Is the Time Right to Start Using Digital Pathology and Artificial Intelligence for the Diagnosis of Lymphoma?

Digital pathology (DP) enables machine and deep learning applications. These image analysis tools make anatomical pathology a quantitative rather than a qualitative science and thereby mark the arrival of the third and likely most impactful revolution in pathology.[1] Digital image acquisition of glass slides has improved substantially in recent years with the implementation of instrumentation capable of digitization at very fast rates, with excellent resolution and even oil magnification if needed.[2] To date, the Food and Drug Administration (FDA) has cleared the marketing of two whole-slide imaging (WSI) systems for DP primary diagnosis. However, these FDA-approved systems specifically exclude nonformalin-fixed paraffin-embedded (FFPE) hematopathology cases. Moreover, published reports on DP validation for primary diagnosis predominantly focus on FFPE surgical pathology cases using a disease-centric approach for solid tumors, with only rare reports including lymphoma cases.[3]

Diagnosing lymphoma by means of WSI is feasible,[4] but with current technology, this remains challenging for several reasons. Lymphoma interpretation requires examination of slides at both low magnification for architecture information (e.g., follicular vs. diffuse growth pattern) and high magnification for cytomorphology (e.g., nuclear chromatin texture and nucleoli). Moreover, there is a high need for examining multiple ancillary studies such as immunohistochemical (IHC) stains. Unfortunately, there is a paucity of publications on this topic to glean how best to overcome these DP challenges when applied to hematopathology. In fact, when performing a PubMed search (January 6, 2020) with the MESH terms "lymphoma" and "whole slide imaging," there are only two publications applicable to routine diagnostic work in clinical practice.

We highlight three recent articles that have broken the mold by expanding the application of WSI to include hematopathology and reflect on our recent relevant work.[5-7] The first article assessed the usefulness of WSI for the diagnosis of lymphoma[5] and the other publications employed deep learning with a convolutional neural network (CNN) algorithm to build diagnostic lymphoma models.[6,7]

In the first article, Amin *et al.* from Japan showed that WSI is potentially a reliable technology for the diagnosis of lymphoma based on 240 cases of nodal and extranodal biopsies and resection slides from FFPE preparation, including H and E, IHC, and other special stains. Slides were scanned at ×20 magnification on a Nanozoomer 2.0 RS, coupled with a 6-month washout period in their validation study to calculate glass versus WSI diagnostic concordance. Their overall concordance between glass slide examination using a light microscope and WSI diagnoses reported in their study was 92.92% (223/240 cases).

They had 15 (6.25%) minor and two (0.83%) major discrepancies, which is consistent with the previously reported concordance rates in other surgical pathology subspecialties.[5] The two major discrepant cases were nodular lymphocyte-predominant Hodgkin lymphoma (NLPHL) and tuberculosis (TB) associated with low-grade non-Hodgkin B-cell lymphoma. In the NLPHL case, the large pleomorphic lymphoid Reed–Sternberg variant cells were less striking on the WSI than glass slides, and in the TB case, mycobacteria with Ziehl–Neelsen staining were missed. This finding suggests that scanning hematopathology slides only at a magnification of ×20 might not be sufficient, and that digitizing slides at a higher magnification (e.g., scanning at ×40 or higher, even with oil magnification) is advantageous. Most of the minor discrepancies were related to grading follicular lymphoma (11 of 15 cases). This is not surprising because grading of follicular lymphoma should be done at high-power magnification according to the World Health Organization (WHO) recommendations.[8] This should further justify the importance of scanning hematopathology slides at high magnification. Interestingly, diffuse large B-cell lymphoma (DLBCL) represented 40.0% of cases in this cohort with a concordance rate of 100%. Perhaps, the high concordance can be attributed to the easily recognizable and striking growth pattern and cell size in these cases, which is likely even evident with WSI at ×20 magnification.

The authors of the first publication also digitized IHC-stained slides, which were used in addition to H and E-based morphology to reach their final diagnoses. For lymphomas with small- to medium-sized cells, the authors indicated that IHC stains made the diagnosis easier, except for follicular lymphoma. Overall, the authors felt that the use of IHC in this setting can compensate for the lower resolution of WSI. Although the evaluators did not find that the diagnosis for T-cell and NK lymphomas was problematic using WSI, they did communicate their difficulty in differentiating between T-cell lymphoma and reactive conditions in cases with small atypical T-cell infiltrates.

The second recent article by Achi *et al.* from Texas in the USA employed deep learning with WSIs to differentiate 128 hematopathology cases into four diagnostic categories based on H and E-stained slides that were digitized by Aperio WSI systems: benign lymph node, DLBCL, Burkitt lymphoma (BL), and small lymphocytic lymphoma (*n* = 32 from each group). WSIs were viewed at ×40 magnification, and

screen capture software, Snag-It® was used to capture 40 × 40 pixel image patches. Four sets of five representative images from the 128 cases (total *n* = 2560 images, 40 × 40 pixels each) were digitally captured from randomly selected fields that were exposed to the CNN algorithm. To train their model, these authors used 1856 images and subsequently used 464 images for validation. The remaining 240 images were used for testing. Out of the 240 test images, a total of 228 images were correctly diagnosed (95% accuracy) for diagnostic prediction. An accuracy of 100% for all 48 sets was reached when three out of the five representative images of the set were used to classify a set.

The authors of both of these recently published articles used lymphoid diseases, representing the more frequent lymphoma types encountered in clinical practice. However, there are almost fifty mature B-cell neoplasms and around thirty mature T- and NK-cell neoplasms according to the 2017 revised WHO classification of tumors of hematopoietic and lymphoid tissues.[8] Therefore, while these applauded articles help drive DP forward, we have a long way to go before DP with artificial intelligence (AI), which will be useful in routine hematopathology clinical practice. Nevertheless, such deep learning algorithms could certainly start to assist pathologists in making intellectual decisions and perhaps reducing the number of ancillary studies needed to diagnose challenging hematopathology cases. One may envision the potential utility of such an AI-based tool to improve laboratory workflow in high-volume practices including automated IHC panel selection or timely case direction to a subspecialist. Another possibility includes the use of such a system as a quality control (e.g., second read) measure.

The third publication gives a preview of the potential value of AI in lymphoma diagnosis.[7] The authors used 10,818 images from BL (*n* = 34) and DLBCL (*n* = 36) cases to either train or apply different CNNs that differed by number of training images, pixels exploited, color, stain augmentation, and how many layers of the network, among other parameters.[7] The best performing optimized CNN with respect to image attributes showed a receiver operating characteristic curve analysis area under the curve of 0.92 for both BL and DLBCL. The results of this publication support the premise that larger number of images/cases with enough morphologic variation, as seen in DLBCL subtypes and in cases with transformation, is required to properly train deep learning algorithms and to avoid results from being overfitted. Overfitting is a common problem in machine learning that refers to a model that learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data in real practice. It is worth highlighting that both the articles

by Achi *et al*.[6] and Mohlman *et al*.[7] emphasized the use of AI to separate lymphoma types rather than focus on the use of WSI. However, both studies suffered from a small number of total cases to be placed into a limited number of diagnostic categories.

The utility of WSI and AI demonstrated in the three articles we chose to write this commentary about only scratches the surface of what is possible in hematopathology. Further validation studies involving multiple institutions from different regions of the world are needed that expand the types of lymphomas tested in order to recapitulate real-life practice and ensure the generalizability of algorithms being developed. In addition, significantly more work is needed in the area of hematopathology to optimize the performance of CNNs with respect to analyzing images at various magnifications and with diverse image attributes and network architecture. Such engagement of DP to augment lymphoma diagnosis is commended and anticipated to further evolve. Clearly, there is also a need for better niche scanners dedicated to hematopathology, including systems that have regulatory approval, to enable advancement in this area. It is likely that DP coupled with AI will contribute to improving the classification of lymphomas one day.

Mohamed E. Salama[1], William R. Macon[1], Liron Pantanowitz[2]

[1]Mayo Clinic, Rochester, Minnesota, USA, [2]University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA

Address for correspondence: Dr. Mohamed E. Salama,
Mayo Clinic, Rochester, Minnesota, USA.
E-mail: salama.mohamed@mayo.edu

REFERENCES

1. Salto-Tellez M, Maxwell P, Hamilton P. Artificial intelligence-the third revolution in pathology. Histopathology 2019;74:372-6.
2. Newitt VN. Whole Slide Imaging for Primary Diagnosis: Now it is happening. CAP Today. Issue; September, 2017. Available from: http://www.captodayonline.com/whole-slide-imaging-primary-diagnosis-now-happening/. [Last accessed on 2020 Apr 10].
3. Orlov NV, Chen WW, Eckley DM, Macura TJ, Shamir L, Jaffe ES, *et al*. Automatic classification of lymphoma images with transform-based global features. IEEE Trans Inf Technol Biomed 2010;14:1003-13.
4. Pantanowitz J, Pantanowitz L. Feasibility of whole slide imaging in hematopathology practice: Experience with 707 international telepathology consultation cases. J Pathol Inform 2019;10:S44-5.
5. Amin S, Mori T, Itoh T. A validation study of whole slide imaging for primary diagnosis of lymphoma. Pathol Int 2019;69:341-9.
6. Achi HE, Belousova T, Chen L, Wahed A, Wang I, Hu Z, *et al*. Automated diagnosis of lymphoma with digital pathology images using deep learning. Ann Clin Lab Sci 2019;49:153-60.
7. Mohlman JS, Leventhal SD, Hansen T, Kohan J, Pascucci V, Salama ME. Improving augmented human intelligence to distinguish burkitt lymphoma from diffuse large b-cell lymphoma Cases. Am J Clin Pathol 2020;153:743-59.

8. Swerdlow SH, Campo E, Harris NH, Elaine JS, Pileri SA, Stein H, *et al*.
WHO classification of tumours of haematopoietic and lymphoid tissues.
In: World Health Organization Classification of Tumours. Revised
4th ed. Lyon, France: IARC; 2017. p. 270.

How to cite this article: Salama ME, Macon WR, Pantanowitz L. Is the
time right to start using digital pathology and artificial intelligence for the
diagnosis of lymphoma? J Pathol Inform 2020;11:16.

Available FREE in open access from: http://www.jpathinformatics.org/text.
asp?2020/11/1/16/288049