

REVIEW

Emerging role of deep learning-based artificial intelligence in tumor pathology

Yahui Jiang¹ | Meng Yang² | Shuhao Wang³ | Xiangchun Li² | Yan Sun¹ 

¹Department of Pathology, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, National Clinical Research Center for Cancer, Tianjin Cancer Institute and Hospital, Tianjin Medical University, Tianjin 300060, P. R. China

²Department Epidemiology and Biostatistics, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, National Clinical Research Center for Cancer, Tianjin Cancer Institute and Hospital, Tianjin Medical University, Tianjin 300060, P. R. China

³Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, P. R. China

Correspondence

Yan Sun, Department of Pathology; Key Laboratory of Cancer Prevention and Therapy; Tianjin's Clinical Research Center for Cancer; National Clinical Research Center for Cancer, Tianjin Cancer Institute and Hospital, Tianjin Medical University, Tianjin 300060, P. R. China.
Email: sunyan@tjmuch.com
Xiangchun Li, Department Epidemiology and Biostatistics; Key Laboratory of Cancer Prevention and Therapy; Tianjin's Clinical Research Center for Cancer; National Clinical Research Center for Cancer, Tianjin Cancer Institute and Hospital, Tianjin Medical University, Tianjin 300060, P. R. China.
Email: lixiangchun2014@foxmail.com

Funding information

National Nature Science Foundation of China, Grant/Award Numbers: 81871990, 81472263

Abstract

The development of digital pathology and progression of state-of-the-art algorithms for computer vision have led to increasing interest in the use of artificial intelligence (AI), especially deep learning (DL)-based AI, in tumor pathology. The DL-based algorithms have been developed to conduct all kinds of work involved in tumor pathology, including tumor diagnosis, subtyping, grading, staging, and prognostic prediction, as well as the identification of pathological features, biomarkers and genetic changes. The applications of AI in pathology not only contribute to improve diagnostic accuracy and objectivity but also reduce the workload of pathologists and subsequently enable them to spend additional time on high-level decision-making tasks. In addition, AI is useful for pathologists to meet the requirements of precision oncology. However, there are still some challenges relating to the implementation of AI, including the issues of algorithm validation and interpretability, computing systems, the unbelieving attitude of pathologists, clinicians and patients, as well as regulators and reimbursements. Herein, we present an overview on how AI-based approaches could be integrated into the workflow of pathologists and discuss the challenges and perspectives of the implementation of AI in tumor pathology.

Abbreviations: AI, artificial intelligence; AR, androgen receptor; ATC, anaplastic thyroid carcinoma; AUC, area under receiver operating characteristic curve; CLIA, Clinical Laboratory Improvement Amendment; CNN, convolutional neural network; CTC, circulating tumor cell; DL, deep learning; DSS, disease-specific survival; EGFR, epidermal growth factor receptor; ER, estrogen receptor; FAT1, FAT atypical cadherin 1; FCN, fully convolutional network; FDA, Food and Drug Administration; FTC, follicular thyroid carcinoma; GAN, generative adversarial network; HE, hematoxylin and eosin; HER2, human epidermal growth factor receptor 2; HGUC, high-grade urothelial carcinoma; HP, hyperplastic polyp; HPF, high power field; HR, hazard ratio; KRAS, Ki-ras2 Kirsten rat sarcoma viral oncogene homolog; ML, machine learning; MSI, microsatellite instability; MSS, microsatellite stability; MTC, medullary thyroid carcinoma; OS, overall survival; PD-L1, programmed death-ligand 1; PTC, papillary thyroid carcinoma; RNN, recurrent neural network; ROI, region of interest; SETBP1, SET binding protein 1; SPOP, speckle-type POZ protein; SSAP, sessile serrated adenoma/polyp; STK11, serine/threonine kinase 11; TCGA, The Cancer Genome Atlas; TSA, traditional serrated adenoma; WSI, whole-slide image.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Cancer Communications* published by John Wiley & Sons Australia, Ltd. on behalf of Sun Yat-sen University Cancer Center

KEYWORDS

artificial intelligence-assisted bioinformatic analysis, artificial intelligence, deep learning, pathology, tumor

1 | BACKGROUND

Artificial intelligence (AI) was termed by McCarthy et al. [1] in the 1950s, referring to the branch of computer science in which machine-based approaches were used to make predictions to mimic what human intelligence might do in the same situation. AI, currently a hot and controversial topic, has been introduced into many aspects of our everyday life, including medicine. Compared with other applications in the treatment of diseases, AI is more likely to enter the diagnostic disciplines based on image analysis such as pathology, ultrasound, radiology, and ophthalmic and skin disease diagnosis [2, 3]. Among these applications, the implementation of AI in pathology presents a special challenge due to the complexity and great responsibility of pathological diagnosis.

The progress of AI in pathology depended on the growth of digital pathology. In the 1960s, Prewitt et al. [4] scanned simple images from a microscopic field of a common blood smear and converted the optical data into a matrix of optical density values for computerized image analysis, which is regarded as the beginning of digital pathology. After the introduction of whole-slide scanners in 1999, AI in digital pathology using computational approaches grew rapidly to analyze the digitized whole-slide images (WSIs). The creation of large-scale digital-slide libraries, such as The Cancer Genome Atlas (TCGA), enabled researchers to freely access richly curated and annotated datasets of pathology images linked with clinical outcome and genomic information, in turn promoting the substantial investigations of AI for digital pathology and oncology [5, 6]. Our group identified an integrated molecular and morphologic signature associated with chemotherapy response in serous ovarian carcinoma using TCGA data in 2012 [7], which contains rudimentary model of machine learning (ML) on WSIs of TCGA.

AI models in pathology have developed from expert systems to traditional ML and then to deep learning (DL). Expert systems rely on rules defined by experts, and traditional ML needs to define features based on expert experience, while DL directly learns from raw data and leverages an output layer with multiple hidden layers (Figure 1) [8]. Compared with expert systems and hand-crafted ML approaches, DL approaches are easier to be conducted and have high accuracy. The increase in computational processing power and blooming of algorithms, such as convolutional neural network (CNN), fully convolutional network (FCN), recurrent neural network (RNN), and generative adversarial network (GAN),

have led to multiple investigations on the usage of DL-based AI in pathology. The application of AI in pathology helps to overcome the limitations of subjective visual assessment from pathologists and integrate multiple measurements for precision tumor treatment [9].

The Food and Drug Administration (FDA) in the USA approved the Philips IntelliSite whole-slide scanner (Philips Electronics, Amsterdam, Netherlands) in 2017, which is a milestone for true digital pathology laboratories. The application of AI in pathology is also promoted by startups, for example, PAIGE.AI [10], Proscia [11], DeepLens [12], PathAI [13] and Inspirata [14], who are using DL-based AI tools for detecting, diagnosing and predicting the prognosis of several types of cancers. Several institutions have decided to digitize their entire pathology workflow [15-17]. Therefore, the application of AI in pathology has been on the road. In this review, we summarized current studies on DL-based AI application in tumor pathology, analyzed the advantages and challenges of AI application, and discussed the optimized workflow and perspectives of human-machine cooperation in tumor pathology.

2 | APPLICATION OF DL-BASED AI IN TUMOR PATHOLOGY

The implementation of AI in tumor pathology refers to almost all kinds of tumors (Table 1) and is involved in tumor diagnosis, subtyping, grading, staging, and prognosis prediction as well as the identification of pathological features, biomarkers, and genetic changes.

2.1 | Tumor diagnosis

It is the most important for pathologists to distinguish tumors from other lesions and distinguish malignant from benign tumors as these can directly affect treatment decisions for different therapeutic strategies. Araújo et al. [18] developed a CNN-based AI algorithm to classify breast WSI images into two categories (carcinoma and non-carcinoma) with an accuracy of 83.3% and into four categories (normal tissue, benign lesions, carcinoma in situ, and invasive carcinoma) with an accuracy of 77.8%. Bejnordi et al. [19] developed a context-aware stacked CNN which was first trained to recognize the lower-level features and then used as input data to train the

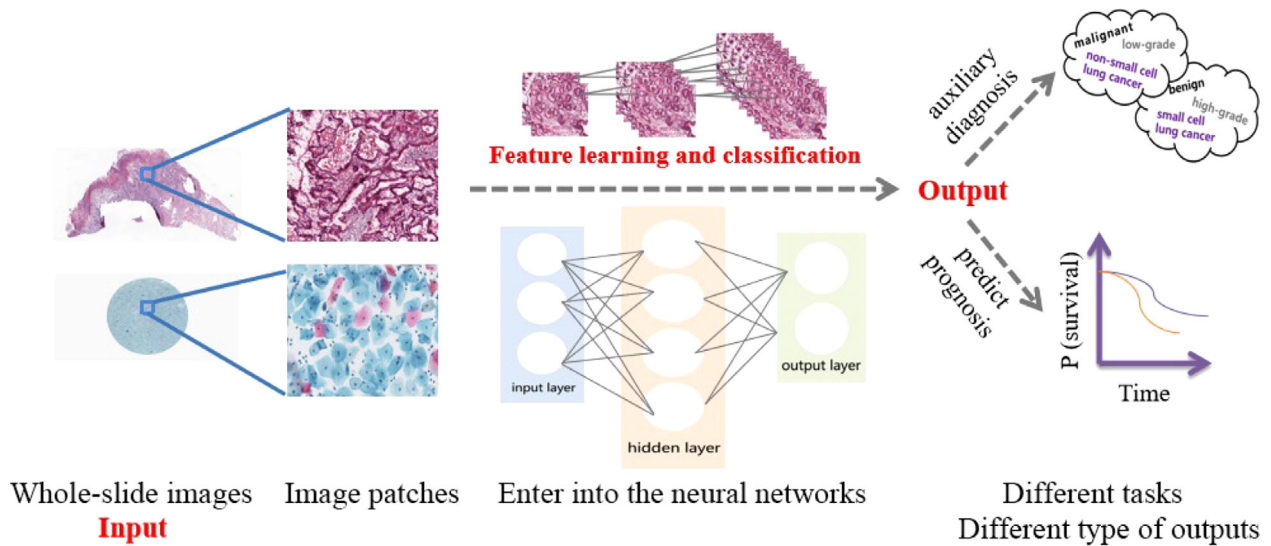


FIGURE 1 General workflow of deep learning-based artificial intelligence in pathology. First, tissue slides are transformed into whole-slide images (WSIs) through digital scanners. Next, various neural networks learn and extract features from the images patch-to-patch. Finally, features are selected and classified to construct different diagnosis- or prognosis- models

next stacked network to recognize the higher-level features. Using this CNN, they obtained an area under the receiver operating characteristic curve (AUC) of 0.962 to distinguish breast malignant tumors from non-malignant lesions and an accuracy of 81.3% to classify benign lesions, ductal carcinoma in situ, and invasive ductal carcinoma. Considering the effect of stroma on tumors, Ehteshami Bejnordi et al. [20] designed a CNN-based model to combine stroma features to distinguish breast invasive cancer from benign lesions. DL-based AI algorithms also gained comparable accuracy with professional pathologists in discriminating between benign and malignant colorectal tumors [21, 22], and melanoma from nevus [23].

Using the weakly supervised DL models that dramatically reduced labeling workload, Mercan et al. [24] classified breast lesions into non-proliferative, proliferative, atypical hyperplasia, carcinoma in situ, and invasive carcinoma using WSIs of breast biopsy specimens with a precision of 81%. Wang et al. [25] classified gastric lesions into normal, dysplasia, and cancer with an accuracy of 86.5% and Tomita et al. [26] classified esophagus lesions into Barrett esophagus, dysplasia, and cancer with an accuracy of 83%.

In addition to biopsy and resection specimens, pathologists should perform cytology diagnosis in routine work. For cervical cytological diagnosis, AI could classify cells as normal or abnormal in smear-based and liquid-based images, reaching an accuracy of 98.3% and 98.6%, respectively [27]. In liquid-based urine cytology, AI could separate high-grade urothelial carcinoma (HGUC) and suspicious HGUC from other lesions based on cell level-features [28] or WSI level-features [29]. AI also performed showed promising ability in the differential diagnosis for thyroid tumors on the basis of cytological

images [30]. In summary, DL-based AI has shown promise in histological and cytology diagnosis for many kinds of tumors.

2.2 | Tumor subtyping

The therapeutic strategies are different in various subtypes of cancers. A CNN-based model automatically differentiated lung adenocarcinoma, squamous cell carcinoma and normal lung tissue on images from biopsies, frozen tissues and formalin-fixed paraffin-embedded tissues with a high AUC (0.83-0.97) [31]. Considering different growth patterns of invasive lung adenocarcinoma cells are related with the clinical outcomes of patients, Gertych et al. [32] and Wei et al. [33] developed CNN algorithms to sort each image tile into individual growth pattern and generate a probability map for a WSI, facilitating pathologists to quantitatively report the major and more malignant components of lung adenocarcinoma, such as micropapillary and solid components. Similarly, DL-based AI were performed for multi-categorization of colorectal polyp [34], ovarian cancer [35], thyroid tumor [36], breast tumor [37], and cervical squamous cell carcinoma [38]. On the basis of cytological image, AI could recognize the histological subtypes of lung cancer with an accuracy of 60%-89% [39].

2.3 | Tumor grading

Pathologists often evaluate tumor grades according to the differentiation of tumor cells, glandular architecture, mitosis,

TABLE 1 Applications of AI in tumor pathology

Applications	The diagnostic points of AI in the studies	Performance	References
Breast cancer			
Diagnosis	Carcinoma/Non-carcinoma	Accuracy: 83.3%	[18]
	Normal/Benign/Carcinoma in situ/Invasive carcinoma	Accuracy: 77.8%	[18]
	Non-malignant/Malignant	AUC: 0.962	[19]
	Benign/Ductal carcinoma in situ/Invasive ductal carcinoma	Accuracy: 81.3%	[19]
	Invasive cancer/Benign lesion based on both tumor cells and stroma	AUC: 0.962	[20]
	Non-proliferative/Proliferative/Atypical hyperplasia/Carcinoma in situ/Invasive carcinoma	Precision: 81%	[24]
Tumor subtyping	Adenosis/Fibroadenoma/Tubular adenoma/Phyllodes tumor/Ductal carcinoma/Lobular carcinoma/Mucinous carcinoma/Papillary carcinoma	Accuracy: 90.66%-93.81%	[37]
Tumor grading	Low/Intermediate/High grade	Accuracy: 69%	[42]
Tumor staging	Heatmap for the region of invasive cancer	Dice coefficient: 75.86% and 76.00%	[44], [45]
	With/Without lymph node metastasis	AUC: 0.994 and 0.996	[46], [47]
	Metastatic regions in lymph node	Sensitivity: 0.807-0.910	[46]-[48]
Evaluation of pathological features	Mitotic count	F-score: 0.611	[57]
	Ki-67 index	F-score: 0.91	[64]
	Proliferation score	Kappa: 0.613	[58]
	Immune cell-rich/Immune cell-poor regions	AUC: 0.99	[62]
Evaluation of biomarkers	HER2 status: Negative (0 and 1+)/ Equivocal (2+)/Positive (3+)	Accuracy: 83% and 87%	[66], [67]
Lung cancer			
Tumor subtyping	Small cell cancer/Non-small cell cancer (Cytology)	Accuracy: 85.6%	[39]
	Adenocarcinomas/Squamous cell carcinomas/Small cell carcinomas (Cytology)	Accuracy: 60%-89%	[39]
	Adenocarcinoma/Squamous cell carcinoma	AUC: 0.83-0.97	[31]
	Solid/Micropapillary/Acinar/Cribriform subtype	F-score: 0.60-0.96	[32]
	Lepidic/Solid/Micropapillary/Acinar/ Cribriform subtype	AUC: 0.961-0.997	[33]
Evaluation of pathological features	Immune cell count	Accuracy: 98.6%	[63]
Evaluation of biomarkers	PD-L1 status: Negative/Positive	AUC: 0.80	[69]
Evaluation of genetic changes	Predict the most commonly mutated genes	AUC: 0.733-0.856	[31]
Prognosis prediction	Low/High risk	Hazard ratio: 2.25	[76]
Colorectal cancer			
Diagnosis	Benign/Malignant	Accuracy: 95%-98%	[21]
	Normal/Cancer	Accuracy: 97%	[22]
Tumor subtyping	HP/SSAP/TSA/Tubular adenoma/Tubulovillous and villous adenoma	Accuracy: 93%	[34]
Tumor grading	Normal /Low-grade cancer/High-grade cancer	Accuracy: 91%	[22]
Evaluation of pathological features	Number of tumor budding	Correlation R ² : 0.86	[59]
Evaluation of genetic changes	MSI/MSS on HE stained images	AUC: 0.77-0.84	[73]
Prognosis prediction	DSS, OS: Low/High risk	Hazard ratio: 1.65-2.30	[73]-[75]

(Continues)

TABLE 1 (Continued)

Applications	The diagnostic points of AI in the studies	Performance	References
Gastric cancer			
Diagnosis	Normal/Dysplasia/Cancer	Accuracy: 86.5%	[25]
Evaluation of genetic changes	HER2 status: Negative (0 and 1+)/Positive (2+ and 3+)	Accuracy: 69.9%	[68]
Prostate cancer			
Tumor grading	Gleason scoring	Accuracy: 75%	[40]
Evaluation of genetic changes	Distinguish SPOP mutant from non-mutant on HE stained images	AUC: 0.74-0.86	[70]
Cervical cancer			
Diagnosis	Normal/Abnormal (Cytology: smear-based and liquid-based)	Accuracy: 98.3% and 98.6%	[27]
Tumor subtyping	Keratinizing/Non-keratinizing/Basaloid squamous cell carcinoma	Accuracy: 93.33%	[38]
Glioma			
Tumor grading	Grade IV/Grade II/III (Glioblastoma multiform)	Accuracy: 96%	[41]
	Grade II/Grade III	Accuracy: 71%	[41]
Prognosis prediction	OS: Low/Intermediate/High risk	Concordance index: 0.754	[77]
Thyroid cancer			
Diagnosis	PTC/Benign nodules (Cytology)	Accuracy: 97.66%	[30]
Tumor subtyping	Normal tissue/Adenoma/Nodular goiter/PTC/FTC/MTC/ATC	Accuracy: 88.33%-100%	[36]
Others			
Diagnosis for esophagus lesion	Barrett esophagus/Dysplasia/Cancer	Accuracy: 83%	[26]
Diagnosis for melanocytic lesion	Nevus/aggressive malignant melanoma	AUC: 0.998	[23]
Diagnosis for urinary tract lesion	HGUC and Suspicious HGUC/other lesions (Cytology)	AUC: 0.88 and 0.92	[28], [29]
Subtyping for ovary cancer	Serous/Mucinous/Endometrioid/Clear cell carcinoma	Accuracy: 78.2%	[35]
Staging for osteosarcoma	Regions of tumor component/Necrosis/non-tumor component	Accuracy: 92.4%	[43]
Biomarker for pancreatic neuroendocrine neoplasm	Segment tumor regions to calculate Ki-67 index	Accuracy: 96.2%	[65]
Multiple tasks in breast cancer	ER status:Negative/Positive	Accuracy: 84%	[78]
	RNA-based molecular subtypes: Basal-like/Non-basal-like	Accuracy: 77%	
	Histological subtypes: Ductal/Lobular	Accuracy: 94%	
	Grade: Low-intermediate grade/High grade	Accuracy: 82%	
	Recurrence-risk: High/Low-medium risk	Accuracy: 76%	
Multiple cancers			
Lung/Breast/Bladder cancer	Different cancer types: Lung/Breast/Bladder cancer	Accuracy: 100%	[79]
	Adenocarcinoma/Squamous cell carcinoma of lung cancer	Accuracy: 92%	
	Identifying biomarkers based on expression pattern	Accuracy: 95%	
	Scoring biomarkers	Accuracy: 69%	
Prostate/Lung cancer	CTC status: Positive/Negative	Accuracy: 84%	[56]

Abbreviations: AUC, area under receiver operating characteristic curve; HE, hematoxylin and eosin; HP, hyperplastic polyp; SSAP, sessile serrated adenoma/polyp; TSA, traditional serrated adenoma; MSI/MSS, microsatellite instability/stability; DSS, disease-specific survival; OS, overall survival; PTC/FTC/MTC/ATC, papillary/follicular/medullary/anaplastic thyroid carcinoma; HGUC, high-grade urothelial carcinoma; ER, estrogen receptor; CTC: circulating tumor cell.

necrosis and more, which are subjective and may possibly skew treatment decision and clinical monitoring. Arvaniti et al. [40] proposed an AI model for the Gleason scoring of prostate cancer; suitable for any Gleason scores. The agreement between the AI model and each of the two pathologists (0.75 and 0.71, respectively) was comparable to that between the two pathologists themselves (0.71). Moreover, the AI model was superior to the pathologists in distinguishing low-risk from intermediate-risk cases based on the Gleason scores calculated by the model. Ertosun et al. [41] developed two separate CNNs to define the grade of gliomas. One of them classified the cases as either glioblastoma multiform (grade IV) or lower grade glioma (grade II and III), with an accuracy of 96%, and the other discriminated grade II glioma from grade III with an accuracy of 71%. For the breast biopsy images, a CNN-based model distinguished low-, intermediate-, and high-grade breast cancers with an accuracy of 69% [42]. DL-based models were also set up to discriminate normal tissue, low-grade and high-grade colorectal adenocarcinoma with an accuracy of 91% [22]. Overall, AI will be helpful to provide objective and reproducible results for tumor grading.

2.4 | Tumor staging

For resection samples, pathologists should provide as much information possible for TNM staging that usually determine therapeutic decisions. In osteosarcoma, a CNN-based model distinguished three types of region of interests (ROIs), namely the tumor, necrotic and non-tumor component (e.g., bone, cartilage), on a patch level (64,000 patches from 82 WSIs) with an accuracy of 92.4% [43]. In addition, the proportion of necrosis, a variable factor for prognosis, could be calculated. Several DL-based networks have also been developed to recognize tumor regions in breast cancer [24, 44, 45].

The assessment of lymph node metastasis is necessary for tumor staging but it is time-consuming and error-prone for pathologists. In the “Cancer Metastases in Lymph Nodes Challenge” (CAMELYON16), a competition between AI and pathologists to assess sentinel lymph nodes of breast cancer, two AI algorithms exceeded pathologists with the best AUC of 0.994 in slide-level detection (only identifying metastasis or not), and two algorithms outperformed pathologists with the best one achieving the mean sensitivity across six false-positive rates of 0.807 in lesion-level detection (recognizing all metastases except for isolated tumor cells) [46]. In the same dataset, Lymph Node Assistant (LYNA), a more optimized algorithm, gained a higher AUC (0.996) in slide-level detection and a higher sensitivity with one false positive per slide (91%) in lesion-level detection through filtering out artifacts. Of note, LYNA corrected two slides misdiagnosed as “normal” by the organizers [47]. Another

study demonstrated that LYNA improved the sensitivity of detecting all micrometastases in lymph nodes from 83% to 91% ($P = 0.02$) with a significantly shorter review time as compared to pathologists alone [48].

In the past decade, numerous studies have shown that circulating tumor cells (CTCs) may be used as a marker to predict disease progression and survival in metastatic [49-51] and possibly even in early-stage cancer patients [52]. High CTC numbers were correlated with aggressive disease, increased metastasis, and decreased time to relapse [53]. Considering the simplicity and minimally invasion of blood collection, CTCs are expected to be used as a marker for monitoring tumor progression and guiding therapeutic management as well as indicating therapy effectiveness [54]. However, technical obstacles, such as small quantity and lack of standardized detection assays and valid marker limit clinical usage [55]. Zeune et al. [56] reported that the manual counting for CTCs of prostate cancer and non-small cell lung cancer on fluorescent images varied among human reviewers and counting platforms, whereas DL-based CTC recognition was relatively stable with higher accuracy than the average level of human reviewers.

Considering the current contribution of AI to identifying tumor regions, detecting lymph node metastasis and CTCs, and the ability of AI to analyze large amounts of data, AI approaches will potentially help pathologists and oncologists perform tumor staging.

2.5 | Evaluation of pathological features

Mitosis represents the proliferation ability of tumor cells. However, mitosis count is time-consuming. The Assessment of Mitosis Detection Algorithms 2013 (AMIDA13) generated a good algorithm recognizing mitoses with the F1-score of 0.611 on 1000 images at high power fields (HPFs) from breast cancers, which was comparable to the performance of inter-observers [57]. The proliferation scores of breast cancers were reported based on the WSIs-level AI detection in the Tumor Proliferation Assessment Challenge 2016 (TUPAC16) [58].

Tumor budding is one of the aggressive behavior of tumors. In colorectal carcinomas, Weis et al. [59] utilized CNN to obtain the absolute number of tumor budding based on cytokeratin-stained WSIs and demonstrated the correlation between the number of budding hotspots and the status of lymph node. It has been reported that the kind and amount of tumor-infiltrating immune cells are related with the sensitivity to immunotherapy and prognostic stratification for tumor patients [60, 61]. A DL method using CD45-annotated digital images could quantify immune cells and distinguish immune cell-rich or -poor regions in breast cancer [62]. The AUC was 0.99 without the limitation of histological types and grades. The AI algorithm had an excellent performance in counting

T-cells and B-cells on both cell patch- and WSI-level images of CD3-, CD8-, and CD20-stained sections from lung tissue, undisturbed by anthracitic pigment [63]. Compared with the hand-crafted AI focused on hand-crafted pathological features, DL-based AI could recognize both hand-crafted features and domain-agnostic features which could be applied across disease and tissue types.

2.6 | Evaluation of biomarkers

Saha et al. [64] used a DL-based approach to automatically detect high-proliferation regions and calculate the Ki-67 index of breast cancer. Niazi et al. [65] developed AI algorithms to segment tumor areas from interstitium and normal pancreatic tissue on the unevenly Ki-67-immunoreactive WSIs and accordingly calculated the Ki-67 index more precisely in pancreatic neuroendocrine neoplasms.

Furthermore, some biomarkers are used for the selection of suitable patients for the related therapies. Trastuzumab could be used on the basis of the status of human epidermal growth factor receptor 2 (HER2) in breast and gastric cancer. In breast cancer, a CNN-based algorithm achieved an overall agreement of 83% with a pathologist which was similar to the agreement among pathologists in predicting HER2 negative (0 and 1+), equivocal (2+), and positive (3+) on WSIs [66]. The prediction results of AI improved when investigators segmented cell membranes as the true expression location of HER2 [67]. Similarly, AI algorithm was also developed to assess HER2-positive areas (2+ and 3+), HER2-negative areas (0 and 1+) and tumor-free areas in gastric cancer with an accuracy of 69.9% [68]. To select potential patients sensitive to pembrolizumab, an AI algorithm identified the expression of programmed death-ligand 1 (PD-L1; negative or positive) on hematoxylin and eosin (HE) stained images of non-small cell lung cancer with AUC of 0.80 which was comparable to the assessment of pathologists based on PD-L1 immunohistochemistry images [69]. Based on the immunohistochemical staining or fluorescent staining WSIs, even on HE staining WSIs, DL-based AI algorithms performed the evaluation of biomarkers, which were involved in diagnosis, prognosis and drug response prediction.

2.7 | Evaluation of genetic changes

The morphological changes shown in WSI are manifestations of underlying genetic changes. To predict whether or not the speckle-type POZ protein (SPOP) gene is mutated in prostate cancer, Schaumberg et al. [70] trained multiple ensembles of residual networks using a cohort of 177 prostate cancer patients from TCGA where 20 had mutant SPOP and validated their findings in an independent cohort from MSK-

IMPACT of 152 patients where 19 had mutant SPOP. Despite the training set was from frozen sections and the validation set was from formalin-fixed paraffin-embedded sections, mutants and non-mutants of SPOP were accurately distinguished (AUC = 0.86). Considering that non-mutant SPOP ubiquitinylates androgen receptor (AR) to mark AR for degradation [71], identification of SPOP mutation state could lead directly to precision medicine. Moreover, SPOP mutation is mutually exclusive with TMPRSS2-ERG gene fusion [72], the prediction of SPOP mutation status provided indirect information of the TMPRSS2-ERG state and potentially others. Coudray et al. [31] trained the DL network to predict the ten most commonly mutated genes in lung adenocarcinoma from TCGA pathology images and found that six of them (serine/threonine kinase 11 [STK11], epidermal growth factor receptor [EGFR], FAT atypical cadherin 1 [FAT1], SET binding protein 1 [SETBP1], Ki-ras2 Kirsten rat sarcoma viral oncogene homolog [KRAS] and TP53) could be predicted with AUCs from 0.733 to 0.856. Additionally, an AI model was developed to identify microsatellite instability (MSI) or microsatellite stability (MSS) based on HE stained images of gastrointestinal cancer without performing microsatellite instability assays, which also showed robustness in snap-frozen samples, Asian populations, and even endometrial cancer with high AUC (0.77-0.84) [73]. Through these AI networks, patients with specific genetic changes were identified based on intrinsic genetic-histologic relationships, which benefited the precision treatment.

2.8 | Prognosis prediction

Bychkov et al. [74] used HE-stained tissue microarray images of colorectal cancer to invent a DL-based method to divide the patients into low- or high-risk groups. It is noticeable that the annotation in the training set was patients' outcome rather than the labeled WSIs. The AI exceeded histological grade (hazard ratio [HR], 2.30 vs. 1.65) and was demonstrated to be an independent prognostic factor with multivariate Cox proportional hazard model analysis. Kather et al. [75] showed that integrated interstitium characteristics (including adipose, debris, lymphocytes, muscle, and desmoplastic stroma) extracted by CNN could independently forecast the overall survival (HR, 2.29) and relapse-free survival (HR, 1.92) of patients with colorectal cancer in multi-centers dataset, regardless of clinical stage. It was demonstrated that DL approaches could predict the prognostic risk by learning histologic features in lung adenocarcinoma [76] and glioma [77]. Kather et al. [73] reported that their algorithm-based prediction of MSI was also fit for predicting overall survival in gastrointestinal cancer. The above investigations indicated that AI algorithms could be used to predict clinical outcome besides pathological diagnosis for cancer patients.

2.9 | Algorithm for multiple tasks and multiple tumors

Most of the above AI algorithms were designed to conduct a specific job in a certain tumor type. As a result, the pathologists need to run many AI algorithms to complete a pathological report, in which the tumor were to be diagnosed, subtyped, graded by different algorithms and the high-risk features were to be assessed by another individual algorithms.

Couture et al. [78] developed a DL-based algorithm to complete multiple tasks in HE-stained tissue microarray images of breast cancer, including identifying histological subtypes (ductal or lobular) with an accuracy of 94%, classifying histological grade (low-intermediate grade or high grade) with an accuracy of 82%, and assessing the status of estrogen receptor (ER; positive or negative) with an accuracy of 84%, discriminating molecular subtypes (basal-like or non-basal-like types) with an accuracy of 77%, as well as stratifying recurrence-risk (high or low-to-medium) with an accuracy of 76%. Another system equipped with six state-of-the-art DL architectures has the potential to determine tumor, subtypes, as well as biomarkers' score in lung, breast, and bladder cancer [79].

3 | CHALLENGES AND PERSPECTIVES

As displayed in the above explorations, DL-based AI has appealing perspective for elevating the efficiency of pathological diagnosis and prognosis. However, there are still some obstacles and challenges in the implementation of AI in tumor pathology.

3.1 | Validation

Current AI algorithms are mainly established on small-scale data and images from single-center. The data from single center were still deviation, although researchers have developed methods to augment the dataset, including but not limited to random rotation and flipping, color jittering, and Gaussian blur [32, 35, 38, 39]. Variations exist in slide preparation, scanner models and digitization among different centers. Zech et al. [80] reported that a CNN for pneumonia detection performed significantly poorer when it was trained using data from one institution and validated independently using data from two other institutions than when it was trained using data from all three institutions ($P < 0.001$). AI algorithms need to be sufficiently validated using multi-institutional data before clinical adoption.

Fortunately, we can find some well-curated, accurate WSI reference datasets across cancer subtypes with annotated cancerous regions, which contribute to standardizing the

evaluation of AI algorithms. Moreover, we can use some large-scale digital-slide libraries, including TCGA, like training or validation datasets. Building comprehensive quality control and standardization tools, data share and validation with multi-institutional data can increase the generalizability and robustness of the AI algorithms. In addition, AI algorithms need to be continually validated and corrected by the diagnosis of expert pathologists.

3.2 | Interpretability

Based on the 'black-box' methods, deep learning-based AI has been questioned about the lack of interpretability which is an obstacle towards the clinical adoption of AI [81-83]. Several studies used post hoc methods or supervised ML models to explain the output of deep learning-based algorithms after DL algorithms made its prediction [47, 48]. However, post hoc analyses of DL methods have been criticized because additional models should not be required to explain how a DL model works [82]. Recently, some researchers integrated DL algorithms and hand-crafted ML approaches to raise the biological interpretability of the model. Wang et al. [84] used a DL approach to segment nuclei in digital HE images of early-stage non-small cell lung cancer to predict tumor recurrence before applying a hand-crafted method involving the interrogation of nuclear shape and texture. More strategies are needed to increase the interpretability of AI algorithms and gain confidence of doctors and patients.

3.3 | Computing system

Histopathological image has large file sizes, being about 1,000 times of an X-ray and 100 times of a CT image. As a result, high specs hardware is required for both storage and processing. It is necessary to design a powerful AI model and build efficient and scalable storage and computing system to analyze the images. With the usage of cloud platforms, there are challenges in the massive bandwidth required to transmit gigapixel-sized WSI images into data clouds as well as managing permanent and uninterrupted communication channels between end-users and the cloud. These challenges would be addressed in the foreseeable future brought by improvements in information technology such as universal adoption of 5G.

3.4 | Attitude of pathologists

In addition to the lack of interpretability of AI, some pathologists are afraid of the change in workflow. When AI is used, the pathologists won't observe the histopathological morphology with a microscope, but on an accelerated parallel

processing (APP). How would the pathologists describe the diagnosis evidence from AI in the diagnosis report? How much responsibility should the pathologists have when they assign a diagnosis report with the help of AI? These issues need to be resolved before the real human-machine cooperation are implemented in clinical practice. In addition, with the development of AI, more and more algorithms/platforms are being developed. Another important issue for pathologists is how to choose an adaptable one and standardize the output from different algorithms/platforms.

3.5 | Attitude of clinicians and patients

All of the diagnostic reports serve clinicians to develop suitable intervention programs for patients. As a result, the output of AI should be understood and trusted by the clinicians. In addition, considering the cost of patients, the clinician needs to decide the minimum diagnostic assays as well as prognostic and predictive assays. The AI-based diagnostic and prognostic/predictive assays should have a high accuracy and be convenient for routine clinical use. Similarly, AI-based tests should be trusted by patients. It is better for AI-based tests to be set up as Clinical Laboratory Improvement Amendments (CLIA)-based tests and enter into clinical guidelines.

3.6 | Regulators and reimbursements

Besides the accreditation of doctors and patients, the clinical adoption of AI digital pathology needs approval by regulatory agencies. The key principle guiding the approval process in most countries is the requirement of an explanation of how the software works [85–87]. The lack of interpretability limits the approval for DL-based AI approaches [82]. Recently, the FDA has started granting approval to DL-based approaches for clinical use in the USA. Philips received an approval for a digital pathology whole-slide scanning solution (IntelliSite) in 2017 [88], and subsequently, the digital pathology solution PAIGE.AI [89] was granted Breakthrough Device designation by the FDA in 2019 [90]. Despite of the above achievements, AI-based devices tend to be assigned to Class II or III in the FDA three-class system for the approval of medical devices, in which Class I devices deem to have the lowest risk and Class III devices to have the highest risk. In the European Union, no AI solutions with prognostic/predictive intent have a Conformité Européenne marking, but the digital pathology solutions developed by Philips, Sectra and OptraSCAN have secured clearance to carry such a designation. Although the FDA apparently intends to regulate CLIA-based tests more stringently, it seems to be a better way for AI-based diagnostic assays to follow the model established by CLIA-based genomic tests to get approval for clinical usage.

Reimbursement of the costs of AI-based diagnostic and prognostic/predictive assays is one of the major issues that affect the application of these assays in clinic [91]. In the USA, insurance companies standardize expenses on the basis of the current procedural terminology codes maintained by the American Medical Association and reported by medical professionals [92–95]. At present, there are no dedicated procedure codes for the use of AI in digital pathology with diagnostic or prognostic intent. AI-based tools probably need to be approved by FDA before they get the new procedure codes and are reimbursable. In China, AI-based tools have not been covered by social medical insurance or insurance companies.

3.7 | Summary

Future pathological diagnosis probably needs to incorporate multimodal measurements, such as proteomics, genomics, and measurements from multiplexed marker-staining platforms, in order to supply a comprehensive patient-specific portrait for tumor precision treatment [96]. Despite the above challenges and obstacles, the potential of DL-based AI approaches for digital pathology is promising since AI has strong feature representation learning capability enabled by improvements in algorithm, accumulation of big data and increased computing power. People will have more confidence in AI algorithms after they are validated using multi-center data and have increased interpretability. The collaboration between pathologists and AI will promote tumor precision treatment.

DECLARATIONS

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FUNDING

This work was partially supported by the National Nature Science Foundation of China (81871990 and 81472263).

AUTHORS' CONTRIBUTIONS

YS and XL conceived this study. YJ, MY, XL and YS drafted the manuscript. SW revised the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

Not applicable.

ORCID

Yan Sun  <https://orcid.org/0000-0001-8916-9598>

REFERENCES

- McCarthy J, Minsky ML, Shannon CE. A proposal for the Dartmouth summer research project on artificial intelligence - August 31, 1955. *Ai Mag*. 2006;27(4):12–4.
- Du XL, Li WB, Hu BJ. Application of artificial intelligence in ophthalmology. *Int J Ophthalmol*. 2018;11(9):1555–61. <https://doi.org/10.18240/ijo.2018.09.21>.
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–8. <https://doi.org/10.1038/nature21056>.
- Prewitt JM, Mendelsohn ML. The analysis of cell images. *Ann N Y Acad Sci*. 1966;128(3):1035–53. <https://doi.org/10.1111/j.1749-6632.1965.tb11715.x>.
- Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*. 2015;19(1A):A68–77. <https://doi.org/10.5114/wo.2014.47136>.
- Gutman DA, Cobb J, Somanna D, Park Y, Wang F, Kurc T, et al. Cancer Digital Slide Archive: an informatics resource to support integrated in silico analysis of TCGA pathology data. *J Am Med Inform Assoc*. 2013;20(6):1091–98. <https://doi.org/10.1136/amiainl-2012-001469>.
- Liu Y, Sun Y, Broadus R, Liu J, Sood AK, Shmulevich I, et al. Integrated analysis of gene expression and tumor nuclear image profiles associated with chemotherapy response in serous ovarian carcinoma. *PLoS One*. 2012;7(5):e36383. <https://doi.org/10.1371/journal.pone.0036383>.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444. <https://doi.org/10.1038/nature14539>.
- Jain RK, Mehta R, Dimitrov R, Larsson LG, Musto PM, Hodges KB, et al. Atypical ductal hyperplasia: interobserver and intraobserver variability. *Mod Pathol*. 2011;24(7):917–923. <https://doi.org/10.1038/modpathol.2011.66>.
- Fuchs TJ, Wild PJ, Moch H, Buhmann JM. Computational pathology analysis of tissue microarrays predicts survival of renal clear cell carcinoma patients. *Med Image Comput Assist Interv*. 2008;11(Pt 2):1–8. https://doi.org/10.1007/978-3-540-85990-1_1.
- Proscia. Proscia digital pathology. Proscia.com. 2019. Available from: <https://proscia.com>.
- Deep Lens. Digital pathology cloud platform. Deeplens.ai. 2019. Available from: <https://www.deeplens.ai>.
- PathAI. PathAI. Pathai.com. 2019. Available from: <https://www.pathai.com/>.
- Aifora. WebMicroscope. Big pictures. Deep Diagnosis. Aiforia.com. 2019. Available from: <https://www.aiforia.com/>.
- Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, et al. 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(12):1710–1722. <https://doi.org/10.5858/arpa.2013-0093-CP>.
- Fraggetta F, Garozzo S, Zannoni GF, Pantanowitz L, Rossi ED. Routine Digital Pathology Workflow: The Catania Experience. *J Pathol Inform*. 2017;8:51. https://doi.org/10.4103/jpi.jpi_58_17.
- Cheng CL, Azhar R, Sng SH, Chua YQ, Hwang JS, Chin JP, et al. Enabling digital pathology in the diagnostic setting: navigating through the implementation journey in an academic medical centre. *J Clin Pathol*. 2016;69(9):784–792. <https://doi.org/10.1136/jclinpath-2015-203600>.
- Araujo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, et al. Classification of breast cancer histology images using Convolutional Neural Networks. *PLoS One*. 2017;12(6):e0177544. <https://doi.org/10.1371/journal.pone.0177544>.
- Bejnordi BE, Zuidhof G, Balkenhol M, Hermesen M, Bult P, van Ginneken B, et al. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. *J Med Imaging (Bellingham)*. 2017;4(4):044504. <https://doi.org/10.1117/1.JMI.4.4.044504>.
- Ehteshami Bejnordi B, Mullooly M, Pfeiffer RM, Fan S, Vacek PM, Weaver DL, et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Mod Pathol*. 2018;31(10):1502–1512. <https://doi.org/10.1038/s41379-018-0073-z>.
- Kainz P, Pfeiffer M, Urschler M. Segmentation and classification of colon glands with deep convolutional neural networks and total variation regularization. *PeerJ*. 2017;5:e3874. <https://doi.org/10.7717/peerj.3874>.
- Awan R, Sirinukunwattana K, Epstein D, Jefferyes S, Qidwai U, Aftab Z, et al. Glandular Morphometrics for Objective Grading of Colorectal Adenocarcinoma Histology Images. *Sci Rep*. 2017;7(1):16852. <https://doi.org/10.1038/s41598-017-16516-w>.
- Wang L, Ding L, Liu Z, Sun L, Chen L, Jia R, et al. Automated identification of malignancy in whole-slide pathological images: identification of eyelid malignant melanoma in gigapixel pathological slides using deep learning. *Br J Ophthalmol*. 2019. <https://doi.org/10.1136/bjophthalmol-2018-313706>.
- Mercan C, Aksoy S, Mercan E, Shapiro LG, Weaver DL, Elmore JG. Multi-Instance Multi-Label Learning for Multi-Class Classification of Whole Slide Breast Histopathology Images. *IEEE Trans Med Imaging*. 2018;37(1):316–325. <https://doi.org/10.1109/TMI.2017.2758580>.
- Wang S, Zhu Y, Yu L, Chen H, Lin H, Wan X, et al. RMDL: Recalibrated multi-instance deep learning for whole slide gastric image classification. *Med Image Anal*. 2019;58:101549. <https://doi.org/10.1016/j.media.2019.101549>.
- Tomita N, Abdollahi B, Wei J, Ren B, Suriawinata A, Hassanpour S. Attention-Based Deep Neural Networks for Detection of Cancerous and Precancerous Esophagus Tissue on Histopathological Slides. *JAMA Netw Open*. 2019;2(11):e1914645. <https://doi.org/10.1001/jamanetworkopen.2019.14645>.
- Zhang L, Le L, Nogues I, Summers RM, Liu S, Yao J. DeepPap: Deep Convolutional Networks for Cervical Cell Classification. *IEEE J Biomed Health Inform*. 2017;21(6):1633–1643. <https://doi.org/10.1109/JBHI.2017.2705583>.
- Vaickus LJ, Suriawinata AA, Wei JW, Liu X. Automating the Paris System for urine cytopathology-A hybrid deep-learning and morphometric approach. *Cancer Cytopathol*. 2019;127(2):98–115. <https://doi.org/10.1002/cncy.22099>.
- Sanghvi AB, Allen EZ, Callenberg KM, Pantanowitz L. Performance of an artificial intelligence algorithm for reporting

- urine cytopathology. *Cancer Cytopathol.* 2019;127(10):658–666. <https://doi.org/10.1002/cncy.22176>.
30. Guan Q, Wang Y, Ping B, Li D, Du J, Qin Y, et al. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. *J Cancer.* 2019;10(20):4876–4882. <https://doi.org/10.7150/jca.28769>.
 31. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med.* 2018;24(10):1559–1567. <https://doi.org/10.1038/s41591-018-0177-5>.
 32. Gertych A, Swiderska-Chadaj Z, Ma Z, Ing N, Markiewicz T, Cierniak S, et al. Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. *Sci Rep.* 2019;9(1):1483. <https://doi.org/10.1038/s41598-018-37638-9>.
 33. Wei JW, Tafe LJ, Linnik YA, Vaickus LJ, Tomita N, Hassanpour S. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Sci Rep.* 2019;9(1):3358. <https://doi.org/10.1038/s41598-019-40041-7>.
 34. Korbar B, Olofson AM, Miraffior AP, Nicka CM, Suriawinata MA, Torresani L, et al. Deep Learning for Classification of Colorectal Polyps on Whole-slide Images. *J Pathol Inform.* 2017;8:30. https://doi.org/10.4103/jpi.jpi_34_17.
 35. Wu M, Yan C, Liu H, Liu Q. Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks. *Biosci Rep.* 2018;38(3). <https://doi.org/10.1042/BSR20180289>.
 36. Wang Y, Guan Q, Lao I, Wang L, Wu Y, Li D et al. Using deep convolutional neural networks for multi-classification of thyroid tumor by histopathology: a large-scale pilot study. *Ann Transl Med.* 2019;7(18):468. <https://doi.org/10.21037/atm.2019.08.54>.
 37. Jiang Y, Chen L, Zhang H, Xiao X. Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module. *PLoS One.* 2019;14(3):e0214587. <https://doi.org/10.1371/journal.pone.0214587>.
 38. Wu M, Yan C, Liu H, Liu Q, Yin Y. Automatic classification of cervical cancer from cytological images by using convolutional neural network. *Biosci Rep.* 2018;38(6). <https://doi.org/10.1042/BSR20181769>.
 39. Teramoto A, Tsukamoto T, Kiriya Y, Fujita H. Automated Classification of Lung Cancer Types from Cytological Images Using Deep Convolutional Neural Networks. *Biomed Res Int.* 2017;2017:4067832. <https://doi.org/10.1155/2017/4067832>.
 40. Arvaniti E, Fricker KS, Moret M, Rupp N, Hermanns T, Fankhauser C, et al. Automated Gleason grading of prostate cancer tissue microarrays via deep learning. *Sci Rep.* 2018;8(1):12054. <https://doi.org/10.1038/s41598-018-30535-1>.
 41. Ertosun MG, Rubin DL. Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of convolutional neural networks. *AMIA Annu Symp Proc.* 2015;2015:1899–1908.
 42. Wan T, Cao JJ, Chen JH, Qin ZC. Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features. *Neurocomputing.* 2017;229:34–44. <https://doi.org/10.1016/j.neucom.2016.05.084>.
 43. Mishra R, Daescu O, Leavey P, Rakheja D, Sengupta A. Convolutional Neural Network for Histopathological Analysis of Osteosarcoma. *J Comput Biol.* 2018;25(3):313–325. <https://doi.org/10.1089/cmb.2017.0153>.
 44. Cruz-Roa A, Gilmore H, Basavanahally A, Feldman M, Ganesan S, Shih NNC, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. *Sci Rep.* 2017;7:46450. <https://doi.org/10.1038/srep46450>.
 45. Cruz-Roa A, Gilmore H, Basavanahally A, Feldman M, Ganesan S, Shih N, et al. High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: Application to invasive breast cancer detection. *PLoS One.* 2018;13(5):e0196828. <https://doi.org/10.1371/journal.pone.0196828>.
 46. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karsssemeijer N, Litjens G, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA.* 2017;318(22):2199–2210. <https://doi.org/10.1001/jama.2017.14585>.
 47. Liu Y, Kohlberger T, Norouzi M, Dahl GE, Smith JL, Mohtashamian A, et al. Artificial Intelligence-Based Breast Cancer Nodal Metastasis Detection: Insights Into the Black Box for Pathologists. *Arch Pathol Lab Med.* 2019;143(7):859–868. <https://doi.org/10.5858/arpa.2018-0147-OA>.
 48. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, et al. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. *Am J Surg Pathol.* 2018;42(12):1636–1646. <https://doi.org/10.1097/PAS.0000000000001151>.
 49. Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, Miller MC, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med.* 2004;351(8):781–791. <https://doi.org/10.1056/NEJMoa040766>.
 50. Cohen SJ, Punt CJ, Iannotti N, Saidman BH, Sabbath KD, Gabrail NY, et al. Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *J Clin Oncol.* 2008;26(19):3213–3221. <https://doi.org/10.1200/JCO.2007.15.8923>.
 51. de Bono JS, Scher HI, Montgomery RB, Parker C, Miller MC, Tissing H, et al. Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin Cancer Res.* 2008;14(19):6302–6309. <https://doi.org/10.1158/1078-0432.CCR-08-0872>.
 52. Rhim AD, Mirek ET, Aiello NM, Maitra A, Bailey JM, McAllister F, et al. EMT and dissemination precede pancreatic tumor formation. *Cell.* 2012;148(1-2):349–361. <https://doi.org/10.1016/j.cell.2011.11.025>.
 53. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. *Science.* 2011;331(6024):1559–1564. <https://doi.org/10.1126/science.1203543>.
 54. Pantel K, Alix-Panabieres C. Real-time liquid biopsy in cancer patients: fact or fiction? *Cancer Res.* 2013;73(21):6384–6388. <https://doi.org/10.1158/0008-5472.CAN-13-2030>.
 55. Strati A, Kasimir-Bauer S, Markou A, Parisi C, Lianidou ES. Comparison of three molecular assays for the detection and molecular characterization of circulating tumor cells in breast cancer. *Breast Cancer Res.* 2013;15(2):R20. <https://doi.org/10.1186/bcr3395>.
 56. Zeune LL, de Wit S, Berghuis AMS, IJzerman MJ, Terstappen L, Brune C. How to Agree on a CTC: Evaluating the Consensus in

- Circulating Tumor Cell Scoring. *Cytometry A*. 2018;93(12):1202–1206. <https://doi.org/10.1002/cyto.a.23576>.
57. Veta M, van Diest PJ, Willems SM, Wang H, Madabhushi A, Cruz-Roa A, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med Image Anal*. 2015;20(1):237–248. <https://doi.org/10.1016/j.media.2014.11.010>.
 58. Veta M, Heng YJ, Stathonikos N, Bejnordi BE, Beca F, Wollmann T, et al. Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge. *Med Image Anal*. 2019;54:111–121. <https://doi.org/10.1016/j.media.2019.02.012>.
 59. Weis CA, Kather JN, Melchers S, Al-Ahmdi H, Pollheimer MJ, Langner C, et al. Automatic evaluation of tumor budding in immunohistochemically stained colorectal carcinomas and correlation to clinical outcome. *Diagn Pathol*. 2018;13(1):64. <https://doi.org/10.1186/s13000-018-0739-3>.
 60. Halama N, Michel S, Kloor M, Zoernig I, Benner A, Spille A, et al. Localization and density of immune cells in the invasive margin of human colorectal cancer liver metastases are prognostic for response to chemotherapy. *Cancer Res*. 2011;71(17):5670–5677. <https://doi.org/10.1158/0008-5472.CAN-11-0268>.
 61. Savas P, Salgado R, Denkert C, Sotiriou C, Darcy PK, Smyth MJ, et al. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. *Nat Rev Clin Oncol*. 2016;13(4):228–241. <https://doi.org/10.1038/nrclinonc.2015.215>.
 62. Turkki R, Linder N, Kovanen PE, Pellinen T, Lundin J. Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples. *J Pathol Inform*. 2016;7:38. <https://doi.org/10.4103/2153-3539.189703>.
 63. Aprupe L, Litjens G, Brinker TJ, van der Laak J, Grabe N. Robust and accurate quantification of biomarkers of immune cells in lung cancer micro-environment using deep convolutional neural networks. *PeerJ*. 2019;7:e6335. <https://doi.org/10.7717/peerj.6335>.
 64. Saha M, Chakraborty C, Arun I, Ahmed R, Chatterjee S. An Advanced Deep Learning Approach for Ki-67 Stained Hotspot Detection and Proliferation Rate Scoring for Prognostic Evaluation of Breast Cancer. *Sci Rep*. 2017;7(1):3213. <https://doi.org/10.1038/s41598-017-03405-5>.
 65. Niazi MKK, Tavolara TE, Arole V, Hartman DJ, Pantanowitz L, Gurcan MN. Identifying tumor in pancreatic neuroendocrine neoplasms from Ki67 images using transfer learning. *PLoS One*. 2018;13(4):e0195621. <https://doi.org/10.1371/journal.pone.0195621>.
 66. Vandenberghe ME, Scott ML, Scorer PW, Soderberg M, Balcerzak D, Barker C. Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer. *Sci Rep*. 2017;7:45938. <https://doi.org/10.1038/srep45938>.
 67. Khameneh FD, Razavi S, Kamasak M. Automated segmentation of cell membranes to evaluate HER2 status in whole slide images using a modified deep learning network. *Comput Biol Med*. 2019;110:164–174. <https://doi.org/10.1016/j.compbiomed.2019.05.020>.
 68. Sharma H, Zerbe N, Klempert I, Hellwich O, Hufnagel 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. <https://doi.org/10.1016/j.compmedimag.2017.06.001>.
 69. Sha L, Osinski BL, Ho IY, Tan TL, Willis C, Weiss H, et al. 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*. 2019;10:24. https://doi.org/10.4103/jpi.jpi_24_19.
 70. Schaumberg AJ, Rubin MA, Fuchs TJ. H&E-stained whole slide deep learning predicts SPOP mutation state in prostate cancer. 2018. Available from: <https://www.biorxiv.org/content/10.1101/064279v9>
 71. An J, Wang C, Deng Y, Yu L, Huang H. Destruction of full-length androgen receptor by wild-type SPOP, but not prostate-cancer-associated mutants. *Cell Rep*. 2014;6(4):657–669. <https://doi.org/10.1016/j.celrep.2014.01.013>.
 72. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet*. 2012;44(6):685–689. <https://doi.org/10.1038/ng.2279>.
 73. Kather JN, Pearson AT, Halama N, Jager D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med*. 2019;25(7):1054–1056. <https://doi.org/10.1038/s41591-019-0462-y>.
 74. Bychkov D, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep*. 2018;8(1):3395. <https://doi.org/10.1038/s41598-018-21758-3>.
 75. Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis CA, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multi-center study. *PLoS Med*. 2019;16(1):e1002730. <https://doi.org/10.1371/journal.pmed.1002730>.
 76. Wang S, Chen A, Yang L, Cai L, Xie Y, Fujimoto J, et al. Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome. *Sci Rep*. 2018;8(1):10393. <https://doi.org/10.1038/s41598-018-27707-4>.
 77. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velazquez Vega JE, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A*. 2018;115(13):E2970–E9. <https://doi.org/10.1073/pnas.1717139115>.
 78. Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, Marron JS, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *NPJ Breast Cancer*. 2018;4:30. <https://doi.org/10.1038/s41523-018-0079-1>.
 79. Khosravi P, Kazemi E, Imielinski M, Elemento O, Hajira-souliha I. Deep Convolutional Neural Networks Enable Discrimination of Heterogeneous Digital Pathology Images. *EBioMedicine*. 2018;27:317–328. <https://doi.org/10.1016/j.ebiom.2017.12.026>.
 80. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med*. 2018;15(11):e1002683. <https://doi.org/10.1371/journal.pmed.1002683>.
 81. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med Image Anal*. 2016;33:170–175. <https://doi.org/10.1016/j.media.2016.06.037>.
 82. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*. 2019;1(5):206–215. <https://doi.org/10.1038/s42256-019-0048-x>.

83. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface*. 2018;15(141). <https://doi.org/10.1098/rsif.2017.0387>.
84. Wang X, Janowczyk A, Zhou Y, Thawani R, Fu P, Schalper K, et al. Prediction of recurrence in early stage non-small cell lung cancer using computer extracted nuclear features from digital H&E images. *Sci Rep*. 2017;7(1):13543. <https://doi.org/10.1038/s41598-017-13773-7>.
85. FDA. Developing a software precertification program: a working model. *Fda.gov*. 2019. Available from: <https://www.fda.gov/media/119722/download>.
86. Daniel, G., Silcox, C., Sharma, I. & Wright, M. Current state and near-term priorities for AI-enabled diagnostic support software in health care. *Duke.edu*. 2019. Available from: https://healthpolicy.duke.edu/sites/default/files/atoms/files/duke_margolisaienableddxss.pdf.
87. Pesapane F, Volonte C, Codari M, Sardanelli F. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. *Insights Imaging*. 2018;9(5):745–753. <https://doi.org/10.1007/s13244-018-0645-y>.
88. Food and Drug Administration. IntelliSite Pathology Solution (PIPS, Philips Medical Systems). *Fda.gov*. 2017. Available from: <https://www.fda.gov/drugs/resources-information-approved-drugs/intellisite-pathology-solution-pips-philips-medical-systems>.
89. PAIGE. PAIGE. *Paige.ai*. 2019. Available from: <https://paige.ai/>.
90. Business Wire. FDA grants breakthrough designation to Paige.AI. *Businesswire.com*. 2019. Available from: <https://www.businesswire.com/news/home/20190307005205/en/FDA-Grants-Breakthrough-Designation-Paige.AI>.
91. Tizhoosh HR, Pantanowitz L. Artificial Intelligence and Digital Pathology: Challenges and Opportunities. *J Pathol Inform*. 2018;9:38. https://doi.org/10.4103/jpi.jpi_53_18.
92. Bueno G, Fernandez-Carrobles MM, Deniz O, Garcia-Rojo M. New Trends of Emerging Technologies in Digital Pathology. *Pathobiology*. 2016;83(2-3):61–69. <https://doi.org/10.1159/000443482>.
93. Montalto MC. An industry perspective: An update on the adoption of whole slide imaging. *J Pathol Inform*. 2016;7:18. <https://doi.org/10.4103/2153-3539.180014>.
94. Lowe, A., et al. Validation of digital pathology in a healthcare environment. *Digitalpathologyassociation.org*. 2011. Available from: https://digitalpathologyassociation.org/_data/files/DPA-Healthcare-White-Paper-FINAL_v1.0.pdf.
95. Digital Pathology Association. Healthcare FAQs. *Digital-pathologyassociation.org*. 2019. Available from: <https://digital-pathologyassociation.org/healthcare-faqs>.
96. 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(11):703–715. <https://doi.org/10.1038/s41571-019-0252-y>.

How to cite this article: Jiang Y, Yang M, Wang S, Li X, Sun Y. Emerging role of deep learning-based artificial intelligence in tumor pathology. *Cancer Communications*. 2020;40:154–166. <https://doi.org/10.1002/cac2.12012>