

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Predicting MYC translocation in HE specimens of diffuse large B-cell lymphoma through deep learning

Swiderska-Chadaj, Zaneta, Hebeda, Konnie, van den Brand, Michiel, Litjens, Geert

Zaneta Swiderska-Chadaj, Konnie Hebeda, Michiel van den Brand, Geert Litjens, "Predicting MYC translocation in HE specimens of diffuse large B-cell lymphoma through deep learning," Proc. SPIE 11320, Medical Imaging 2020: Digital Pathology, 1132010 (16 March 2020); doi: 10.1117/12.2549650

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

Predicting MYC translocation in HE specimens of diffuse large B-cell lymphoma through deep learning

Zaneta Swiderska-Chadaj^a, Konnie Hebeda^a, Michiel van den Brand^a, and Geert Litjens^a

^a Department of Pathology, Radboud University Medical Center, the Netherlands

ABSTRACT

Diffuse large B-cell lymphoma (DLBCL) is the most common type of B-cell lymphoma. It is characterized by a heterogeneous morphology, genetic changes and clinical behavior. A small specific subgroup of DLBCL, harbouring a MYC gene translocation is associated with worse patient prognosis and outcome. Typically, the MYC translocation is assessed with a molecular test (FISH), that is expensive and time-consuming. Our hypothesis is that genetic changes, such as translocations could be visible as changes in the morphology of an HE-stained specimen. However, it has not proven possible to use morphological criteria for the detection of a MYC translocation in the diagnostic setting due to lack of specificity.

In this paper, we apply a deep learning model to automate detection of the MYC translocations in DLBCL based on HE-stained specimens. The proposed method works at the whole-slide level and was developed based on a multicenter data cohort of 91 patients. All specimens were stained with HE, and the MYC translocation was confirmed using fluorescence in situ hybridization (FISH). The system was evaluated on an additional 66 patients, and obtained AUROC of 0.83 and accuracy of 0.77. The proposed method presents proof of a concept giving insights in the applicability of deep learning methods for detection of a genetic changes in DLBCL. In future work we will evaluate our algorithm for automatic pre-screen of DLBCL specimens to obviate FISH analysis in a large number of patients.

Keywords: Digital Pathology, DLBCL, Lymphoma, Deep Learning, CNN

1. INTRODUCTION

Lymphoid malignancies are diagnosed in more than 280000 people every year worldwide.¹ Diffuse Large B-cell Lymphoma (DLBCL) is one of the most common lymphoma subtypes, that is characterized by clinical, histological and molecular heterogeneity. The occurrence of this disease increases with age, and most patients are over the age of 60 at diagnosis.^{2,3} Although the heterogeneity in morphology has long been recognized and also been associated with outcome, it has not proven possible to use morphological criteria for prognosis in the diagnostic setting due to inter- and intra-observer variability.^{4,5}

MYC gene rearrangement is identified in 5-12% of DLBCL. Patients with an MYC-translocation-positive DLBCL have a poorer prognosis when they are treated with standard chemotherapy regimens.

In recent years, deep learning (DL)⁶ has brought a revolution to the field of pattern analysis and machine learning, by providing algorithms with the capacity to learn complex representations from the raw data itself, achieving human and even a super-human level of performance in many fields, including medical image analysis and computational pathology.⁷⁻⁹ New tools to solve complex tasks in the domain of computational pathology, such as lung adenocarcinoma detection and classification,² detection of lymph node metastases in breast cancer⁷ or prostate cancer detection and grading¹⁰ have been developed. Application of DL methods opens new doors in computational pathology and allows the solving of complex problems alongside achieving high performance. Convolutional Neural Networks (CNN) such as AlexNet,¹¹ U-Net⁶ or ResNet¹² are successfully applied to various problems in computer visions. These architectures are most popular in a large variety of image challenges as well as in biomedical challenges. However, a successful application of the DL method to solve the complex task from the biomedical domain is not trivial. A simple application of one of the DL methods does not guarantee good results because biomedical tasks are much more complicated than simple image classification from MNIST or ImageNet datasets.

Further author information: (Send correspondence to Z.S.C.)

Z.S.C.: E-mail: zaneta.swiderska@gmail.com

Medical Imaging 2020: Digital Pathology, edited by John E. Tomaszewski,
Aaron D. Ward, Proc. of SPIE Vol. 11320, 1132010 · © 2020 SPIE
CCC code: 1605-7422/20/\$21 · doi: 10.1117/12.2549650

1.1 Description of purpose

We expect that the morphology of the DLBCL is correlated with other features of the disease, such as protein expression profile, MYC translocation status, and clinical outcome.

One of the methods proposed to detect MYC rearrangement in clinical practices is based on the detection of a myc protein expression by immunohistochemistry (IHC). However, results presented in¹³ show a sensitivity of 0.88 with a specificity of only 0.36. As such, the IHC procedure is not recommended to detect the MYC-gene rearrangement. Currently, the fluorescence in-situ hybridization (FISH) technique is considered the gold standard in detecting genetic abnormalities in DLBCL, but it is not available in all pathology laboratories and requires expertise, time and money. Since in future genetic changes will increasingly guide lymphoma treatment, availability of a (prescreening) test for MYC translocation would be very helpful, especially in case of limited resources.

In this work, we will, for the first time, use deep learning to detect morphological characteristics of MYC-translocation in HE specimens. Deep Learning methods have been successfully applied to many complex tasks.^{2,3,14} The resultant automated detection of MYC negative lymphoma could reduce the number of FISH tests.

We present work on automatic detection of MYC translocated DLBCL on a whole-slide level. The performed research is based on a multi-center data cohort that includes 91 WSIs, and was evaluated on a test set of 66 WSIs. As far as we know, this is the first study on automatic detection of MYC genetic abnormalities in lymphoma based on the morphology of HE-stained specimens.

2. MATERIALS

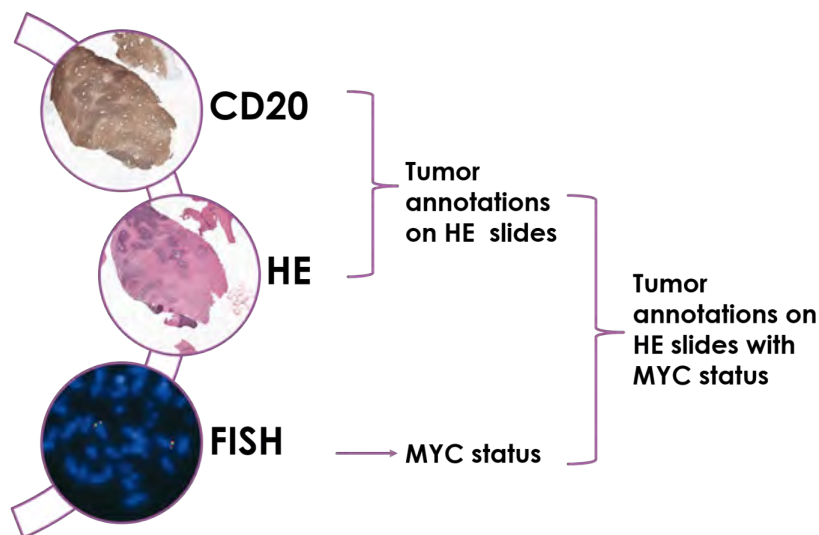


Figure 1. The data types used in the study.

For this study, we collected HE-stained glass slides, FISH test results, and, if available, CD20 stained glass slides from 157 patients with DLBCL that were analyzed by FISH for MYC translocation during the last 5 years (fig. 1). Specimens were sent from 11 medical centers in the Netherlands. All specimens were cut and the slides were HE stained in our lab. The HE staining is used by pathologists to assess the morphological characteristics of the tissue, whereas CD20 an immunohistochemical stain for B-cells, and as such, can be used to highlight the DLBCL. Slides were digitized using a Pannoramic 250 Flash II scanner (3DHistech, Hungary), resulting in WSIs with a pixel size of $0.24 \mu\text{m}/\text{px}$ (objective magnification 20x).

The ground truth information about the MYC translocation status was provided by pathologists based on the FISH test. As a result, each slide has one label: MYC positive or MYC negative. The data cohort includes 157 WSIs, where 53 WSIs were MYC positive and 104 WSIs were MYC negative.

In order to train deep learning method, annotations of tumor areas were prepared by medical students trained specifically for this task. If CD20-stained specimen were available, these were used to support the annotation procedure.

The 157 whole-slide images from independent patients were divided into a training (n=73), validation (n=18) and test set (n=66). The test set contained 66 images, with 33 MYC positive and 33 MYC negative slides.

3. METHODS

The proposed method is based on a patch-classification convolutional neural network and a rule-based slide-level classification (figure 2).

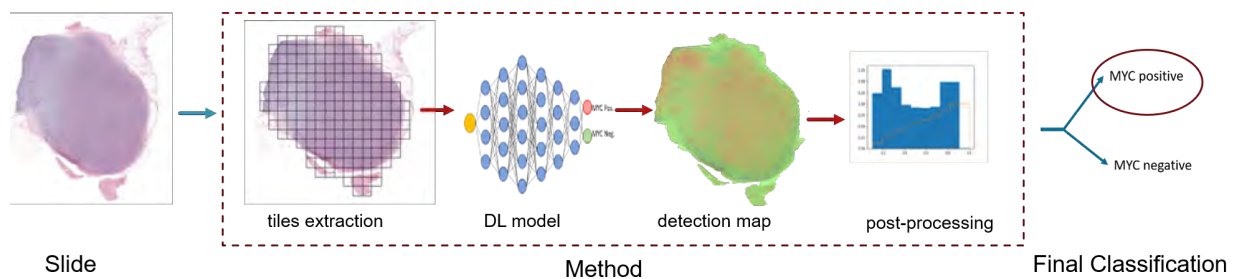


Figure 2. The schema of proposed method.

The training dataset was extracted from HE stained specimens based on supporting the analysis of CD20 stained specimens, that show tumor areas. The MYC status was evaluated on a whole slide level based on the FISH study. A single patch has size 512x512 pixels and was extracted from slides with 5x magnification (pixel size 1 μm). The resolution and size of patches were established after consultation with pathologists. An application of patches with lower resolution can reduce capability to recognize cells, whereas patches with higher resolution can include insufficient tissue areas to present structural pattern (fig. 3).

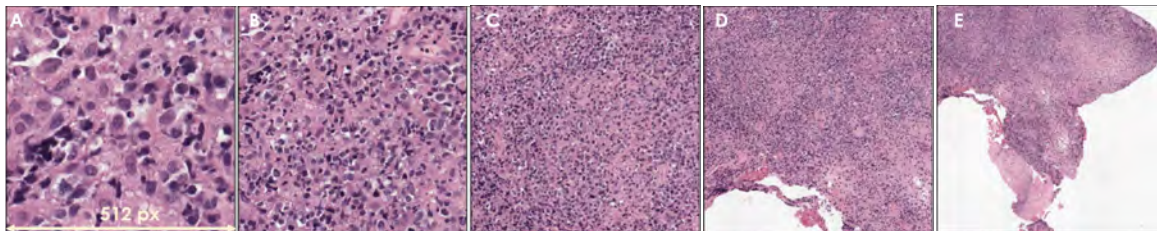


Figure 3. Example of patches with different resolution, where a single patch has a size 512x512px and: A- resolution 20x (pixel size 0.25 μ), B- resolution 10x (pixel size 0.5 μ), C- resolution 5x (pixel size 1 μ), D- resolution 2,5x (pixel size 2 μ) and E-resolution 1,25(pixel size 4 μ).

For each extracted tile, a target map with one class (MYC positive or MYC negative) was created based on the FISH examination. It should be noted that the method development dataset is unbalanced (20 MYC positive WSIs and 71 MYC negative WSIs), which means that MYC negative classes included more data variability than the MYC possessive class. Given the limited number of whole-slide images in the training set, we applied extensive data augmentation based on a modification of brightness, contrast, saturation, and rotation, as well as additive Gaussian noise and Gaussian blur augmentation. The image augmentation reduces the effort needed to acquire additional training data, improves the robustness and ability of CNN to generalize, and decreases the risk of overfitting.¹⁵ Figure 4 present example of training patches for MYC positive and MYC negative classes.

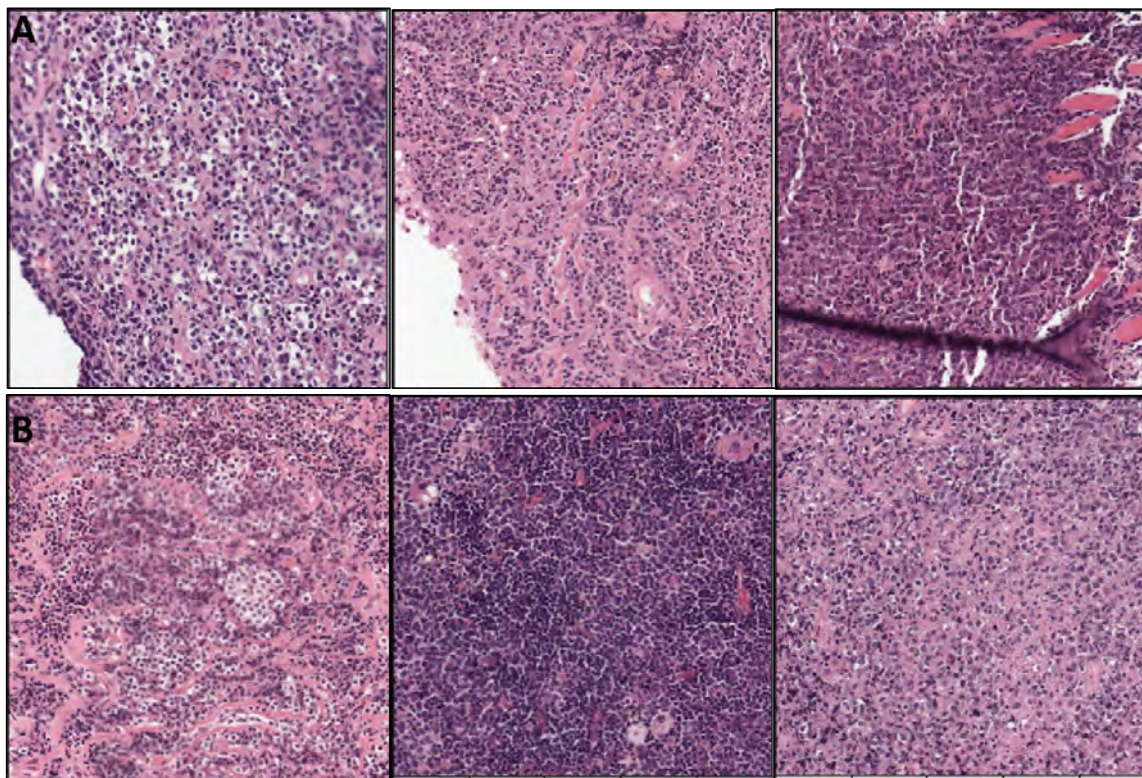


Figure 4. Example of training patches, where: A- MYC positive patches, B- MYC negative patches.

We investigated a deep learning strategy based on a semantic segmentation (pixel classification) by U-Net.⁶ This architecture consists of two paths: a contracting path to capture context and a symmetric expanding path that enables precise localization. The contraction part is the component that is mainly responsible for learning data representation, whereas the expansion part is mostly responsible for producing a fine-grained segmentation. In our study, we adapted the original U-Net architecture⁶ by increasing network depth to 5 levels to increase the context used for segmentation, as well as by adding spatial dropout layers with factor 0.25 between convolutional layers, with the aim of reducing overfitting. The model was trained with learning rate 0.0005, Adam optimizer and categorical cross entropy loss function by 500 epochs with batch size 200 and mini-batch size 1. The validation loss was monitored during the training, and the best model was used in the final classification.

After training the network is used to generate a MYC positive likelihood map for every slide in the test set. In order to reduce the number of calculations, only tiles containing tissue are classified. The tissue mask was created based on a simple thresholding operation. The final slide classification as MYC positive or MYC negative is based on cumulative histogram analysis of the likelihood map. The classification criterion was established based on the validation set, and point 0.7 of the cumulative histogram was selected as a most differentiable for both classes.

The method is evaluated as a binary whole-slide classification task (MYC positive or negative) using receiver-operating characteristic (ROC) analysis (fig.2).

4. RESULTS

The trained model achieved a patch classification accuracy of 0.67 in the validation set. Next, we analyzed the classification results on the whole slide level for the test set using ROC analysis (fig. 5). The test set includes 66 WSIs with an equal proportion between MYC positive and MYC negative slides. The method achieved 0.83 in a term of the area under the ROC curve, and the accuracy was 0.77 with sensitivity 0.88 and specificity 0.66.

Table 1. Results for whole slide classification for three cut-off values.

cut-off value	False Positive	False Negative	Accuracy	Sensitivity	Specificity	AUC
t1	18	2	0.70	0.94	0.44	0.83
t2	11	4	0.77	0.88	0.66	
t3	11	7	0.73	0.78	0.66	

Table 1 presents detailed results. Figure 6 presents detection results with zoomed areas and Figure 7 shows an example of likelihood maps for MYC positive and MYC negative cases.

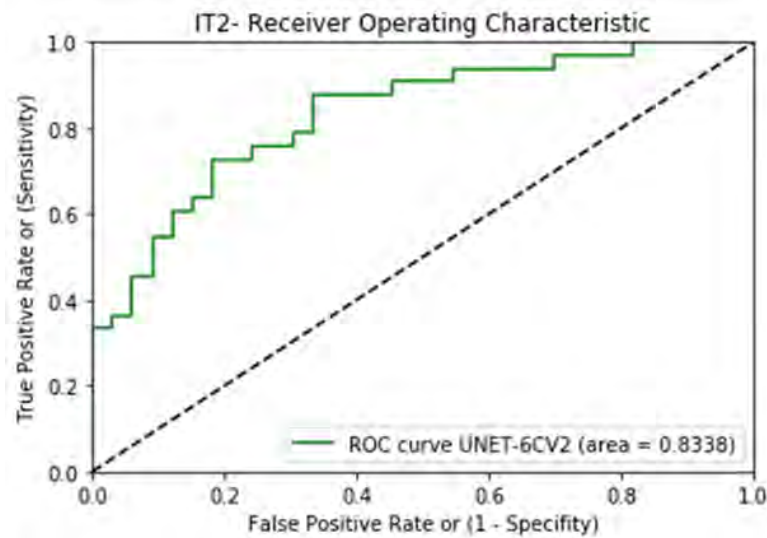


Figure 5. The ROC curve results for the test set.

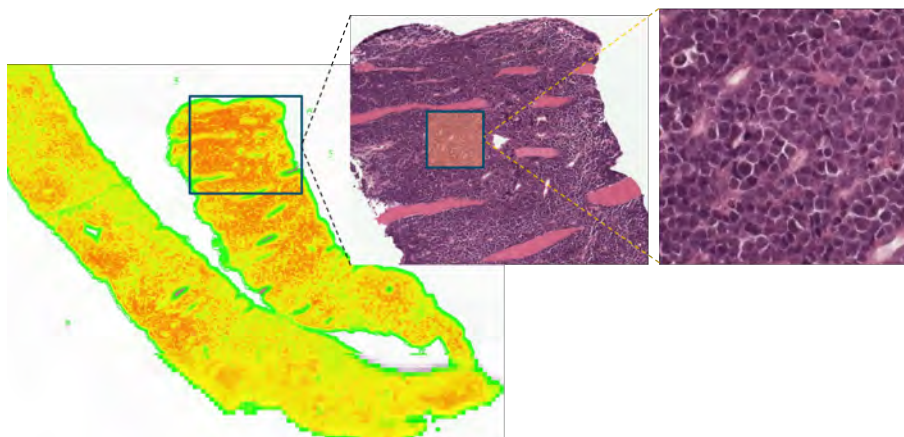


Figure 6. Example of MYC probability map, where a hot-map was applied to presents probability.

5. DISCUSSION AND CONCLUSIONS

We investigated a deep learning strategy based on a semantic segmentation (pixel classification) using U-Net to detect MYC positive slides base on a whole slide analysis.

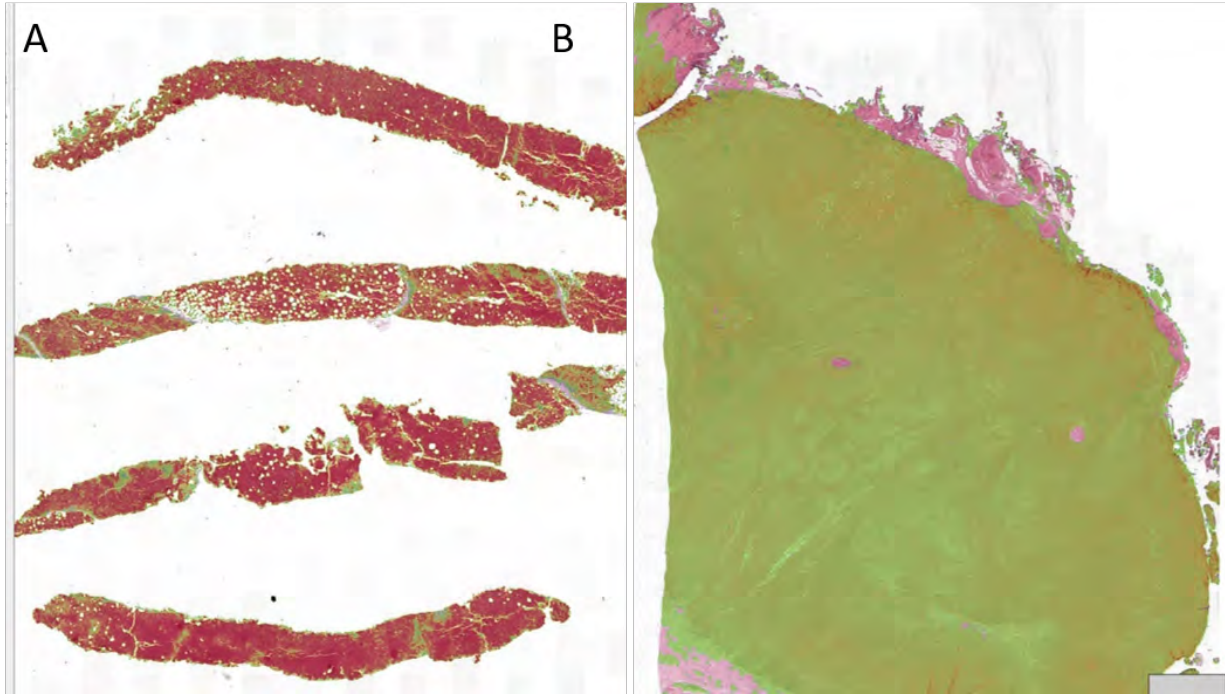


Figure 7. Example of results for whole slides, where a MYC negative tumor is depicted in red (A) and a MYC positive tumor is green (B).

The results show that we are able to detect MYC positive tumors based only on the evaluation of HE-stained slides. The accuracy equals 0.77, with a sensitivity 0.88 and specificity of 0.66. This allows room for future improvement that can be done by application of more MYC positive data for training and extended analysis of current wrong classifications. An important role has multicenter training set and validation on an independent test set, that allows for assessment of method robustness.

MYC rearrangement detection based on immunohistochemistry, which in some pathology labs is used as a surrogate marker or prescreening test for FISH, presented in¹³ achieved a sensitivity of 0.88 with specificity of 0.36. As such, our results show that deep learning methods could outperform immunohistochemistry as a screening method for MYC translocations.

The visual inspection of MYC positive probability maps shows that artifact areas were classified as false positive areas, which finally impacted the whole slide classification. In future work, we would like to extend the proposed method by adding a dedicated step for an artifact elimination.

Analysis of features distinguished MYC positive slides from MYC negative can help understanding better biological bases of the disease. In future work, we are going to investigating enhancing the interpretability of algorithm decisions.

Detection of morphological characteristics as a marker for specific genetic abnormalities and clinical features by using deep learning is expected to support pathology not just in DLBCL diagnosis, but in oncology in general.

REFERENCES

- [1] Sabattini, E., Bacci, F., Sagromoso, C., and Pileri, S., "Who classification of tumours of haematopoietic and lymphoid tissues in 2008: an overview," *Pathologica* **102**(3), 83–87 (2010).
- [2] Gertych, A., Swiderska-Chadaj, Z., Ma, Z., Ing, N., Markiewicz, T., Cierniak, S., Salemi, H., Guzman, S., Walts, A. E., and Knudsen, B. S., "Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides," *Scientific reports* **9**(1), 1483 (2019).

- [3] Litjens, G., Sánchez, C. I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen-Van De Kaa, C., Bult, P., Van Ginneken, B., and Van Der Laak, J., “Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis,” *Scientific reports* **6**, 26286 (2016).
- [4] Cook, J. R., Goldman, B., Tubbs, R. R., Rimsza, L., Leblanc, M., Stiff, P., and Fisher, R., “Clinical significance of myc expression and/or “high grade” morphology in non-burkitt, diffuse aggressive b-cell lymphomas: a swog s9704 correlative study,” *The American journal of surgical pathology* **38**(4), 494 (2014).
- [5] Scott, D. W., King, R. L., Staiger, A. M., Ben-Neriah, S., Jiang, A., Horn, H., Mottok, A., Farinha, P., Slack, G. W., Ennishi, D., Schmitz, N., Pfreundschuh, M., Nowakowski, G. S., Kahl, B. S., Connors, J. M., Gascoyne, R. D., Ott, G., Macon, W. R., and Rosenwald, A., “High-grade B-cell lymphoma with MYC and BCL2 and/or BCL6 rearrangements with diffuse large B-cell lymphoma morphology,” *Blood* **131**, 2060–2064 (05 2018).
- [6] Ronneberger, O., Fischer, P., and Brox, T., “U-net: Convolutional networks for biomedical image segmentation,” in *[International Conference on Medical image computing and computer-assisted intervention]*, 234–241, Springer (2015).
- [7] Bejnordi, B. E., Veta, M., Van Diest, P. J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J. A., Hermsen, M., Manson, Q. F., Balkenhol, M., et al., “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” *Jama* **318**(22), 2199–2210 (2017).
- [8] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S., “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature* **542**(7639), 115 (2017).
- [9] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al., “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *Jama* **316**(22), 2402–2410 (2016).
- [10] Bulten, W., Pinckaers, H., van Boven, H., Vink, R., de Bel, T., van Ginneken, B., van der Laak, J., Hulsbergen-van de Kaa, C., and Litjens, G., “Automated deep-learning system for gleason grading of prostate cancer using biopsies: a diagnostic study,” *The Lancet Oncology* (2020).
- [11] 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).
- [12] Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A., “Inception-v4, inception-resnet and the impact of residual connections on learning,” in *[Thirty-First AAAI Conference on Artificial Intelligence]*, (2017).
- [13] Kluk, M. J., Ho, C., Yu, H., Chen, B. J., Neuberg, D. S., Dal Cin, P., Woda, B. A., Pinkus, G. S., and Rodig, S. J., “Myc immunohistochemistry to identify myc-driven b-cell lymphomas in clinical practice,” *American journal of clinical pathology* **145**(2), 166–179 (2016).
- [14] Arvaniti, E., Fricker, K. S., Moret, M., Rupp, N., Hermanns, T., Fankhauser, C., Wey, N., Wild, P. J., Rueschoff, J. H., and Claassen, M., “Automated gleason grading of prostate cancer tissue microarrays via deep learning,” *Scientific reports* **8** (2018).
- [15] Tellez, D., Litjens, G., Bandi, P., Bulten, W., Bokhorst, J.-M., Ciompi, F., and van der Laak, J., “Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology,” *arXiv preprint arXiv:1902.06543* (2019).