Application of Artificial Intelligence Technology in Pathological Image Analysis of Breast Tissue

Mengmeng Jia 1, Xinjian Guo 2 and Fang Tian 1*

1 Department of Computer Technology and Application, Qinghai University, Xining, China
2 Department of Pathology, Affiliated Hospital of Qinghai University, Xining, China

Abstract. Breast histopathological examination usually relies on experienced pathologists. In recent years, artificial intelligence technology has been applied to digital pathological images to help doctors diagnose breast tumor. This paper mainly introduces the research progress of artificial intelligence technology in pathological classification and histological grade of breast cancer. In view of the simplification of the pathological image analysis object of breast tissue, a multimodal breast cancer aided pathological diagnosis model is proposed. The proposed model combines the patient’s immunohistochemical diagnosis data and pathological image data with artificial intelligence to analyze and diagnose. To achieve a deeper evaluation of the patient's condition on the basis of simple tumor screening, optimize the clinical treatment plan to make it more targeted. Discuss the main problems faced by artificial intelligence technology in the field of breast pathological image analysis, and prospect its development prospects.

1. Introduction
According to 2018 global cancer statistics research, the highest incidence of cancer is lung cancer and female breast cancer, both of which are tied for first place, accounting for 11.6%. Breast cancer seriously threatens women's health [1]. At present, there is still a lack of effective etiological prevention methods for breast cancer. Breast cancer screening is an important secondary prevention method for early detection, early diagnosis and early treatment of breast cancer. Pathological examination is the gold standard for the diagnosis of abnormal lesions by various screening methods [2]. At present, histopathological examination of breast tumors is usually carried out by experienced pathologists. This inspection task is a very time-consuming task and relies heavily on the experience of pathologists. With the in-depth development of personalized medicine, the workload of each medical port has increased, the ratio of medical resources is uneven, and the qualifications of doctors are uneven, which makes the complexity of pathological diagnosis work significantly increase [3].

With the rise of digital pathology and the development of Artificial Intelligence (AI) technology, the combination of pathological images and AI has emerged. This promotes the intelligent diagnosis of pathology, liberates the pathologist from the traditional microscope working environment with high intensity, low efficiency and high occupational injury, which effectively improves the diagnosis efficiency of the doctor and weakens the subjectivity and instability of the doctor’s diagnosis [4].

2. Digital pathology and artificial intelligence
Digital pathology refers to the application of computers and networks in the field of pathology. With the help of digital scanning technology, high-resolution digital images are acquired by automatic microscope scanning, and these images are automatically processed with high precision, multiple fields of view, and seamless joints, and then obtain Whole Slide Images (WSIs). Each WSI consists of hundreds of millions of image pixels, and each pixel contains XY coordinate position information, colour information, and grayscale information. WSIs are usually organized in a pyramid structure. As shown in Figure 1, the higher the level of the pyramid, the smaller the image magnification. Each layer is compressed by the data of the bottom image according to an efficient algorithm.

![Figure 1. Pyramid structure of WSIs [5].](image)

WSI technology has broad prospects in the pathology digitization, avoiding many limitations brought by the early photomicrography technology, including poor image quality, pathologists unable to view the high-resolution overview of the whole slide, and it takes a long time to fully examine the slide. WSI can achieve permanent annotation, straight-line distance measurement, perimeter and area measurement, and higher-level image analysis. At the same time, image pixels have features such as coordinates, colour, and grayscale, which can be used by computers for object segmentation and automatic quantization [6-9].

Since the birth of AI, the theory and technology have become increasingly mature, and the application field has also been expanding. The standardization and digitization of pathological slides have provided a background of big data for the development of artificial intelligence, making AI have made revolutionary progress in the field of pathological diagnosis. The scope of application is focused on benign and malignant tumor diagnosis, disease classification, survival analysis and prognosis analysis [10].

3. Research status of AI in the field of breast pathological diagnosis

3.1. Application of AI in histopathologic classification of breast cancer

The automatic classification of breast cancer pathological images is an early work. In the early stage, scholars' research [11-13] was mostly based on traditional machine learning algorithms, which required manual feature extraction methods, which not only required professional domain knowledge, but also consumed a lot of time and effort, and the most difficult thing was to extract differentiated high-quality features [14]. In this case, even if they reported better results in the study, they had less clinical significance.

In 2015, Spanhol et al. [15] published BreaKHIs, a large breast cancer pathological image data set, which greatly promoted the research of breast cancer pathological image classification. It consists of 7,909 microscopic imaging pathological images collected from 82 different patients with different
magnifications. With the rise of deep learning and the support of large data sets, scholars [16-19] gradually used deep learning algorithms for pathological image analysis, and the classification ability has been greatly improved, and the recognition rate of certain tumor subtypes even as high as 100%. On the one hand, deep learning can automatically learn features from the data, avoiding the complexity and limitations of manual design and feature extraction in traditional algorithms. On the other hand, Convolutional Neural Network (CNN) has been widely used in target detection, image recognition and other fields [20-21], and many complex model architectures have been developed one after another. At present, the research on the automatic classification of pathological images, almost all of them are implemented by Deep Convolutional Neural Networks.

At present, the research of pathological image analysis based on patch-level image has been relatively mature, but it cannot fully meet the requirements of daily pathological consultation, so the pathological analysis of WSIs is the focus of scholars in recent years. After 2016, the challenges of CAMELYON16 [22] and CAMELYON17 [23-24] were successively held. The CAMELYON dataset contains thousands of breast pathological WSIs, where the training set provides the label of the lesion area and the patient's pN stage (regional lymph node staging in the international TNM staging system). Bejnordi et al. [22] tested 32 algorithms submitted to the competition for automatic detection and classification of breast cancer metastases in WSIs of histological lymph node sections, and found that nearly 80% of the participants used deep learning algorithms based on convolutional neural networks. Bandi et al. [24] discussed the details of the 12 algorithms submitted to the competition and found that the participants used convolutional neural networks in combination with pre-processing and post-processing steps. The best results of the competition were obtained using a pre-trained model. Tian et al. [25] used three popular convolutional neural network architectures based on the information of morphological heterogeneity of tumors and their adjacent regions, and then used random forest method to achieve feature extraction and classification based on whole breast pathology images. The AUC value of the WSI-level classification model reached 0.94, but the model's performance in identifying the lesion area was moderate, and the FROC score was only 0.52. Li et al. [26] proposed a neural conditional random field tumor recognition algorithm, which greatly improved the localization accuracy of the lesion area and increased the FROC score to 0.8096.

3.2. Application of AI in histological grade of breast cancer

Early grade diagnosis of breast cancer plays an important role in predicting the degree of disease invasion and the prognosis of patients. The number of mitosis is an important indicator for evaluating the grade of breast cancer. The variability and complexity of mitosis increase the difficulty of mitotic recognition. Existing computer-aided recognition methods use either hand-made features or features learned by convolutional neural networks.

Dan et al. [27] used a deep convolutional neural network with the largest pooling layer as the detection algorithm. This method won the champion of mitosis detection in the ICPR 2012 competition. Although this method has high accuracy, it consumes too much training time because of the large model structure. Wang et al. [28] proposed a cascaded mitotic detection method, which intelligently combines the CNN model and manual features (morphology, colour, and texture features). It requires less computing resources, but the detection is more accurate and faster, and the F-score tested on ICPR2012 data set is 0.7345. Liu et al. [29] proposed a pathological image processing method based on low-rank representation, which does not require cell segmentation and model training, and the recognition process is faster and more convenient. Abdulkadir et al. [30] adopted the feature extraction method based on convolutional neural networks, reduced the dimension of the extracted high-dimensional features, and used support vector machine to achieve the final classification of cells, which also achieved good classification results. Beevi et al. [31] proposed a multi-classifier system based on deep confidence network. Compared with the existing technology, this framework has better classification performance and higher sensitivity, and is more practical for clinical application. Qi [32] proposed three mitotic detection and classification methods, which greatly
improved the classification effect while optimizing the feature extraction process, and the recognition accuracy was as high as 95.44%.

4. Multimodal breast cancer aided pathological diagnosis model

By investigating the research status of artificial intelligence in the analysis of breast pathological images, we find that most of the current research work only involves the analysis of labeled pathological images. The singleness of the research object makes the research direction narrow, so it is difficult to make breakthroughs in other application fields. Recently, some scholars [33] have tried to combine multi-omics data and pathological images, and proposed a method of breast cancer survival prediction based on multimodal data fusion. This method constructs deep neural network and convolution neural network to process omics data and pathological image data respectively, which greatly improves the performance of breast cancer survival prediction.

In routine pathological diagnosis of tumors, it is difficult to make an accurate diagnosis only by conventional paraffin sections in some cases. As an irreplaceable auxiliary diagnostic technique in routine pathological diagnosis, immunohistochemistry technique plays an important role in judging the benign and malignant tumors, ensuring the accuracy of pathological diagnosis and guiding the use of tumor drugs. By observing the tumor tissue morphology and combining with specific antibodies expressed by tumor cells, the pathologist can diagnose and identify most of the conventional types of breast cancer. In summary, we propose a computer-aided diagnosis model CDMIP (Computer-aided Diagnostic Model by Integrating Immunohistochemical Diagnostic Data and Pathological Images, CDMIP), which combines immunohistochemical diagnosis data and pathological image data, is shown in figure 2. The model fully considers the difference between the two kinds of data, respectively adopts different methods for data processing and feature extraction, fuses features and uses appropriate machine learning classifiers, and the comprehensive results of pathological classification, histological grade and pathological stage of breast tumor were obtained finally.

![Figure 2. Schematic diagram of CDMIP.](image)

The feature extraction process for pathological image data is complicated. The diagnosis of the degree of differentiation of breast cancer mainly involves nuclear segmentation and feature extraction. For the pathological classification of breast cancer, the following three steps are involved: (1) Pre-processing the labeled pathological images. First, the appropriate segmentation algorithm is used to extract breast tissue, and the image patches are extracted from the normal tissue area and the tumor tissue area respectively, and the data enhancement method is used to expand the number of training
samples to balance the positive and negative sample ratio and enhance the robustness of the algorithm. (2) Design the neural network algorithm architecture. The classification performance of the algorithm directly affects the morphological features of the diseased tissue. Therefore, timely adjustment of the network structure and parameters according to the learning situation of the algorithm is very important to improve the accuracy of the algorithm. (3) Extract features of tumor tissue. According to the classification results, the lesion area is automatically located, and the relevant features such as the shape, texture, and spatial layout of the tumor tissue are extracted, such as the proportion of the lesion area, the number of lesion areas, the maximum radius of the tumor and so on. For immunohistochemical diagnosis data, the expression results of immunohistochemical markers of breast cancer were pre-processed directly, such as missing value processing, data transformation and attribute creation, and then the data features which are meaningful to the diagnostic results were obtained.

Then, the different features extracted above are fused to form the tumor feature data set. The data set combines the characteristics of tumor expression at image level and protein level, and it contains information such as the size of the tumor, the number of lymph node metastases, the malignant degree of cancer cells, the degree of differentiation, and the way of invasion, which can meet the needs of daily pathological diagnosis of breast cancer. Finally, select an appropriate classifier, predict the risk of recurrence and metastasis of the patient and formulate a more targeted treatment plan for the patient according to the comprehensive performance of the tumor tissue in pathological classification, clinical stage, molecular subtype classification and so on.

The computer-aided diagnosis model CDMIP that fuses immunohistochemical diagnosis data and pathological image data proposed in this study, through the information fusion of multimodal data, comprehensively assess the patient's condition. On the basis of simple tumor screening, further exploring the molecular subtypes of breast cancer can improve the accuracy of cancer recognition, especially the recognition ability of complex or rare cases, which is of great significance in guiding the clinical treatment and judging the prognosis.

5. Summary and prospect
The wide application of artificial intelligence technology in pathological analysis greatly lightens the burden of pathologists, weakens the subjectivity and instability of diagnosis, and greatly improves the diagnostic efficiency of doctors. At the same time, on the premise of ensuring accuracy, computer-aided technology can also provide personalized treatment suggestions and disease prognosis judgment for patients.

At present, the application of artificial intelligence technology in the field of breast tissue pathological image classification and grading is more mature, in addition, this technology has also been successfully applied in the fields of patient survival analysis and prognosis analysis. Although computer-aided diagnosis has gradually been widely used in the field of pathology, there are still many problems that need to be solved to achieve a technical breakthrough and truly popularize into clinical practice: First, at the current stage, scholars focus on the improvement of the classification model, and generally get high scores on the selected test set, but due to the lack of a very large public database, the generalization ability of the model has yet to be verified. Second, because pathological diagnosis is the gold standard for breast cancer diagnosis, higher diagnostic accuracy is required. The study found that the model's recognition rate for common cancer types can reach almost 100%, but the sensitivity for pathological slides containing rare and mixed cancer types is not high. It is possible to improve the accuracy and specificity of diagnosis by finely labeling some slides or directly adding difficult samples to the database to make the training samples more comprehensive and diverse. Third, in addition to immunohistochemical diagnostic data, it is also possible to consider combining clinical case data or other breast screening image data for multimodal joint diagnosis, reducing the misdiagnosis rate and missed diagnosis rate from multiple angles, and further improving the diagnostic performance of the model.
The technological innovation brought about by the rapid development of intelligent algorithms in the era of big data makes the application of artificial intelligence technology in the field of medical better and better. The lack of professional pathologists and the increase in workload provide a good opportunity for the application and development of computer-aided technology in breast histopathology. I believe that with the deepening and close cooperation between algorithm scientists and pathologists, it will greatly promote the rapid progress of intelligent pathological diagnosis.

Acknowledgments
The work is supported by the foundational application research of Qinghai Province Science and Technology Fund (No. 2016-ZJ-743) and also supported by the Qinghai Provincial Computing Test Center Project (No. k151949).

References
[1] Freddie B, Jacques F, Isabelle S, et al. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 2018.
[2] Mo Miao, Liu Guangyu, et al. Advances in breast cancer screening research. Tumor, 2012, 32 (09): 748-754.
[3] Chen Yao, Lyu Qing. Research status of artificial intelligence in breast cancer screening and diagnosis. Chinese Journal of Bases and Clinics in General Surgery, 2019, 26 (05): 119-124
[4] Deng Yang, Bao Ji. Prospects of computer-aided diagnosis research in digital pathology. Practical Journal of Clinical Medicine, 2017, 14 (05): 10-12.
[5] The CAMELYON16 Challenge; 2017. https://camelyon16.grand-challenge.org. Accessed 13 November 2017.
[6] Zheng Zhongxi. Embracing the era of digital pathology. Practical Journal of Clinical Medicine, 2017 (05): 14-17.
[7] Ho J, Parwani A V, Jukic D M, et al. Use of whole slide imaging in surgical pathology quality assurance: design and pilot validation studies. Human Pathology, 2006, 37(3):0-331.
[8] Pantanowitz, Liron. Digital images and the future of digital pathology. Journal of Pathology Informatics, 2010, 1(1):15.
[9] Ghaznavi F, Evans A, Madabhushi A, et al. Digital Imaging in Pathology: Whole-Slide Imaging and Beyond. Annual Review of Pathology: Mechanisms of Disease, 2013, 8(1):331-359.
[10] Yu Guanzhen, Wei Peilian, Chen Ying, et al. Application and thinking of artificial intelligence in tumor pathology diagnosis and evaluation. Academic Journal of Second Military Medical University, 2017 (11): 6-11.
[11] Kowal M, Filipczuk P, Obuchowicz A, et al. Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. Computers in Biology and Medicine, 2013,43(10):1563-1572.
[12] Zhang Y, Zhang B, Coenen F, et al. One-class kernel subspace ensemble for medical image classification. EURASIP Journal on Advances in Signal Processing, 2014,2014(1):17.
[13] Wang P, Hu X, Li Y, et al. Automatic cell nuclei segmentation and classification of breast cancer histopathology images. Signal Processing, 2016,122:1-13.
[14] He Xueying, Han Zhongyi, Wei Benzheng. Automatic classification of breast cancer pathological images based on deep learning. Computer Engineering and Applications, 2018, 054 (012): 121-125.
[15] Spanhol F A, Oliveira L S, Petitjean C, et al. A dataset for breast cancer histopathological image classification. IEEE Transactions on Biomedical Engineering, 2016,63(7):1455-1462.
[16] Spanhol F A, Oliveira L S, Petitjean C, et al. Breast cancer histopathological image classification using Convolutional Neural Networks. 2016 International Joint Conference on Neural Networks (IJCNN), 2016:2560-2567.
[17] Bayramoglu N, Kannala J, Heikkilä J. Deep learning for magnification independent breast cancer histopathology image classification. International Conference on Pattern Recognition. IEEE, 2017:2440-2445.

[18] Ling Yu, Sun Ziqiang. Breast pathology image recognition algorithm based on convolutional neural network. Journal of Jiangsu University (Natural Science Edition), 2019 (5).

[19] Liu Qiaoli. Research on Automatic Diagnosis of Diseases Based on Machine Learning. Yantai University, 2019.

[20] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015,521(7553):436-444.

[21] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:770-778.

[22] 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(22):2199–2210.

[23] Geert L, Peter B, Ehteshami B B, et al. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. GigaScience, 2018.

[24] 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 Transactions on Medical Imaging. 2018.

[25] Tian Y, Yang L, Wang W, et al. Computer-aided Detection of Squamous Carcinoma of the Cervix in Whole Slide Images. 2019.

[26] Li Y, Ping W. Cancer Metastasis Detection with Neural Conditional Random Field. 2018.

[27] Cirean D C, Giusti A, Gambardella L M, et al. Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. International Conference on Medical Image Computing and Computer-Assisted Intervention. 2013.

[28] Wang, H. et al. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. Journal of Medical Imaging.2014,1(3):034003.

[29] Liu Ya, Xu Jun, Hang Renlong et al. Breast cancer pathological image mitosis detection based on low rank representation. Application Research of Computers, 2015, 32 (1): 280-284.

[30] Albayrak A, Bilgin G. Mitosis detection using convolutional neural network based features.2016 IEEE 17th International Symposium on Computational Intelligence and Informatics (CINTI). IEEE, 2016.

[31] Beevi K S, Nair M S, Bindu G R. A Multi-Classifier System for Automatic Mitosis Detection in Breast Histopathology Images Using Deep Belief Networks. IEEE Journal of Translational Engineering in Health &Medicine, 2017, 5(99):1-11.

[32] Qi Ying. Study on Mitosis Recognition in Breast Histopathology Images. Guilin University of Electronic Technology, 2018.

[33] Dongdong Sun. Study of breast cancer survival prediction with multimodal data fusion. University of Science and Technology of China, 2018.