Review Article

Progress on deep learning in digital pathology of breast cancer: a narrative review

Jingjin Zhu[1], Mei Liu[2], Xiru Li[3]

[1]School of Medicine, Nankai University, Tianjin, China; [2]Department of Pathology, Chinese People's Liberation Army General Hospital, Beijing, China; [3]Department of General Surgery, Chinese People's Liberation Army General Hospital, Beijing, China

Contributions: (I) Conception and design: X Li, J Zhu; (II) Administrative support: X Li; (III) Provision of study materials or patients: X Li; (IV) Collection and assembly of data: J Zhu; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Prof. Xiru Li. Department of General Surgery, Chinese People's Liberation Army General Hospital, Beijing, China. Email: 2468li@sina.com; Prof. Mei Liu. Department of Pathology, Chinese People's Liberation Army General Hospital, Beijing, China. Email: liumei301@126.com.

Background and Objective: Pathology is the gold standard criteria for breast cancer diagnosis and has important guiding value in formulating the clinical treatment plan and predicting the prognosis. However, traditional microscopic examinations of tissue sections are time consuming and labor intensive, with unavoidable subjective variations. Deep learning (DL) can evaluate and extract the most important information from images with less need for human instruction, providing a promising approach to assist in the pathological diagnosis of breast cancer. To provide an informative and up-to-date summary on the topic of DL-based diagnostic systems for breast cancer pathology image analysis and discuss the advantages and challenges to the routine clinical application of digital pathology.

Methods: A PubMed search with keywords ("breast neoplasm" or "breast cancer") and ("pathology" or "histopathology") and ("artificial intelligence" or "deep learning") was conducted. Relevant publications in English published from January 2000 to October 2021 were screened manually for their title, abstract, and even full text to determine their true relevance. References from the searched articles and other supplementary articles were also studied.

Key Content and Findings: DL-based computerized image analysis has obtained impressive achievements in breast cancer pathology diagnosis, classification, grading, staging, and prognostic prediction, providing powerful methods for faster, more reproducible, and more precise diagnoses. However, all artificial intelligence (AI)-assisted pathology diagnostic models are still in the experimental stage. Improving their economic efficiency and clinical adaptability are still required to be developed as the focus of further researches.

Conclusions: Having searched PubMed and other databases and summarized the application of DL-based AI models in breast cancer pathology, we conclude that DL is undoubtedly a promising tool for assisting pathologists in routines, but further studies are needed to realize the digitization and automation of clinical pathology.

Keywords: Breast cancer; digital pathology; deep learning (DL); artificial intelligence (AI)

Submitted Jan 06, 2022. Accepted for publication Mar 04, 2022.
doi: 10.21037/gs-22-11
View this article at: https://dx.doi.org/10.21037/gs-22-11

Introduction

Breast cancer is the most prevalent cancer diagnosis for women. According to the latest Global Cancer (GLOBOCAN) statistics from the World Health Organization (WHO) in December 2020, there were 2,261,419 new cases of breast cancer in women worldwide in 2020, accounting for 11.7% of the incidence and 15.5% of the mortality, ranking first of all cancers (1).

Pathology is the gold standard criteria for breast cancer diagnosis (2), which can not only identify the nature

of lesions but also provide detailed information for the treatment and prognosis of invasive cancer, such as tumor size, histological type and grade, presence or absence of ductal carcinoma in situ (DCIS), lymphovascular invasion (LVI) and lymph node metastasis, and resection margins status (3-5). In the meantime, individualized medical treatment and precision medical care have been constantly modified (6-8). For breast cancer, the main focus is on endocrine therapy for hormone receptor positivity and anti-human epidermal growth factor receptor 2 (HER2) targeted treatment (8-11). Therefore, accurate biomarker assessment has become particularly vital in the clinical laboratory (12-14).

However, conventional manual microscopy procedures are usually time-consuming and laborious, and the lack of pathologists is an evident issue in most parts of the world (15-17), preventing the large amount of clinically relevant information contained in histopathology images from being deeply explored and effectively utilized.

In recent years, with the establishment of public databases and the development of artificial intelligence (AI) technology, the digital pathology workflow is emerging (16,18,19). Digital microscopy technology based on whole slide imaging enables the preservation of the entire glass slides in the form of digital images as well as provides a platform for the application of AI (20-22). In particular, deep learning (DL) methods, using biologically-inspired networks to represent data, have made groundbreaking improvements in computer-aided diagnosis (23,24). This paper introduced the development of digital pathology and reviewed the current research status of DL-based AI models in the diagnosis, classification, grading, staging, and prognostic prediction of breast cancer, and analyzed the advantages and challenges of digital pathology in routine clinical applications. We present the following article in accordance with the Narrative Review reporting checklist (available at https://gs.amegroups.com/article/view/10.21037/gs-22-11/rc).

Methods

A PubMed search with keywords ("breast neoplasm" or "breast cancer") and ("pathology" or "histopathology") and ("artificial intelligence" or "deep learning") was conducted. Relevant publications in English published from January 2000 to October 2021 were screened manually for their title, abstract, and even full text to determine their true relevance. Articles proposed for the development of digital pathology image-based AI models to assist in the diagnosis and prognostic assessment of breast cancer were identified. References from the searched articles and other supplementary articles were also studied. Final database search was conducted on October 20th, 2021 (*Table 1*).

Digital pathology

Digital pathology refers to the process of acquiring high-resolution images of stained tissue slides using whole-slide scanner equipment and then training AI models based on different algorithms to perform objective analysis of the digitized images, which can assist pathologists in their routine work (19,25,26).

There are four main processes in whole slide imaging to produce a complete digital image: image acquisition, storage, splicing processing, and visualization (27). Several studies have shown that diagnoses derived from digital images of frozen sections or paraffin sections are highly consistent with those from microscopic field interpretation (28-32). However, each whole slide image (WSI) contains enormous amount of information, relying only on the pathologist's visual inspections for cancer detection, tumor staging and grading, and other analyses would take a lot of time and effort. Especially for quantitative metrics, the subjective measurements and the low reproducibility lead to a huge demand for automated systems (33,34). The advancement of AI technology provides an efficient tool to automate or assist in the diagnosis of pathology and to improve the current dilemma of the lack of pathologists.

AI models in digital pathology have evolved from expert systems to traditional machine learning (ML) to DL (35,36). Both expert systems and traditional ML models rely on the rules or features defined by experts on the basis of their experience. They take data and explicitly program logical rules to generate narrow, specialized outcomes, thereby outperforming humans (37). In contrast, a key differentiating feature of DL is its autodidactic quality (15). DL enables to input image data directly and learn feature representations automatically without feature engineering, achieving end-to-end result output (35,38). It follows that the unique characteristics of deep neural networks allow them to extract information from highly dense and complex histopathological images more straightforwardly and suitably (39). Several studies have proved that DL methods have higher accuracy than traditional methods (40-42).

At present, the commonly used types of DL include convolutional neural network (CNN), recurrent neural network (RNN), deep belief nets (DBN), generative

Table 1 Search strategies of this study

Items	Specification
Date of search	2021.10.20
Databases and other sources searched	PubMed
Search terms used	#1 ("breast neoplasm" [Mesh] OR "breast cancer" [Mesh]) #2 ("pathology" [Mesh] OR "histopathology" [tiab]) #3 ("artificial intelligence" [Mesh] OR "deep learning" [Mesh]) #1 AND #2 AND #3
Timeframe	2000.01–2021.10
Inclusion and exclusion criteria	Not in English
Selection process	All retrieved articles will be uploaded to the database management software Endnote X9 with the duplicate studies deleted
	Two authors will independently screen the literature based on its title and abstract and initially remove literature that is not relevant to the topic
	Finally, the full text will be read in detail to confirm the included studies. The disagreement between the two authors in the process of selection will be resolved through discussion or discussion with the third author

adversarial networks (GAN), and autoencoders (24,37,39). Of these, CNN is the most widely used network in digital pathology, and UNet, VGGNet, GoogleNet, ResNet and DenseNet are all common basic models of CNNs. A typical CNN contains three types of network layers: convolutional layer, pooling layer, and fully connected layer (43,44). The convolution layer is used to convolve the image to extract features, the pooling layer to minimize the quantity of convolved features to lower the amount of computing power required, and the fully-connected layer to give the classification results (45). Furthermore, some scholars have selected other networks or integrated multiple networks to improve diagnostic performance (24).

AI in breast cancer pathology

Qualitative diagnosis

Morphological observation on histopathological sections to distinguish tumors from other types of lesions and to differentiate benign from malignant tumors can directly guide the clinical treatment strategies (39,46). Since the publication of the BreaKHis dataset, several methods have been proposed for the classification of breast histopathology images. Spanhol *et al.* (47) and Bayramoglu *et al.* (48) used CNN to classify breast cancer pathology images for both benign and malignant categories, respectively. The experimental evaluation, tested on the BreaKHis dataset and evaluated in comparison with previous studies, showed that the CNN-based models achieved better results than the traditional ML classification algorithms, with a classification accuracy higher than 80%. However, developing such a DL-based system from scratch requires the developer to have extensive pathology expertise, sufficient samples, and a long model training time to tune the system for good performance.

Transfer learning has been demonstrated that can achieve comparable or superior performance to the neural networks trained from scratch in a relatively short training period (49). Based on this perspective, Spanhol *et al.* (50) used DeCAF as an alternative scheme, which made use of a pre-trained CNN as a feature extractor to extract the feature vector from different layers of the network, and the output was used as the input to another classifier to train on problem-specific data. This method developed a high-accuracy system very fast, which obtained a comparable identification rate compared with the above-proposed method of training CNNs from scratch. In addition, the method allowed the comparison of the features learned from the CNN with hand-crafted features, verifying that CNN can extract image features effectively.

Later, Araújo *et al.* (51) refined the classification in further detail and classified the images into four categories: normal tissue, benign lesions, carcinoma *in situ*, and invasive carcinoma, with an accuracy of 77.8%. It should be noticed that, unlike invasive carcinoma, the

identification of carcinoma *in situ* is dependent on tumor location. Therefore, this CNN architecture was designed to retrieve information at different scales, including both nuclei and overall tissue organization, making it suitable for histological classification not only at the patch level but also at the slide image level. Furthermore, with the goal of advancing the state-of-the-art in automatic classification, the grand challenge on breast cancer histology images (BACH) was organized in conjunction with the 15th International Conference on Image Analysis and Recognition (ICIAR 2018). The majority of submitted methods are based on DL, demonstrating its dominant tendency in computer-aided analysis (52). Among them, the model combining Resnet-101 and Densenet-161 proposed by Chennamsetty *et al.* (53) and the Inception-Resnet-v2 model proposed by Kwok *et al.* (54) alleviated the feature redundancy and gradient vanishing problems due to the increasing depth of the network, increasing the overall four-category classification accuracy to 87%. Besides, the methods of Yan *et al.* (55) integrated the advantages of CNN and RNN to preserve the short-term and long-term dependencies between the patches to retain contextual information. This method achieved state-of-the-art results with an average accuracy of 91.3% for a 4-category classification task.

Subclass identification

Identifying the pathological subclasses of benign and malignant breast lesions is of equal significance to assist in assessing the potential risk of deterioration of benign lesions and guiding the selection of surgical procedures (56), as well as predicting the postoperative recurrence rate of malignant lesions (57). For breast pathology, the changes in tissue structure range from non-proliferative changes to proliferative changes, such as usual ductal hyperplasia (UDH), atypical ductal hyperplasia (ADH), DCIS, and invasive ductal carcinoma (IDC) (58).

ADH is a low-grade neoplastic intraductal hyperplasia with the same histologic and immunophenotypic features as low-grade DCIS and the differential diagnosis between them is based on size only. According to the consensus recommendation, for benign proliferative lesions with ADH, open surgical excision (OE) is preferred rather than a vacuum-assisted biopsy (VAB) and followed up for 5 years (56). Thus, automated multi-class breast cancer classification has higher clinical values than binary classification.

Gecer *et al.* (59) presented a CNN system that classifies WSIs of breast biopsies into five diagnostic categories, including non-proliferative changes, proliferative changes, ADH, DCIS, and IDC. The overall slide-level classification accuracy of 55% was comparable to the performances of the 45 pathologists that practice breast pathology in their daily routines. Han *et al.* (60) proposed a class structure-based deep convolutional neural network (CSDCNN) to classify BACH from the BreakHis dataset into eight sub-classes, including adenosis, fibroadenoma, phyllodes tumor, tabular adenoma, ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma for the first time. Notably, different classes have subtle differences and cancerous cells have high coherency (61,62). Thus the researchers took into account the relation of feature space among intra-class and inter-class and formulated some feature space distance constraints for controlling the feature similarities of different classes of the histopathological images in the design process. The average accuracy is 93.2% at the patient level and 93.8% at the image level for all magnification factors (60). Likewise, Alom *et al.* (63) proposed a classification model based on the Initial Recurrent Residual Convolutional Neural Network (IRRCNN). To facilitate comparison of results, they applied the same experimental setup as in (60). The IRRCNN model showed 97.95% average testing accuracy for ×40 magnification that is 2.15% better compared to the CSDCNN. And for the patient-level performance analysis, IRRCNN has achieved 96.84% average highest testing accuracy for eight classes breast cancer classification which is around 2.14% higher testing accuracy compared to the CSDCNN (63).

The proposed eight-classification models were all trained and tested on publicly available datasets, and it is necessary to verify whether their performance would remain robust when applied in a real clinical environment. In addition, it should be taken into account that a complete histological slide often contains multiple types of lesions which cannot be simply categorized into one type, diagnosing only the most obvious subtype would lead to a loss of information. Therefore, the results obtained from the above researches can be used as a baseline for future researches to design models that can quantify the percentage of pathological subtypes of each breast lesion and develop more practical assisted diagnostic systems.

Invasive region division

Assessment of tumor size is typically confined to areas containing invasive cancer (64). For the resected breast tissue, an initial distinction is made between the areas corresponding to invasive and non-invasive lesions or

normal tissues. The accurate description of their region is a prerequisite for the correct staging of the tumor (65). Cruz-Roa *et al.* (66) used the CNN-based classifier to automatically detect the presence and extent of invasive breast cancer at WSIs. The results showed that the cancer regions detected by the model in positive cases overlapped with the manually labeled cancer regions by pathologists at least 80%. Following that, to improve the efficiency, they proposed a High-throughput Adaptive Sampling for whole-slide Histopathology Image analysis (HASHI), which can estimate the probability of the presence of invasive breast cancer within a WSI. Compared to applying the tile classifier densely over the entire WSI, the newly proposed method takes less than 1 min to run on each WSI and achieves an average Dice coefficient of 0.76, showing great potential to be a clinical decision support tool (67). For the same dataset, Romero *et al.* (68) proposed a DL model derived from inception architecture. They placed a multi-level batch normalization module between each convolutional step for feature extraction and obtained a balanced accuracy of 89% and an F1 score of 90%.

The above methods all sample large-sized WSI into smaller patches for analysis, which leads to an exponential increase in the processing required. Patil *et al.* (69) also used the HASHI strategy to take down-sampled low-resolution WSI in combination with a skip connection-based auto-encoder model of U-Net for image segmentation. Instead of passing the samples through CNN for all the samples, the proposed architecture can perform computations directly on a scaled WSI and decrease the number of computations exponentially. In the study of Celik *et al.* (70), two popular network architectures, ResNet-50 and DenseNet-161, trained on large image datasets, were employed using the transfer learning technique. Without the need to redesign the deep network architecture, only the last layers of these networks are trained for IDC detection. Compared to the state-of-art techniques, their developed system obtained the highest classification performance with an F-sore of 92.38% for DenseNet-161 and 94.11% for ResNet-50 (70).

These algorithms have demonstrated reliable automated detection of infiltrating cancers and could serve as a basis for future research to implement a reliable system for immunophenotypic characterization as quantitative measurements of biomarkers should only be analyzed for invasive malignant epithelial cells. However, current methods mostly use manually defined regions of interest (ROI), which generally contain different proportions of mesenchymal fibroblasts and inflammatory cells and these cells cannot be completely removed through the digital image analysis. In addition, according to the segmentation area of invasive cancer, it is possible to assess tumor responsiveness to neoadjuvant therapy through the determination of the relative percentage of tumor epithelium and stroma in tumor volume before and after chemotherapy.

Histological grading

In 2003, the WHO adopted the Nottingham grading system as the standard histological grading system for invasive breast cancer. According to this system, the following three factors should be evaluated: (I) degree of tubular formation, (II) nuclear pleomorphism, (III) mitotic activity (71,72). The differences and misunderstandings among pathologists in their interpretations of the criteria will definitely weaken the guidance of histological grading for clinical prognostic assessment (73,74).

Dalle *et al.* (75) developed the first grading system that combines the three criteria, detecting tubule formation in low-resolution images, selecting individual cells and classifying them in high-resolution images for nuclear pleomorphism and mitotic count scoring. Although this system tended to score at a slightly lower level than pathologists, it can remind pathologists of widely varying cases by providing a second opinion. Wan *et al.* (76) conjuncted the semantic-level features extracted by CNN with pixel-level (texture) and object-level (architecture) features to create an integrated set of image attributes and utilized a cascaded approach to train a multiple support vector machine (SVM) in distinguishing between low, intermediate, and high Nottingham grade images from breast histopathology with an overall accuracy of 0.69. Couture *et al.* (77) used the VGG16 architecture that was pre-trained on the ImageNet dataset to classify low-intermediate *vs.* high tumor grade images, obtained an accuracy of 82%.

In addition, several DL-based methods are showing good performance in the assessment of single criteria for the histological grading of breast cancer:

(I) In the assessment of degree of tubular formation: Romo-Bucheli *et al.* (78) dropped the traditional assessment method of identifying ductal lumen. Instead, they used a deep neural network for the identification of tubule cell nuclei in WSIs and used the ratio of tubule nuclei to the overall number of nuclei as an indicator for the assessment of

glandular duct formation, with an optimal F-score of 71%. Similarly, Whitney *et al.* (79) also focused on identifying tubule nuclei, with the difference that they used a larger number of nuclear-specific features to assess tubule-forming structures.

(II) In the assessment of nuclear pleomorphism: Das *et al.* (80) performed a comparative analysis of four breast cancer grading techniques based on a common dataset of breast cancer nuclear heterogeneity scoring algorithms. Wherein, Rezaeilouyeh *et al.* (81) used phase values of shearlet coefficients as a key feature for breast cancer grading and CNN to learn the most relevant feature representations with a classification accuracy of 75%. In contrast, the Multi-Resolution Convolutional Network (MR-CN) with Plurality Voting (MR-CN-PV) model proposed by Xu *et al.* (82) gave a better result for nuclear atypia scoring with a classification accuracy of 80%.

(III) In the assessment of mitotic activity: Ciresan *et al.* (83) applied DL to mitosis detection for the first time and won the ICPR 2012 mitosis detection competition with an F1 value of 0.78%. Subsequently, in the assessment of mitosis detection algorithms 2013 (AMIDA13) challenge, the IDSIA model based on Maximum Pool Convolutional Neural Network (MMPCNN) showed comparable agreement with pathologists, reaffirming that DL has good performance in image recognition. Moreover, relabeling experiments showed that a large part of the "false positives" generated by the IDSIA model can be considered as true mitosis (84). It can be inferred that the complexity of the task and observer variability could lead to missed mitotic detection, while AI has a greater advantage over visual assessment.

Biomarker quantification

Investigating subtypes of breast cancer at the molecular level is routinely analyzed for planning specific treatments and exploring new therapeutic techniques. Due to the impracticality of clinical diagnosis of breast cancer by genetic phenotyping in the current stage, immunohistochemistry (IHC) for protein expression is often used as an alternative (85). According to expert consensus, four biomarkers should be analyzed by IHC in pathological examination of breast cancer specimens: estrogen receptor (ER), progesterone receptor (PR), HER2, and Ki67 (12,86,87). However, international standardization for the quantitative indicators is still missing and the measured inter-laboratory variability is rather high (88). The traditional method of visual assessment by pathologists and manual calculation of the percentage of positively stained nuclei has significant sampling and counting bias. It is estimated that these biases resulted in about 10% of patients not being treated adequately (89,90). Several studies have demonstrated that digital image analysis is superior to manual biomarker assessment in breast cancer (91,92).

Vandenberghe *et al.* (93) developed a computational approach based on the CNN that automatically scores HER2. In a cohort of 71 breast tumor resection samples stained by IHC, the automated method showed a concordance of 83% with a pathologist. The 12 discordant cases were independently reviewed, leading to a modification of diagnosis from initial pathologist assessment for 8 cases, 7 of which were consistent with the AI diagnostic opinion (93). Saha *et al.* (94) proposed a deep neural network (HscoreNet) to compute the score of ER and PR based on the staining intensity, the color expression, and the number of immunopositive and immunonegative nuclei of IHC stained images, achieving excellent performance, with 95.87% precision and 94.53% classification accuracy.

In addition, considering the difference in localization of the positive signal of IHC staining in different biological markers, Feng *et al.* (95) proposed a novel model based on the DenseNet to recognize both nuclear staining and cell membrane staining and grade the staining intensity as a sequential learning task. The scoring consistency of ER/PR, Ki67 and HER2 between this model and expert interpretation was 92.79%, 97.12% and 80.46% respectively (95). Other scholars have improved the UNet model and used staining intensity as well as membrane connectivity for hyperpixel-based tissue region classification and cell membrane segmentation to achieve HER2 assessment at the WSI level, more in line with the guideline scoring criteria (96,97).

Since the labeling index should only analyze invasive malignant epithelial cells, other tumor markers, such as cytokeratin, are used to help define tumor regions and accurate proliferation index calculations. But the overlapping pigments make visual analysis more difficult (98). Valkonen *et al.* (99) developed a DL-based digital mask for automated epithelial cell detection using fluoro-chromogenic cytokeratin-Ki67 double staining and

sequential hematoxylin-IHC staining as training material. The results showed that the effect of epithelial cell masking on the Ki67 labeling index was substantial; 52 tumor images initially classified as low proliferation (Ki67 <14%) without epithelial cell masking were re-classified as high proliferation (Ki67 ≥14%) after applying the DL mask (99). Shamai *et al.* (100) applied CNN to a process they termed morphological-based molecular profiling (MBMP) for robust determination of molecular expression based on hematoxylin and eosin (H&E) stained tissue section images. MBMP escapes technical issues such as fixation or antigen retrieval, obsoletes the need for subjective human interpretation, and avoids false-negative findings due to splice variants missing the antibody binding site and the accuracy in prediction of ER expression is more than 90% (100). He *et al.* (101) proposed ST-Net, combining DL with spatial transcriptomics to predict the spatial expression differences of 102 genes including the above four biomarkers from the H&E stained images directly.

Lymph node status assessment

Lymphatic metastasis is the most common way for breast cancer metastasis and the sentinel lymph node (SLN) is the first site of lymphatic metastasis from the tumor *in situ* (102). In recent years, the conception of SLN biopsy has significantly changed the way of treating axillary lymph nodes during surgery. Patients with SLN of 1–2 metastases and axillary descending stage after neoadjuvant therapy can be conditionally exempt from axillary lymph node dissection (ALND), which can reduce the patient's risk of upper extremity limitation, pain, and edema after surgery (103,104). Thus, the correct assessment of SLN status is not only an important part of the clinical staging of breast cancer but also an essential basis for the selection of patient treatment strategies (103). However, the accuracy of SLN assessment by pathologists is not satisfactory, especially in the diagnosis of micro-metastatic lesions with an average sensitivity of 38.3% only (105). DL has been demonstrated to identify metastases in SLN slides with 100% sensitivity and rectify nearly 40% of the underdiagnosed cases (106).

Aiming to investigate the potential of AI for the detection of metastases in SLN slides, Ehteshami Bejnordi *et al.* (107) organized the Cancer Metastases in Lymph Nodes Challenge 2016 (CAMELYON16) competition. In the submitted methods, the GoogleNet-based deep neural network outperformed the pathologist with the best AUC of 0.994. Later, Steiner *et al.* (108) proposed a more optimized algorithm, Lymph Node Assistant (LYNA), which can obtain higher sensitivity for lesion detection by filtering image artifacts, and demonstrated that algorithm-assisted pathologists have higher accuracy than pathologists alone.

According to the results of the NSABP B-32 trial, patients with SLN biopsies suggestive of occult metastases, including micro-metastases and isolated tumor cells (ITC), showed significant differences in overall survival and disease free survival compared to patients without occult metastases (109). Therefore, in the CAMELYON17 competition, ITC and the smallest type of metastasis had been included for the classification setting of SLN metastases. Moreover, to improve clinical relevance, the CAMELYON17 competition focused on patient-level pN-stage prediction including multiple WSIs per patient (110). Overall, the kappa metric ranged from 0.89 to −0.13 across all submissions. The best results were obtained with pre-trained architectures such as ResNet. It performed well on slides containing macroscopic metastases and metastasis-free tumors but poorly in identifying ITC with an accuracy of 11.4%. In addition, most of the methods took hundreds of minutes to run, which created a barrier to clinical application. To improve computational efficiency, Kong *et al.* (111) and Zhao *et al.* (112) used transfer learning to accelerate model convergence, reducing the time for a single WSI review to 5.6 and 7.2 min. Afterwards, Campanella *et al.* (113) trained a weakly supervised learning model based on 44,732 full-slice scanned images, avoiding the manual process of extensive annotation, and obtained an AUC value of 0.965 in a test of identifying axillary lymph node metastases in breast cancer. Their results showed the clinical application of the proposed model would allow pathologists to exclude 65–75% of slides while retaining 100% sensitivity, laying the foundation for the deployment of computational decision support systems in clinical practice (113).

Surgical margin assessment

Breast-conserving surgery (BCS) followed by radiation therapy (RT) is the standard procedure of early-stage breast cancer treatment. If clear margins are obtained, it could provide similar survival rates as total mastectomy while better cosmetic results (114,115).

The current standard for margin assessment is a histologic review provided by the pathologist of the tissue embedded in paraffin and stained with H&E. Unfortunately, time requirements for this process don't allow for its use

intraoperatively (116). Frozen-section analysis (FSA) is an alternative method that can be performed in a relatively short time. Two common sampling methods for the frozen section are the surgeon taking small samples of tissue from the defect cavity after removal of tumor majority and the pathologist taking a sample directly from the primary resection specimen for evaluation (117,118). These methods have a low sampling percentage, resulting in a sensitivity of 81% and a mean reoperation rate of 5.9%, ranging from 0% to 23.9% (115). Sampling and analyzing larger amounts of tissue may increase the sensitivity of detecting small tumor lesions, but practical issues such as surgical time and cost should also be taken into consideration.

Several researches demonstrated that the use of X-rays for specimen imaging can improve the targeting of sampling and lead to a significant reduction in positive margins (119-122). For example, Zhang *et al.* (122) used a breast pathology cabinet X-ray system (CXS) to assist in breast cancer tumor bed identification. Compared to visual observation, CXS can significantly improve the accuracy of measurement and the efficiency of tumor collection.

Large format histopathology is another efficient way for visualization of the tumor and resection margins as it eliminates the process of slicing the tissue into multiple blocks, avoiding undersampling of cancer specimens (123-126). However, due to the limitations of the frozen technique, large sections are not available for intraoperative evaluation at present. It is expected that if the frozen large section technique can be implemented in the future, we can combine it with the AI algorithms for identifying cancer areas, enabling rapid intraoperative margin assessment.

Prognosis prediction

Prognostic model refers to the use of statistical methods to determine the quantitative relationship between the risk factors and the probability of clinical outcomes based on the patient's disease state. Breast cancer prognostic models can help clinicians and healthcare providers make more informed medical decisions on chemotherapy exemption (127). In recent decades, the most popular model is the COX proportional hazards model, which has been extensively studied in the fields of statistical learning (128). These methods based on the traditional COX proportional hazards model mostly utilized structural characteristics of the patient's information, tumor staging and characteristics and combined these variables linearly (129-132).

With the development of medical imaging technology, more and more unstructured medical images are available for diagnosis, treatment and survival analysis. Previous studies showed that some computational methods had been introduced to predict cancer clinical outcomes based on pathological images by assuming that pathological images may provide complementary information related to tumor characteristics and achieved good performance for lung cancer (133-135). However, there are few studies that use pathological images for clinical outcome analysis of breast cancer due to its high degree of complexity and heterogeneity.

Sun *et al.* (136) conducted a powerful method named GPMKL based on multiple kernel learning (MKL) for breast cancer survival prediction by integrating genomic data and features distilled from pathological images. The result showed that compared with the use of single-dimensional data namely the genomic data, the joint use of genomic data and pathological images increased the AUC from 0.794 to 0.821, which demonstrated that the pathological image information plays a critical part in accurately predicting the survival time (136). Klimov *et al.* (137) also developed an ML approach to identify prognostically relevant features obtained from the texture of H&E slides to predict DCIS recurrence risk. It was verified that the model was able to identify a high-risk group of patients that had almost a 50% chance of recurring within 10 years and provide predictive value for the long-term outcome of radiotherapy after BCS in patients with different risk groups (137).

Recently, DL-based approaches for the integration of data from different modalities have been proposed and successfully applied in cancer prognosis prediction, which are highly flexible and can interpret the complexity of data in a non-linear manner (138-140). Wang *et al.* (141) presented a novel unified framework named genomic and pathological deep bilinear network (GPDBN) for prognosis prediction by integrating both genomic data and pathological images. Their findings also suggested that prognosis prediction methods based on data from different modalities outperformed those using single modality data. More importantly, GPDBN outperformed all non-DL methods, indicating that sophisticated DL-based methods are advantageous in integrating data from different modalities.

Challenges in the clinical application of digital pathology

With the development of AI technology, pathology analysis

is no longer limited to the traditional qualitative analysis but transitioned to quantitative analysis gradually (142). Obtaining the pathological diagnosis through conducting data statistics, establishing mathematical models, and calculating the parameters related to lesions can effectively reduce the mistakes caused by subjective factors and improve the overall level and efficiency of medical services. But there are challenges to fully realize the digitalization and automation of clinical pathology as follows:

(I) Financial investment. In general, tissue sections are usually scanned at ×20 or ×40 objective magnification, and the images obtained from ×40 objective scans are converted into files of 0.5 to 4 GB in size, which would occupy a large amount of memory space (143). Hence, the storage of images requires high-specification hardware. To date, most computational programs are executed on the CPU of a computer, while DL is better performed on a graphics processing unit (GPU) (144,145). As a result, more expensive GPUs may need to be purchased to improve work performance.

(II) Data Sharing. Compared to traditional analysis methods where image features are selected manually, DL is highly data-dependent, as it must identify these features automatically. In order to make the model have good generalization characteristics, the training samples should be comprehensive and representative. In addition, most present AI methods still require pathologists to label the training set images manually when training the models, which is a tedious and time-consuming task. Although weakly supervised learning methods can avoid the implementation of this step, they also require the support of large datasets. The classification accuracy of models trained on small data sets is not satisfactory (146). Therefore, data share worldwide to obtain numerous different datasets can improve model stability and be sufficient to deal with clinical complexity without fine labeling (113).

(III) Image preprocessing. In surgical pathology, there are no recognized standards for tissue processing, staining, and slide preparation (147,148). As a result, an AI models that performed well on a set of WSIs may not work well in generalization and utilization due to a series of biases. Applying an image pre-processing step for color normalization to reduce the effect of coloring and processing could maintain the model with good performance to some extent (149).

(IV) Standardized training. Hanna *et al.* (150) demonstrated that pathologists who lacked training or experience in the technical application of digital pathology platforms had an increase in average reading time per slide of 19 s and a 19% decrease in efficiency per case assessment. Consequently, additional courses or seminars to provide relevant training will improve the adaptation of pathologists to digital systems and the efficiency in their application, facilitating the safe and efficient use of digital pathology platforms (151).

(V) Responsibility and regulation. To accomplish tasks more stably, DL models are becoming more and more complex in structure, which leads to a loss of interpretability of the inner workings of the models, creating a "black box" problem (152). Activation maps, or heatmaps, are methods that attempt to address the "black box" issue by highlighting areas of images with the output classification label (153-155). However, these methods still require human interpretation to verify whether the features identified by DL models are the same as those used by physicians to diagnose the disease. The ultimate goal should be the information provided by the user about the decision-making process of the DL model to build trust and the facilitate adoption and deployment of DL technologies in clinical scenarios. In addition, it is crucial to clarify the status of AI in the healthcare system and to regulate the relevant laws to improve the liability and regulatory system. In this way, legal liability can be defined if medical disputes occur when applying AI for diagnosis.

Discussion

Digital image analysis methods have been widely used in many fields of modern medicine and the FDA has approved a variety of AI-based diagnostic systems for radiology clinical diagnosis, performing manual-like or even more than manual tasks, such as tumor region identification and segmentation (156-158). In contrast to imaging methods such as CT and MRI, histopathology images have much larger pixels. The morphology and spatial disposition of millions of cells in a slide contain much more dense and complex information that cannot be analyzed effectively by

visual recognition alone. The integration of DL technology into pathology diagnosis can not only compensate for the unpredictable factors due to the pathologist's subjective experience but also improve diagnostic accuracy in less time, fundamentally changing the way detect and treat breast cancer in the near future. In addition, integrating pathology with other types of information, such as genomics and radiomics, contributes to a deeper exploration of image information and further understanding of the mechanisms of disease development. To date, all AI-assisted pathology diagnostic models are still in the experimental stage. How to improve the economic efficiency and clinical adaptability of the models is still the focus of research for the long-term future.

In conclusion, having searched PubMed and other databases and summarized the application of DL-based AI models in breast cancer pathology, we conclude that DL is undoubtedly a promising tool for assisting pathologists in routines, but further studies are needed to realize the digitization and automation of clinical pathology.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the Narrative Review reporting checklist. Available at https://gs.amegroups.com/article/view/10.21037/gs-22-11/rc

Peer Review File: Available at https://gs.amegroups.com/article/view/10.21037/gs-22-11/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://gs.amegroups.com/article/view/10.21037/gs-22-11/coif). XL serves as an Editor-in-Chief of *Gland Surgery* from May 2017 to April 2022. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Akram M, Iqbal M, Daniyal M, et al. Awareness and current knowledge of breast cancer. Biol Res 2017;50:33.
3. Gradishar WJ, Anderson BO, Abraham J, et al. Breast Cancer, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2020;18:452-78.
4. Solanki M, Visscher D. Pathology of breast cancer in the last half century. Hum Pathol 2020;95:137-48.
5. Cardoso F, Senkus E, Costa A, et al. 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4)†. Ann Oncol 2018;29:1634-57.
6. Akechi T, Yamaguchi T, Uchida M, et al. Smartphone problem-solving and behavioural activation therapy to reduce fear of recurrence among patients with breast cancer (SMartphone Intervention to LEssen fear of cancer recurrence: SMILE project): protocol for a randomised controlled trial. BMJ Open 2018;8:e024794.
7. Ho D, Quake SR, McCabe ERB, et al. Enabling Technologies for Personalized and Precision Medicine. Trends Biotechnol 2020;38:497-518.
8. Harbeck N, Gnant M. Breast cancer. Lancet 2017;389:1134-50.
9. Meric-Bernstam F, Johnson AM, Dumbrava EEI, et al. Advances in HER2-Targeted Therapy: Novel Agents and Opportunities Beyond Breast and Gastric Cancer. Clin Cancer Res 2019;25:2033-41.
10. The Lancet. Breast cancer targeted therapy: successes and challenges. Lancet 2017;389:2350.
11. Modi S, Park H, Murthy RK, et al. Antitumor Activity and Safety of Trastuzumab Deruxtecan in Patients With HER2-Low-Expressing Advanced Breast Cancer: Results From a Phase Ib Study. J Clin Oncol 2020;38:1887-96.
12. Coates AS, Winer EP, Goldhirsch A, et al. Tailoring therapies--improving the management of early breast

cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol 2015;26:1533-46.
13. Sheffield BS, Kos Z, Asleh-Aburaya K, et al. Molecular subtype profiling of invasive breast cancers weakly positive for estrogen receptor. Breast Cancer Res Treat 2016;155:483-90.
14. Caselli E, Pelliccia C, Teti V, et al. Looking for more reliable biomarkers in breast cancer: Comparison between routine methods and RT-qPCR. PLoS One 2021;16:e0255580.
15. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44-56.
16. Fuyong Xing, Yuanpu Xie, Hai Su, et al. Deep Learning in Microscopy Image Analysis: A Survey. IEEE Trans Neural Netw Learn Syst 2018;29:4550-68.
17. Metter DM, Colgan TJ, Leung ST, et al. Trends in the US and Canadian Pathologist Workforces From 2007 to 2017. JAMA Netw Open 2019;2:e194337.
18. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med 2018;24:1559-67.
19. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. Lancet Oncol 2019;20:e253-61.
20. Pantanowitz L, Valenstein PN, Evans AJ, et al. Review of the current state of whole slide imaging in pathology. J Pathol Inform 2011;2:36.
21. Evans AJ, Bauer TW, Bui MM, et al. US Food and Drug Administration Approval of Whole Slide Imaging for Primary Diagnosis: A Key Milestone Is Reached and New Questions Are Raised. Arch Pathol Lab Med 2018;142:1383-7.
22. Bian Z, Guo C, Jiang S, et al. Autofocusing technologies for whole slide imaging and automated microscopy. J Biophotonics 2020;13:e202000227.
23. Chang MC, Mrkonjic M. Review of the current state of digital image analysis in breast pathology. Breast J 2020;26:1208-12.
24. Robertson S, Azizpour H, Smith K, et al. Digital image analysis in breast pathology-from image processing techniques to artificial intelligence. Transl Res 2018;194:19-35.
25. Bera K, Schalper KA, Rimm DL, et al. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 2019;16:703-15.
26. Pantanowitz L, Sharma A, Carter AB, et al. Twenty Years of Digital Pathology: An Overview of the Road Travelled, What is on the Horizon, and the Emergence of Vendor-Neutral Archives. J Pathol Inform 2018;9:40.
27. Kumar N, Gupta R, Gupta S. Whole Slide Imaging (WSI) in Pathology: Current Perspectives and Future Directions. J Digit Imaging 2020;33:1034-40.
28. Bencze J, Szarka M, Kóti B, et al. Comparison of Semi-Quantitative Scoring and Artificial Intelligence Aided Digital Image Analysis of Chromogenic Immunohistochemistry. Biomolecules 2021;12:19.
29. Huang Y, Lei Y, Wang Q, et al. Telepathology consultation for frozen section diagnosis in China. Diagn Pathol 2018;13:29.
30. Fallon MA, Wilbur DC, Prasad M. Ovarian frozen section diagnosis: use of whole-slide imaging shows excellent correlation between virtual slide and original interpretations in a large series of cases. Arch Pathol Lab Med 2010;134:1020-3.
31. Pantanowitz L, Sinard JH, Henricks WH, 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;137:1710-22.
32. Mukhopadhyay S, Feldman MD, Abels E, et al. Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology: A Multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study). Am J Surg Pathol 2018;42:39-52.
33. Gavrielides MA, Gallas BD, Lenz P, et al. Observer variability in the interpretation of HER2/neu immunohistochemical expression with unaided and computer-aided digital microscopy. Arch Pathol Lab Med 2011;135:233-42.
34. Gavrielides MA, Conway C, O'Flaherty N, et al. Observer performance in the use of digital and optical microscopy for the interpretation of tissue-based biomarkers. Anal Cell Pathol (Amst) 2014;2014:157308.
35. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. Med Image Anal 2016;33:170-5.
36. Kim SK, Huh JH. Consistency of Medical Data Using Intelligent Neuron Faster R-CNN Algorithm for Smart Health Care Application. Healthcare (Basel) 2020;8:185.
37. Choi RY, Coyner AS, Kalpathy-Cramer J, et al. Introduction to Machine Learning, Neural Networks, and Deep Learning. Transl Vis Sci Technol 2020;9:14.
38. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J Pathol Inform 2016;7:29.

39. Jiang Y, Yang M, Wang S, et al. Emerging role of deep learning-based artificial intelligence in tumor pathology. Cancer Commun (Lond) 2020;40:154-66.
40. Deng S, Zhang X, Yan W, et al. Deep learning in digital pathology image analysis: a survey. Front Med 2020;14:470-87.
41. Li W, Manivannan S, Akbar S, et al. Gland segmentation in colon histology images using hand-crafted features and convolutional neural networks. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). Prague: IEEE, 2016.
42. Zhang L, Le Lu, Nogues I, et al. DeepPap: Deep Convolutional Networks for Cervical Cell Classification. IEEE J Biomed Health Inform 2017;21:1633-43.
43. Sharma S, Ball JE, Tang B, et al. Semantic Segmentation with Transfer Learning for Off-Road Autonomous Driving. Sensors (Basel) 2019;19:2577.
44. Stamoulakatos A, Cardona J, McCaig C, et al. Automatic Annotation of Subsea Pipelines using Deep Learning. Sensors (Basel) 2020;20:674.
45. Sinha RK, Pandey R, Pattnaik R. Deep Learning For Computer Vision Tasks: A review. arXiv 2018;arXiv:1804.03928.
46. Ellis IO, Galea M, Broughton N, et al. Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. Histopathology 1992;20:479-89.
47. Spanhol FA, Oliveira LS, Petitjean C, et al. Breast cancer histopathological image classification using Convolutional Neural Networks. 2016 International Joint Conference on Neural Networks (IJCNN). Vancouver: IEEE, 2016.
48. Bayramoglu N, Kannala J, Heikkilä J. Deep learning for magnification independent breast cancer histopathology image classification. 2016 23rd International Conference on Pattern Recognition (ICPR). Cancun: IEEE, 2016.
49. Abdolahi M, Salehi M, Shokatian I, et al. Artificial intelligence in automatic classification of invasive ductal carcinoma breast cancer in digital pathology images. Med J Islam Repub Iran 2020;34:140.
50. Spanhol FA, Oliveira LS, Cavalin PR, et al. Deep features for breast cancer histopathological image classification. 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Banff: IEEE, 2017.
51. Araújo T, Aresta G, Castro E, et al. Classification of breast cancer histology images using Convolutional Neural Networks. PLoS One 2017;12:e0177544.
52. Aresta G, Araújo T, Kwok S, et al. BACH: Grand challenge on breast cancer histology images. Med Image Anal 2019;56:122-39.
53. Chennamsetty SS, Safwan M, Alex V. Classification of Breast Cancer Histology Image using Ensemble of Pre-trained Neural Networks. In: Image Analysis and Recognition. Póvoa de Varzim: Springer International Publishing, 2018:804-11.
54. Kwok SS. Multiclass Classification of Breast Cancer in Whole-Slide Images. In: Campilho A, Karray F, ter Haar Romeny B. editors. Image Analysis and Recognition. Cham: Springer, 2018.
55. Yan R, Ren F, Wang Z, et al. Breast cancer histopathological image classification using a hybrid deep neural network. Methods 2020;173:52-60.
56. Rageth CJ, O'Flynn EAM, Pinker K, et al. Second International Consensus Conference on lesions of uncertain malignant potential in the breast (B3 lesions). Breast Cancer Res Treat 2019;174:279-96.
57. Pozzi G, Castellano I, D'Anna MR, et al. B3-lesions of the breast: Risk of malignancy after vacuum-assisted breast biopsy versus core needle biopsy diagnosis. Breast J 2019;25:1308-9.
58. Elmore JG, Longton GM, Carney PA, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA 2015;313:1122-32.
59. Gecer B, Aksoy S, Mercan E, et al. Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks. Pattern Recognit 2018;84:345-56.
60. Han Z, Wei B, Zheng Y, et al. Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model. Sci Rep 2017;7:4172.
61. Zhang X, Zhou F, Lin Y, et al. Embedding Label Structures for Fine-Grained Feature Representation. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016.
62. Wang J, Song Y, Leung T, et al. Learning Fine-grained Image Similarity with Deep Ranking. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014.
63. Alom MZ, Yakopcic C, Nasrin MS, et al. Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network. J Digit Imaging 2019;32:605-17.
64. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991;19:403-10.
65. Genestie C, Zafrani B, Asselain B, et al. Comparison of

the prognostic value of Scarff-Bloom-Richardson and Nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems. Anticancer Res 1998;18:571-6.
66. Cruz-Roa A, Gilmore H, Basavanhally A, 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.
67. Cruz-Roa A, Gilmore H, Basavanhally A, 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:e0196828.
68. Romero FP, Tang A, Kadoury S. Multi-Level Batch Normalization in Deep Networks for Invasive Ductal Carcinoma Cell Discrimination in Histopathology Images. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). Venice: IEEE, 2019.
69. Patil SM, Tong L, Wang MD. Generating Region of Interests for Invasive Breast Cancer in Histopathological Whole-Slide-Image. Proc COMPSAC 2020;2020:723-8.
70. Celik Y, Talo M, Yildirim O, et al. Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. Pattern Recognit Lett 2020;133:232-9.
71. Giuliano AE, Edge SB, Hortobagyi GN. Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer. Ann Surg Oncol 2018;25:1783-5.
72. BLOOM HJ, RICHARDSON WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer 1957;11:359-77.
73. Boiesen P, Bendahl PO, Anagnostaki L, et al. Histologic grading in breast cancer--reproducibility between seven pathologic departments. South Sweden Breast Cancer Group. Acta Oncol 2000;39:41-5.
74. Theissig F, Kunze KD, Haroske G, et al. Histological grading of breast cancer. Interobserver, reproducibility and prognostic significance. Pathol Res Pract 1990;186:732-6.
75. Dalle JR, Leow WK, Racoceanu D, et al. Automatic breast cancer grading of histopathological images. Annu Int Conf IEEE Eng Med Biol Soc 2008;2008:3052-5.
76. Wan T, Cao J, Chen J, et al. Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features. Neurocomputing 2017;229:34-44.
77. Couture HD, Williams LA, Geradts J, 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.
78. Romo-Bucheli D, Janowczyk A, Gilmore H, et al. Automated Tubule Nuclei Quantification and Correlation with Oncotype DX risk categories in ER+ Breast Cancer Whole Slide Images. Sci Rep 2016;6:32706.
79. Whitney J, Corredor G, Janowczyk A, et al. Quantitative nuclear histomorphometry predicts oncotype DX risk categories for early stage ER+ breast cancer. BMC Cancer 2018;18:610.
80. Das A, Nair MS, Peter SD. Computer-Aided Histopathological Image Analysis Techniques for Automated Nuclear Atypia Scoring of Breast Cancer: a Review. J Digit Imaging 2020;33:1091-121.
81. Rezaeilouyeh H, Mollahosseini A, Mahoor MH. Microscopic medical image classification framework via deep learning and shearlet transform. J Med Imaging (Bellingham) 2016;3:044501.
82. Xu J, Zhou C, Lang B, et al. Deep Learning for Histopathological Image Analysis: Towards Computerized Diagnosis on Cancers. In: Lu L, Zheng Y, Carneiro G, et al. editors. Deep Learning and Convolutional Neural Networks for Medical Image Computing: Precision Medicine, High Performance and Large-Scale Datasets. Cham: Springer International Publishing, 2017:73-95.
83. Cireşan DC, Giusti A, Gambardella LM, et al. Mitosis detection in breast cancer histology images with deep neural networks. Med Image Comput Comput Assist Interv 2013;16:411-8.
84. Veta M, van Diest PJ, Willems SM, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. Med Image Anal 2015;20:237-48.
85. Guiu S, Michiels S, André F, et al. Molecular subclasses of breast cancer: how do we define them? The IMPAKT 2012 Working Group Statement. Ann Oncol 2012;23:2997-3006.
86. Goldhirsch A, Wood WC, Coates AS, et al. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol 2011;22:1736-47.
87. Goldhirsch A, Winer EP, Coates AS, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol 2013;24:2206-23.
88. Polley MY, Leung SC, McShane LM, et al. An international Ki67 reproducibility study. J Natl Cancer

Inst 2013;105:1897-906.
89. Dobson L, Conway C, Hanley A, et al. Image analysis as an adjunct to manual HER-2 immunohistochemical review: a diagnostic tool to standardize interpretation. Histopathology 2010;57:27-38.
90. Di Palma S, Collins N, Faulkes C, et al. Chromogenic in situ hybridisation (CISH) should be an accepted method in the routine diagnostic evaluation of HER2 status in breast cancer. J Clin Pathol 2007;60:1067-8.
91. Rizzardi AE, Johnson AT, Vogel RI, et al. Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring. Diagn Pathol 2012;7:42.
92. Stålhammar G, Fuentes Martinez N, Lippert M, et al. Digital image analysis outperforms manual biomarker assessment in breast cancer. Mod Pathol 2016;29:318-29.
93. Vandenberghe ME, Scott ML, Scorer PW, et al. Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer. Sci Rep 2017;7:45938.
94. Saha M, Arun I, Ahmed R, et al. HscoreNet: A Deep network for estrogen and progesterone scoring using breast IHC images. Pattern Recognition 2020;102:107200.
95. Feng M, Chen J, Xiang X, et al. An Advanced Automated Image Analysis Model for Scoring of ER, PR, HER-2 and Ki-67 in Breast Carcinoma. IEEE Access 2021;9:108441-51.
96. 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;110:164-74.
97. Brügmann A, Eld M, Lelkaitis G, et al. Digital image analysis of membrane connectivity is a robust measure of HER2 immunostains. Breast Cancer Res Treat 2012;132:41-9.
98. Nielsen PS, Riber-Hansen R, Jensen TO, et al. Proliferation indices of phosphohistone H3 and Ki67: strong prognostic markers in a consecutive cohort with stage I/II melanoma. Mod Pathol 2013;26:404-13.
99. Valkonen M, Isola J, Ylinen O, et al. Cytokeratin-Supervised Deep Learning for Automatic Recognition of Epithelial Cells in Breast Cancers Stained for ER, PR, and Ki-67. IEEE Trans Med Imaging 2020;39:534-42.
100. Shamai G, Binenbaum Y, Slossberg R, et al. Artificial Intelligence Algorithms to Assess Hormonal Status From Tissue Microarrays in Patients With Breast Cancer. JAMA Netw Open 2019;2:e197700.
101. He B, Bergenstråhle L, Stenbeck L, et al. Integrating spatial gene expression and breast tumour morphology via deep learning. Nat Biomed Eng 2020;4:827-34.
102. Mamounas EP, Kuehn T, Rutgers EJT, et al. Current approach of the axilla in patients with early-stage breast cancer. Lancet 2017. [Epub ahead of print]. doi: 10.1016/S0140-6736(17)31451-4.
103. Maguire A, Brogi E. Sentinel lymph nodes for breast carcinoma: an update on current practice. Histopathology 2016;68:152-67.
104. Krag DN, Anderson SJ, Julian TB, et al. Technical outcomes of sentinel-lymph-node resection and conventional axillary-lymph-node dissection in patients with clinically node-negative breast cancer: results from the NSABP B-32 randomised phase III trial. Lancet Oncol 2007;8:881-8.
105. Vestjens JHMJ, Pepels MJ, de Boer M, et al. Relevant impact of central pathology review on nodal classification in individual breast cancer patients. Ann Oncol 2012;23:2561-6.
106. Litjens G, Sánchez CI, Timofeeva N, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci Rep 2016;6:26286.
107. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA 2017;318:2199-210.
108. Steiner DF, MacDonald R, Liu Y, et al. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. Am J Surg Pathol 2018;42:1636-46.
109. Krag DN, Anderson SJ, Julian TB, et al. Sentinel-lymph-node resection compared with conventional axillary-lymph-node dissection in clinically node-negative patients with breast cancer: overall survival findings from the NSABP B-32 randomised phase 3 trial. Lancet Oncol 2010;11:927-33.
110. Bandi P, Geessink O, Manson Q, et al. From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge. IEEE Trans Med Imaging 2019;38:550-60.
111. Kong B, Sun S, Wang X, et al. Invasive Cancer Detection Utilizing Compressed Convolutional Neural Network and Transfer Learning. In: Frangi A, Schnabel J, Davatzikos C, et al. editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Cham: Springer, 2018:156-64.
112. Zhao Z, Lin H, Chen H, et al. PFA-ScanNet: Pyramidal Feature Aggregation with Synergistic Learning for Breast Cancer Metastasis Analysis. In: Shen D, Liu T, Peters TM, et al. editors. Medical Image Computing and Computer

Assisted Intervention – MICCAI 2019. Cham: Springer, 2019:586-94.
113. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019;25:1301-9.
114. Veronesi U, Cascinelli N, Mariani L, et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. N Engl J Med 2002;347:1227-32.
115. Garcia MT, Mota BS, Cardoso N, et al. Accuracy of frozen section in intraoperative margin assessment for breast-conserving surgery: A systematic review and meta-analysis. PLoS One 2021;16:e0248768.
116. Maloney BW, McClatchy DM, Pogue BW, et al. Review of methods for intraoperative margin detection for breast conserving surgery. J Biomed Opt 2018;23:1-19.
117. Black C, Marotti J, Zarovnaya E, et al. Critical evaluation of frozen section margins in head and neck cancer resections. Cancer 2006;107:2792-800.
118. Yahalom R, Dobriyan A, Vered M, et al. A prospective study of surgical margin status in oral squamous cell carcinoma: a preliminary report. J Surg Oncol 2008;98:572-8.
119. Park KU, Kuerer HM, Rauch GM, et al. Digital Breast Tomosynthesis for Intraoperative Margin Assessment during Breast-Conserving Surgery. Ann Surg Oncol 2019;26:1720-8.
120. Mariscotti G, Durando M, Pavan LJ, et al. Intraoperative breast specimen assessment in breast conserving surgery: comparison between standard mammography imaging and a remote radiological system. Br J Radiol 2020;93:20190785.
121. Lin C, Wang KY, Xu F, et al. The application of intraoperative specimen mammography for margin status assessment in breast-conserving surgery: A single-center retrospective study. Breast J 2020;26:1871-3.
122. Zhang M, Ma Y, Geng C, et al. Assisted computer and imaging system improve accuracy of breast tumor size assessment after neoadjuvant chemotherapy. Transl Cancer Res 2021;10:1346-57.
123. Bryant P, Haine N, Johnston J, et al. Application of large format tissue processing in the histology laboratory. J Histotechnol 2019;42:150-62.
124. Tot T. The role of large-format histopathology in assessing subgross morphological prognostic parameters: a single institution report of 1000 consecutive breast cancer cases. Int J Breast Cancer 2012;2012:395415.
125. Tot T. Cost-benefit analysis of using large-format histology sections in routine diagnostic breast care. Breast 2010;19:284-8.
126. Monica MAT, Morandi L, Foschini MP. Utility of large sections (macrosections) in breast cancer pathology. Transl Cancer Res 2018;7:S418-23.
127. Min N, Wei Y, Zheng Y, et al. Advancement of prognostic models in breast cancer: a narrative review. Gland Surg 2021;10:2815-31.
128. Cox DR. Regression models and life-tables. J Royal Stat Soc (B) 1972;34:187-202.
129. Phung MT, Tin Tin S, Elwood JM. Prognostic models for breast cancer: a systematic review. BMC Cancer 2019;19:230.
130. Olivotto IA, Bajdik CD, Ravdin PM, et al. Population-based validation of the prognostic model ADJUVANT! for early breast cancer. J Clin Oncol 2005;23:2716-25.
131. Mook S, Schmidt MK, Rutgers EJ, et al. Calibration and discriminatory accuracy of prognosis calculation for breast cancer with the online Adjuvant! program: a hospital-based retrospective cohort study. Lancet Oncol 2009;10:1070-6.
132. Wishart GC, Azzato EM, Greenberg DC, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res 2010;12:R1.
133. Zhu X, Yao J, Xin L, et al. Lung cancer survival prediction from pathological images and genetic data — An integration study. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). Prague: IEEE, 2016.
134. Yu KH, Zhang C, Berry GJ, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun 2016;7:12474.
135. Wang H, Xing F, Su H, et al. Novel image markers for non-small cell lung cancer classification and survival prediction. BMC Bioinformatics 2014;15:310.
136. Sun D, Li A, Tang B, et al. Integrating genomic data and pathological images to effectively predict breast cancer clinical outcome. Comput Methods Programs Biomed 2018;161:45-53.
137. Klimov S, Miligy IM, Gertych A, et al. A whole slide image-based machine learning approach to predict ductal carcinoma in situ (DCIS) recurrence risk. Breast Cancer Res 2019;21:83.
138. Cheerla A, Gevaert O. Deep learning with multimodal representation for pancancer prognosis prediction. Bioinformatics 2019;35:i446-54.
139. Mobadersany P, Yousefi S, Amgad M, et al. Predicting cancer outcomes from histology and genomics using

convolutional networks. Proc Natl Acad Sci U S A 2018;115:E2970-9.
140. Yao J, Zhu X, Zhu F, et al. Deep Correlational Learning for Survival Prediction from Multi-modality Data. In: Descoteaux M, Maier-Hein L, Franz A, et al. editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017. Cham: Springer, 2017:406-14.
141. Wang Z, Li R, Wang M, et al. GPDBN: deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction. Bioinformatics 2021;37:2963-70.
142. Gurcan MN, Boucheron LE, Can A, et al. Histopathological image analysis: a review. IEEE Rev Biomed Eng 2009;2:147-71.
143. Hanna MG, Parwani A, Sirintrapun SJ. Whole Slide Imaging: Technology and Applications. Adv Anat Pathol 2020;27:251-9.
144. Nobile MS, Cazzaniga P, Tangherloni A, et al. Graphics processing units in bioinformatics, computational biology and systems biology. Brief Bioinform 2017;18:870-85.
145. Lee JG, Jun S, Cho YW, et al. Deep Learning in Medical Imaging: General Overview. Korean J Radiol 2017;18:570-84.
146. Mercan C, Aksoy S, Mercan E, et al. Multi-Instance Multi-Label Learning for Multi-Class Classification of Whole Slide Breast Histopathology Images. IEEE Trans Med Imaging 2018;37:316-25.
147. Wick MR. The hematoxylin and eosin stain in anatomic pathology-An often-neglected focus of quality assurance in the laboratory. Semin Diagn Pathol 2019;36:303-11.
148. Feldman AT, Wolfe D. Tissue processing and hematoxylin and eosin staining. Methods Mol Biol 2014;1180:31-43.
149. Salvi M, Michielli N, Molinari F. Stain Color Adaptive Normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology. Comput Methods Programs Biomed 2020;193:105506.
150. Hanna MG, Reuter VE, Hameed MR, et al. Whole slide imaging equivalency and efficiency study: experience at a large academic center. Mod Pathol 2019;32:916-28.
151. Sarwar S, Dent A, Faust K, et al. Physician perspectives on integration of artificial intelligence into diagnostic pathology. NPJ Digit Med 2019;2:28.
152. Hayashi Y. The Right Direction Needed to Develop White-Box Deep Learning in Radiology, Pathology, and Ophthalmology: A Short Review. Front Robot AI 2019;6:24.
153. Hägele M, Seegerer P, Lapuschkin S, et al. Resolving challenges in deep learning-based analyses of histopathological images using explanation methods. Sci Rep 2020;10:6423.
154. Holzinger A, Malle B, Kieseberg P, et al. Towards the Augmented Pathologist: Challenges of Explainable-AI in Digital Pathology. arXiv 2017;arXiv:1712.06657.
155. Jaume G, Pati P, Foncubierta-Rodriguez A, et al. Towards Explainable Graph Representations in Digital Pathology. arXiv 2020;arXiv:2007.00311.
156. Echle A, Rindtorff NT, Brinker TJ, et al. Deep learning in cancer pathology: a new generation of clinical biomarkers. Br J Cancer 2021;124:686-96.
157. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019;25:954-61.
158. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys 2019;29:102-27.

Cite this article as: Zhu J, Liu M, Li X. Progress on deep learning in digital pathology of breast cancer: a narrative review. Gland Surg 2022;11(4):751-766. doi: 10.21037/gs-22-11