Requirements for implementation of artificial intelligence in the practice of gastrointestinal pathology

Hiroshi Yoshida, Tomoharu Kiyuna

ORCID number: Hiroshi Yoshida 0000-0002-7569-7813; Tomoharu Kiyuna 0000-0003-3050-6718.

Author contributions: Yoshida H and Kiyuna T contributed equally to this work.

Conflict-of-interest statement: All authors have no competing interests to be declared.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/Licenses/by-nc/4.0/

Manuscript source: Invited manuscript

Specialty type: Gastroenterology and hepatology

Country/Territory of origin: Japan

Peer-review report’s scientific quality classification

Hiroshi Yoshida, Department of Diagnostic Pathology, National Cancer Center Hospital, Tokyo 104-0045, Japan

Tomoharu Kiyuna, Digital Healthcare Business Development Office, NEC Corporation, Tokyo 108-8556, Japan

Corresponding author: Hiroshi Yoshida, MD, PhD, Staff Physician, Department of Diagnostic Pathology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. hiroyosh@ncc.go.jp

Abstract

Tremendous advances in artificial intelligence (AI) in medical image analysis have been achieved in recent years. The integration of AI is expected to cause a revolution in various areas of medicine, including gastrointestinal (GI) pathology. Currently, deep learning algorithms have shown promising benefits in areas of diagnostic histopathology, such as tumor identification, classification, prognosis prediction, and biomarker/genetic alteration prediction. While AI cannot substitute pathologists, carefully constructed AI applications may increase workforce productivity and diagnostic accuracy in pathology practice. Regardless of these promising advances, unlike the areas of radiology or cardiology imaging, no histopathology-based AI application has been approved by a regulatory authority or for public reimbursement. Thus, implying that there are still some obstacles to be overcome before AI applications can be safely and effectively implemented in real-life pathology practice. The challenges have been identified at different stages of the development process, such as needs identification, data curation, model development, validation, regulation, modification of daily workflow, and cost-effectiveness balance. The aim of this review is to present challenges in the process of AI development, validation, and regulation that should be overcome for its implementation in real-life GI pathology practice.

Key Words: Artificial intelligence; Deep learning; Digital image analysis; Digital pathology; Clinical implementation; Gastrointestinal cancer

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.
Core Tip: The advances in artificial intelligence (AI) will revolutionize medical practice, as well as other areas of medicine. Deep learning algorithms have shown promising benefits in various areas of diagnostic histopathology. Despite this, AI technology is not widely used as a medical device and is not approved by a regulatory authority. Thus, implying that certain improvements in the development process are still necessary for the implementation of AI in the real-life histopathology-practice. This paper aims to provide a review of recent AI developments in gastrointestinal pathology and the challenges in their implementation.

INTRODUCTION

The integration of artificial intelligence (AI) will cause a revolution in various areas of medicine[1], including gastrointestinal (GI) pathology, in the next decade. Advances in slide scanner technology have made it possible to quickly digitalize histological slides at high resolution, which could be used in clinical practice, research, and education [2-4]. The drastic increase in computing capacity and improvement in information technology (IT) infrastructure has allowed rapid and efficient processing of large-sized data such as whole slide images (WSIs). In recent years, there has been an increase in computer applications utilizing AI to analyze images[5].

AI is an umbrella terminology for the different strategies a computer can employ to think and learn like a human. Pathological AI models have progressed from expert systems to conventional machine learning (ML) and deep learning (DL)[6]. Both expert systems and conventional ML use expert knowledge and expert-defined rules about objects. On the contrary, DL directly extracts features from the raw data and leverages multiple hidden layers of data for the output[7](Figure 1). Compared to conventional ML, DL is simpler to conduct, performs with high-precision, and is cost-effective[5,8]. Its implementation enhances the reproducibility of the subjective visual assessment by human pathologists and integrates multiple parameters for precision medicine[9,10]. Currently, DL algorithms have shown promising benefits in different facets of diagnostic histopathology, such as tumor identification, classification, prognosis prediction, and biomarker/genetic alteration prediction[5,11]. In addition, various AI applications have been developed for GI pathology[12-14].

AI applications using DL algorithms have demonstrated various benefits in the field of GI pathology. Recent reviews (gastric and colorectal) provide an overview of the rapid and extensive progress in the field[5,11-14]. In 2017, the Philips IntelliSite (Philips Electronics, Amsterdam, The Netherlands) whole-slide scanner was approved by the Food and Drug Administration (FDA) in the United States. The implementation of AI in pathology is also promoted by various startups such as DeepLens[15] and PathAI[16]. Some institutions have agreed to digitize their pathology workflow[17,18]. Although these advances are promising, unlike in the field of radiology or cardiology imaging[19], no histopathology-related AI application has been approved by a regulatory authority or for public reimbursement. This indicates that there are still many obstacles to be resolved before the introduction of AI applications in real-life histopathology practice (Figure 2).

In this review, we aim to present and summarize challenges in the process of development, validation, and regulation that should be overcome for the implementation of AI in real-life GI pathology practice. The complete and comprehensive review of the literature on GI pathology-related AI applications is beyond the scope of this paper and is well described elsewhere[12-14]. Here, we focused on how we can adopt these recent advancements in our daily practice.
AI-APPLICATIONS IN GI PATHOLOGY

AI applications in tumor pathology, including GI cancers[4,5] have been developed for tumor diagnosis, subtyping, grading, staging, prognosis prediction, and identification of biomarkers and genetic alterations. In the current decade, the implementation of DL technologies has dramatically improved the accuracy of digital image analysis[6]. DL is one of the ML methods that are particularly effective for digital image analysis[6]. DL is based on the use of convolutional neural networks (CNNs), consisting of millions of artificial neurons, assembled in several layers that are capable of translating its input data (pixel value matrix for an image) into a more abstract representation (Figure 1). The various layers of mathematical computation are fed into a dataset of digitized images annotated with a specific label (e.g., carcinoma or benign lesion); ultimately, the CNN learns how to categorize images according to their respective labels. They automatically identify the most distinctive and common characteristics of each type of object. CNNs outperform hand-crafted or conventional ML techniques (using support vector machines or random forests), by a substantial margin, in image classification[8,20]. In GI pathology, the prediction targets also include tumor classification, the clinical outcome of the patient, and genetic alterations within the tumor (Tables 1 and 2).

In addition, a variety of ML methods have been developed. The strengths and weaknesses of typical ML methods are summarized in Table 3. All of the current ML methods have their advantages and disadvantages, and it is necessary to select an appropriate method according to the purpose of image analysis. DL-based methods are most commonly used in current image analysis of GI pathology; however, they have limitations of requiring substantial data sets and insufficient interpretability. In the future, the development of new ML methods that can compensate for the disadvantages of current ML methods will further accelerate the development of AI-models.

Histopathological AI-applications in gastric cancer

Several attempts have been made to classify pathological images of gastric cancer using AI (Table 1). Before we go into details of AI research review, it should be noted that the comparison of performances should not rely only on accuracy; we should pay attention to the task difficulty in the research framework, i.e., (1) dataset size (results...
Table 1 Artificial intelligence applications in gastric cancer pathology

| Ref.          | Task                          | No. of cases/data set | Machine learning method         | Performance                                           |
|---------------|-------------------------------|-----------------------|---------------------------------|------------------------------------------------------|
| Bollschweiler et al[29] | Prognosis prediction      | 135 cases             | ANN                             | Accuracy (93%)                                        |
| Duraipandian et al[62]     | Tumor classification         | 700 slides            | GastricNet                      | Accuracy (100%)                                       |
| Cosatto et al[65]          | Tumor classification         | > 1200 WSIs           | MIL                             | Accuracy (93%)                                        |
| Sharma et al[21]           | Tumor classification         | 454 cases             | CNN                             | AUC (up to 0.97), accuracy (81% for necrosis detection) |
| Jiang et al[81]            | Prognosis prediction         | 786 cases             | SVM classifier                  | AUCs (up to 0.83)                                     |
| Qu et al[82]               | Tumor classification         | 9720 images           | DL                              | AUCs (up to 0.97)                                     |
| Yoshida et al[23]          | Tumor classification         | 3062 gastric biopsy specimens | ML                             | Overall concordance rate (55.6%)                       |
| Kather et al[34]           | Prediction of microsatellite instability | 1147 cases (gastric and colorectal cancer) | Deep residual learning | AUC (0.81 for gastric cancer; 0.84 for colorectal cancer) |
| Garcia et al[30]           | Tumor classification         | 3257 images           | CNN                             | Accuracy (96.9%)                                      |
| Leon et al[63]             | Tumor classification         | 40 images             | CNN                             | Accuracy (up to 89.7%)                                |
| Fu et al[32]               | Prediction of genomic alterations, gene expression profiling, and immune infiltration | > 1000 cases (gastric, colorectal, esophageal, and liver cancers) | Neural networks. | AUC (0.9) for BRAF mutations prediction in thyroid cancers |
| Liang et al[34]            | Tumor classification         | 1900 images           | DL                              | Accuracy (91.1%)                                      |
| Sun et al[85]              | Tumor classification         | 500 images            | DL                              | Accuracy (91.6%)                                      |
| Tomita et al[24]           | Tumor classification         | 502 cases (esophageal adenocarcinoma and Barret esophagus) | Attention-based deep learning | Accuracy (83%)                                        |
| Wang et al[86]             | Tumor classification         | 608 images            | Recalibrated multi-instance deep learning | Accuracy (86.5%)                                      |
| Iizuka et al[22]           | Tumor classification         | 1746 biopsy WSIs      | CNN, RNN                        | AUCs (up to 0.98), accuracy (95.6%)                   |
| Kather et al[33]           | Prediction of genetic alterations and gene expression signatures | > 1000 cases (gastric, colorectal, and pancreatic cancer) | Neural networks | AUC (up to 0.8)                                        |

ANN: Artificial neural network; GastricNet: The deep learning framework; WSIs: Whole slide images; MIL: Multi-instance learning; AUC: Area under the curve; CNN: Convolutional neural networks; SVM: Support vector machine; DL: Deep learning; ML: Machine learning; RNN: Recurrent neural networks.

for small sample size are less reliable), (2) resolution of detection (tissue level or region level), (3) number of categories to be classified, (4) multi-site validation (sources of training and test dataset are from the same site or not), and (5) constraints on target lesion (e.g., adenocarcinoma only, or any lesions except lymphoma). Sharma and colleagues documented the detection of gastric cancer in histopathological images using two DL-based methods: one analyzed the morphological features of the whole image, while the other investigated the focal features of the image independently. These models showed an average accuracy of up to 89.7%[21]. Iizuka et al[22] reported an AI algorithm, based on CNNs and recurrent neural networks, to classify gastric biopsy images into gastric adenocarcinoma, adenoma, and non-neoplastic tissue. Within three independent test datasets, the algorithm demonstrated an area under the curve (AUC) of 0.97 for the classification of gastric adenocarcinoma. Yoshida et al[23], using gastric biopsy specimens, contrasted the classification outcomes of experienced pathologists with those of the NEC Corporation-built ML-based program “e-Pathologist”. While the total concordance rate between them was only 55.6 percent (1702/3062), the concordance rate was as high as 90.6 percent (1033/1140) for the biopsy specimens negative for a neoplastic lesion. Tomita et al[24] attempted to automate the identification of pre-neoplastic/neoplastic lesions in Barrett esophagus or gastric adenomas/adenocarcinomas.

The above tumor classification studies have shown that AI can be used for histopathological image analysis. However, other obstacles are hindering its use in real-life practice. For example, although the workload of pathologists can be...
### Table 2 Artificial intelligence applications in colorectal cancer pathology

| Ref.          | Task                                                                 | No. of cases/data set | Machine learning method                  | Performance                  |
|--------------|----------------------------------------------------------------------|-----------------------|------------------------------------------|-----------------------------|
| Xu et al.[38] | Tumor classification: 6 classes (NL/AD/MC/SC/PC/CCTA)                | 717 patches           | AlexNet                                  | Accuracy (97.5%)            |
| Awans et al.[37] | Tumor classification: Normal/Low-grade cancer/High-grade cancer       | 454 cases             | Neural networks                          | Accuracy (97%, for 2-class; 91%, for 3-class) |
| Haj-Hassan et al.[37] | Tumor classification: 3 classes (NL/AD/ADC)                        | 30 multispectral image patches | CNN                                      | Accuracy (99.2%)             |
| Kainz et al.[38] | Tumor classification: Benign/Malignant                               | 165 images            | CNN (LeNet-5)                            | Accuracy (95%-98%)           |
| Korbar et al.[36] | Tumor classification: 6 classes (NL/HP/SSP/TSA/TA/TVA-VA)         | 697 cases             | ResNet                                   | Accuracy (93.0%)             |
| Yoshida et al.[35] | Tumor classification                                                 | 1328 colorectal biopsy WSIs | ML                                       | Accuracy (90.1%, adenoma)    |
| Alom et al.[39] | Tumor microenvironment analysis: Classification, Segmentation and Detection | 21135 patches        | DCRN/R2U-Net                             | Accuracy (91.1%, classification) |
| Bychkov et al.[42] | Prediction of colorectal cancer outcome (5-yr disease-specific survival). | 420 cases         | Recurrent neural networks                 | HR of 2.3, AUC (0.69)        |
| Weiss et al.[50] | Evaluation of tumor budding                                           | 401 cases             | CNN                                      | Correlation R (0.86)         |
| Ponzio et al.[91] | Tumor classification: 3 classes (NL/AD/ADC)                          | 27 WSIs (13500 patches) | VGG16                                    | Accuracy (96 %)              |
| Kather et al.[34] | Tumor classification: 2 classes (NL/Tumor)                           | 94 WSIs               | ResNet18                                 | AUC (> 0.99)                 |
| Kather et al.[34] | Prediction of microsatellite instability                             | 360 TCGA-DX (93408 patches), 378 TCGA- KR (60894 patches) | ResNet18                         | AUC: TCGA-DX—(0.77, TCGA-DX; 0.86, TCGA- KR) |
| Kather et al.[26] | Tumor microenvironment analysis: classification of 9 cell types      | 86 WSIs (100000)      | VGG19                                    | Accuracy (94%-99%)           |
| Kather et al.[26] | Prognosis predictions                                                | 1296 WSIs             | VGG19                                    | Accuracy (94%-99%)           |
| Kather et al.[26] | Prognosis prediction                                                  | 934 cases             | Deep learning (comparison of 5 networks) | HR for overall survival of 1.99 (training set) and 1.63 (test set) |
| Geessink et al.[29] | Prognosis prediction, quantification of intratumoral stroma           | 129 cases             | Neural networks                          | HRS of 2.04 for disease-free survival |
| Sena et al.[40] | Tumor classification: 4 classes (NL/HP/AD/ADC)                      | 393 WSIs (12,565 patches) | CNN                                      | Accuracy (80%)               |
| Shapcott et al.[92] | Tumor microenvironment analysis: detection and classification         | 853 patches and 142 TCGA images | CNN with a grid-based attention network | Accuracy (84%, training set; 65%, test set) |
| Sirinukunwattana et al.[31] | Prediction of consensus molecular subtypes of colorectal cancer      | 1206 cases            | Neural networks with domain-adversarial learning | AUC (0.84 and 0.95 in the two validation sets) |
| Swiderska-Chadaj et al.[93] | Tumor Microenvironment Analysis: Detection of immune cell, CD3+, CD8+ | 28 WSIs               | FCN/LSM/U-Net                            | Sensitivity (74.0%)          |
| Yoon et al.[39] | Tumor classification: 2 classes (NL/Tumor)                           | 57 WSIs (10280 patches) | VGG                                      | Accuracy (93.5%)             |
| Echle et al.[46] | Prediction of microsatellite instability                             | 8836 cases            | ShuffleNet Deep learning                 | AUC (0.92 in development cohort; 0.96 in validation cohort) |
| Iizuka et al.[22] | Tumor classification: 3 classes (NL/AD/ADC)                          | 4036 WSIs             | CNN/RNN                                  | AUCs (0.96, ADC; 0.99, AD)   |
| Skrede et al.[28] | Prognosis predictions                                                | 2022 cases            | Neural networks with multiple instance learning | HR (3.04 after adjusting for established prognostic markers) |

NL: Normal mucosa; ADC: Adenocarcinoma; MC: Mucinous carcinoma; SC: Serrated carcinoma; PC: Papillary carcinoma; CCTA: Cribriform comedo-type adenocarcinoma; AD: Adenoma; CNN: Convolutional neural network; HP: Hyperplastic polyp; SSP: Sessile serrated polyp; TSA: Traditional serrated adenoma; TA: Tubular adenoma; TVA: Tubulovillous adenoma; VA: Villous adenoma; WSI: Whole slide images; ML: Machine learning; DCRN: Densely connected recurrent convolutional network; R2U-Net: Recurrent residual U-Net; HR: Hazard ratio; AUC: Area under the curve; TCGA: The Cancer
Table 3 Advantages and disadvantages of representative machine-learning methods in the development of artificial intelligence-models for gastrointestinal pathology

| AI model                        | Advantages                                                      | Disadvantages                                      |
|---------------------------------|----------------------------------------------------------------|---------------------------------------------------|
| Conventional ML (supervised)    | User can reflect domain knowledge to features                  | Requires hand-crafted features; Accuracy depends heavily on the quality of feature extraction |
| Conventional ML (unsupervised)  | Executable without labels                                       | Results are often unstable; Interpretability of the results |
| Deep neural networks (CNN)      | Automatic feature extraction; High accuracy                     | Requires a large dataset; Low explainability (Black box) |
| Multi-instance learning         | Executable without detailed labels                              | Requires a large dataset; High computational cost  |
| Semantic segmentation (FCN, U-Net) | Pixel-level detection gives the position, size, and shape of the target | High labeling cost                                 |
| Recurrent neural networks       | Learn sequential data                                           | High computational cost                            |
| Generative adversarial networks | Learn to synthesize new realistic data                         | Complexity and instability in training             |

AI: Artificial intelligence; ML: Machine learning; CNN: Convolutional neural network; FCN: Fully convolutional network

minimized, by defining cases for no further review by a pathologist, even in 'negative' gastric biopsies, other findings, in addition to neoplastic lesions, such as Helicobacter pylori infection, need to be reviewed and recorded. Therefore, AI application cannot be functional until it sufficiently represents diagnostic procedures of real-life practice.

The prediction of prognosis from histopathological images of GI cancers is also an attractive area for AI application. Considering the many types of histopathological prognostic features of cancer, such as tumor differentiation or lymphovascular involvement, the unveiling of hidden morphological features may be expected from AI for better prediction of clinical outcomes from the histopathological images alone[25-27]. After ingesting a sufficient number of histopathological images from patients with known outcomes, AI may comprehensively predict the patient's future outcomes. Recently, an exponentially increasing number of studies conducted for major GI cancers have demonstrated the feasibility of this concept[26,28,29]. Additionally, according to a recent study, tumor-infiltrating lymphocytes were associated with the prognosis of patients with gastric cancer[30]. CNN model may detect tumor-infiltrating lymphocytes on histopathological specimens with an acceptable accuracy of 96.9%[30]. The development of DL models that incorporate clinical and multi-omics data is also a promising approach for predictive purposes[19]. Prognosis prediction by AI applications might be more accurate than that by the conventional pathological method; however, these AI-based predictions alone seem not to be accepted in clinical practice due to lack of interpretability. If doctors and patients cannot understand the reason for prediction, they will not recognize misprediction by AI. We cannot provide patients' care based on prediction as in “fortune-telling.” Biological and clinical reasons for the prediction by AI application must be understood prior to its implementation into clinical practice.

Some researchers have also attempted to predict biomarker status from histopathological images alone using AI applications. Specimens of various GI cancers can be processed to identify molecular markers that may predict responses to targeted therapies. Research has shown that certain clinically relevant molecular alterations in GI cancers are associated with specific histopathological features detected on hematoxylin-eosin (HE) slides; there have been some successful attempts to adopt AI applications for HE sections as surrogate markers for these alterations[31-34].

**Histopathological AI-applications in colorectal cancer**

As in gastric cancer, various AI applications have recently been developed for colorectal cancer (Table 2). Regarding tumor classification, several AI algorithms have been trained to classify the dataset into two to six specific classes, such as normal, hyperplasia, adenoma, adenocarcinoma, and histological subtypes of polyps or...
Yoshida H et al. AI implementation to GI pathology practice

Figure 2 Challenges for implementation in the development process of an artificial intelligence application. The process of development and implementation of an artificial intelligence (AI) application is composed of multiple steps from needs identification to use in real-life (left). In each step, various challenges keep AI applications from being implemented into clinical practice (right). AI: Artificial intelligence; IT: Information technology.

Further, a few studies have predicted prognosis using pathological images for colorectal cancer[26,34,42]. Bychkov et al[42] used 420 tissue microarray-WSIs to predict the 5-year disease-specific survival of patients and obtained an AUC of 0.69. Kather et al[26] used more than 1000 histological images, collected from three institutions, to predict the prognosis of the patient; they observed accuracy of 99%. Another study, using the ResNet model for direct identification of microsatellite instability (MSI) on histological images, demonstrated an AUC of 0.77 for both FFPE and frozen specimens from The Cancer Genome Atlas (TCGA)[34]. The identification of colorectal cancer with MSI is crucial; these tumors are reportedly highly responsive to immunomodulating therapies[43,44]; moreover, the MSI could be a clue for the diagnosis of Lynch syndrome[45]. MSI is usually identified by polymerase chain reaction (PCR), but not all patients are screened for MSI in clinical practice. Echle et al[46] recently developed a DL model to detect colorectal cancer with MSI using more than 8800 images. The DL algorithm demonstrated an AUC of 0.96 in the multi-institutional validation cohort. Furthermore, the consensus molecular subtype of...
colorectal cancer could be predicted from the images of colorectal surgical specimens using a CNN-based model[31]. Although prediction of molecular alterations by AI application might seem attractive, as clinically relevant biomarkers cannot be identified using HE stained slides and conventional PCR assay are both expensive and time-consuming, AI can neither achieve complete concordance with the gold standard test nor replace it. Thus, users must consider how to employ AI for predicting biomarkers with an appropriate, cost-effective balance in real-life practice.

A ROAD TO IMPLEMENTATION OF AI APPLICATIONS INTO REAL-LIFE PRACTICE

To achieve clinical implementation of the AI, several steps should be considered (Figure 2). Colling et al[47] presented an expected roadmap for the routine use of AI in pathology practice. They highlighted the main aspects of designing and applying AI in daily practice. The steps concerning design creation, ethics, financing, development, validation and regulation, implementation, and effect on the workforce were closely reviewed. For pathological image analysis, various problems exist in the execution of these steps, which would prevent the AI from being implemented in the clinical practice for GI cancers.

Identification of the true needs in daily practice

AI applications can either conduct routine tasks, usually performed by pathologists, or offer novel insights into diseases that are not possible by human pathologists[12]. The applications are needed to fill gaps and address unmet needs without impacting the daily workflow in the pathology department. The needs include mitosis detection, tumor-percentage calculation, lymph node metastasis, and other activities that are considered monotonous, repetitive, or vulnerable to higher interobserver variability.

The initial step in the development of the AI application is to recognize the true clinical need and define a possible solution. The novel AI applications can be developed by various stakeholders, including pathologists, physicians, computer scientists, engineers, IT companies, and drug companies. However, viewpoints between the professionals in academia and industry differ. For example, individuals in academia and businesses have different goals, such as grant funding, academic publications, and profitable commercial products.

Even if there is a problem that pathologists are eager to solve, the market size of the problem could be small. If the cost of developing an AI application to solve the problem cannot be recovered by the subsequent profit from the sale of the application, the company may not develop it. There is a wide range of classification tasks in diagnostic pathology, and it is difficult to secure an appropriate market for an AI application specializing only in a single task. For example, an AI algorithm can detect lymph node metastases in breast cancer as reliably as human pathologists[48,49]. Still, this tool has not been widely used or approved by the regulatory authorities. Although there could be many reasons, one is the imbalance between the overall cost of its implementation and the benefit of detecting only breast cancer lymph node metastases in real-life pathology practice.

Another significant concern is obtaining consent for the use of patient data in AI-model development[50]. Although the consent for research use could be obtained in most studies, patients might not consent to commercial use of their data required for product development, which could be an obstacle when developing products for clinical implementation. Therefore, consent should be obtained at the beginning of the research, conveying the possibility of its commercial use for product development; a framework for global data sharing should be developed.

For the development of AI algorithms, at least three parties need to collaborate, which include pathologists who know the true needs, academic professionals who can develop technology, and companies that will promote AI applications as products. In addition, to obtain a sufficiently sized market, it may be vital to develop global networks and online services using the cloud.

Development

After a concept of AI has been conceived and collaboratively established, the development of AI is carried out through the following steps: defining the output, designing the algorithm, collection of a pilot or larger follow-up sample, annotation and processing of data, and performing statistical analysis of the data.
High-quality data set curation is one of the major hurdles in the development of AI applications. Generally, CNNs require hundreds or thousands of data sets of pathological images to achieve significant performance and sufficient generalizability\cite{51}. For rare tumors, researchers can obtain a very limited number of images; thus, it requires efficient data augmentation techniques and learning methods to resolve this issue. Conversely, in the case of transfer learning, small-scale datasets consisting of $< 100$ digital slides may suffice\cite{52}.

In addition, publicly available datasets should be developed for global data sharing. However, few such datasets are available in pathology, partly due to confidentiality, copyright, and financial problems\cite{53}. Even under such circumstances, TCGA provides many WSIs and associated molecular data\cite{54}. However, even TCGA data does not include sufficient numbers of cases for training AI applications for clinical implementation. Another potential source of datasets could be the public challenges provided for developing DL algorithms\cite{55}.

The development of AI applications with sufficient performance needs training on huge datasets demonstrating scanning\cite{36} and staining protocol variability\cite{56,57}. The major challenges for its implementation into practice are laboratory infrastructure and reproducibility and robustness of the AI model. Recently, automated methods for reducing blur in images have been developed. Automated algorithms (for example, HistoQC\cite{58} and DeepFocus\cite{59}) can reportedly standardize the quality of WSIs; these AI applications automatically detects optimum quality regions and eliminates out-of-focus or artifact-related regions. Standardization of the color, displayed by histopathological slides, is important for the accuracy of AI; the color variations are often produced due to differences in batches or manufacturers of staining reagents, variations in the thickness of tissue sections, the differences in staining protocols, and disparity in scanning characteristics. These variations lead to inadequate classification by AI applications\cite{56,60}. AI algorithms have been developed to standardize the data\cite{61}, including staining\cite{62} and color characteristics\cite{63}.

After data set curation, the annotation of the dataset is required. Histopathological image annotation is not a simple task. The extent of annotation detail depends on the application of AI, which could vary from classification at the slide level to labeling at the pixel level. The annotation task, for many images, by human experts is time-consuming and tedious. In addition, variability in annotation performance, especially when the task is difficult, may affect the accuracy of the trained models. Moreover, for manufacturers, this task could be often expensive. Among GI pathologies, many lesions, such as intramucosal gastric carcinoma, do not have high interobserver reproducibility. When developing an AI application to assist pathologists in making a diagnosis, if the target disease shows significant interobserver variability, the correctness of the annotation of the dataset cannot be guaranteed, and the trained algorithm may not be able to reproduce performance in the dataset when used in other facilities, which may hinder its clinical implementation.

The problem of annotation in AI is an important research area. The majority of the AI models are trained using images of small tissue patches collected from WSIs. Since the patches, cropped from positive tissue, may not contain a tumor unless the tissue is filled with tumors, it is challenging to construct a high-accuracy model, particularly when pixel-level labeling is unavailable. To conduct patch-based training, without detailed annotation, multi-instance learning (MIL) algorithm can be used\cite{64,65}. Cosatto et al\cite{65} employed MIL for gastric cancer detection; they used over 12000 cases, $2/3$rd for training and $1/3$rd for the test, and achieved an AUC of 0.96. MIL is especially effective when there is a large dataset, and detailed annotations are impossible to obtain\cite{51}.

After the preparation of the annotated dataset, the model development process is usually composed of the following steps: preparation of the datasets for training, testing, and validation; selecting the ML framework, ML technique, and learning method. Once the learning process is completed, the output of the model is evaluated through performance metrics, and the hyperparameters are fine-tuned to improve performance. Considering the exponential increase in AI research for image analysis, this step does not seem to be a major obstacle to the implementation of AI in clinical practice.

**Validation and regulation**

As AI-based technologies grow increasingly, an evidence-based approach is required for their validation. Colling et al\cite{47} presented summarized guidance by the current in vitro device regulation and their recommendations for the main components of validation. In laboratory medicine, apart from clinical evaluation, analytical validation should be considered\cite{66}. The establishment of steps and criteria for the validation of
new tests against existing gold standards is essential. For image analysis validation, the technique is often compared with the "ground truth" (for example, comparing an AI-technology analyzing HER2 expression within the tumor to a detailed tumor assessment performed manually). It would be appropriate to compare the digital pathology technique with the performance of human pathologists. However, considering inter- and intra-observer variability in visual assessments of human pathologists, it is difficult to identify the ground truth; thus, it involves careful designing of the study and acceptance of the limitations of the present gold standard. Currently, most AI applications seem to have difficulty in establishing absolute ground truth. Therefore, the robustness and reproducibility of AI applications should be repeatedly validated in large and variable patient cohorts.

The relative lack of a validation cohort is an urgent issue in the development of AI-based applications. Histopathological slides, with detailed clinical data linked to them, cannot be often shared widely for reasons such as privacy protection. Annotations by pathologists, which are usually considered the "ground truth", are still controversial. Inter-observer variability and subjectivity in assessments by a pathologist indicate that a certain amount of uncertainty is inherent to ground truth. However, where the pathologist’s assessment is the only available ground truth, it is important to enhance accuracy through validation as the next best measure. Efficient validation and testing require multicenter assessments involving multiple pathologists and datasets. If the AI application is intended to be used in real-life practice, it should be robust against pre-analytical variations within the target images, such as differences in staining conditions and WSI scanners, and its performance should be reproducible. With respect to this, a significant proportion of currently published AI research in GI cancers has not been externally validated.

### Regulatory challenges

Appropriate regulations are required for the safe and effective use of AI in pathological practice. Unlike other laboratory tests, it is difficult to understand how predictions are made in AI applications; therefore, they are often viewed as black boxes. While various visualization techniques, including gradient saliency maps and filter visualization methods, have been developed, it may not be possible for users to fully understand all the parameter changes causing erroneous performance or misprediction. Regulatory approval should be structured to minimize potential harm, define the risk-benefit balance, develop appropriate validation standards, and promote innovation.

Regulatory authorities, such as the FDA, the Centers for Medicare and Medicaid Services (CMS), and the European Union Conformité Européenne (EUCE) are not yet completely prepared for the implementation of AI applications in clinical medicine. As a result, AI-based devices are being controlled by prior and potentially obsolete guidelines for testing medical devices.

In the United States, the FDA is devising novel regulations for AI-based devices to make them safer and more effective. CMS controls laboratory testing through the Clinical Laboratory Improvement Amendments (CLIA). CLIA stipulates that appropriate validation must be performed for all laboratory tests using human tissue before clinical implementation, regardless of their FDA approval. Currently, CLIA has no specific regulations for validating AI applications. The EUCE will replace the medical device directive in May 2021, and in vitro diagnostic medical device directives will be replaced by in vitro diagnostic regulation in May 2022. Successful clinical implementation of AI-based applications will be assisted by the global market, and those clinically enforcing the applications will need to pay particular attention to the regulatory trends in their own country as well as in the US and EU. For AI applications to be approved by the FDA and EUCE, they should be established based on the updated details on FDA and EUCE regulations.

### Implementation

Before implementing an AI application in real-life pathology practice, several obstacles must be addressed. Established business-use cases and a guarantee from pathologists for the use of the AI system should be accounted for before investing substantial time, energy, and funds on AI applications and required IT infrastructure.

The changes required for shifting daily workflow in the pathology department, from glass slides to WSIs, must be addressed. The department would require new digital pathology-related devices, a specific data management system, data storage facilities, and additional personnel to handle these changes. Simultaneously, an institutional IT infrastructure is required to enable users to operate through both on-site and cloud-based computing systems. Therefore, in the real-world, digital
pathology systems, requiring substantial investment, may hamper the implementation of these technologies[71]. Notably, augmented microscopy, connected directly to the cloud network service, might solve the issue of whole slide scanner installation. Chen and colleagues reported the augmented reality microscope, overlaying AI-based information onto the sample-view in real-time, may enable a seamless integration of AI into the routine workflow[72]. According to Hegde et al[73], the cloud-based AI application (SMILY, Similar image search for histopathology), developed by GOOGLE, irrespective of its annotation status, allows the search for morphologically similar features in a target image.

In addition, one must consider the relative inexperience of pathologists with AI-based technologies and acknowledge the range of issues the department would encounter prior to the implementation of AI. Second, a pathologist must buy-in to make significant improvements in a conventional century-old workflow. In view of the fact that progress does not happen immediately, the pathologist’s management concerns should be dealt with separately from the technological hurdles. Initially, pathologists must commit to the installation of both digital pathology systems and AI applications to a pathology department. They have to understand the long-term risk-benefit balance of AI implementation. The present DL-based AI applications lack interpretability, which may contribute to patients’ and clinicians’ reluctance. Developing AI solutions that can be interpreted by end-users, thereby providing them with detailed descriptions of how their predictions are made, could be useful[74]. For lack of interpretability of DL model, various solutions, such as generating attention heat map[75], constructing interpretable model[76], creating external interpretive model[77], have been reported. However, this black box problem is not yet fully resolved.

On the downside, dependence on AI assistance for diagnoses can result in fewer opportunities for trainees to learn diagnostic skills. Although AI can be used as an auxiliary method to improve the quality and precision of clinical diagnoses, resident pathologists should be trained and encouraged to understand the utility, limitations, and pitfalls of AI application[78]. As molecular pathologists have become necessary, since the advent of genomic medicine, “computational pathologists”[47] will become necessary in the near future.

As with other clinical tests, ongoing post-marketing quality assurance is also essential for the safe and effective use of AI in clinical practice. Apart from laboratory testing processes, laboratory staff should understand the quality management system. As in conventional laboratory tests, a novel scheme of external quality assurance for AI applications in pathology should be urgently prepared for its implementation.

The use of AI applications in diagnostic practice poses complex new issues around the legal ramifications of signing a report prepared using AI by a pathologist. In order to incorporate their output into a pathological report, a pathologist should be confident in the performance of the algorithm; further, any algorithms used should be validated and regulated correctly. Although AI applications may not replace pathologists in view of this legal issue, they can be employed to support the pathologists in their clinical work. In particular, AI researchers are attempting to provide their predictions/results with confidence estimates and localize pathology-related features. This could help mitigate interpretability and confidence-building concerns.

**CONCLUSION**

The immense potential of AI in pathological practice can be harnessed by improving workflows, eliminating simple mistakes, increasing diagnostic reproducibility, and revealing predictions that are impossible with the use of conventional visual methods by human pathologists. The clinically implemented AI applications are expected to be user-friendly, explainable, robust, manageable, and cost-effective. Considering the current limited clinical awareness and uncertainty about how AI tools can be introduced into real-life practice, caution should be paid to their deployment. Eventually, AI applications may be implemented and used appropriately, provided they are supported by human pathologists, standardized usage recommendations, and harmonization of AI applications with present information systems.

AI can play a pivotal role in the practice of pathologists and the development of precision medicine for GI cancers. However, there are various barriers to its effective implementation. To overcome these barriers and implement AI at the practice level, it is necessary to work with a range of stakeholders, including pathologists, clinicians,
developers, regulators, and device vendors, to establish a strong network to grab true needs, expand the market, and use the application safely and efficiently.

REFERENCES

1 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019; 25: 44-56 [PMID: 30617339 DOI: 10.1038/s41591-018-0300-7]

2 Andras I, Mazzone E, van Leeuwen FWB, De Nayer G, van Oosterom MN, Beato S, Buckle T, O'Sullivan S, van Leeuwen PJ, Beulens A, Crisan A, D'Hondt F, Schatteman P, van Der Poel H, Dell'Oglio P, Mottrie A. Artificial intelligence and robotics: a combination that is changing the operating room. World J Urol 2020; 38: 2359-2366 [PMID: 31776737 DOI: 10.1007/s00345-019-03037-6]

3 Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Belgafia S, Cacciaveghe NG, Cathro HP, Cheng J, Cooper K, Dickey GE, Gill RM, Heaton RP Jr, Kerstens R, Lindberg GM, Malhotra RK, Mandell JW, Manlucu ED, Mills AM, Mills SE, Moskaluk CA, Nelis M, Patil DT, Przybycien CG, Reynolds JP, Rubin BP, Saboorian MH, Salicru M, Samols MA, Sturgis CD, Turner KO, Wick MR, Yoon YJ, Zhao P, Taylor CR. Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology: A Multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study). J Surg Pathol 2018; 42: 39-52 [PMID: 28961557 DOI: 10.1007/PAS.0000000000004685]

4 Nazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. Lancet Oncol 2019; 20: e253-e261 [PMID: 31044723 DOI: 10.1016/S1470-2045(19)30154-8]

5 Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 2019; 16: 703-715 [PMID: 31399609 DOI: 10.1038/s41571-019-0252-7]

6 Rashidi HH, Tran NK, Betts EV, Howell LP, Green R. Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods. Acad Pathol 2019; 6: 2374289519873088 [PMID: 31523704 DOI: 10.1177/2374289519873088]

7 LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521: 436-444 [PMID: 26017442 DOI: 10.1038/nature14539]

8 Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006; 313: 504-507 [PMID: 16873662 DOI: 10.1126/science.1127647]

9 Jain RK, Mehta R, Dimitrov R, Larsson LG, Musto PM, Hodges KB, Ulbright TM, Hattab EM, Agaram N, Idrees MT, Badve S. Atypical ductal hyperplasia: interobserver and intraobserver variability. Mod Pathol 2011; 24: 917-925 [PMID: 21525546 DOI: 10.1038/modpathol.2011.66]

10 Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson AN, Nelson HD, Pepe MS, Allison KH, Schnitt SJ, O'Malley FP, Weaver DL. Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA 2015; 313: 1122-1132 [PMID: 25781441 DOI: 10.1001/jama.2015.1405]

11 Jiang Y, Yang M, Wang S, Li X, Sun Y. Emerging role of deep learning-based artificial intelligence in tumor pathology. Cancer Commun (Lond) 2020; 40: 154-166 [PMID: 32277744 DOI: 10.1002/cac.21202]

12 Calderaro J, Kather JN. Artificial intelligence-based pathology for gastrointestinal and hepatobiliary cancers. Gut 2020 [PMID: 33214163 DOI: 10.1136/gutjnl-2020-322880]

13 Niu PH, Zhao LL, Wu HL, Zhao DB, Chen YT. Artificial intelligence in gastric cancer: Application and future perspectives. World J Gastroenterol 2020; 26: 5408-5419 [PMID: 33024393 DOI: 10.3748/wjg.v26.i36.5408]

14 Thakur N, Yoon H, Chong Y. Current Trends of Artificial Intelligence for Colorectal Cancer Pathology Image Analysis: A Systematic Review. Cancers (Basel) 2020; 12: [PMID: 32668721 DOI: 10.3390/cancers12071884]

15 Khan A, Nawaz U, Ullaq A, Robinson RW. Real-time plant health assessment via implementing cloud-based scalable transfer learning on AWS DeepLens. PLoS One 2020; 15: e0243243 [PMID: 33332376 DOI: 10.1371/journal.pone.0243243]

16 PathAI. PathAI: Present Machine Learning Models that Predict the Homologous Recombination Deficiency Status of Breast Cancer Biopsies at the 2020 SABCS, [cited 7 January 2021]. In: PathAI [Internet]. Available from: https://www.pathai.com/news/pathai-sabcs2020

17 Pantanowitz L, Sinard JH, Henricks WH, Fatherree LA, Carter AB, Contis L, Beckwith BA, Evans AJ, Lal A, Parwani AV; College of American Pathologists Pathology and Laboratory Quality Center. 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 2013; 137: 1710-1722 [PMID: 23636907 DOI: 10.5888/apra.2013-0093-CP]

18 Cheng CL, Azhar R, Sng SH, Chua YQ, Hwang JS, Chin JP, Seah WK, Loke JC, Ang RH, Tan PH. Enabling digital pathology in the diagnostic setting: navigating through the implementation journey in an academic medical centre. J Clin Pathol 2016; 69: 784-792 [PMID: 26878390 DOI: 10.1136/jclinpath-2015-203600]

19 Hamamoto R, Suvarna K, Yamada M, Kobayashi K, Shinkai N, Miyake M, Takahashi M, Jinmai S, Shinnoyama R, Sakai A, Takasawa K, Bolatkan A, Shouz K, Dozen A, Machino H, Takahashi S, Asada K, Komatsu M, Sese I, Kaneko S. Application of Artificial Intelligence Technology in
Oncology: Towards the Establishment of Precision Medicine. *Cancers (Basel)* 2020; 12 [PMID: 33256107 DOI: 10.3390/cancers12123532]

20 de Groof AJ, Struyvenberg MR, van der Putten J, van der Sommen F, Fockens KN, Curvers WL, Zinger S, Pouw RE, Coron E, Baldaque-Silva F, Pech O, Weusten B, Meining A, Neuhauß H, Bisschops R, Dent J, Schoon EJ, de With PH, Bergman JJ. Deep-Learning System Detects Neoplasia in Patients With Barrett’s Esophagus With Higher Accuracy Than Endoscopists in a Multistep Training and Validation Study With Benchmarking. *Gastroenterology* 2020; 158: 915-929. e4 [PMID: 31759929 DOI: 10.1053/j.gastro.2019.11.030]

21 Sharma H, Zerbe N, Klempt I, Hellwich O, Hufnagl P. Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. *Comput Med Imaging Graph* 2017; 61: 2-13 [PMID: 28676295 DOI: 10.1016/j.compmedimag.2017.06.001]

22 Iizuka O, Kanavati F, Kato K, Rambeau M, Arihiro K, Tsuneki M. Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours. *Sci Rep* 2020; 10: 1504 [PMID: 32001752 DOI: 10.1038/s41598-020-58467-9]

23 Yoshida H, Shimazu T, Kiyuna T, Marugame A, Yamashita Y, Cosatto E, Taniguchi H, Sekine S, Ochiai A. Automated histological classification of whole-slide images of gastric biopsy specimens. *Gastric Cancer* 2018; 21: 249-257 [PMID: 28577229 DOI: 10.1016/s1002-0079(17)30317-x]

24 Tomita N, Abdollahi B, Wei J, Ren B, Suriawinata A, Hassanpour S. Attention-Based Deep Neural Networks for Detection of Cancers and Precancerous Esophagus Tissue on Histopathological Slides. *JAMA Netw Open* 2019; 2: e1914645 [PMID: 31693124 DOI: 10.1001/jamanetworkopen.2019.14645]

25 Courtiol P, Maussion C, Morari M, Pronier E, Pilcer S, Seffa M, Manceron P, Toldo S, Zaslavskiy M, Le Stang N, Girard N, Elemento O, Nesheim J, Albregtsen F, Farstad IN, Domingo E, Nelson TE, Lawer CA, Zerbe N, Klempert I, Hellwich O, Hufnagl P. Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. *Gastric Cancer* 2020; 21: 1519-1525 [PMID: 31591589 DOI: 10.1016/j.gastr.2019.04.083-3]

26 Kather JN, Krismam J, Charoentong P, Luedde T, Herrpel E, Weiss CA, Gaiser T, Marx A, Valousov NA, Ferber D, Jansen L, Reyes-Aldasoro CC, Zörnig I, Jüger D, Brenner H, Chang-Claude J, Hoffmeister M, Halama N. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLOs Med* 2019; 16: e1002730 [PMID: 30677016 DOI: 10.1371/journal.pmed.1002730]

27 Luo X, Yin S, Yang L, Fujimoto J, Yang Y, Moran C, Kalhorn N, Weisserdt A, Xie Y, Gazdar A, Minna J, Wistuba II, Mao Y, Xiao G. Development and Validation of a Pathology Image Analysis-based Predictive Model for Lung Adenocarcinoma Prognosis - A Multi-cohort Study. *PLoS Med* 2019; 16: e1002730 [PMID: 30677016 DOI: 10.1371/journal.pmed.1002730]

28 Geessink OGF, Boidoulivi A, Klaase JM, Esheshtani Bejnordi B, Litjens GJS, van Pelt GW, Mekser WE, Nagtegaal ID, Cimpoi F, van der Laak JAWM. Computer aided quantification of intratumoral stroma yields an independent prognosticator in rectal cancer. *Cell Oncol (Dordr)* 2019; 42: 331-341 [PMID: 30825182 DOI: 10.3390/cancers12123532]

29 García E, Hermozra R, Beltran-Castanon C, Cano L, Castillo M, Castaneda C. Automatic Lymphocyte Detection on Gastric Cancer IHC Images Using Deep Learning. *IEEE* 2017; 200-204 [DOI: 10.1109/CBMS.2017.94]

30 Sirinukunwattana K, Domingo E, Richman SD, Redmond KL, Blake A, Verrill C, Leedham SJ, Chatzizipis A, Hardy C, Whalley CM, Wu CH, Begg MS, McDermott U, Dunne PD, Meade A, Walker SM, Murray GI, Samuel L, Seymour M, Tomlinson I, Quirke P, Maughan T, Rittscher J, Koelzer VH; S:CORT consortium. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut* 2021; 70: 544-554 [PMID: 32690604 DOI: 10.1136/gutjnl-2019-319866]

31 Fu Y, Jung AW, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, Yates LR, Jimenez-Linan M, Moore L, Gerstung M. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Cancer* 2020; 1: 800-810 [DOI: 10.1038/s41555-020-0085-8]

32 Kather JN, Heij LR, Grabhsch HI, Loeflter C, Echle A, Muti HS, Krause J, Niehues JM, Sommer KA, Bankhead P, Kooreman LFS, Schulte JJ, Cipriani NA, Buelow RD, Boor P, Ortiz-Brüchle N, Hanby AM, Spears V, Kochanny S, Patnaik A, Aksuwasanukorn A, Brenner H, Hoffmeister M, van den Brandt PA, Jüger D, Trautwein C, Pearson AT, Luedde T. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat Cancer* 2020; 1: 789-799 [DOI: 10.1038/s43018-020-0087-6]

33 Kather JN, Pearson AT, Halama N, Jüger D, Krause J, Loosen SH, Marx A, Boor P, Tacke F, Neumann UP, Grabhsch HI, Yoshiokawa T, Brenner H, Chang-Claude J, Hoffmeister M, Trautwein C, Luedde T. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* 2019; 25: 1054-1056 [PMID: 31160815 DOI: 10.1038/s41591-019-0462-y]

34 Yoshida H, Yamashita Y, Shimazu T, Cosatto E, Kiyuna T, Taniguchi H, Sekine S, Ochiai A. Automated histological classification of whole slide images of colorectal biopsy specimens.

---

**Yoshida H et al. AI implementation to GI pathology practice**
deep learning on whole slide images. Reuter VE, Klimstra DS, Fuchs TJ. Clinical-grade computational pathology using weakly supervised learning. *Nat Med* 2017; 23: 1089-1094 DOI: 10.1038/nm.4287.

40. Venâncio R. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *J Am Coll Radiol* 2017; 14: 1104-1110 DOI: 10.1016/j.jacr.2017.02.004.

41. Kalinovsky A, Liauchuk V, Bueno G, Fernandez-Carrobles MM, Serrano I, Deniz O, Racoceanu D, Seno S, Navab N, Watanabe S, Chang EL. Large scale tissue histophotometry image classification, segmentation, and visualization via deep convolutional activation features. *BMC Bioinformatics* 2017; 18: 281 DOI: 10.1186/s12859-017-1685-x.

42. Bychkov D, Linder N, Turki R, Nordling S, Kovanes PE, Verrill C, Walliander M, Lundin M, Haglund C, Lundin J. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep* 2018; 8: 3395 DOI: 10.1038/s41598-018-1758-3.

43. Mandal R, Samstein RM, Lee KW, Havel JJ, Wang H, Krishna C, Sabio EY, Makarov V, Kuo F, Bleca P, Ramassamy AT, Durham JN, Bartlett B, Ma X, Srivastava R, Middha S, Zehir A, Hechtman JF, Morris LG, Weinhold N, Riaz N, Le DT, Diaz LA Jr, Chri TA. Genetic diversity of tumors with mismatch repair deficiency influences anti-PD-1 immunotherapy response. *Science* 2019; 364: 485-491 DOI: 10.1126/science.aau0447.

44. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, Biedrzycki B, Donehower RC, Zeaheer A, Fisher GA, Croenzen TS, Lee JJ, Duffy SM, Goldberg RM, de la Chapelle A, Koshiji M, Bhaije F, Huebner T, Hruban RH, Wood LD, Cuka N, Pardom DM, Papadopoulos N, Kinzler KW, Zhou S, Cornish TC, Taube JM, Anders RA, Eshleman JR, Vogelstein B, Diaz LA Jr. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* 2015; 372: 2509-2520 DOI: 10.1056/NEJMoa1500596.

45. Lynch HT, Snyder CL, Shaw TG, Heind CD, Hitchins MP. Milestones of Lynch syndrome: 1895-2015. *Nat Rev Cancer* 2015; 15: 181-194 DOI: 10.1038/nrc3878.

46. Echle A, Grabsch HI, Quirke P, van den Brandt PA, West NP, Hitchens GGA, Heij LR, Tan X, Richman SD, Krause J, Alwers E, Jenniskens J, Offermans K, Gray R, Brenner H, Chang-Claude J, Graff JP, Darrow M, Borowsky A, Olson KA, Gandour-Edwards R, Datta Mitra A, Wei H, Kather JN. Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning. *J Pathol* 2015; 236: 284-293 DOI: 10.1002/path.4370.

47. Colling R, Pitman H, Oien K, Rajpoot N, Macklin P; CM-Path AI in Histopathology Working Group, Snead D, Sackville T, Verrill C. Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. *J Pathol Inform* 2019; 10: 1 DOI: 10.18632/oncotarget.21819.
Yoshida H et al. AI implementation to GI pathology practice

D, Gao G, Durbin-Johnson B, Rashidi HH. Impact of pre-analytical variables on deep learning accuracy in histopathology. Histopathology 2019; 75: 39-53 [PMID: 30801768 DOI: 10.1111/his.13844]

Hipp JD, Sica J, McKenna B, Monaco J, Madabhushai A, Cheng J, Balis UJ. The need for the pathology community to sponsor a whole slide imaging repository with technical guidance from the pathology informatics community. J Pathol Inform 2011; 2: 31 [PMID: 21845229 DOI: 10.4103/2153-5539.83191]

Cooper LA, Demicco EG, Saltz JH, Powell RT, Rao A, Lazar AJ. PanCancer insights from The Cancer Genome Atlas: the pathologist’s perspective. J Pathol 2018; 244: 512-524 [PMID: 29288495 DOI: 10.1002/path.5028]

Hartman DJ, Van Der Laak JAWM, Gurcan MN, Pantanowitz L. Value of Public Challenges for the Development of Pathology Deep Learning Algorithms. J Pathol Inform 2020; 11: 7 [PMID: 32318315 DOI: 10.4103/jpi.jpi_64_19]

Yoshida H, Yokota H, Singh R, Kiyuna T, Yamaguchi M, Kikuchi S, Yagi Y, Ochiai A. Meeting Report: The International Workshop on Harmonization and Standardization of Digital Pathology Image, Held on April 4, 2019 in Tokyo. Pathobiology 2019; 86: 322-324 [PMID: 31707388 DOI: 10.1159/0005052718]

Inoue T, Yagi Y. Color standardization and optimization in whole slide imaging. Clin Diagn Pathol 2020; 4 [PMID: 3308926 DOI: 10.15761/cdp.100019]

Janowczyk A, Zuo R, Gilmore H, Feldman M, Madabhushi A. HistQC: An Open-Source Quality Control Tool for Digital Pathology Slides. JCO Clin Cancer Inform 2019; 3: 1-7 [PMID: 30990737 DOI: 10.1200/CCJ.18.00157]

Senaras C, Niazi MKK, Lozanski G, Gurcan MN. DeepFocus: Detection of out-of-focus regions in whole slide digital images using deep learning. PLoS One 2018; 13: e0205387 [PMID: 30359393 DOI: 10.1371/journal.pone.0205387]

Komura D, Ishikawa S. Machine Learning Methods for Histopathological Image Analysis. Comput Struct Biotechnol J 2018; 16: 34-42 [PMID: 32075936 DOI: 10.1016/j.csbj.2018.01.001]

Yagi Y, Gilbertson JR. Digital imaging in pathology: the case for standardization. J Telemed Telecare 2005; 11: 109-116 [PMID: 15901437 DOI: 10.1258/1357630053687070]

Janowczyk A, Basavanahally A, Madabhushi A. Stain Normalization using Sparse AutoEncoders (StaNoSA): Application to digital pathology. Comput Med Imaging graph 2017; 57: 50-61 [PMID: 27373749 DOI: 10.1016/j.compmedimag.2016.05.003]

Vahadade A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, Steiger K, Schlitter AM, Esposito I, Navab N. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. IEEE Trans Med Imaging 2016; 35: 1962-1971 [PMID: 27164577 DOI: 10.1109/TMI.2016.2529665]

Dietterich TG, Latthrop RH, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence 1997; 89: 31-71 [DOI: 10.1016/S0004-3702(96)00034-3]

Cosatto E, Laquerre PF, Malon C, Graf HP, Saito A, Kiyuna T, Marugane A, Kamijo K. Automated gastric cancer diagnosis on H and E-stained sections; training a classifier on a large scale with multiple instance machine learning. Proceedings of SPIE - Progress in Biomedical Optics and Imaging, MI; 2013

Mattocks CJ, Morris MA, Matthijs G, Swinnen G, Corveleyn A, Dequeker E, Müller CR, Pratt V, Wallace A; EuroGentest Validation Group. A standardized framework for the validation and verification of clinical molecular genetic tests. Eur J Hum Genet 2010; 18: 1276-1288 [PMID: 20666432 DOI: 10.1038/ejhg.2010.101]

Pasa F, Golkov V, Pfeiffer F, Cremers D, Pfeiffer D. Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization. Sci Rep 2019; 9: 6268 [PMID: 31000728 DOI: 10.1038/s41598-019-42557-4]

Allen TC. Regulating Artificial Intelligence for a Successful Pathology Future. Arch Pathol Lab Med 2019; 143: 1175-1179 [PMID: 31268373 DOI: 10.5858/arpa.2019-0229-ED]

U.S. Food and Drug Administration. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD). [cited 7 January 2021]. In: U.S. Food and Drug Administration [Internet]. Available from: https://www.fda.gov/media/122535/download

European Commission. Medical Devices – Sector. [cited 7 January 2021]. In: European Commission [Internet]. Available from: https://ec.europa.eu/growth/sectors/medical-devices_en

Retamero JA, Aneiros-Fernandez J, Del Moral RG. Complete Digital Pathology for Routine Histopathology Diagnosis in a Multicenter Hospital Network. Arch Pathol Lab Med 2020; 144: 221-228 [PMID: 31295015 DOI: 10.5858/arpa.2018-0541-0A]

Chen PC, Gadepalli K, MacDonald R, Liu Y, Kadoski S, Nagpal K, Kohlberger T, Dean J, Corrado GS, Hipp JD, Mermal CH, Stumpe MC. An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis. Nat Med 2019; 25: 1453-1457 [PMID: 31406351 DOI: 10.1038/s41591-019-0359-7]

Hegde N, Hipp JD, Liu Y, Emmert-Buck M, Reif E, Smilkov D, Terry M, Cai CJ, Amin MB, Mermal CH, Nelson PQ, Peng LH, Corrado GS, Stumpe MC. Similar image search for histopathology: SMILY. NPJ Digit Med 2019; 2: 56 [PMID: 31304402 DOI: 10.1038/s41746-019-0131-z]

Tosun AB, Pullara F, Becich MJ, Taylor DL, Fine JL, Chernubhotla SC. Explainable AI (xAI) for Anatomic Pathology. Adv Anat Pathol 2020; 27: 241-250 [PMID: 32541594 DOI: 10.1097/PAP.0000000000000840]
Learning to detect lymphocytes in immunohistochemistry with deep learning. 

Manson Q, Sherman M, Polonia A, Parry J, Abubakar M, Litjens G, van der Laak J, Ciompi F. Swiderska-Chadaj Z

Bioeng Biotechnol Shapcott M

Convolutional Networks - An Experimental Study. Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: Bioimaging, 2018: 58-66

DOI: 10.1109/ACCESS.2019.2918800

Shabbir A, Huang Z. Real-time Raman spectroscopy for in vivo, online gastric cancer diagnosis during clinical endoscopic examination. J Biomed Opt 2012; 17: 081418 [PMID: 23224179 DOI: 10.1117/1.JBO.17.8.081418]

Jiang Y, Xie J, Han Z, Liu W, Xi S, Huang L, Huang W, Lin T, Zhao L, Hu Y, Yu J, Zhang Q, Li T, Cai S, Li G. Immunomarker Support Vector Machine Classifier for Prediction of Gastric Cancer Survival and Adjuvant Chemotherapeutic Benefit. Clin Cancer Res 2018; 24: 5574-5584 [PMID: 30042208 DOI: 10.1158/1078-0432.CCR-18-0848]

Qu J, Hiruta N, Terai K, Nosato H, Murakawa M, Sakanashi H. Gastric Pathology Image Classification Using Stepwise Fine-Tuning for Deep Neural Networks. J Healthc Eng 2018; 2018: 8961781 [PMID: 30346677 DOI: 10.1155/2018/8961781]

León F, Gélvez M, Jaimes Z, Gelvez T, Arguello H. Supervised Classification of Histopathological Images Using Convolutional Neuronal Networks for Gastric Cancer Detection. 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA). IEEE, 2019: 1-5

Liang Q, Nan Y, Coppola G, Zou K, Sun W, Zhang D, Wang Y, Yu G. Weakly Supervised Biomedical Image Segmentation by Reiterative Learning. IEEE J Biomed Health Inform 2019; 23: 1205-1214 [PMID: 29994489 DOI: 10.1109/JBHI.2018.2850040]

Sun M, Zhang G, Hang D, Qi X, Zhou X, Chang Q. Accurate Gastric Cancer Segmentation in Digital Pathology Images Using Deformable Convolution and Multi-Scale Embedding Networks. IEEE Access 2019; 7: 75530-75541 [DOI: 10.1109/ACCESS.2019.2918800]

Wang S, Zhu Y, Yu L, Chen H, Lin H, Wan X, Fan X, Heng PA. RMDL: Recalibrated multi-instance deep learning for whole slide gastric image classification. Med Image Anal 2019; 58: 101549 [PMID: 31499320 DOI: 10.1016/j.media.2019.101549]

Awan R, Sirinukunwattana K, Epstein D, Jefferyes S, Qidwai U, Aftab Z, Mujeeb I, Snead D, Rajpoot N. Glandular Morphometrics for Objective Grading of Colorectal Adenocarcinoma Histology Images. Sci Rep 2017; 7: 16852 [PMID: 29203775 DOI: 10.1038/s41598-017-16516-w]

Kainz P, Pföffer M, Urschler M. Segmentation and classification of colon glands with deep convolutional neural networks and total variation regularization. PeerJ 2017; 5: e3874 [PMID: 29018612 DOI: 10.7717/peerj.3874]

Alom M, Yakopcic C, Taha T, Asari V. Microscopic Nuclei Classification, Segmentation and Detection with improved Deep Convolutional Neural Network (DCNN) Approaches. 2018 Preprint. Available from: arXiv:1811.03447

Weis CA, Kather JN, Melchers S, Al-Ahmdi H, Pollheimer MJ, Langner C, Gaiser T. Automatic evaluation of tumor budding in immunohistochemically stained colorectal carcinomas and correlation to clinical outcome. Diagn Pathol 2018; 13: 64 [PMID: 30153844 DOI: 10.1186/s13000-018-0739-3]

Ponzio F, Maci E, Ficarra E, Di Cataldo S. Colorectal Cancer Classification Using Deep Convolutional Networks - An Experimental Study. Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: Bioimaging, 2018: 58-66

Shapcott M, Hewitt KJ, Rajpoot N. Deep Learning With Sampling in Colon Cancer Histology. Front Bioeng Biotechnol 2019; 7: 52 [PMID: 30972333 DOI: 10.3389/fbioe.2019.00052]

Swiderska-Chadaj Z, Pinkcahrs H, van Rijthoven M, Balkenhol M, Melnikova M, Geessink O, Manso Q, Sherman M, Polonia A, Parry J, Abubakar M, Litjens G, van der Laak J, Cimopi F. Learning to detect lymphocytes in immunohistochemistry with deep learning. Med Image Anal 2019; 58: 101547 [PMID: 31476576 DOI: 10.1016/j.media.2019.101547]
