Quantifying the effect of color processing on blood and damaged tissue detection in Whole Slide Images

Neel Kanwal*, Saul Fuster*, Farbod Khoraminia†, Tahlita C.M. Zuiverloon†, Chunming Rong*, Kjersti Engan*
*Department of Electrical Engineering and Computer Science, University of Stavanger, Norway
†Department of Urology, Erasmus Cancer Institute, Rotterdam, Netherlands

***Abstract*—Histological tissue examination has been a long-standing practice for cancer diagnosis where pathologists identify the presence of tumors on glass slides. Slides acquired from laboratory routine may contain unintentional artifacts due to complications in surgical resection. Blood and damaged tissue artifacts are two common problems associated with transurethral resection of the bladder tumor. Differences in histotechnical procedures among laboratories may also result in color variations and minor inconsistencies in outcome. A digitized version of a glass slide known as a whole slide image (WSI) holds enormous potential for automated diagnostics. The presence of irrelevant areas in a WSI undermines diagnostic value for pathologists as well as computational pathology (CPATH) systems. Therefore, automatic detection and exclusion of diagnostically irrelevant areas may lead to more reliable predictions. In this paper, we are detecting blood and damaged tissue against diagnostically relevant tissue. We gauge the effectiveness of transfer learning against training from scratch. Best models give 0.99 and 0.89 F1 scores for blood and damaged tissue detection. Since blood and damaged tissue have subtle color differences, we assess the impact of color processing methods on the binary classification performance of five well-known architectures. Finally, we remove the color to understand its importance against morphology on classification performance.**

***Keywords*—Digital Pathology, Whole Slide Image, Preprocessing, Artifacts, Deep Learning, Blood, Damaged Tissue**

I. Introduction

Histopathological examination has been a gold standard practice for cancer diagnosis [1]. Bladder cancer diagnosis is performed by retrieving a tissue sample by transurethral resection of the bladder tumor. Preparing a glass slide for microscopic inspection requires several tissue processing steps. Producing slides of good quality require histotechnical competence and extensive care in handling the extracted tissue [2]. However, extraneous factors could introduce unintentional artifacts and variations in a small or large portion of the slide. Blood and tissue artifacts such as damaged areas result from complicated specimen collection procedures [2]. Blood clots on a slide appear as a result of hemorrhage and do not reveal cancer information. Damaged and burnt tissue is often unreliable to observe histologic features, and pathologists do not extract any diagnostic or prognostic information from these areas [3].

Digital pathology (DP) brings us to unveil the benefits of deep learning (DL) in pathology. A digitized high-resolution version of a glass slide is known as a Whole Slide Image (WSI). A WSI is a pyramid container pre-stored at several magnification levels to emulate conventional microscopy and facilitates rapid zoom in and out. WSIs are information-rich and contain the potential for automating different tasks. Despite the benefits of automated WSI analysis, artifacts transferred to digitized glass slides may adversely affect the overall performance of computational pathology (CPATH) systems [4], [5]. Manual quality control (QC) can ensure the diagnostic relevance of a WSI but can be subjective and laborious at the same time. Hence, WSI preprocessing can play a vital role in automatically excluding irrelevant areas before running a diagnostic or prognostic algorithm. In addition, artifacts detection methods can find defects at regional levels in a WSI and benefit existing QC methods [6], [7] that focus only on metadata, contrast, overall sharpness, and noise for evaluating WSI quality.

Damaged tissue artifacts and blood can be detected due to notable differences in color and morphology. Some recent works in DL for CPATH systems involve manual selection of a region of interest (ROI) for developing diagnostic models [8], [9]. Although manual selection provides diagnostically relevant regions, it curbs the objectivity and speed of the process [10]. Bahlmann *et al.* [11] used color deconvolution to separate epithelial regions from cytoplasm tissue, benefiting from the difference in absorption of Hematoxylin and Eosin (H&E) stains. Extending a similar idea, Mercan *et al.* [12] developed binary logistic models and color histograms to localize diagnostically relevant areas based on texture features. These methods did not consider the effects of artifacts on their final goal. There are, however, some works on detecting blood and damaged areas; Chadaj *et al.* [13] proposed a method for detecting blood clots using decision tree classifiers and color deconvolution on hemorrhoid specimens. Clymer *et al.* [14] in their three-stage framework, detected blood vessels at low magnification using hierarchical Convolutional Neural Networks (CNN). In another study, Chadaj *et al.* [15] segmented damaged tissue areas in immunohistochemically stained brain tumor slides using the U-Net model. The limitation of that work with immunohistochemical staining was overcomed by Wetteland *et al.* [3], where damaged tissue and blood patches from HES WSIs were detected using a multi-scale VGG16.

Color variations are common due to inconsistencies in the staining procedure among laboratories and the color response of WSI scanning hardware [16]. Color normalization is a common preprocessing step in the CPATH system for reducing the altered color appearance [17]. Since color

This work is financially supported by CLARIFY Project. CLARIFY is European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 860627.

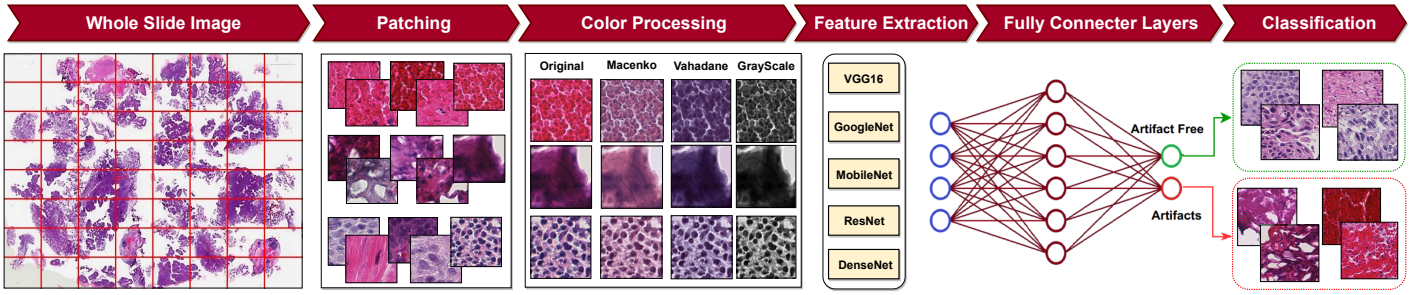

Figure 1. An overview of our proposed blood and damaged tissue detection method. Patches are extracted from a WSI using a predefined size. Various color processing schemes are applied to the obtained dataset before extracting features. A classifier with three FC layers is added after each CNN to evaluate their ability of finding hidden representations. Finally, blood and damaged tissue are classified against diagnostically relevant tissue.

processing techniques change the appearance, their position in the preprocessing pipeline is debatable. Color processing applied before artifacts detection may ultimately influence the appearance of blood and damaged areas. Several CPATH studies perform color normalization on a controlled dataset to improve overall classification performance [16], [18], [19]. However, the effect of color processing on *artifacts* is often overlooked. Color normalization methods are reported to yield better performance, but how they affect artifacts detection is not yet discussed in the literature.

In this paper, we are performing binary blood and damage tissue classification on H&E WSI using five well-known CNN architectures. We report the performance of models trained from scratch against transfer learning setting. Color might be an important factor for detecting blood and damaged tissue artifacts; therefore, we assess the effect of color processing methods on model performance. Finally, we remove color to quantify the importance of color against morphology in blood and damaged tissue detection. An overview of our pipeline can be seen in Figure 1.

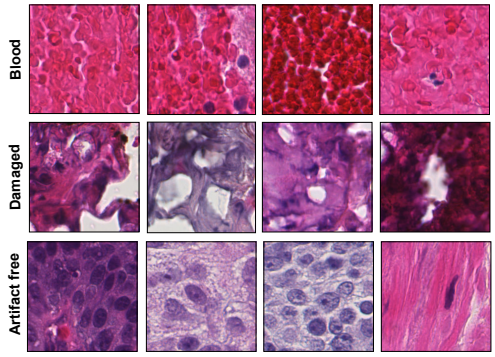

Figure 2. Examples of blood, damaged, and artifact-free patches in our dataset. Blood and damaged patches have subtle color and morphological differences compared to normal (artifact-free) tissue.

Table I. Number of patches for blood, damaged and artifact-free classes.

Region	Training	Validation	Test	Total
Blood	16,743	4,86	1,409	*22,338*
Damaged	2,577	332	1,023	*3,932*
Artifact-free	5,249	1,441	2,144	*8,834*

II. Data Materials

We analyzed 55 WSIs of bladder tumor resection specimens from Erasmus MC, Rotterdam, The Netherlands. These images are H&E stained and scanned at 40x using a Hamamatsu Nanozoomer. WSIs are in *ndpi* format with pixel size of 0.227 µm x 0.227 µm. A split of 35/10/10 WSIs is used for a training, validation, and test set. WSIs were manually annotated by non-pathologist for air bubbles, blood, damage, fold, and blur, and the rest of the tissue is considered artifact-free. Since WSIs are gigapixel images, they are too large to fit into the memory at once; thus, we split the images into sub-images (patches) of 224×224 pixels at 40x magnification. Figure 2 shows some examples of extracted patches for three classes. Table I further shows the breakdown of the total number of patches extracted for each subset.

III. Method

Figure 1 provides a graphical overview of our proposed blood and damaged tissue detection method. Background and foreground are separated using Otsu thresholding. Patches from a WSI foreground are extracted for desired classes using an 80% overlap between the grid and annotation mask. The Histolab [20] library is used to extract patches from a WSI. We have created two color normalized and a gray-scale version of the obtained dataset to train CNN architectures for various experimental setups. VGG16 [21], GoogleNet [22], MobileNet [23], ResNet [24], and DenseNet [25] are chosen CNN candidates based on their benchmark performance on ImageNet [26] dataset. Our proposed system uses these CNN as backbones by removing fully connected (FC) layers and using them as feature extractors. A custom classifier with three FC layers is added to all architectures.

A. Color Processing

Three different color processing schemes are tested; Normalization by Macenko [27] and Vahadane [16], two popular color deconvolution-based normalization methods, and reducing color influence by converting RGB patches to gray-scale. The latter can be used to investigate if the models relies more on morphology instead of colorimetric features for discriminating blood and damaged patches from artifact-free regions. The color processing stage in Figure 1 depicts example outcome of each scheme on three classes.

Let $I \in R^{m \times n}$ be a vectorized image normalized to [0,1], where m =3 (color channels) and n=number of pixels. The

Table II. Results from experiment no. 1: our proposed method is tested with different feature extractors (as architecture backbone) with no color processing applied. We compare architectures' performance when feature extractors are trained with random weights (scratch) against ImageNet weights (transfer learning).

Feature Extractors		**Damaged**					**Blood**				
		Validation			*Test*		*Validation*			*Test*	
		Acc.	F1	MCC	F1	MCC	Acc.	F1	MCC	F1	MCC
Scratch	VGG16	94.07	0.83	0.80	0.8	0.66	**98.41**	**0.98**	**0.95**	**0.99**	**0.97**
	MobileNet	92.66	0.8	0.76	0.82	0.69	94.95	0.96	0.87	0.99	0.97
	GoogleNet	94.24	0.82	0.80	**0.84**	**0.70**	94.06	0.95	0.86	0.99	0.97
	ResNet	92.49	0.79	0.75	0.75	0.6	94.01	0.95	0.85	0.99	0.96
	DenseNet	**94.36**	**0.83**	**0.81**	0.76	0.62	96.39	0.97	0.93	0.99	0.93
Transfer Learning	VGG16	95.6	0.87	0.85	0.86	0.77	**99.39**	**0.995**	**0.98**	0.997	0.98
	MobileNet	**95.71**	**0.88**	**0.86**	**0.89**	**0.81**	98.91	0.992	0.97	0.997	0.98
	GoogleNet	94.25	0.84	0.81	0.89	0.79	98.64	0.990	0.96	0.994	0.97
	ResNet	95.37	0.86	0.84	0.85	0.75	99.04	0.993	0.97	0.994	0.96
	DenseNet	93.68	0.81	0.78	0.88	0.79	98.56	0.990	0.96	0.997	0.98

Table III. Results from experiments no. 2 and 3: our proposed method is tested with different feature extractors (as architecture backbone) with three color processing applied. Outcomes are compared with experiment no. 1 to quantify the effect of color processing in blood and damaged tissue detection.

Feature Extractors		**Damaged**					**Blood**				
		Validation			*Test*		*Validation*			*Test*	
		Acc.	F1	MCC	F1	MCC	Acc.	F1	MCC	F1	MCC
Macenko Normalization	VGG16	**95.43**	**0.87**	0.84	**0.93**	**0.86**	**98.89**	0.99	**0.97**	0.99	0.92
	MobileNet	94.81	0.85	0.82	0.91	0.83	98.50	0.98	0.96	0.99	0.94
	GoogleNet	94.02	0.81	0.79	0.88	0.78	97.81	0.98	0.94	0.99	0.94
	ResNet	95.38	0.87	**0.85**	0.90	0.82	98.73	0.99	0.96	0.99	0.94
	DenseNet	94.07	0.81	0.79	0.83	0.72	98.66	0.99	0.96	0.99	0.93
Vahadane Normalization	VGG16	93.45	0.80	0.77	0.91	0.83	**99.09**	0.99	0.97	0.99	0.97
	MobileNet	**95.54**	**0.87**	0.84	0.88	0.80	98.17	0.99	0.96	0.99	0.97
	GoogleNet	93.01	0.78	0.84	0.88	0.79	98.93	0.99	0.97	0.99	0.95
	ResNet	94.36	0.84	**0.85**	**0.95**	**0.90**	98.73	0.99	0.96	0.99	0.96
	DenseNet	92.7	0.77	0.77	0.82	0.72	98.66	0.99	0.96	0.99	0.96
Gray-scale Transformation	VGG16	93.90	0.81	0.79	**0.89**	**0.80**	**99.07**	0.99	**0.98**	0.99	0.95
	MobileNet	**96.05**	**0.88**	**0.86**	0.88	0.79	98.93	0.99	0.97	0.99	0.94
	GoogleNet	92.72	0.77	0.74	0.88	0.79	98.25	0.98	0.95	0.99	0.96
	ResNet	93.28	0.78	0.76	0.86	0.77	99.04	0.99	0.97	0.99	0.95
	DenseNet	93.11	0.78	0.76	0.87	0.78	98.91	0.99	0.97	0.99	0.96

Beer-Lambert law states $I = I_0 exp(WH)$, where $W \in R^{m \times s}$ is the stain color apperance matrix (3×2), and $H \in R^{s \times n}$ is the stain density map for each pixel ($2\times$ number of pixels). The optical density (OD) space makes the linear relationship with the combination of stains, and can be defined as:

$$OD = -\log_{10}(I) = WH \tag{1}$$

When color normalization is performed based on a reference image, the W_{ref} is used instead of the W when transforming back to RGB space. The main difference in many color deconvolution and/or normalization methods relies in how W and H are estimated from the OD image.

1) Macenko Color Normalization: The Macenko's method [27] is based on singular value decomposition (SVD) of the OD transformed image. The two largest singular values are used to create a plane representing the W, and data is represented and normalized on this plane.

2) Vahadane Color Normalization: Vahadane's structure-preserving technique [16] is also based on the OD space. It adds a sparseness constraint on each row of H from the assumption that only either eosin or hematoxylin is present at a given pixel.

3) Gray-scale Transformation: Input RGB patch (I) is converted to grays-scale (L) mode using ITU-R 601-2 luma transform, defined as $L = 0.2989R + 0.5870G + 0.1140B$.

B. Evaluation Metrics

Let TP, TN, FP, FN denote true positive, true negative, false positive, and false negative, respectively. Then, Precision = $TP/(TP + FP)$ and Recall = Sensitivity = $TP/(TP + FN)$. We report F1 score defined as; $F1 = 2 \cdot (\text{precision} \cdot \text{recall})(\text{precision} + \text{recall})$ and the Mather Correlation Coefficient (MCC) as defined in Eq 2. MCC returns a higher value $\in \{-1, 1\}$ when a binary predictor detects majority of positive and negative instances correctly. MCC provides a reliable statistical measure in a binary classification task with imbalanced dataset which reflects most of the medical scenarios [28]. Finally, we provide the receiver operating characteristics (ROC) curve, which highlights classification performance on all classification thresholds.

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \tag{2}$$

C. Experimental Setup

Our experimental setup was built on Pytorch [29] DL framework. Color normalization was implemented on extracted patches using Histocartography library [30]. Several hyperparameters were explored using grid-search. Final parameters were fixed to a batch size of 32, *SGD* optimizer, *ReduceLROnPlateau* scheduler initialized with the learning rate of 0.01, dropout of 0.2, focal loss [31] to cope with the class imbalance and early stopping with the patience of 10 epochs on validation loss to avoid over-fitting.

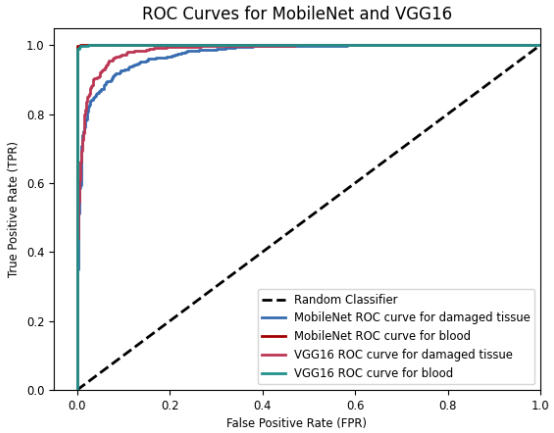

Figure 3. Receiver operating characteristics (ROC) curves for MobileNet and VGG16 for blood and damaged tissue detection.

IV. Experimental Results

We have conducted several experiments to quantify CNN performance in detecting blood and damaged tissue artifacts. We use model weights with the lowest validation loss during the training process to report evaluation metrics on validation and test sets.

A. Experiment 1

In this experiment, we compare all architectures' classification performance when trained from scratch against transfer learning. Table II shows learning with different feature extractors (as architecture backbone) initialized with random (scratch setting) and ImageNet weights (transfer learning setting). MobileNet yields the best results for damaged tissue detection in the transfer learning setting. It improves MCC and F1 against its counterpart in the scratch setting by 10% and 8% on the validation set, respectively. For blood detection, VGG16 is outperforming all other feature extractors in both settings. F1 and MCC scores are higher in blood detection than damaged tissue detection, which can be due to visible morphological and color differences of blood against artifact-free regions. Figure 3 further manifests the ability of MobileNet and VGG16 for damaged tissue and blood detection.

Training from scratch might give good results for some feature extractors (i.e., DenseNet for damaged tissue detection) since the training set is not large enough, and architectures can be over-trained on validation data but fail to perform well on unseen data. Overall, architectures trained with transfer learning setting converge better and take less training time than their counterparts in the scratch setting. Therefore, we use transfer learning setting to continue with our next experiment.

B. Experiment 2

The purpose of experiment 2 is to quantify the effect of color processing on detecting blood and damaged tissue, which we know has different colors than normal tissue. We have reported the classification performance of all architectures using Vahadane and Macenko color normalized versions of the dataset in Table III. For damaged tissue, architectures do not improve with color normalized datasets compared to the original dataset. Except for VGG16, which generalizes well for the test set with both color processing schemes. In the case of blood detection, there is a slight performance drop with all feature extractors in both cases. Overall, Vahadane color normalization yields better results in the blood detection task than Macenko color normalization.

Color normalization schemes increase the computational complexity of the task and result in little to no improvement compared to experiment 1. It could be due to the fact that color processing heavily affects blood and damaged regions as shown in the color processing stage of Figure 1. Therefore, architectures might rely on morphology to distinguish blood and damaged regions more than the color.

C. Experiment 3

In order to further explore the importance of color, we removed the color of the dataset by transforming RGB patches to grayscale. In the last row of Table III, we can see more or less similar classification performance for damaged tissue detection. For blood detection, most of the feature extractors yield little detrimental effect on MCC values compared to the transfer learning setting in Table II. Therefore, color is far less critical than tissue morphology in detection of blood and damaged tissue artifacts.

V. Conclusion and Future Work

WSI preprocessing is promising for CPATH systems as well as QC tools. Abnormalities in the data may add irrelevant morphological features and pose a greater risk of inaccurate predictions; thus, automatically excluding artifacts and diagnostically irrelevant areas can be important for CPATH systems. We have demonstrated the ability of several CNN architectures in detecting blood and damaged tissue from H&E WSI. MobileNet and VGG16 have shown to be effective backbones for detecting damaged tissue artifacts and blood, respectively.

Classification of blood against diagnostically relevant areas is less challenging than damaged tissue. CNN architectures rely more upon morphological features than color to differentiate between these regions. Therefore, instead of improving detection performance for artifacts, color normalization methods only enhance the computational complexity of the task. Artifact detection can be performed before color normalization in preprocessing pipeline of CPATH systems. This work can be further extended to improve QC methods and detect other artifacts such as blur, folded regions, and air bubbles.

REFERENCES

[1] S. A. Taqi, S. A. Sami, L. B. Sami, and S. A. Zaki, "A review of artifacts in histopathology," *Journal of oral and maxillofacial pathology: JOMFP*, vol. 22, no. 2, p. 279, 2018.

[2] V. Rastogi, N. Puri, S. Arora, G. Kaur, L. Yadav, and R. Sharma, "Artefacts: a diagnostic dilemma–a review," *Journal of clinical and diagnostic research: JCDR*, vol. 7, no. 10, p. 2408, 2013.

[3] R. Wetteland, K. Engan, T. Eftestøl, V. Kvikstad, and E. A. Janssen, "A multiscale approach for whole-slide image segmentation of five tissue classes in urothelial carcinoma slides," *Technology in Cancer Research & Treatment*, vol. 19, p. 1533033820946787, 2020.

[4] S. Morales, K. Engan, and V. Naranjo, "Artificial intelligence in computational pathology – challenges and future directions," *Digital Signal Processing*, vol. 119, p. 103196, 2021.

[5] G. Campanella, M. G. Hanna, L. Geneslaw, A. Miraflor, V. Werneck Krauss Silva, K. J. Busam, E. Brogi, V. E. Reuter, D. S. Klimstra, and T. J. Fuchs, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature medicine*, vol. 25, no. 8, pp. 1301–1309, 2019.

[6] M. S. Hossain, T. Nakamura, F. Kimura, Y. Yagi, and M. Yamaguchi, "Practical image quality evaluation for whole slide imaging scanner," in *Biomedical Imaging and Sensing Conference*, vol. 10711. International Society for Optics and Photonics, 2018, p. 107111S.

[7] P. Shrestha, R. Kneepkens, J. Vrijnsen, D. Vossen, E. Abels, and B. Hulsken, "A quantitative approach to evaluate image quality of whole slide imaging scanners," *Journal of pathology informatics*, vol. 7, 2016.

[8] B. M. Priego-Torres, D. Sanchez-Morillo, M. A. Fernandez-Granero, and M. Garcia-Rojo, "Automatic segmentation of whole-slide h&e stained breast histopathology images using a deep convolutional neural network architecture," *Expert Systems With Applications*, vol. 151, p. 113387, 2020.

[9] J. Urdal, K. Engan, V. Kvikstad, and E. A. Janssen, "Prognostic prediction of histopathological images by local binary patterns and rusboost," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 2349–2353.

[10] S. Kothari, J. H. Phan, and M. D. Wang, "Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade," *Journal of pathology informatics*, vol. 4, 2013.

[11] C. Bahlmann, A. Patel, J. Johnson, J. Ni, A. Chekkoury, P. Khurd, A. Kamen, L. Grady, E. Krupinski, A. Graham *et al.*, "Automated detection of diagnostically relevant regions in h&e stained digital pathology slides," in *Medical Imaging 2012: Computer-Aided Diagnosis*, vol. 8315. International Society for Optics and Photonics, 2012, p. 831504.

[12] E. Mercan, S. Aksoy, L. G. Shapiro, D. L. Weaver, T. Brunye, and J. G. Elmore, "Localization of diagnostically relevant regions of interest in whole slide images," in *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 1179–1184.

[13] Z. Swiderska-Chadaj, T. Markiewicz, S. Cierniak, and R. Koktysz, "Automatic quantification of vessels in hemorrhoids whole slide images," in *2016 17th International Conference Computational Problems of Electrical Engineering (CPEE)*. IEEE, 2016, pp. 1–4.

[14] D. Clymer, S. Kostadinov, J. Catov, L. Skvarca, L. Pantanowitz, J. Cagan, and P. LeDuc, "Decidual vasculopathy identification in whole slide images using multiresolution hierarchical convolutional neural networks," *The American Journal of Pathology*, vol. 190, no. 10, pp. 2111–2122, 2020.

[15] Z. Swiderska-Chadaj, T. Markiewicz, J. Gallego, G. Bueno, B. Grala, and M. Lorent, "Deep learning for damaged tissue detection and segmentation in ki-67 brain tumor specimens based on the u-net model," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, pp. 849–856, 2018.

[16] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE transactions on medical imaging*, vol. 35, no. 8, pp. 1962–1971, 2016.

[17] F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. van der Laak, and P. H. de With, "Stain normalization of histopathology images using generative adversarial networks," in *2018 IEEE 15th International symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 573–577.

[18] F. Pérez-Bueno, M. López-Pérez, M. Vega, J. Mateos, V. Naranjo, R. Molina, and A. K. Katsaggelos, "A tv-based image processing framework for blind color deconvolution and classification of histological images," *Digital Signal Processing*, vol. 101, p. 102727, 2020.

[19] D. Tellez, G. Litjens, P. Bándi, W. Bulten, J.-M. Bokhorst, F. Ciompi, and J. Van Der Laak, "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology," *Medical image analysis*, vol. 58, p. 101544, 2019.

[20] R. Colling, H. Pitman, K. Oien, N. Rajpoot, P. Macklin, C.-P. A. in Histopathology Working Group, V. Bachtiar, R. Booth, A. Bryant, J. Bull *et al.*, "Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice," *The Journal of pathology*, vol. 249, no. 2, pp. 143–150, 2019.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[23] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan *et al.*, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1314–1324.

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

[25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[26] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[27] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *2009 IEEE international symposium on biomedical imaging: from nano to macro*. IEEE, 2009, pp. 1107–1110.

[28] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation," *BMC genomics*, vol. 21, no. 1, pp. 1–13, 2020.

[29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035.

[30] G. Jaume, P. Pati, V. Anklin, A. Foncubierta, and M. Gabrani, "Histocartography: A toolkit for graph analytics in digital pathology," in *MICCAI Workshop on Computational Pathology*. PMLR, 2021, pp. 117–128.

[31] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.