DOI: https://doi.org/10.47275/2332-4864-108
Volume 5 Issue 2
Case Report

Scholit Publishers

Journal of Clinical & Anatomic Pathology

The Revival of the H&E with Artificial Intelligence

Nikolay Burlutskiy[1], Paul Waring[2] and Jason Hipp[3]*

[1]*AI Data Scientist, AstraZeneca, Cambridge, UK*
[2]*Executive Director Molecular Pathology, AstraZeneca, Cambridge, UK*
[3]*Senior Director, Translational Medicine, Head of Pathology Data Science & Innovation, AstraZeneca, USA*

Abstract

The Hematoxylin and Eosin (H&E) stained tissue microscopic glass slide is a 100-year-old technology that remains the universal standard for histopathological diagnosis by pathologists. With the advancement of Artificial Intelligence (AI), and especially Deep Learning (DL), the H&E stained tissue slide is beginning to show even greater value.

For decades, pathologists have applied natural intelligence and deep thinking to morphological features highlighted by H&E staining to categorize and classify diseases into reproducibly recognizable entities. In some instances, AI has been shown to match the ability of expert pathologists and exceed the ability of non-expert pathologists in tasks that require precise quantitation. Furthermore, by examining thousands of morphological features on hundreds or thousands of cases, AI algorithms can quickly identify subtle patterns that have eluded pathologists, allowing AI to 'see' things that pathologists cannot.

In this paper, we will discuss some of the recent advancements in AI/DL and H&E pathology by demonstrating key applications to the clinical practice of pathology, followed by a discussion about new avenues that are being created by AI for H&E pathology.

Keywords: Artificial Intelligence (AI); Deep Learning (DL); Hematoxylin and Eosin (H&E); H&E pathology.

Introduction

Since most cells are colorless and transparent, there is a need for the examined tissue to be stained. Hematoxylin and Eosin (H&E) is the most common dye and more than a 100 years old technique; Eosin is an acidic dye what stains basic structures, such as the cytoplasm, in pink or red color; Hematoxylin can be considered as a basic dye staining acidic structures, for example, nuclei, a purplish-blue color [1]. H&E remains the most useful and versatile stain in histology analysis. Despite the extensive use of immunohistochemistry (IHC) to reveal protein expression and the growing use of genetic analysis to identify changes in DNA and RNA, the simple H&E stain remains the backbone of histopathological diagnosis [1].

Recently, many pathology labs have been digitized through scanning H&E and IHC stained slides using high-resolution scanners. As a result, the digitized gigapixel images, Whole Slide Images (WSIs), enable large scale analysis with computerized image analysis methods and tools, including artificial intelligence (AI), Deep Learning (DL), and Machine Learning (ML) - based methods.

Deep learning (DL), a subfield of AI, that utilizes the full potential of artificial neural networks especially convolutional neural networks (CNN) with multiple, non-linear 'deep' layers, has demonstrated great success in solving image-based problems [2]. For instance, CNN has been successful in identifying and recognizing faces, traffic signs, and objects, cars, animals, enabling vision in robots, self-driving cars, and more recently, enabling vision in medical applications. These neural networks showed human-level performance in several classification and segmentation tasks. The performance of CNN is usually evaluated by accuracy, sensitivity, specificity, precision, recall which is also known as a true positive rate (TPR), false-positive rate (FPR) which is equal to 1 – specificity. Practically, there is always a trade-off between TPR and FPR, like the trade-off between sensitivity and specificity very often faced in the medical domain, one must find a balance between false positives and false negatives in the performance of a trained model. For instance, one can plot a receiver operating characteristic curve (ROC curve), a curve where TPR versus FPR are plotted at different classification thresholds, and then one can find an operating point (TPRi, FPRi) with a satisfactory balance between TPRi and FPRi. In order to evaluate and compare the performance of trained AI models using a single performance metric, an Area Under the Curve (AUC), the entire

Correspondence to: Jason Hipp, Senior Director, Translational Medicine, Head of Pathology Data Science & Innovation, AstraZeneca, USA; E-mail: jason.hipp@astrazeneca.com

Citation: Burlutskiy N, Waring P, Hipp J (2020) The Revival of the H&E with Artificial Intelligence. J Clin Anat Pathol, 5(2): 108. DOI: https://doi.org/10.47275/2332-4864-108

Received: September 19, 2020; **Accepted:** October 04, 2020; **Published:** October 10, 2020

two-dimensional area underneath the ROC curve, is calculated and then used to report the performance of the model; AUC ranges from 0 to 1 where 1 is an ideal model classifying all instances correctly. Another performance metric used for evaluation of models is F1-score, the harmonic mean of precision and recall; F1 score varies from 0 to 1 where 1 corresponds to an ideal model which has no false positives neither false negatives.

In the few last years, there has been a steady growth in the community of researchers applying CNN based methods to medical images, ranging from ultrasound, computerized tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) as well as WSI images including IHC, multiplex immunofluorescence (mIF), and H&E images [3,4] enabling the AI to 'see' important information in the images. One catalyzer for such a breakthrough in AI-based pathology applications was the release of a few large H&E datasets through the organization of a few computational challenges, such as TUPAC [24], CAMELYON16 [5] and CAMELYON17 [6], more recently PANDAS [7] and HEROHE [8] and many more others. In these challenges, teams competed in developing the most accurate model for different tasks such as classifying H&E breast tissue into different cancer types [9], predicting metastases in lymph nodes for breast cancer patients [5,6] predicting Gleason score [10], and even HER2 scoring of WSI images directly from H&E images, without any IHC support [11,12]. As a result, various groups of researchers and industry scientists developed methods for common pathologists' tasks such as directing the pathologist to a region of the slide [13,14], determine the presence of a particular malignancy [15,16], provide tumor's grading [10,17-20], quantify biomarker expression [11,12,21-23]. The researchers have also tackled more challenging tasks that are not routinely performed by histopathologists, such as predicting gene profiling and mutation status [12,24-29] predicting 'risk' category as determined by gene expression [30,31], survival and drug response [32-34]. The focus of this paper is about the application of AI to H&E stained images, which are routinely produced, cost-effective, ubiquitous, and mature technique in clinical pathology practice. Recently, these applications have revived the potential of H&E and attracted interest across academia, resulting in a growing number of related impactful publications as well as sparked activity within the industry creating a few successful startups specializing in applying AI/DL to H&E images. All these promising approaches will be discussed in this paper, as well as the limitations of current AI/DL solutions.

How AI can Help the Pathologist with H&E Analysis

AI-based approaches have demonstrated a pathologist level performance in several routine clinical tasks. These results encourage the development of tools for both research and clinical usage, for example, decision support tools and tools to facilitate the development of new biomarkers [11,13,14, and 35] providing the second opinion to the pathologist and reducing the pathologists' time in performing tedious tasks [9,15, and 16], providing support for quantification of different phenomena observed in the clinical practice [5,10, and 17-20]. Even more encouraging are the results for tasks that are not what the pathologist is trained to perform but potentially useful in the pathologist's practice [24-27,29]. We will discuss such tasks, looking from two different perspectives, what pathologists can and cannot do as well.

Case 1: Directing the Pathologist to a Region of the Slide

The digitized H&E images can be the size of 100k by 100k pixels which can make them challenging to analyze without automatic support or guidance to find 'interesting' regions. Furthermore, one may be interested to find a similar region in the literature. A few papers [11,13] addressed this issue by providing the pathologist with functionality to select a region of interest and then search through examples applying both supervised, with annotations, and unsupervised, annotation-free, deep learning methods [13].

A retrieval system for searching regions that are visually like manually defined regions of interest in various data sources including both proprietary and public datasets, and scientific literature, was developed by Schaer R, et al. (2019) [14]. The system was tested by pathologists, highlighting its capabilities and suggesting ways to improve potential use in clinical practice, including H&E image search.

Researchers also developed (i) tissue-type models trained to segment tumors into cancer cells, cancer-associated stroma, and necrotic regions, and (ii) cell-type models trained to detect lymphocytes, plasma cells, fibroblasts, macrophages, and cancer cells [35]. Then these segmentations and detected cells can be overlaid with the H&E image that will help the pathologist to direct their decisions for diagnostic and prognostic purposes. Even more, the tissue-type and cell-type information can be used for training other predictive models by combining and quantifying this information into human-interpretable features (HIFs).

Case 2: Identity, Classify, and Quantify Malignancy in H&E

Pathology practice often implies analyzing several H&E slides per patient to identify, classify, and quantify malignancy in the tissue. Several AI-based applications demonstrated outstanding performance for such tasks.

For example, tumor proliferation rate is an important biomarker for estimating the prognosis of breast cancer patients [36]. An automatic method to quantify the tumor proliferation score was developed by Paeng K, et al. (2017) as a part of the TUPAC competition [19].

The CAMELYON16 challenge, which assessed automated solutions for detecting lymph node metastases, evaluated 32 DL algorithms of which 7 outperformed a panel of 11 pathologists, with an AUC of 0.994 (best algorithm) vs 0.884 (best pathologist) [5].

A model trained to help pathologists detect, outline, and quantify potentially cancerous areas in H&E stained prostate tissue was evaluated on H&E stained needle biopsies achieving an F1-score of 0.80 and Precision-Recall AUC of 0.89 at the pixel-level [15]. Also, the model demonstrated performance on par with pathologists [15]. A similar model for lung cancer H&E slides reached a precision of 0.80 and a recall of 0.86 [16]. A few models trained to classify WSIs into adenocarcinoma, adenoma, and non-neoplastic for gastric and colon cancers were evaluated on three independent test sets achieving AUCs up to 0.97 and 0.99 for gastric adenocarcinoma and adenoma, respectively, and 0.96 and 0.99 for colonic adenocarcinoma and adenoma [9].

One problem of training DL models is the requirement for expensive and time-consuming pixel-wise manual annotation by pathologists to

define and highlight the morphological features. A DL system that uses only the reported diagnoses as labels for training, thereby avoiding the annotation process, was presented by Campanella G, et al. (2019) [37]. The system was trained on a dataset of 44,732 whole slide images from 15,187 patients without any form of data curation. The trained models were tested on prostate cancer, basal cell carcinoma, and breast cancer metastases to axillary lymph nodes resulted in AUC above 0.98 for all cancer types. Its clinical application enables pathologists to exclude up to 75% of all slides while retaining 100% sensitivity [37].

Case 3: Perform Tumor Grading

Gleason grading, as many tumor gradings, suffers from subjectivity and high intra and inter-user variability. The developed DL system for Gleason grading of H&E biopsies by Bulten W, et al. (2020) [10] achieved a high agreement with the reference standard (quadratic Cohen's kappa 0·918, 95% CI 0.891–0.941) and scored highly at clinical decision thresholds: benign versus malignant with AUC of 0.990, grade group of 2 or more (AUC of 0.978), and grade group of 3 or more (AUC of 0.974). The trained algorithm outperformed 10 of 15 pathologist observers as well as obtained a high agreement on an independent dataset annotated by two pathologists with quadratic Cohen's kappa 0.723 and 0.707 [10].

The application of convolutional neural networks (CNN) for tumor grading as well as for decomposing tumor microenvironment (TME) from H&E stained kidney specimen was explored by Khoshdeli M, et al. (2018) [17].

The mitotic count is part of the Bloom & Richardson grading system [20], and a well-recognized prognostic factor in breast cancer. Currently, pathologists count the number of mitoses which is a tedious and subjective procedure suffering from reproducibility. Automatic methods for mitosis detection and counting were developed in the TUPAC competition [19].

An AI-based approach to match an unseen H&E slide with the pathology of already diagnosed and curated cases was presented in the work of Kalra S, et al. (2020) [18]. In this work, data from The Cancer Genome Atlas (TCGA) study from almost 11,000 patients, 30,000 slides covering 25 anatomic sites, and 32 cancer subtypes was used to build such a system. The results were assessed with conservative "majority voting" to build consensus for subtype diagnosis and demonstrated high accuracy values for H&E slides [18].

AI Doing What Pathologists Cannot Do

Case 1: Predict Gene Expression and Mutational Status from H&E

Chen M, et al. (2020) [24] presented a DL model was trained to predict the ten most common and prognostic mutated genes in Hepatocellular carcinoma (HCC) with external AUCs of 0.71 to 0.89. These findings demonstrated that a DL approach can assist pathologists in the classification and detection of gene mutation in liver cancer.

Nuclear BAP1 (nBAP1) staining is a close surrogate for both monosomy 3 and BAP1 mutation which are strong prognostic factors predicting metastatic risk in uveal melanoma (UM), the most common primary intraocular malignancy in adults. Zhang H, et al. (2020) [28] trained a few models for predicting nBAP1 with the best model AUC of 0.90 (95% CI: 0.901-0.908).

Tumor mutational burden (TMB) is a relatively new promising biomarker for predicting response to immunotherapy in cancer patients. The measurement of TMB requires whole-exome sequencing (WES), which is an expensive and time-consuming process. Jain M, et al. (2020) [25] predicted TMB from lung adenocarcinoma (LUAD) H&E images with the precision-recall curve (PR-ROC) of 0.92 and an average precision of 0.89. Another paper provided patient-level TMB predictions with AUC scores of 0.752 and 0.742 for TCGA bladder cancer (BLCA) and LUAD cohorts [26]. TEMPUS achieved an AUC of 0.854, an average precision of 0.723, and an accuracy of 0.889 on colorectal cancer H&E images [27].

Recently, Schmauch B, et al. (2020) [29] used H&E images and RNA-Seq data (FPKM-UQ values) for 28 different cancer types and 8,725 TCGA subjects for training a DL network capable to predict transcriptomic profile from the corresponding high-definition WSIs [29].

Case 2: Predict Protein (IHC) Expression from H&E

Predicting HER2 protein expression using only H&E stained images is a promising, time, and cost-saving initiative. For instance, Rishi R, et al. (2020) [11] explored whether a DL approach could predict the clinical subtypes of breast cancer, as assessed by immunostaining for estrogen, progesterone, and HER2 receptors (ER/PR/HER2) directly from H&E images. The models reached 0.89 AUC (ER), 0.81 AUC (PR), and 0.79 AUC (HER2) on a large, independent test set (n = 2531). Another group of researchers trained a weakly supervised DL model for automatic HER2 scoring from H&E images and reached 83.3% and 53.8% classification accuracy on two independent datasets [12].

Another important biomarker is tumor programmed death-ligand 1 (PD-L1) status to stratify patients who may benefit from programmed death-1 (PD-1)/PD-L1 inhibitors. Sha L, et al. (2019) [23] built a DL model to correlate the association between PD-L1 status and tumor histopathological patterns in H&E images was trained. The model showed promising performance of AUC of 0.67–0.81, $p \leq 0.01$ over a range of PD-L1 cutoff thresholds.

Ki-67 is a widely used cell proliferation and prognostic biomarker [22]. A method that can predict Ki-67 positive cells directly from H&E stained slides by CNN was proposed by Liu Y, et al. (2020) [21]. The model achieved an average accuracy of 0.9371 in discrimination of Ki-67 negative cell images, positive cell images, and background images.

Case 3: Predict 'Risk' Category as Determined by Gene Expression from an H&E

A DL approach based on H&E stained images to identify patients who could benefit from molecular testing was developed by Couture H, et al. (2019) [30]. The trained classifiers predicted tumor grade, ER status, PAM50 intrinsic subtype, histologic subtype, and risk of recurrence score.

Oncotype DX, a 21-gene assay test, is a gene-expression test to assess the risk of early-stage ER+ breast cancer used to guide clinicians on the usage of chemotherapy. However, these tests are expensive and require the sample to be sent to a reference laboratory for testing. Whitney J, et al. (2018) [31] employed H&E images for building a model to predict corresponding risk categories with reported per-patient accuracies ranging from 75 to 86%.

Case 4: Predict Survival and Drug Response

A deep learning model for predicting disease-specific survival across ten cancer types from TCGA based on weakly-supervised approach without pixel-level annotations was developed by Wulczyn E, et al. (2020) [32]. The model, adjusted for cancer type, stage, age, and sex, showed significant association with disease-specific survival (hazard ratio of 1.58, 95% CI 1.28–1.70, $p<0.0001$) [32].

Prediction of colorectal cancer outcome from H&E stained histopathology slides allowed to extract and learn salient, discriminative, and clinically meaningful content predicted outcome with accuracy of 0.70 and F1-score of 0.81 [33]. A model to distinguish between patients who would respond to Nivolumab immunotherapy for NSCLC was trained by Wang X, et al. (2018) [34]. This ML classifier used H&E imaged nuclear features, such as variance in nuclear shape and chromatin structure, yielded an AUC of 0.65 on the training set (n = 32) and an AUC of 0.60 on the independent validation set from a separate institution (n = 24) [34].

What AI Cannot Do Now but Pathologists Can Do

So far, AI has demonstrated impressive value in solving specific tasks. This type of AI is called 'narrow AI'. The challenge of developing 'general AI' (GAI) tools that can be generalized to perform unguided tasks performed by pathologists is still in its infancy. For example, validation and quality assessment of WSIs for diagnostic purposes require manual inspection by pathologists [38]. Training DL models also requires pathologists to manually annotate the slides for a specific task, the performance of such algorithms is evaluated against the ground truth which is provided by pathologists [5,9,10,20,36]. Extensive validation of AI-based solutions is crucial for establishing trust in the algorithms. Now, the intended use of most developed AI-based tools is more for decision support purposes, where the tools need thorough validation before being widely adopted. Another obstacle for wider adoption is legislation as well as regulatory approval around these newly built tools. Nevertheless, the FDA is actively working on developing a framework for AI-based software algorithms to serve as medical devices [39]. In Europe, health institutions using a medical device can be exempt from some of the provisions of in vitro diagnostic medical device regulations (2017/746) in cases when the tools meet the relevant General Safety and Performance Requirements. For such an exemption, health institutions will need to have a proper quality management system along with appropriate technical documentation in place [40]. So far, most WSI/H&E systems have been certified via the self-certification route. Despite the uncertainties in future regulatory changes around AI-based systems, it is strongly recommended that all AI tools should undergo Conformité Européene in vitro diagnostic devices (CE-IVD) marking. This process will ensure additional clinical evidence, rigor, and assessment by Notified Bodies [40].

We anticipate that AI tools will soon become incorporated into specific workflows to augment the skill and expertise of pathologists that require quantitation, accuracy, and precision. Combining AI with large data sets also promises to reveal subtle or overlooked morphological features associated with particular molecular events, providing new insights that will allow pathologists and AI scientists to further classify diseases into discrete entities with biological, pathological, clinical, and therapeutic significance.

Discussion

To summarize, AI tools demonstrate great value for the pathologist in saving time, improved performance, and provide better-informed decisions. There are even cases where AI demonstrated impressive ability in tasks doing what pathologists are not trained to do (e.g. predicting HER2 score from H&E images). Nevertheless, there is still a long way to go to make these tools mature enough, reliable, interpretable, and safe for wide integration into the clinical workflow. There are a few challenges to address for further development in this field.

First, despite the numerous AI applications using H&E slides, there is a need to push the frontiers further to unleash the full potential of H&E images, to learn the boundaries that can and cannot be done utilizing rich information in H&E images.

Second, further automation of H&E-based algorithms requires the appropriate quality of H&E stained slides. Unfortunately, quality control of H&E images is still semi-automatic since sometimes there are artifacts present in the slides, such as air bubbles, tissue folding, pen-marking, dust particles, and blurry areas. There has been researched to detect these areas and to determine whether the image passes the quality control or whether it must be re-stained or rescanned. For example, Janowczyk A, et al. (2019) [38] developed open-source software for quality control of digitized slides. Nevertheless, this step must be performed with help from pathologists. To add to this point, one needs to understand the requirements for H&E quality from an AI algorithmic point of view, how differently AI algorithms 'see' H&E slides compared to the pathologist.

Third, many DL algorithms suffer from the generalizability problem [41]. An algorithm trained on one dataset might not generalize to another H&E dataset especially if the staining of the new dataset is far from what the algorithm was trained upon. There is vast variability in scanner settings and scanning artifacts as well. Researchers have already pushed the generalizability of H&E based algorithms via data augmentations, applying generative adversarial neural networks, and multi-task learning [42] that allows one to optimize the usage of already existing datasets without scanning new slides. Potentially, soon there will be a large and diverse H&E dataset available, like general images 'ImageNet' dataset widely used in the computer vision community, which will help to build more generalizable H&E algorithms.

Fourth, the challenge of explainability and interpretability of AI/DL models has been also recently addressed by the research community [35,43]. There is a huge need to transform the 'black boxes' of trained DL models into 'white boxes' which can be trusted and safely deployed into the clinical workflow.

Finally, an active field of research is combining H&E with other imaging modalities such as CT, MRI, PET as well as other biological modalities including genomics, transcriptomics, proteomics, electronic health records [44-46]. Potentially, such an approach will lead to better performance of AI algorithms as well as will allow us to fully uncover the potential of H&E stained images.

Acknowledgments

Thanks to Miika Ahdesmaki for his feedback and comments on the paper.

Conflict of Interest

The authors are AstraZeneca employees and have AstraZeneca stocks.

References

1. Titford M (2005) The long history of hematoxylin. Biotech Histochem 80: 73-78. https://doi.org/10.1080/10520290500138372
2. Wang X, Zhao Y, Pourpanah F (2020) Recent advances in deep learning. Int J Mach Learn Cybern 11: 747-750. https://doi.org/10.1007/s13042-020-01096-5
3. Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, et al. (2017) A survey on deep learning in medical image analysis. Med Image Anal 42: 60-88. https://doi.org/10.1016/j.media.2017.07.005
4. Srinidhi CL, Ciga O, Martel AL (2021) Deep neural network models for computational histopathology: A survey. Med Image Anal 67. https://doi.org/10.1016/j.media.2020.101813
5. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, et al. (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 318: 2199-2210. https://doi.org/10.1001/jama.2017.14585
6. Bandi P, Geessink O, Manson Q, Van Dijk M, Balkenhol M, et al. (2019) From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge, IEEE Transact Med Imag 38: 550-560. https://doi.org/10.1109/TMI.2018.2867350
7. Prostate cANcer graDe Assessment (PANDA) Challenge (2020) Prostate cancer diagnosis using the Gleason grading system.
8. HEROHE (2020) Grand challenge, European Congress on Digital Pathology 2020.
9. Iizuka O, Kanavati F, Kato K, Rambeau M, Arihiro K, et al. (2020) Deep learning models for histopathological classification of gastric and colonic epithelial tumours, Sci Rep 10: 1504. https://doi.org/10.1038/s41598-020-58467-9
10. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, et al. (2020) Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet 21: 233-241. https://doi.org/10.1016/S1470-2045(19)30739-9
11. Rawat RR, Ortega I, Roy P, Sha F, Shibata D, et al. (2020) Deep learned tissue "fingerprints" classify breast cancers by ER/PR/Her2 status from H&E images. Sci Rep 10: 7275. https://doi.org/10.1038/s41598-020-64156-4
12. Oliveira SP, Ribeiro Pinto J, Gonçalves T, Canas-Marques R, Cardoso MJ, et al. (2020) Weakly-supervised classification of HER2 expression in breast cancer H&E stained slides. App Sci 10: 4728. https://doi.org/10.3390/app10144728
13. Hegde N, Hipp JD, Liu Y, Emmert-Buck M, Reif E, et al. (2019) Similar image search for histopathology: SMILY. Dig Med 2: 1-9. https://doi.org/10.1038/s41746-019-0131-z
14. Schaer R, Otálora S, Jimenez-del-Toro O, Atzori M, Müller H (2019) Deep learning-based retrieval system for gigapixel histopathology cases and the open access literature, J Pathol Inform 10: 19. https://doi.org/10.4103/jpi.jpi_88_18
15. Burlutskiy N, Pinchaud N, Gu F, Hägg D, Andersson M, et al. (2019) Segmenting potentially cancerous areas in prostate biopsies using semi-automatically annotated data. Med Imag Deep Learn 102: 92-108.
16. Burlutskiy N, Gu F, Wilen LK, Backman M, Micke P. et al. (2018) A deep learning framework for automatic diagnosis in lung cancer. Med Imag Deep Learn.
17. Khoshdeli M, Borowsky A, Parvin B (2018) Deep learning models differentiate tumor grades from H&E stained histology sections. Annu Int Conf IEEE Eng Med Biol Soc 2018: 620-623. https://doi.org/10.1109/EMBC.2018.8512357
18. Kalra S, Tizhoosh HR, Shah S, Choi C, Damaskinos S, et al. (2020) Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence. Digit Med 3:31. https://doi.org/10.1038/s41746-020-0238-2
19. Paeng K, Hwang S, Park S, Kim M (2017) A unified framework for tumor proliferation score prediction in breast histopathology, In: Cardoso M. et al. (eds) Deep learning in medical image analysis and multimodal learning for clinical decision support, DLMIA 2017. Lect Notes Comp Sci 10553, Springer, Cham. https://doi.org/10.1007/978-3-319-67558-9_27
20. Elston CW, Ellis IO (1991) Pathological prognostic factors in breast cancer. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathol 19: 403-410.
21. Liu Y, Li X, Zheng A, Zhu X, Liu S (2020) Predict Ki-67 positive cells in H&E-stained images using deep learning independently from IHC-stained images. Front Mol Biosci 7: 183. https://doi.org/10.3389/fmolb.2020.00183
22. Scholzen T, Gerdes J (2000) The Ki-67 protein: from the known and the unknown. J Cell Physiol 182: 311-322. https://doi.org/10.1002/(sici)1097-4652(200003)182:3<311::aid-jcp1>3.0.co;2-9
23. Sha L, Osinski BL, Ho IY, Tan TL, Willis C, et al. (2019) Multi-field-of-view deep learning model predicts non-small cell lung cancer programmed death-ligand 1 status from whole-slide hematoxylin and eosin images. J Pathol Inform 10: 24. https://doi.org/10.4103/jpi.jpi_24_19
24. Chen M, Zhang B, Topatana W, Cao J, Zhu H, et al. (2020) Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. Precis Oncol 4: 14. https://doi.org/10.1038/s41698-020-0120-3
25. Jain MS, Massoud TF (2020) Predicting tumour mutational burden from histopathological images using multiscale deep learning. Nat Mach Learn 2: 356–362. https://doi.org/10.1038/s42256-020-0190-5
26. Xu H, Park S, Clemenceau JR, Radakovich N, Lee SH, et al. (2020) Deep learning approach to predict tumor mutation burden (TMB) and delineate its spatial heterogeneity from whole slide images. Biorxiv. https://www.biorxiv.org/content/10.1101/554527v2

27. oshi RP, Kruger AJ, Sha L, Kannan M, Khan AA, et al. (2020) Learning relevant H&E slide morphologies for prediction of colorectal cancer tumor mutation burden using weakly supervised deep learning, J Clin Oncol 38. https://doi.org/10.1200/JCO.2020.38.15_suppl.e15244

28. Zhang H, Kalirai H, Acha-Sagredo A, Yang X, Zheng Y, et al. (2020) Coupland, Piloting a deep learning model for predicting nuclear BAP1 immunohistochemical expression of uveal melanoma from hematoxylin-and-eosin sections. Trans Vision Sci Tech 9: 50. https://doi.org/10.1167/tvst.9.2.50

29. Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, et al. (2020) A deep learning model to predict RNA-Seq expression of tumours from whole slide images. Nat Commun 11: 3877.

30. Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, et al. (2019) Niethammer, Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. Breast Cancer 4: 30. https://www.nature.com/articles/s41523-018-0079-1

31. Whitney J, Corredor G, Janowczyk A, Ganesan S, Doyle S, et al. (2018) Quantitative nuclear histomorphometry predicts oncotype DX risk categories for early stage ER+ breast cancer, BMC Cancer 18: 610. https://doi.org/10.1186/s12885-018-4448-9

32. Wulczyn E, Steiner DF, Xu Z, Sadhwani A, Wang H, et al. (2020) Deep learning-based survival prediction for multiple cancer types using histopathology images. PLoS ONE 15: e0233678. https://doi.org/10.1371/journal.pone.0233678

33. Yue X, Dimitriou N, Arandjelovic O (2019) Colorectal cancer outcome prediction from H&E whole slide images using machine learning and automatically inferred phenotype profiles. ArXiv. https://arxiv.org/abs/1902.03582

34. Wang X, Barrera C, Velu P, Bera K, Prasanna P, et al. (2018) Computer extracted features of cancer nuclei from H&E stained tissues of tumor predicts response to nivolumab in non-small cell lung cancer. Tumor Biol 36: 12061-12061.

35. Diao JA, Chui WF, Wang JK, Mitchell RN, Rao SK, et al. (2020) High-resolution mapping of cells and tissues from pathology images for the interpretable prediction of molecular phenotypes in cancer, bioRxiv.https://doi.org/10.1101/2020.08.02.233197

36. van Diest PJ, van der Wall E, Baak JP (2004) Prognostic value of proliferation in invasive breast cancer: a review. J Clin Pathol 57: 675-681.

37. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Silva VW, et al. (2019) Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 25: 1301-1309. https://doi.org/10.1038/s41591-019-0508-1

38. Janowczyk A, Zuo R, Gilmore H, Feldman M, Madabhushi A (2019) HistoQC: An open-source quality control tool for digital pathology slides, JCO Clin Cancer Info 3. https://doi.org/10.1200/CCI.18.00157

39. Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, et al. (2013) Validating whole slide imaging for diagnostic purposes in pathology: guideline from the college of American Pathologists Pathology and Laboratory Quality Center. Arch Pathol Lab Med 137: 1710-1722.

40. Colling R, Pitman H, Oien K, Rajpoot N, Macklin P, et al. (2019) Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. J Pathol 249: 143-150. https://doi.org/10.1002/path.5310

41. Kawaguchi K, Kaelbling LP, Bengio Y (2020) Generalization in deep learning, Mathematics of Deep Learning, Cambridge University Press. https://arxiv.org/abs/1710.05468v6

42. Mormont R, Geurts P, Marée R (2020) Multi-task pre-training of deep neural networks for digital pathology. IEEE J Biomed Health Info. https://arxiv.org/abs/2005.02561v2

43. Schaumberg AJ, Juarez-Nicanor WC, Choudhury SJ, Pastrián LG, Pritt BS, et al. (2020) Interpretable multimodal deep learning for real-time pan-tissue pan-disease pathology search on social media. Mod Pathol 33: 2169-2185. https://doi.org/10.1038/s41379-020-0540-1

44. Tan X, Su AT, Hajiabadi H, Tran M, Nguyen Q, et al. (2020) Applying machine learning for integration of multi-modal genomics data and imaging data to quantify heterogeneity in tumour tissues. Art Neural Net 2190: 209-228. https://doi.org/10.1007/978-1-0716-0826-5_10

45. Chen RJ, Lu MY, Wang J, Williamson DF, Rodig SJ, et al. (2020) Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis, IEEE Trans Med Imag.https://doi.org/10.1109/TMI.2020.3021387

46. Food and Drug Administration (2019) Proposed regulatory framework for modifications to artificial intelligence and machine learning based software as a medical device. US Food & Drug Administration.