&E dye stained WSI have been published[13]. For example, Xu et al.[14], proposed using deep convolutional neural network activation features applied to classification, segmentation, and visualization in large-scale tissue histopathology images. They used a pre-trained AlexNet[15] architecture to visualize deep features to show that their network was able to learn complex clinical characteristics. The input images were H&E-stained patch-level images extracted from whole slide images. With a growing number of machine learning-based image analysis tools and software, that utilize dye-stained tissue images to perform, e.g., automated diagnosis, image segmentation, or classification for diagnosis and treatment evaluation but there are still technical barriers that prevent the translation of these advances into clinical applications. One of the main challenges is the lack of model generalization due to variability in H&E slide preparation and digital scanning process[16]. Heterogeneity of H&E images can affect automatic image analysis algorithms. However, standardization approaches like staining normalization based on color deconvolution and deep learning[17] and color augmentation techniques[18] have shown improvements in the generalization of the classification tasks for several tissue types. Recent works have tackled the staining variability problem by performing digital (i.e., virtual) staining of the non-stained histology images to match the images of histologically stained samples. Ozcan et al.[19] trained a deep neural network via a generative adversarial approach to stain an autofluorescence image of an unstained tissue section, thus bypassing the need for histological staining and bright-field microscopy. Recent work from our laboratory showed that a deep-learning model can computationally generate histological stains (H&E) from non-stained images of prostate core biopsy, and that these computational stains were clinically indistinguishable from the corresponding histologically stained tissue[9] images. In this study, we augment previously described automated virtual H&E staining models with classification, detection and localization of prostate tumors without the need for pixel-level annotations of WSI as an end-to-end deep learning system.

3. DATASET AND PREPROCESSING

3.1 Study population and data

Partners Human Research Committee approved study protocol (2014P002435) for utilization of excess material from prostate core biopsies performed during routine clinical care between 2014 and 2017, at Brigham and Women's Hospital, Boston, Massachusetts. Informed consent was waived because data were deidentified and samples were obtained as part of routine clinical care. Deidentified WSIs were transferred to the Massachusetts Institute of Technology for processing and analysis, and was exempt from institutional review board review per the Committee on the Use of Humans as Experimental Subjects guidelines. A dataset with WSIs of 46 histopathological slide image pairs at 20X magnification was obtained from 38 patients (mean age 66.2 years and a standard deviation of 8.92) with demographic diversity. Since each WSI contained 1-6 core biopsy tissue cores, single prostate tissue needle core biopsy images were extracted that resulted in 112 high-resolution native non-stained and H&E dye-stained image registered pairs. All WSI were examined by four board-certified/trained expert pathologists as previously described by us[8]. Sedeen Viewer software was used for detailed clinical labels indicating tumor regions with Gleason grades and other atypical manifestations on the H&E dye-stained images were provided. For training deep learning models on WSIs, patches instead of the whole slide images were used as input. Clinical evaluation showed that patches of 1024×1024 pixel size provided enough tissue for contextual representation to detect tumors. Patches with more than

85% white background were discarded. This process resulted in 95,605 patches that were further split into training, validation, and test sets (Table 1).

Table 1. Data splits of computationally H&E-stained prostate core biopsy images for training deep learning models in this study

Class label	Train	Val	Test	Total
No-tumor	47,435	6,107	3,813	57,355
Tumor	31,329	4,175	2,746	38,250
Total	78,764	10,282	6,559	95,605

3.2 Annotations

All WSI were manually labeled by a board-certified/trained expert pathologist. Detailed labels in the form of free-form outlines encompassing tumors regions with Gleason grades and other atypical manifestations on the H&E dye-stained images were provided using the Sedeen Viewer (PathCore Inc., Toronto, Ontario, Canada) on notebook computer screens (Dell Computers, Round Rock, TX). Annotations were classified - Gleason grade 3 (G3), Gleason grade 4 (G4), Gleason grade 5 (G5) as described previously[9].

3.3 Optimum patch size

It is computationally impractical to perform classification and segmentation with deep learning-based algorithms using the extremely large size of a single histopathology image (~100,000✕ 100,000 pixels). And downscaling the entire histopathology image to an acceptable size can reduce detailed information for accurate tumor detection and segmentation tasks. To overcome this problem, most histopathology classification and segmentation studies adopt a patch sampling technique[20]. The contribution of minimum sufficient patch size to cover the essential tissue structures in histopathological images on classification and segmentation performance of neural network models is not well understood. In one study[14], authors reported 336 ×336 pixels for 20X magnification and 672 ×672 pixels for 40X magnification scale yielded the highest accuracy in binary and multiclass tissue histopathology image classification tasks[21,22]. In this study, we systematically evaluated classification performance by optimization of different patch sizes with 50% overlapping rate and distributions of no-tumor and tumorous patches (Figure 1). Patches or regions of larger sizes provided higher classification (tumor/no-tumor) accuracy for neural networks (Table 2). A board-certified pathologist independently evaluated the patches with different sizes and reported that sizes of 1024×1024 provide enough tissue for contextual representation to detect a tumor. Based on these results, each histopathology image was divided into a set of overlapping square patches with a size of 1024 ×1024 pixels at 20X magnification. These patches formed a rectangular grid with 512-pixel stride, i.e., distance between adjacent patches. To further reduce the number of patches and misclassification, patches with more than 85% white background were discarded.
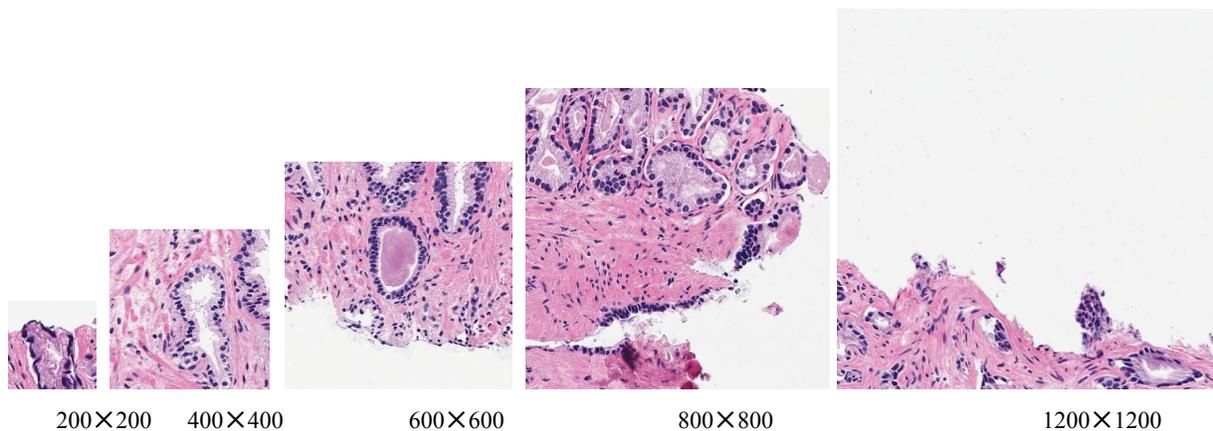

200✕200 400✕400   600✕600   800✕800   1200✕1200

Figure 1. A sample set of patches with various pixel sizes used for optimization of deep learning models

Table 2. Classification accuracy of different H&E dye-stained prostate core biopsy patch sizes with and without tumors

Patch size	Patches without tumors	Patches with tumors	Accuracy (%)
200×200	151,067	74,005	84.14
400×400	48,284	24,722	86.53
600×600	26,198	13,548	86.75
800×800	15,124	8,398	87.65
1024×1024	9,833	5,650	88.21

3.4 Distribution of tumor and no-tumor patches

Another factor that affects the classification performance of deep neural networks is the distribution of image patches with and without tumors in training and validation data. We used random undersampling of patches in different ratios to evaluate the classification accuracy where the ratio of patches without tumors to those with tumors varies between 0.98 to 2. Figure 2 shows the box plot of classification accuracy in multiple experiments 1-6 using different random undersampling of 200×200 pixels patches without tumors. The ratio of 1.4 (no-tumor: tumor) achieves higher classification performance with low variation. And increasing or decreasing this ratio negatively impacts the accuracy of the classification. Based on these experiments, each image was cropped with 50% overlapping regions into multiple patches of size 1024× 1024 pixels resulting in 95,605 registered pair patches. A sampled patch from a WSI was assigned a tumor label if at least ten percent of patch size contains cancerous tissue, otherwise it was assigned a no-tumor label. The 95,605 generated patches were further split into training, validation, and test sets reported in Table 1 showing the ratio of no-tumor (57,355) to tumor patches (38,250) was ~1.4. The test set consisted of 13 single-core WSI, that contain one or more of five Gleason grade tumors.
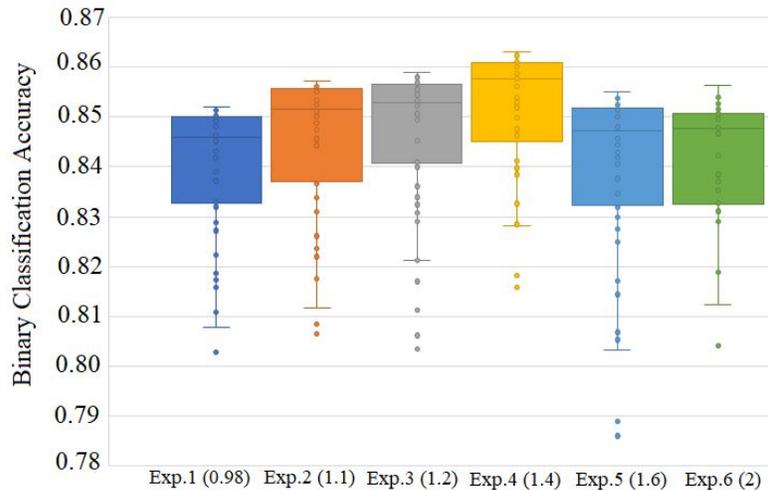

Figure 2. Box plot of classification accuracy in multiple experiments from different random undersampling of H&E dye stained prostate core biopsy image patches of 200×200 pixels. The values specified for each experiment in parenthesis represent the ratio of patches without-tumors to those with tumor. Each dot shows the value of accuracy in each iteration through the full training dataset.

4. CLASSIFICATION OF IMAGE PATCHES AND TUMOR LOCALIZATION

4.1 Prostate tumor classification network

Deep convolutional neural networks have been progressively applied to computer vision applications like medical image analysis due to their significant performance advantages provided by high-performance computing and the

availability of large datasets[23]. We used a deep residual network (ResNet), a type of convolutional neural network with residual blocks that has shown promising results on both natural image[24] and medical image classification and segmentation tasks[25]. ResNet was implemented in a supervised manner to take in WSI patches of sizes 1024×1024 as input and output a prediction probability for tumor with any histological patterns or no-tumor tissue. ResNet architecture was also initialized on a large dataset of natural images such as ImageNet[26]. This fine-tuning process uses the generic image features from natural images that are fundamental for all images and optimizes them for the variety of medical imaging modalities[27]. Figure 3 shows a boxplot diagram to compare the performance of the pre-trained and fine-tuned ResNet on ImageNet data and the same architecture trained using only WSI images. The distribution of accuracies obtained from the ImageNet pre-trained models have higher minimum and maximum accuracy compared to the models trained on WSI. They also have fewer outliers in their distributions across the full course of training. In summary, pre-trained models using ImageNet dataset outperformed the models trained on only histopathology images.

4.2 Input data preprocessing and model parameters

To offset color differences between WSI slides, all color channel values were normalized to the mean and standard deviation of the entire training set before using WSI patches for training classifiers. We also augmented the training set by performing color jittering on the brightness, contrast, saturation, and hue of each image. The final WSI patches for training consisted of ~60% no-tumor and ~40 % tumor classes. The ResNet network was trained for 50 epochs on the augmented training set, starting with an initial learning rate of 0.001 and decaying by a factor of 0.9 every two epochs. Models used multiclass cross-entropy loss function while training[28]. ResNet-18 was selected over alternative architectures (ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152) since it had the smallest number of parameters that led to the fastest training time and achieved a similar performance.
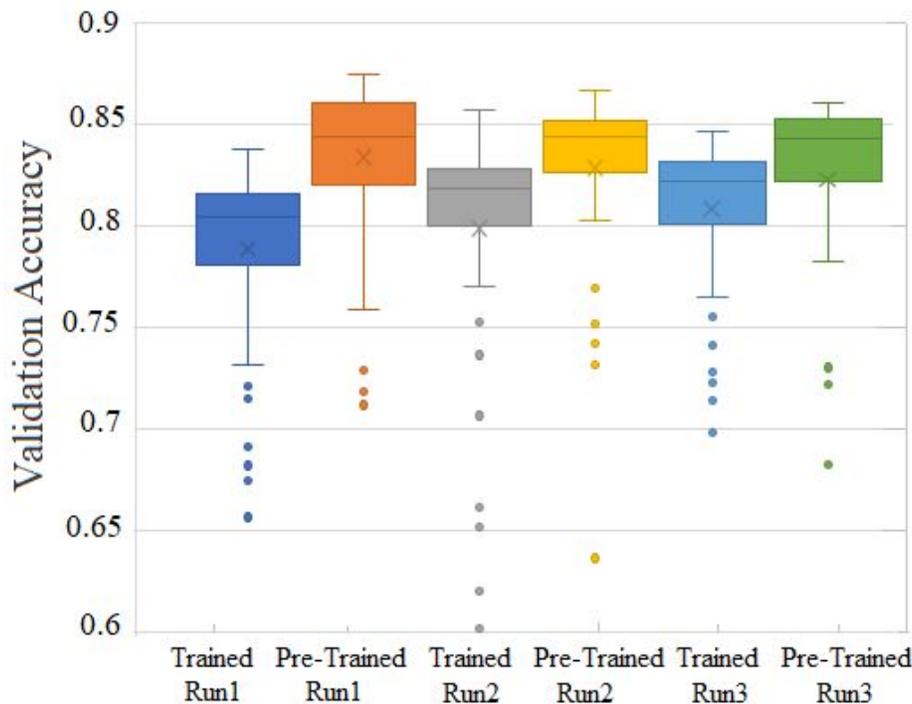

Figure 3. Visual comparison of the performance of ImageNet pre-trained ResNet-18 versus ResNet-18 for tumor classification that has been trained on only prostate histopathology images. In each box, the central mark is the median, the box edges are the 25th and 75th percentiles, the whiskers extend to the most extreme data points, and the outliers show training accuracies in initial iterations. The ResNet-18 classifier network trained for 37 iterations (each dot shows the value of accuracy in each iteration) in three replicas (Run1, Run2, and Run3).

4.3 Classification using H&E dye-stained and computationally H&E stained images

In the first experiment, we used a pre-trained ResNet-18 and retrained it on ground truth 78,764 H&E dye-stained patches with tumor annotations -Classifier I. The mean and standard deviation of the entire training set were used to normalize the color channel values. All patches had fixed length and width of 1024×1024, except patches generated at the edges. Resizing to fixed size was performed by zero-padding rather than a conventional approach of scaling up patches using interpolation. Hashemi et al.[29] have shown that zero-padding has no effect on the classification accuracy while resulting in better computational efficiency because neighboring zero input pixels do not activate their corresponding convolutional filters in a deeper layer. Wang et al.[30] demonstrated the clinical advantage of zero-padding to prevent the risk of deforming the histological patterns in the border patches that may occur with scaling. In the second experiment, illustrated in Figure 4, we trained another Resnet-18 model -Classifier II, using computationally H&E stained images generated by our previously published GAN-CS model[9]. For both classifiers single core non-stained WSIs and their registered H&E dye stained or computationally H&E stained were patched into 1024×1024 sizes, registered and fed into the GAN-CS model. Table 3 lists Classifier I and Classifier II performances using accuracy, precision, recall and F1-score metrics. Classifier II outperformed Classifier I when deployed in the end-to-end framework. This result can be explained based on the fact that the data for training the Classifier II in the end-to-end system comes from the same distribution. Classifier I had better recognition of no-tumor tissue patches than tumor tissue patches (precision>>recall), however, recognition of tumor patches increased significantly in Classifier II. Classifier II was finally used in the end-to-end system.

Table 3. Accuracy, precision, recall, and F1-score over test data with computationally H&E stained images using different configurations where Classifier I and Classifier II integrated in the end-to-end system.

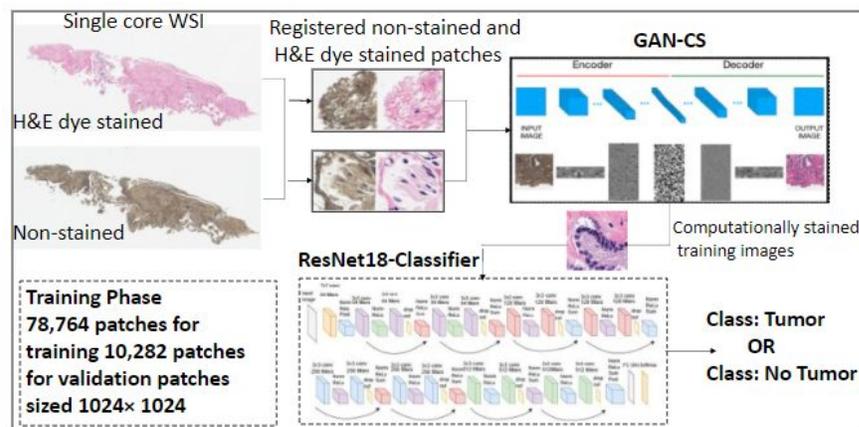| Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Classifier I	79.60	87.63	59.76	71.06
Classifier II	86.37	85.05	81.90	83.44

Figure 4. Block diagram of the Resnet18 Classifier II in the training phase. Single core non-stained WSIs and their registered dye-stained WSI were patched and fed into the GAN-CS model. The computational H&E-stained images were created and used as training and validation inputs to classify each patch as tumor or no-tumor.

4.5 Weakly supervised tumor localization

Gradient Backpropagation (GBP) approach[31] was used to compute a saliency map for localizing the tumor regions. Using this method, by visualizing the activation of the higher layer neurons in the network we could identify which input image pixels are important to output tumor classifications from the ResNet18-model as shown in Figure 4. This was done by adding an additional guidance signal from the higher layers to usual backpropagation. For evaluation of pixel-wise localization of regions of interest using the GBP algorithm, we employed a standard segmentation metric called Dice index[32] (Dice, 1945). Dice index is a measure of agreement or similarity between two sets of samples. Given True Positive pixels (TP) as a set of common pixels belonging to both the ground truth mask and segmentation

mask, False Positive pixels (FP) as a set of pixels belonging to the segmentation mask but ground truth mask, and False Negative pixels (FN) as a set of pixels belonging to the ground truth but segmentation mask, we use a final Dice index (Eq.1) with a small modification as follows:

$$DiceIndex = \frac{TP}{TP + \alpha FP + (1 - \alpha)FN} \tag{1}$$

Dice index ranged in the interval [0, 1], where the higher the value, the more concordant the segmentation and the ground truth. α denotes a coefficient in the interval of [0, 1] where it is usually set to 0.5 that represents a balance between FP and FN. However, this was not aligned with our weakly supervised context and the coarse ground truth annotation of the data that can introduce more FN pixels. To remedy this, we assigned a greater value of 0.7 to α, decreasing the impact of FN pixels for evaluation of the localization performance of the model. The evaluation of the GBP localization technique over our H&E dye stained and computationally stained test datasets are listed in Table 4. The mean Dice index was computed over all tumor patches (aDice index) and Dice index on correctly classified tumor patches (cDice index) was used to measure how well it predicted positive tumor regions. The weakly localization algorithm achieved similar dice indices on both test datasets used for training Classifier I (ground truth H&E dye stained images) and Classifier II (computational H&E dye stained images). The localization results for a sample image that was classified correctly as tumor is shown in Figures 6 and 7. We also trained a U-Net architecture[33] on our H&E dye stained dataset in a fully supervised setting to obtain an upper bound performance in terms of pixel-wise localization. The learning rate was set to 0.1, the batch size was set to 16. The model was trained for 100 epochs and the learning rate was divided by two every five epochs.

Table 4. Mean Dice index across all tumor patches (aDice index) and across all correctly classified tumor patches (cDice index) obtained with weakly supervised localization models and the Gradient Backpropagation for H&E dye stained and computationally stained test datasets.

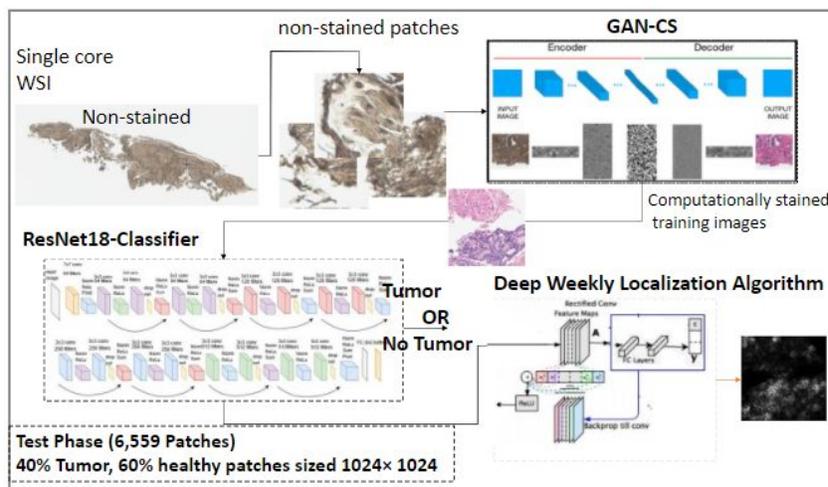| Test Data	aDice index	cDice index
H&E dye stained	60.59±2.20	66.61±2.03
H&E computationally stained	60.51±2.17	65.07±1.99
U-Net	67.31± 2.01	70.24±1.86

Figure 5. Automated end-to-end deep learning system for classification and tumor localization from native non-stained pathology images. (a) The cGan pix2pix-based model (GAN-CS) accepts a native non-stained patch and generates a computationally H&E-stained patch. (b) The classifier takes the computational stained patch image from the GAN-CS model and classifies it as tumor or no-tumor. (c) The deep weakly localization algorithm was then applied to the Resnet-18 classifier model to extract regions of interest (localization map)

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
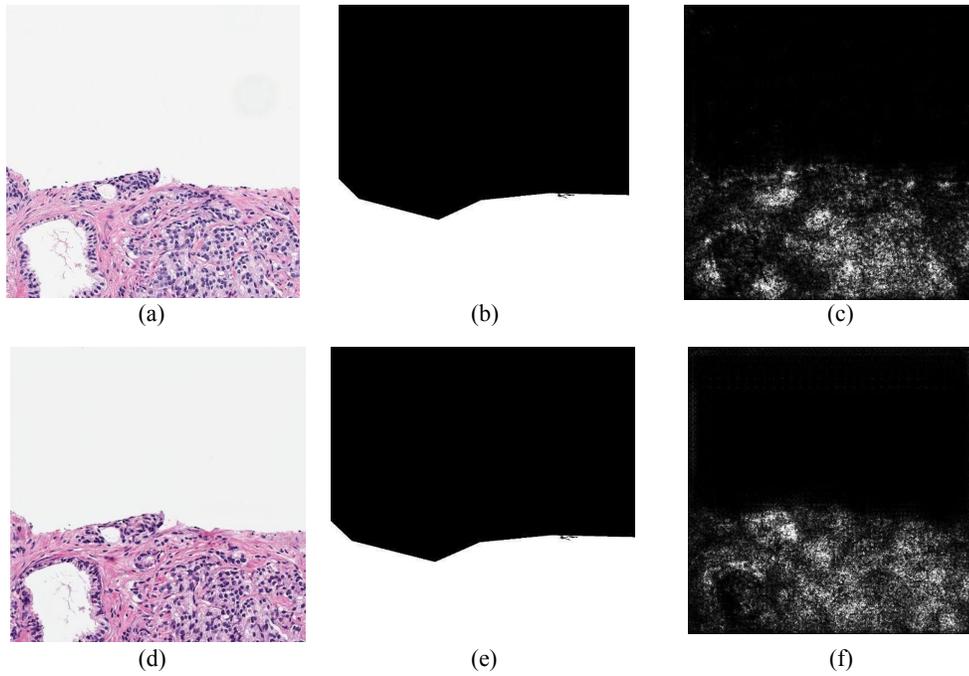
(d)　　　　　　　　　　(e)　　　　　　　　　　(f)

Figure 6. Localization results for sample ground truth H&E dye and computationally H&E stained images that were classified correctly as tumor. Panel (a) H&E dye-stained image, (b) ground truth tumor annotations by pathologist, (c) localization map by using the Gradient Backpropagation. Panels (d), (e) and (f) follow similar order for computationally H&E stained images (bottom row)
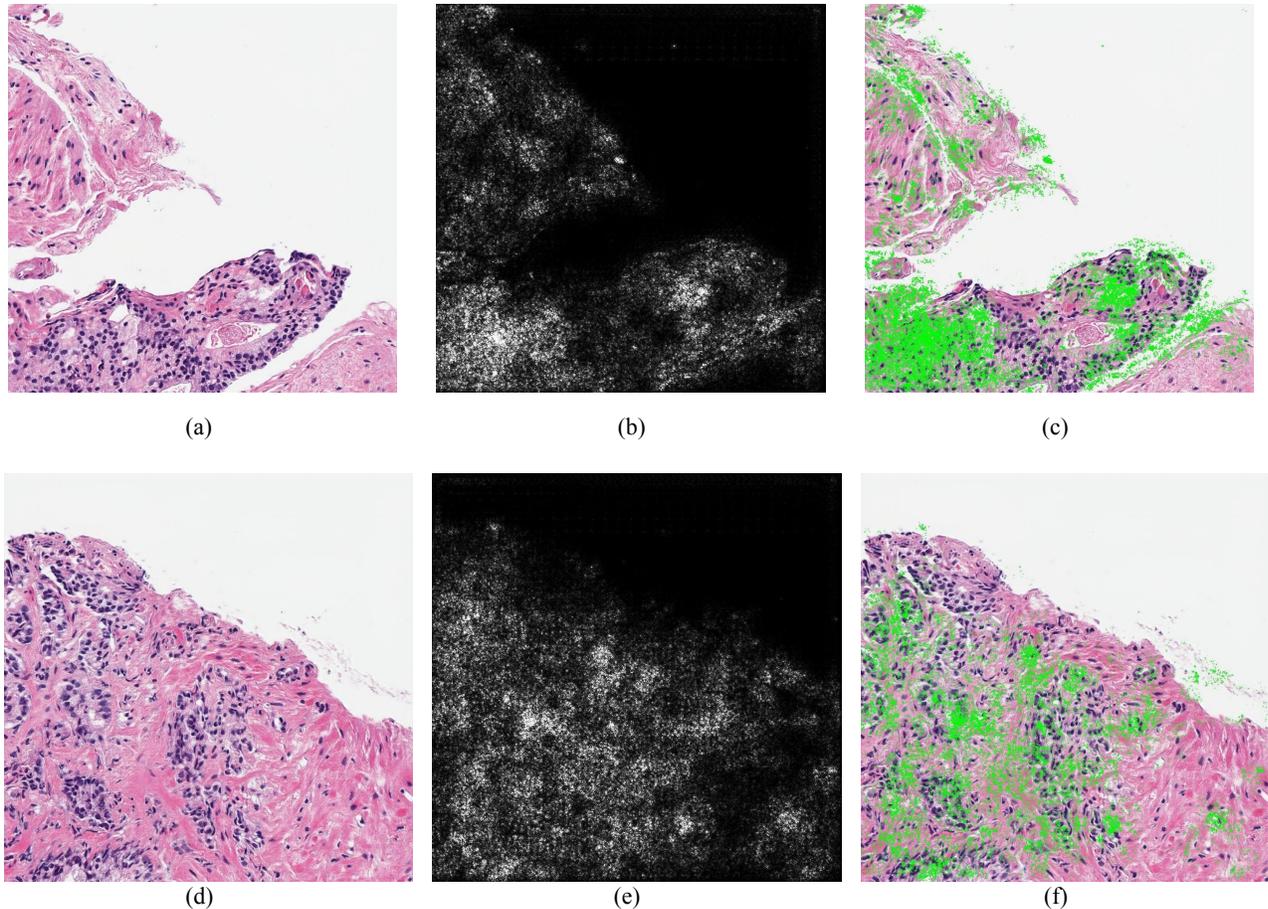
Figure 7. Localizing tumors in prostate core biopsy images by the proposed end-to-end framework described in this study. (a) computational H&E stained images, (b) localization by the Gradient Backpropagation, (c) superimposed image where the regions with strong activations, shown in green, represent tumor regions. Panels (d), (e) and (f) follow similar order as another sample image.

5. RESULTS AND DISCUSSION

We propose an automated end-to-end framework for histopathology image analysis that takes a non-stained image, provides H&E contrast with computational staining technique, and then detects and localizes the histologic patterns of prostate tumors. Inspired by the gold standard diagnosis by pathologist that often look at WSI at low magnification to find potentially important regions and then zoom in to the high magnification to perform more accurate analysis of the tissue structure, we leveraged patch level tumor detection and localization that can provide deep learning models with higher resolution for tumor recognition. We also explored different patch sizes (clinically and computationally) to find suitable sizes for crucial diagnostic information and report that 1024×1024 patch size is the most optimal. We also explored the efficacy of transferring features extracted from the Resnet-18 classifier trained by a large natural image database, ImageNet, to pathology images. Transfer learning made our network activation features more suitable for pathology images by learning subtle features that capture complex clinical representation. Our final strategy involved using a pre-trained network as a feature extractor and then fine-tuning on H&E dye and computationally stained WSI. Furthermore, we investigated the optimal class distribution for tumor and no-tumor classes in our classification task by using random under-sampling technique that involves randomly selecting examples from the no-tumor class as a majority for training. Our results showed that a ratio of no-tumor patches to tumor patches when equal to 1.4 led to the best performance. Increasing and decreasing this ratio negatively impacted the accuracy of the classification. In our final model, data distribution was ~1.49 and we expect our training samples to ideally be representative of the population that the model will be applied to. The ResNet-18 classifier network trained on ~79,000 patches of size

1024×1024 with magnification factor 20X for 50 iterations in three replicas. After augmenting the training dataset by applying flipping and rotation, we achieved performance of 86.37% accuracy and 85.05% precision. The Classifier II outperformed the Classifier I when integrated with Gan-CS in an end-to-end pipeline (Table 3). It had higher sensitivity (Recall) compared to Classifier I that shows lower numbers of tumor patches that are classified as no-tumor (FN). Furthermore, the higher F1 score and accuracy in Classifier II reveal better quality of computational images than H&E dye-stained images in detecting neoplasm regions. The low recall values obtained in all three experiments suggests that coarse/ambiguous labels of WSIs according to the diagnosis without dense annotation generate more false negative samples. The deep weakly supervised GBP model integrated into our end-to-end framework and produces an activation map where high activation responses correspond to image region of interest without the need for pixel-level annotation. From a clinical perspective, pixel-wise region localization can provide a more accurate and visual explanatory factor for classification of images to tumor or no-tumor. Extracting regions of interest as highly desirable property in our end-to-end pipeline, for instance, can be used by the pathologist or be further inspected if they indicate cancerous regions.

6. CONCLUSIONS

This study reports an end-to-end deep learning framework for virtual H&E staining, automatic classification, and localization of prostate tumors from non-stained core biopsy images. Three deep learning methods, the pix2pix, ResNet, and GBP, were trained and validated for high performance. The proposed end-to-end system consists of the GAN-CS model, the ResNet-18 classifier, and the deep weekly supervised learning model. A computationally H&E stained patch was first generated from a non-stained input image using the GAN-CS model, and then was fed into a Resnet-18 classifier for classification as tumor or no-tumor. A deep weekly-supervised learning GBP algorithm was used to localize class-specific (tumor) regions on images outputted from the Resnet-18 classifier. If an input image patch was classified as tumor, the GBP localization module would generate a saliency map locating the tumor regions on computationally stained images. The proposed end-to-end framework design makes optimal use of the modular design with the simple strategy to train its modules independently and structure their connectivity which is far more efficient than end-to-end central learning that can push network architecture to keep growing more and more complex in our task. In this way, the system maintains the valuable information contained in each individual component that could be ignored during end-to-end training. From the clinical perspective, fulfilling automated diagnosis requirements (staining, classification, segmentation) in one place, can help the diagnosis and treatment of prostate cancer in the early stages by improving consistency and speed in diagnosis. Our pixel-wise region localization capability provides accurate visual explanations for the classification of images as tumor or no-tumor. To the best of our knowledge, this is the first work attempting to discriminate tumors as small as 1024×1024 pixels from non-cancerous ones in the prostate using a non-stained whole slide image. Our core contributions are to extend the utility and performance of generative virtual H&E staining deep learning methods and models. We also extend the utility of computationally H&E stained images for the medical imaging community to use them for tumor localization and classification.

REFERENCES

[1] Cornish, T.C., Swapp, R.E. and Kaplan, K.J., "Whole-slide imaging: routine pathologic diagnosis," Advances in anatomic pathology 19(3), 152-159 (2012).

[2] Webster, J.D. and Dunstan, R.W., "Whole-slide imaging and automated image analysis: considerations and opportunities in the practice of pathology," Veterinary pathology, 51(1), 211-223 (2014).

[3] Mosquera-Lopez, C., Agaian, S., Velez-Hoyos, A. and Thompson, I., "Computer-aided prostate cancer diagnosis from digitized histopathology: a review on texture-based systems," IEEE reviews in biomedical engineering, 8,

98-113 (2014).

[4] Bayramoglu, N., Kaakinen, M., Eklund, L. and Heikkila, J., "Towards virtual H&E staining of hyperspectral lung histology images using conditional generative adversarial networks," In Proceedings of the IEEE International Conference on Computer Vision Workshops, 64-71 (2017).

[5] Howat, W.J. and Wilson, B.A., "Tissue fixation and the effect of molecular fixatives on downstream staining procedures," Methods, 70(1), 12-19 (2014).

[6] Blutke, A. Wanke, R. "Sampling strategies and processing of Biobank tissue samples from porcine biomedical models," JoVE (Journal of Visualized Experiments), 6(133): e57276 (2018).

[7] Alturkistani, H.A. Tashkandi, F.M. Mohammedsaleh, Z.M., "Histological stains: a literature review and case study," Global journal of health science, 8(3):72 (2016).

[8] Rana, A. Yauney, G. Lowe, A. Shah, P., "Computational Histological Staining and Destaining of Prostate Core Biopsy RGB Images with Generative Adversarial Neural Networks," 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 828-834 (2018).

[9] Rana, A., Lowe, A., Lithgow, M., Horback, K., Janovitz, T., Da Silva, A., Tsai, H., Shanmugam, V., Bayat, A. and Shah, P., "Use of deep learning to develop and analyze computational hematoxylin and eosin staining of prostate core biopsy images for tumor diagnosis," JAMA network open, 3(5), e205111-e205111 (2020).

[10] Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M. and Yener, B., "Histopathological image analysis: a review," IEEE reviews in biomedical engineering,2: 147–71 (2009).

[11] Bayat, A. and Pomplun, M., "Deriving high-level scene descriptions from deep scene cnn features," In 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), IEEE, 1-6 (2017).

[12] da Fonseca, P.C., He, J. and Morris, E.P., "Molecular model of the human 26S proteasome," Molecular cell," 46(1), 54-66 (2012).

[13] Dimitriou N., Arandjelović O., Caie PD., "Deep Learning for Whole Slide Image Analysis: An Overview," Frontiers in Medicine, 6 (2019).

[14] Xu, Y., Jia, Z., Wang, L.B., Ai, Y., Zhang, F., Lai, M., Eric, I. and Chang, C., "Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features," BMC bioinformatics, 18(1), 1-17 (2017).

[15] Krizhevsky, A., Sutskever, I. and Hinton, G.E., "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, 25,1097-1105 (2012).

[16] Otálora, S., Atzori, M., Andrearczyk, V., Khan, A. and Müller, H., "Staining invariant features for improving generalization of deep convolutional neural networks in computational pathology," Frontiers in bioengineering and biotechnology, 7, 198 (2019).

[17] Ciompi, F., Geessink, O., Bejnordi, B.E., De Souza, G.S., Baidoshvili, A., Litjens, G., Van Ginneken, B., Nagtegaal, I. and Van Der Laak, J., "The importance of stain normalization in colorectal tissue classification with convolutional networks," IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) 160-163 (2017).

[18] Van Eycke, Y.R., Balsat, C., Verset, L., Debeir, O., Salmon, I. and Decaestecker, C., "Segmentation of glandular epithelium in colorectal tumours to automatically compartmentalise IHC biomarker quantification: A deep learning approach," Medical image analysis, 49, 35-45 (2018).

[19] Venson, Y., Wang, H., Wei, Z., de Haan, K., Zhang, Y., Wu, Y., Günaydın, H., Zuckerman, J.E., Chong, T., Sisk, A.E. and Westbrook, L.M., Ozcan, A., "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nature biomedical engineering, 3(6), 466-477(2019).

[20] Cruz-Roa, A., Basavanhally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., Shih, N., Tomaszewski, J. and Madabhushi, A., "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," Medical Imaging 2014: Digital Pathology Vol. 9041, 904103 (2014).

[21] Bayat, A., Do Koh, H., Kumar Nand, A., Pereira, M. and Pomplun, M., "Scene grammar in human and machine recognition of objects and scenes," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1992-1999 (2018).

[22] Bayat, A., Bayat, A.H. and Ghasemi, A.S., "Classifying human walking patterns using accelerometer data from smartphone," Int. J. Comput. Sci. Mobile Comput, 3, 1-6 (2017).

[23] Shah, P., Kendall, F., Khozin, S. et al. Artificial intelligence and machine learning in clinical development: a translational perspective. npj Digit. Med. 2, 69 (2019)

[24] He, K., Zhang, X., Ren, S. and Sun, J., "Deep residual learning for image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition 770-778 (2016).

[25] Wei, J.W., Tafe, L.J., Linnik, Y.A., Vaickus, L.J., Tomita, N. and Hassanpour, S., "Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks," Scientific reports, 9(1), 1-8 (2019).

[26] ImageNet, http://imagenet.stanford.edu/about-overview.

[27] Kumar, A., Kim, J., Lyndon, D., Fulham, M. and Feng, D., "An ensemble of fine-tuned convolutional neural networks for medical image classification," IEEE journal of biomedical and health informatics, 21(1), 31-40 (2016).

[28] Brownlee, J., "A Gentle Introduction to Cross-Entropy for Machine Learning," https://machinelearningmastery.com/cross-entropy-for-machine-learning/ (22 December 2020)

[29] Hashemi, M., "Enlarging smaller images before inputting into convolutional neural networks: zero-padding vs. interpolation," Journal of Big Data, 6(1), 1-13 (2019).

[30] Wang, S., Yang, D.M., Rong, R., Zhan, X. and Xiao, G., "Pathology image analysis using segmentation deep learning algorithms," The American journal of pathology, 189(9), 1686-1698 (2019).

[31] Springenberg, J.T., Dosovitskiy, A., Brox, T. and Riedmiller, M., "Striving for simplicity: The all convolutional net," arXiv preprint arXiv:1412.6806 (2014).

[32] "Dice coefficient," https://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%93Dice_coefficient.

[33] Ronneberger, O., Fischer, P. and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," International Conference on Medical image computing and computer-assisted intervention 234-241. Springer, Cham (2015).