Toward Artificial Intelligence–Driven Pathology Assessment for Hematologic Malignancies

Olivier Elemento

Summary: In this issue of Blood Cancer Discovery, Brück and colleagues applied unsupervised and supervised machine learning to bone marrow histopathology images from patients with myelodysplastic syndrome (MDS). Their study provides new insights into the pathobiology of MDS and paves the way for increased use of artificial intelligence for the assessment and diagnosis of hematologic malignancies.

See related article by Brück et al., p. 238 (3).

Myelodysplastic syndrome (MDS) is a group of hematologic disorders characterized by ineffective and dysplastic hematopoiesis in the bone marrow. For decades, the diagnosis of MDS has relied almost solely on visual morphologic analysis of stained bone marrow samples, involving assessment of cellular dysplasia, blast count, cellularity, monocytes, ring sideroblasts, and a few other features. MDS have been further classified into subtypes based on the World Health Organization (WHO) classification scheme, which combines morphology with blood count and cytogenetics (1). The number and complexity of features used for diagnosis and the subjectivity of their assessment can make a precise diagnosis difficult to achieve in some cases, even for the highly trained hematopathologist. As a result, bone marrow assessments are known to be associated with substantial inter-pathologist variability. More recently, genomic analysis of MDS has allowed for further characterization of MDS cells. A number of frequently mutated genes were identified, for example, TET2, ASXL1, SF3B1, SRSF2, DNMT3A, and RUNX1 (2). With a few exceptions such as the megakaryocyte dysmorphology in 5q-deleted MDS or presence of ring sideroblasts in SF3B1-mutated MDS, the relationship between morphology and mutations remains poorly characterized. This is unfortunate, because the concordance between normally correlated features can reinforce the confidence in each individual feature and may contribute to enhancing diagnostic accuracy.

Can machine learning (ML; a form of artificial intelligence, or AI) and automated image analysis provide additional insights for MDS? This is the problem that Brück and colleagues sought to address in this issue of Blood Cancer Discovery (3). Histopathology slides are now routinely digitized and therefore amenable to automated or semiautomated image analysis. Large datasets combining these images with genomic and clinical information are being collected at a number of medical centers. In other malignancies, the application of ML to histopathology slides has been shown to provide diagnostic and prognostic accuracy on par with trained diagnosticians. For example, ML can diagnose skin malignancies from digital images with performance rivaling trained dermatologists (4). ML can also distinguish accurately between select subtypes of cancers (5), predict prognosis in mesothelioma (6), and even predict the presence of specific mutations, for example, EGFR in lung adenocarcinomas (7).

Brück and colleagues leveraged some relatively recent key developments in the field of ML applied to images. One of them is called transfer learning. Transfer learning refers to the reutilization of deep neural networks that have already been trained to recognize a multitude of nonmedical images and objects from huge image databases such as the well-known ImageNet database (8). By mimicking the way the human visual cortex works, these networks have learned to recognize basic image features such as shapes and textures. These features are encoded within the networks during training, but also an opportunity to capture intratissue heterogeneity. Moreover, the pretrained network features are automatically generated and not selected on the basis of prior knowledge of bone marrow morphology. To make them interpretable for a pretrained neural network. Therefore, there remain many obstacles to applying these pretrained neural networks to pathology images. One is that entire slides of variable and sometimes large sizes cannot typically be directly used as input to trained network. Instead, digitized slides have to be split into smaller and uniformly sized tiles. This is a limitation, as tiles need to somehow be combined, but also an opportunity to capture intratissue heterogeneity. Moreover, the pretrained network features are automatically generated and not selected on the basis of prior knowledge of bone marrow morphology. To make them interpretable for a trained hematopathologist, they must be decoded.

Brück and colleagues analyzed bone marrow samples from 143 patients with MDS, 51 patients with myelodysplastic/myeloproliferative neoplasm (MDS/MPN), and 11 control subjects. In total, 500 tiles were analyzed. Upon extracting features using the widely used VGG16 and Xception convolutional...
neural networks (CNN), Brück and colleagues performed an unsupervised analysis by mapping each tile onto a two-dimensional space using the uniform manifold approximation and projection (UMAP) technique, most frequently used for analyzing single-cell transcriptomic data. In this representation, tiles with similar image feature profiles are located close to each on the two-dimensional space. Five clusters emerged, and visual analysis of tiles within each cluster shed light on their content. These clusters captured high abundance of lipid droplets, red blood cells, stroma, hypocellular tiles, and hypercellular tiles. This served as validation that the deep learning–extracted features do in fact reflect biologically relevant features, a nonnegligible observation given the fact that the CNNs were trained on ImageNet, a database that contains few if any histopathologic images. Brück and colleagues then averaged feature profiles for tiles belonging to the same sample and then reprojected the samples in two dimensions using UMAP, followed by unsupervised clustering. Five clusters emerged again. Healthy subjects formed a distinct cluster and were not only homogenous but also highly distinguishable from the four MDS clusters. Each of the four clusters were enriched in specific MDS WHO subtypes. However, MDS sample groupings only partially overlapped with the WHO subtypes. This is not unexpected because WHO subtypes are defined not only by bone marrow cytology but also by cytogenetics, blast proportion, and blood cell counts.

Brück and colleagues then moved on to supervised analyses, seeking to train statistical models based on elastic net–regularized regression to predict a variety of tumor and patient characteristics based on image features. Certain characteristics, such as mutations in TET2, ASXL1, and STAG2, chromosome 7 monosomy, and 7q deletion, turned out to be highly predictable from morphologic features. This demonstrates that the links between morphologic features and mutations are more extensive than previously thought in MDS. Brück and colleagues were also able to predict the risk stratifying IPSS-R score (Revised International Prognostic Scoring System), overall survival, and progression to acute myeloid leukemia (AML) with high accuracy by solely employing features extracted from stained slides. Strikingly, progression to AML was best predicted using a model that combined histopathologic features and conventional IPSS-R scores. A model was able to discern patients with MDS from patients with MDS/MPN with an accuracy of 0.81 (1.0 being the maximum achievable accuracy). Analysis of direct cell and image segmentation and classification showed, as expected, that predicted MDS samples were hypoplastic with higher stromal involvement.

This is not the first time that ML has been used to assess MDS. For example, Nagata and colleagues successfully used morphologic features from pathology reports to predict mutational profiles (9). Mori and colleagues were able to accurately assess dysplasia in bone marrow smear using a CNN (10). Brück and colleagues demonstrate how much the field of AI has advanced in the past few years, especially the concept of transfer learning. One can now use off-the-shelf, pretrained deep neural networks and deploy them on a new medical image dataset of modest size (a few hundred images) without need for extensive retraining (which would take a much larger dataset to achieve).

Altogether Brück and colleagues’ work represents an exciting step toward AI-driven pathologic assessment in MDS and more generally in hematologic malignancies. Although not the central goal of Brück and colleagues, their work clearly suggests that MDS can be distinguished from normal bone marrow with minimal to no ambiguity using ML. One would imagine that once extended to a larger dataset, validated with data from other independent centers, the AI behind Brück and colleagues may be used to help improve or refine MDS diagnosis, especially in cases that are difficult to diagnose or classify using conventional methods. There may be additional applications for approaches like that of Brück and colleagues. For example, the unbiased assessment that AI provides can also be used either in preclinical animal models or in clinical trials to assess the effect of investigational therapies on the bone marrow of patients with MDS.

Despite the advances described by Brück and colleagues and in many other recent articles applying AI to medical images, much work remains to be done to enable routine use of AI in pathology and other fields. AI models ideally need to be trained on diverse data from multiple medical centers to minimize biases associated with individual centers. Unfortunately, data sharing across medical centers remains infrequent and difficult due to competition between centers and sometimes unnecessary regulatory requirements. Prospective validation of AI models using randomized clinical trials remains too rare. One of the key barriers to adopting medical AI is the limited interpretability of AI models. By opting to use transfer learning instead of directly applying convolutional neural networks to pathology images, Brück and colleagues’ approach has the advantage of generating features whose importance for classification can rather easily be assessed. That does not mean that they can be readily understood, even by a trained hematopathologist. As shown by Brück and colleagues, these features need to be decoded. AI models also need to further integrate the complexity of tumors. For example, averaging feature profiles for the same sample as performed by Brück and colleagues works but also loses information on potential intratumor heterogeneity. Such heterogeneity, if better captured, may help improve predictive models based on histopathology slides.

Will conventional morphologic assessment and the ubiquitously applied WHO classification soon be replaced by AI? Probably not, but Brück and colleagues outline an exciting future, where cooperation between hematopathologists and AI improves the quality and depth of pathology assessments for MDS and other malignancies.

Author’s Disclosures
O. Elemento reports other support from OneThree Biotech, Owkin, and Freenome; grants, personal fees, and other support from Volastra Therapeutics; and personal fees from Champions Oncology during the conduct of the study.

Acknowledgments
O. Elemento acknowledges grant support from the NIH (UL1TR002384 and R01CA194547).
REFERENCES


