frontiers in Oncology

REVIEW
published: 10 November 2021
doi: 10.3389/fonc.2021.763527

Evaluating Cancer-Related Biomarkers Based on Pathological Images: A Systematic Review

Xiaoliang Xie[1,2†], Xulin Wang[3†], Yuebin Liang[4,5], Jingya Yang[4,5,6], Yan Wu[4,5], Li Li[7], Xin Sun[8], Pingping Bing[9], Binsheng He[9], Geng Tian[4,5,10] and Xiaoli Shi[4,5*]*

[1] Department of Colorectal Surgery, General Hospital of Ningxia Medical University, Yinchuan, China, [2] College of Clinical Medicine, Ningxia Medical University, Yinchuan, China, [3] Department of Oncology Surgery, Central Hospital of Jia Mu Si City, Jia Mu Si, China, [4] Geneis Beijing Co., Ltd., Beijing, China, [5] Qingdao Geneis Institute of Big Data Mining and Precision Medicine, Qingdao, China, [6] School of Electrical and Information Engineering, Anhui University of Technology, Ma'anshan, China, [7] Beijing Shanghe Jiye Biotech Co., Ltd., Bejing, China, [8] Department of Medical Affairs, Central Hospital of Jia Mu Si City, Jia Mu Si, China, [9] Academician Workstation, Changsha Medical University, Changsha, China, [10] IBMC-BGI Center, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, China

OPEN ACCESS

Edited by:
Min Tang,
Jiangsu University, China

Reviewed by:
Min Chen,
Hunan Institute of Technology, China
Li Peng,
Hunan University of Science and Technology, China

***Correspondence:**
Geng Tian
tiang@geneis.cn
Xiaoli Shi
shixl@geneis.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *24 August 2021*
Accepted: *18 October 2021*
Published: *10 November 2021*

Citation:
Xie X, Wang X, Liang Y, Yang J, Wu Y, Li L, Sun X, Bing P, He B, Tian G and Shi X (2021) Evaluating Cancer-Related Biomarkers Based on Pathological Images: A Systematic Review. Front. Oncol. 11:763527. doi: 10.3389/fonc.2021.763527

Many diseases are accompanied by changes in certain biochemical indicators called biomarkers in cells or tissues. A variety of biomarkers, including proteins, nucleic acids, antibodies, and peptides, have been identified. Tumor biomarkers have been widely used in cancer risk assessment, early screening, diagnosis, prognosis, treatment, and progression monitoring. For example, the number of circulating tumor cell (CTC) is a prognostic indicator of breast cancer overall survival, and tumor mutation burden (TMB) can be used to predict the efficacy of immune checkpoint inhibitors. Currently, clinical methods such as polymerase chain reaction (PCR) and next generation sequencing (NGS) are mainly adopted to evaluate these biomarkers, which are time-consuming and expansive. Pathological image analysis is an essential tool in medical research, disease diagnosis and treatment, functioning by extracting important physiological and pathological information or knowledge from medical images. Recently, deep learning-based analysis on pathological images and morphology to predict tumor biomarkers has attracted great attention from both medical image and machine learning communities, as this combination not only reduces the burden on pathologists but also saves high costs and time. Therefore, it is necessary to summarize the current process of processing pathological images and key steps and methods used in each process, including: (1) pre-processing of pathological images, (2) image segmentation, (3) feature extraction, and (4) feature model construction. This will help people choose better and more appropriate medical image processing methods when predicting tumor biomarkers.

Keywords: histopathological image analysis, cancer biomarker, deep learning, color normalization, feature extraction

INTRODUCTION

Biomarkers are critical in cancer diagnosis, treatment, and prognosis. They can be used for patient's evaluation in a variety of clinical settings, such as risk assessment, early diagnosis, drug effect evaluation, and prognosis prediction (1–3). With the development of immunology, molecular biology and genomics, studies of cancer biomarkers have attracted a lot of attention in recent years (4). Currently, biomarker identification usually employs technologies such as PCR, NGS and gene expression arrays (5). However, the data generated by these technologies need to be analyzed and interpreted manually. In addition, this kind of test usually costs a lot of money. For example, the test of tumor mutation burden (TMB) usually costs more than one thousand dollars. Thus, it will be of great value to develop a more intelligent and economical method in tumor biomarker identification (6).

Pathological image analysis is used to solve problems related to medical images which were applied in biomedical research and diagnosis. Its main objective is to extract clinically relevant physiological and pathological information or knowledge from images, and its main research direction is image segmentation, classification, and retrieval (7). With the rapid development and popularization of medical imaging technology, the amount of medical image data is growing rapidly. It will provide important and beneficial support for nursing and medical research to extract useful knowledge and information automatically from massive medical image data for clinical diagnosis and treatment (8). Recently, researchers have paid much attention to the analysis and study of tumor patients through pathological images and morphology (9). Mobadersany (10) proposed that the morphological characteristics of tumor tissue images could reflect the genetic and molecular characteristics and predict the degree of tumor deterioration, and the deep learning method could be used to integrate the morphological characteristics of tumor tissue images and genomics to predict the survival rate of glioma patients. Xu (11) proposed a method based on deep tissue network to automatically distinguish 10 tissue components in the colorectal full-scan tissue image. Yu (12) for the first time constructed the recurrence risk prediction model of LUAD and LUSC by automatically extracting morphological features from the full-scan histopathological images of lung cancer to provide prognostic information for patients. Vaidya (13) proposed to combine radiology and pathology to predict the risk of early lung cancer recurrence, with an accuracy rate of 70%. Wu (14) and others constructed a deep convolutional neural network framework to evaluate the risk of lung cancer recurrence and metastasis from histopathology images, with the area under the receiver operating characteristic (ROC) curve (AUC) in the test dataset of 0.79. Jain and Massoud explored a machine learning algorithm named Image2TMB to predict TMB from lung adenocarcinoma histopathological images. Its average precision was 0.89 and achieved predictive level of a panel of ~100 genes. Microsatellite instability (MSI) was another immunotherapy biomarker (15) which requires additional immunohistochemical or genetic analyses in clinical practice (16). Kather et al. developed a deep residual learning method that can predict MSI status directly from hematoxylin and eosin (H&E) stained histology slides (17). These findings suggest that inferring genomic features from histopathological images is possible and analyzing histopathological images is important for studying cancer treatments, mutated gene expression status, cancer prognosis and risk of recurrence.

However, full-scan histopathological images are highly complex, with large image size and about 2 GB of storage space after compression. It is a big challenge for hardware and image analysis algorithm to use computer to process image directly in this kind of high resolution and large size image. At the same time, the histopathological structure types in the images are disordered, and the histological morphology is very different, so it is difficult to describe with fixed features. All these factors bring great difficulty to the processing of full scan histopathological images. Based on the above problems, this paper summarizes the whole process and key steps of current pathological image processing, including image preprocessing, image segmentation, feature extraction and model construction, to help researchers choose more suitable medical image processing methods and predict biomarkers more accurately. We summarized the overall flow of pathological image processing in **Figure 1**.

IMAGE PREPROCESSING

The biggest obstacle to histopathological image analysis is the difference in image morphology due to high heterogeneity of the disease itself. At the same time, improper tissue treatment or staining during the slice preparation will result in morphological changes of cells and tissues, making it difficult to identify its original structure. In addition, the background noise and the lack of contrast caused by the different light source conditions were also important factors. Proper preprocessing method can correct images by eliminating irrelevant information, and filter out interference and noise, which can improve the detectability of target information and simplify the calculation to the maximum extent.

Common preprocessing methods such as using spatial filtering techniques to enhance the main structure in the image, image enhancement can improve the contrast between the region of interest and the background, and color normalization can reduce the effect of staining batches (18, 19). Among these, color normalization is the most commonly used image preprocessing methods for evaluating cancer-related biomarkers based on histopathological images.

Color Normalization

In response to the problem of color change, Reinhard (20) and others proposed a method of color normalization, that is, in the lαβ color space, the mean and standard deviation of each channel in the image are compared with the target by a set of linear transformations. Then, match the mean and standard deviation. However, if multiple patches are used, the assumption of a unimodal distribution of pixels in each channel of the lαβ color space is not valid. Therefore, this may cause the background area to be mapped as a colored area and

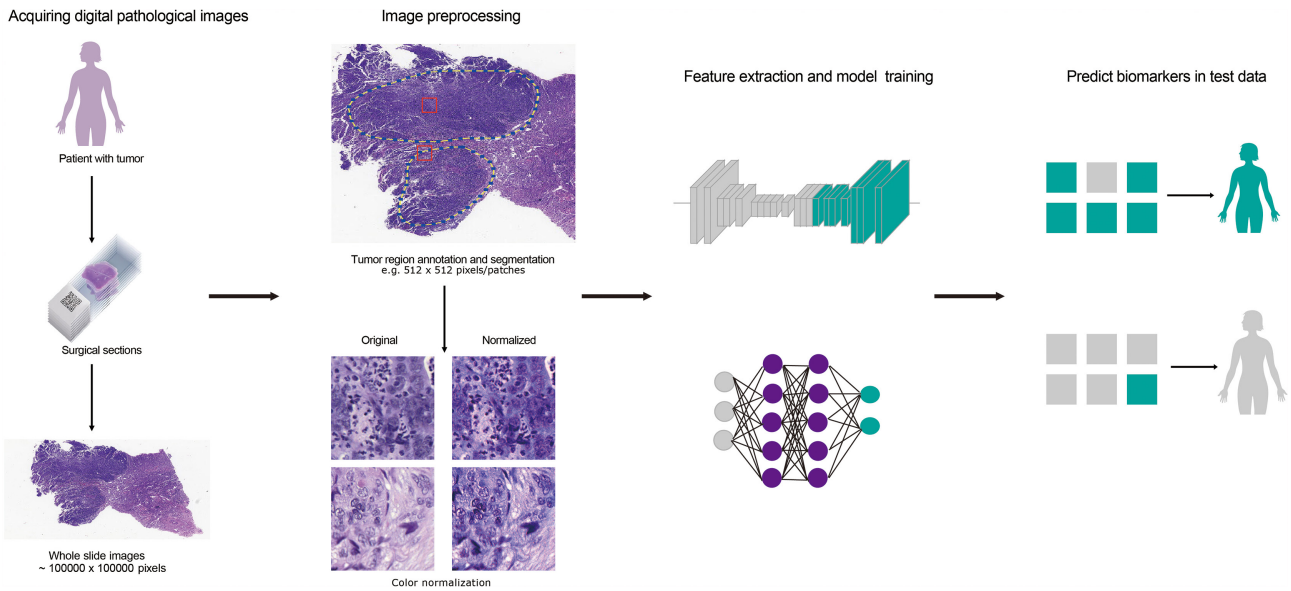

FIGURE 1 | The flow chart for predicting cancer-related biomarkers based on digital pathological images. Firstly, H&E stained histology slides of patients were obtained and whole slide images (WSIs) was obtained after scanning. Secondly, tumor regions were annotated by pathologists or through CNN model. Then, the regions were segmented to patches and color-normalized. Thirdly, feature extraction and model training were carried out according to biomarker labels. Finally, biomarker prediction was implemented in test dataset.

the foreground to be incorrectly mapped. As shown in **Table 1** below, some methods of color normalization were summarized.

IMAGE SEGMENTATION

Medical image segmentation is a complex and critical step in the field of medical image processing and analysis. The purpose of this process is to segment certain parts of the medical image with specific meaning, extract relevant features, and then provide reliable information for clinical diagnosis and pathological research. The two most common types of medical image segmentation are tissue segmentation and cell segmentation.

Tissue Segmentation

Pathologists have identified that degree of structural differentiation of the tissue is one of the earliest prognostic factors for breast cancer patients. Cancer destroys the ability of the nucleus to communicate with each other and causes it to organize itself into structures such as tubules, thereby making the tubules lack of indicators of advanced malignant tumors. Tubules are usually round or oval in structure and consist of a lumen surrounded by a layer of epithelial cells. The main challenge of tubule segmentation is that it has a similar appearance to other structures, such as the tearing of adipose tissue formed during tissue preparation, and the outer layer of well-arranged epithelial cell with nuclei missing.

TABLE 1 | A summary of color normalization methods.

Authors	Methods	Characteristics	References
Magee	A method based on supervised pixel classification	Estimate the color of the coloring.	(21)
Macenko	A method based on singular value decomposition (SVD)	Direct estimation matrix.	(22)
Niethammer	An improved method based on singular value decomposition (SVD)	By expanding (22), a priori estimation of staining matrix is used to improve the stability of each staining.	(23)
Khan	Nonlinear mapping based on source image to target image	An improvement is proposed on the method of (21), using the representation method of color deconvolution.	(18)
Vahadane	A technique of dye separation and color normalization (SPCN)	It does a good job of maintaining the quality of biological structure and the number of stains.	(24)
Ramakrishnan	The improved SPCN	In the SPCN technology, some improvements are proposed for the occasional errors in estimating color bases, which lead to artifacts.	(25)

For glandular segmentation, most of the early attempts used hand-made features for segmentation. Wu (26) identified the initial seed region based on large cavity regions and extended the seed to the surrounding epithelial nuclear chain. Farjam (27) proposed using a variance filter to compute cluster texture features for segmentation. However, robust segmentation requires more domain knowledge, and calculating texture features only using the variance filter may not provide sufficient information for the local structure of the organization. Naik (28) used a Bayesian classifier to detect the lumen region, and then used the kernel-based level set to stop the curve and refine it. Although this method has been reported to work well in benign cases, it may fail in malignant cases with fairly complex glands. Nguyen (29) used color space analysis to group the nucleus, cytoplasm, and lumen, and increased the lumen area to achieve segmentation under constraints. Gunduz-Demir (30) represented each tissue component as a disc and connected nearby discs with an edge to construct a graph. They performed area growth on a cavity disc bounded by a line connected to the nuclear disc. Nosrati, Hamarneh (31) and Cohen (32) first divided the tissue area into different components, and then used a constrained level set algorithm to segment the glands. Sirinukunwattana (33) identified epithelial superpixels and used the epithelial region as the vertex of a polygon, which approximated the boundary of a gland. Most of the methods discussed above first distinguish tissue regions and then use region growth or level sets to segment glandular regions. Recently, a slightly different approach that first used background information to identify potential epithelial regions, and then used multi-resolution cell localization descriptors to identify connected epithelial cells to segment glands was proposed by Li (34).

Cell Segmentation

The morphology of cells in histopathological images provides important information for the diagnosis and prognosis of cancer. Researchers at home and abroad have tried a variety of algorithms to solve the problem of cell segmentation in H&E images (34–36). The algorithms generally divided into two categories, one is to detect single cells accurately and the other is to segment cells. The algorithms in **Table 2** is commonly used to detect the appropriate seed point or contour of the nucleus.

The other type detects the candidate area of the cell and then divides it into individual nuclei. The first step in morphological analysis of a cell is the segmentation of individual nuclei, which is usually performed manually in current clinical practice. However, due to the large volume of histopathological images and complex cell structures, manual examination is a time-consuming and labor-intensive task. It is necessary to study computerized methods to reduce the workload of pathologists and improve the analysis efficiency (45). Nuclear segmentation tasks still have some major challenges. First, different types of organs or cells are highly heterogeneous in appearance. Therefore, the method based on prior knowledge of geometric features cannot be directly applied to different images. Second, some other structures, such as the cytoplasm and cell matrix, may have similar characteristics to the nucleus, making it difficult to distinguish the nucleus from the background. Third, the cells are often stacked together. In order to find the exact location and boundary of each nucleus, it is usually necessary to perform the next step to separate the clustered or overlapped nuclei.

In view of the importance of nuclear distribution and morphology, the task of using computer algorithms for accurate nuclear segmentation provides a logical starting point for computer-aided tissue image analysis. The precise segmentation of the nucleus can not only perform deeper level feature extraction and classification in the nucleus, but also serve as a relatively simple distribution of basal cells and acellular cells. Many techniques have been applied to the task of nuclear segmentation, but in some cases they have only achieved partial success. For example, the intensity threshold method usually fails due to image noise and nucleus clustering. Label-based watershed segmentation requires accurate parameter selection, while the computational cost of active contours and deformable models is too high (24, 42, 46–50). Machine learning-based kernel segmentation methods are generally better at meeting these challenges because they can learn to recognize changes in nuclear morphology and staining patterns. More precisely, convolutional neural networks (CNNs) have recently demonstrated their latest performance in kernel segmentation (51, 52). Ciregan (53) applied deep CNN to the automatic detection of mitotic cells in breast cancer histological images. Using the original intensity of the test image, CNN provides a probability map where each pixel value is the probability of the mitotic cell centroid. Then using the disk to check the probability map for smoothing, and non-maximum suppression to get the final centroid detection. Xing (54) and others respectively learned three different CNN models corresponding to pathological images of brain tumors, pancreatic neuroendocrine tumors, and breast cancer, and

TABLE 2 | A summary of methods on segmentation after detection of individual cells.

Methods	Characteristics	References
Based on different voting rules	Simple and suitable for segmentation of most images	(11, 37–39)
Based on Laplace operator and gaussian filter	Accurately detect the edge of the cell	(40)
Based on H-minima transformation	Effectively restrain oversegmentation and reduce undersegmentation	(41)
Based on Morphologic manipulation	Could output an image by acting a structure element on the input image	(42, 43)
Based on back propagation with MRF	Good at dealing with the problems of image local volume and artifacts	(34)
Based on the active contour model	Could convert pixels to a distance field	(43)
Based on the level set	A numerical method based on the theory of geometric active contour model	(37, 44)

applied them to automatic nuclear detection. Liu and Yang (55) did not use simple non-maximum suppression to refine the detection, but converted the detection problems of pancreatic neuroendocrine and lung cancer cell nuclear into optimization problems. Xing (47), Sirinukunwattana (51) and Song (55) have proposed some advanced techniques in nuclear detection and segmentation, which estimate the probability of nuclear and non-nuclear regions (both types) based on the learned nuclear phenomena graphs and rely on complex post-processing methods to obtain the final core shape and the separation between contacting nuclei. Song et al. used a graph partitioning method (55) and Xing et al. used a kernel mapping distance transformation, followed by H-minima thresholding and region growth (47). Although different methods have been developed for the problem of overlapping and clustering nuclei in many literatures, and have achieved varying degrees of success, this problem has not been completely solved, as there is a large amount of overlap contact specimens of nuclei.

In addition, a special type of nucleus, mitosis, has attracted much attention in the field of image analysis. Mainly because the mitotic index is used to evaluate the cell proliferation rate of cancer cells, it could predict the prognosis of invasive breast cancer well, but its evaluation process is extremely time-consuming (56). On the H&E image, mitosis has specific morphological features: dense nuclear staining, enlarged nuclei, less clear nuclear membrane, and burr-like edges. Researchers such as Belien (57) proposed image processing technology for semi-automatic segmentation of mitotic images in the 1990s. Due to the limitations of the image quality and machine learning algorithms at the time, the algorithm proposed by Belien et al. (57) required fourgen staining to display chromosomes, and the false positive rate is 19-42%. With the digitization of pathological images, two H&E tissue datasets of breast cancer have been published internationally, and pathological experts have annotated mitotic images in the images, which has greatly promoted the development of algorithms in mitotic image segmentation. Then, the International Conference on Pattern Recognition (ICPR) (58) held a competition for mitotic detection in breast cancer tissue images in 2012, providing different types of images, allowing participants to analyze classic images of H&E stained sections, and use 10 bands multispectral microscope images, which may be more discriminatory for detecting mitosis. Deep learning maximizing CNN significantly outperforms other manual feature-based methods and paves the way for future use of CNNs (53).

The biggest challenge for mitosis detection is that apoptosis, necrosis or squeezed nuclei and lymphocyte nuclei have similar morphology to mitosis, which is difficult for even experienced pathologists to identify. In addition, pathologists need to observe suspicious split images on multiple focal planes, while currently digital images are single focal plane imaging. Although some scanners can acquire multifocal plane images, their storage capacity is large and cannot be widely used. We expect that in the future, as storage costs decrease and new image compression technologies emerge, this limitation will be eliminated (59). Therefore, the automatic segmentation of mitotic images in H&E images at this stage is more challenging than general nuclear segmentation and is far from being applicable to pathological work.

MODEL CONSTRUCTION

After the ideal segmentation results were obtained from the tissue segmentation and nuclear segmentation modules, the morphological features of histopathological images were extracted, and the correlation between the morphological features and biomarkers of the full-scan histopathological images was found and the feature model was established.

Beck et al. constructed a computer pathologist system to extract 6,642 dimensional features from H&E histopathological images of breast cancer (60). Some of the features are based on the existing knowledge system, such as the formation degree of counting glandular tube after automatic segmentation (61) and automatic grading (62), but most of the features go beyond the existing descriptive semantics of pathology. Computer-aided diagnosis is also based on the prognosis of characteristic models, modeling based on object characteristics, and then estimating the prognosis of model parameters. Tutac (63) proposed a semi-automatic grading system based on knowledge model for the first time, which automatically detected and measured the three components of histological grading, namely nucleus, adenotuine and mitosis, through semantic retrieval. The consistency of the scoring results of this model was higher than that of manual evaluation. Dalle (64) further improved the above work based on multi-resolution method and Gaussian model function, realized automatic histological classification, and the automatic classification results were highly consistent with the manual evaluation results.

Pathology is morphology-based, but the classification and assessment of disease is not limited to morphology, and requires reference to immunological, molecular, and clinical characteristics of patients. Based on the genome, Wang (65) mined prognostic features in H&E histopathological images of triple negative breast cancer (TNBC), and selected 48 pairs of significantly correlated image features and gene clusters through the TNBC genome map and H&E images of 44 cases, among which 4 pairs were significantly correlated with prognosis. Basavanhally (66) showed that H&E morphological characteristics and IHC molecular characteristics can replace expensive Oncotype DX risk assessment for the invasiveness of ER negative breast cancer. Yuan (67) proposed a mathematical statistical model to evaluate the proportion of lymphocytes in TNBC tumors, and the results showed that lymphocytes were related to the survival of TNBC, and the image-based evaluation results were similar to the results of gene expression spectrum detection. According to the prognostic model theory of Steyerberg (61), we can further utilize the results of image characteristics and molecular characteristics, and construct a prediction model by integrating complementary prognostic factors, which can be used to comprehensively and accurately predict the prognosis of breast cancer. Currently, integrating information from different dimensions to construct multimodal

fusion models for predicting cancer biomarkers or prognosis of patients have been studied in several laboratories. The main process of building multimodal fusion models is shown in **Figure 2**. Making full use of multidimensional information for fusion modeling is of great help to improve the prediction accuracy, which will also be a direction of the development of digital pathology. Chen et al. used CNNs and GCNs to extract morphological features from digital histology images and SNNs to extract genomic signatures (68). Then they employed the Kronecker Product and a gating-based attention mechanism to fuse these deep features and further validated the approach on glioma and clear cell renal cell carcinoma (CCRCC) data from TCGA. Mobadersany et al. presented a novel method to predict outcomes of patients from histopathological images and proved that the accuracy was comparable to the traditional manual histological grading. To further improve performance, they combined histopathological images and genomic data to develop a comprehensive model called GSCNN. And its performance was significantly better than that of SCNN model and WHO paradigm based on genomic subtype and histological grading (69).

LIMITATIONS AND FUTURE WORK

Cancer histology contains rich phenotypic information and can reflect underlying molecular mechanisms and disease progression. A large number of studies have shown that deep learning of digital pathological images of tumor tissue samples can be used for cancer diagnosis, classification, drug efficacy evaluation and prognosis prediction. This method has the advantage of fast and low cost. In this work, we summarized the overall process and key steps of processing full-scan section images to help people choose better and more appropriate medical image processing methods when predicting tumor biomarkers.

However, the application of artificial intelligence (AI) technology in precision medicine has some limitations currently. Firstly, the diagnosis process of deep learning model is fuzzy and the interpretability is limited, and the lack of interpretability is unacceptable to the Medical Association (70–72). So this problem is an important obstacle to its verification and application in clinical practice. Heat map analysis provides an in-depth understanding of the histological patterns related to the prediction target, which is helpful for the interpretation of the deep learning model. Chen et al. had used this method to locate and interpret features in the study of multimodal fusion for predicting survival outcome of cancer patients (68). It can also be used as a practical tool to lead pathologists to discover the tissue regions related to biomarkers. For example, the presence of edema in glioma was not previously considered as an adverse marker by pathologists, but was detected as a recognition feature in the model of predicting cancer prognosis (69). Associated with this finding, the degree of edema may be correlated to the growth rate of cancer in previous study (73). Cao et al. verified the reliability of the deep learning model in two independent cohorts when predicting MSI with pathological images, and explained the interpretability of the model by exploring the correlation between pathological features and multi-omics signatures. This is also a method to promote clinicians to accept the application of AI in digital pathological images (74). It can be predicted that improving the interpretability of the model or establishing interpretable machine learning methods will be an important topic to be explored in the future.

Secondly, a substantive problem limiting its clinical application is the frequent workflow switching due to the limited integration of computer-aided pathological diagnosis in the current pathological workflow (70–72). Currently, the

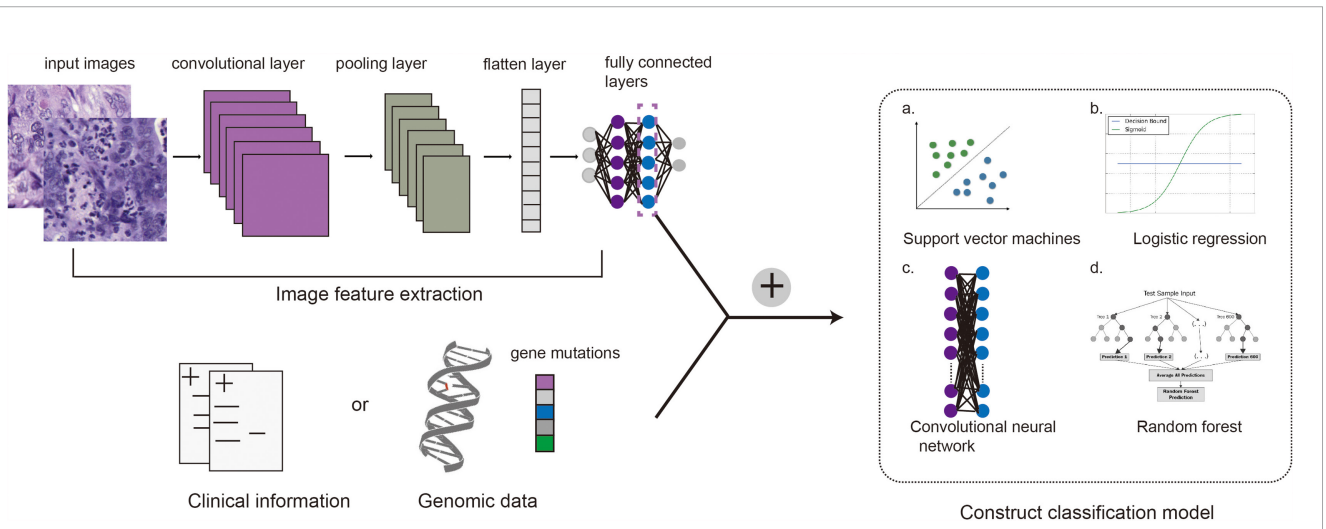

FIGURE 2 | Main process of constructing a compound framework by combining pathological images with genomic data or clinical information. Convolutional neural networks are commonly used to extract image features, and then genomic features or clinical information are integrated into the full connection layer. Support vector machine (a), logistic regression (b), convolutional neural network (c) or random forest (d) can be used to establish the final multimodal fusion model.

research on diagnosis and subtyping of cancer through digital pathological images is relatively mature. Some latest studies on predicting cancer prognosis, treatment response and disease progress monitoring through pathological images have been reported. Kather et al. developed a deep learning model that can directly predict microsatellite instability from H&E histological images of stomach and colorectal cancer and the AUC values ranged from 0.69 to 0.84 in independent validation datasets (17). Cao et al. explored an EPLA model with AUC of 0.8504 (95% CI: 0.7591-0.9323) in the external validation set (74). However, more histological images of patients are needed to optimize the model and improve accuracy. If a complete pathological diagnosis and prediction process through extensive analysis of various data can be established and verified clinically, it will contribute to the application of AI in precision medicine (71, 75).

Thirdly, it is difficult to unify the staining and imaging process of tissue section in different laboratories, which leads to a large number of variables in pathological images and further makes it difficult to establish models with high stability and good generalization performance. Just as molecular diagnosis relies on qualified samples and sequencing data, digital image analysis also requires strict control of sample quality, clear quality requirements for input files, and adequate training for pathologists. These requirements of digital pathological image analysis will also drive to improve the volume and accuracy of histomorphological evaluation. On the other hand, in order to promote clinical transformation, a roadmap and regulatory framework for the routine use of AI in pathology have been published (76).

Other literatures also list possible practical problems: slow implementation time of computer-aided pathology, insufficient clinical validation of computer-aided pathology, and limited impact on health economics (9, 71). The ability to overcome these limitations will determine the future of digital pathology.

AUTHOR CONTRIBUTIONS

GT and XShi designed the project. XX, XW, and XShi searched literatures and wrote the manuscript. YL, JY, YW, LL, XSun, PB, and BH revised the manuscript. All authors have approved the final version of the manuscript.

FUNDING

This study was supported by Natural Science Foundation of Hunan, China (Grant No. 2018JJ3570), Major Project for New Generation of AI (Grant No. 2018AAA0100400), the National Natural Science Foundation of Hunan (Grant Nos. 2018JJ2098), the National Natural Science Foundation of China (Grant No. 11571052, 11731012).

REFERENCES

1. Henry NL, Hayes DF. Cancer Biomarkers. *Mol Oncol* (2012) 6(2):140–6. doi: 10.1016/j.molonc.2012.01.010
2. Liu H, Qiu C, Wang B, Bing P, Tian G, Zhang X, et al. Evaluating DNA Methylation, Gene Expression, Somatic Mutation, and Their Combinations in Inferring Tumor Tissue-Of-Origin. *Front Cell Dev Biol* (2021) 9:619330. doi: 10.3389/fcell.2021.619330
3. He B, Lang J, Wang B, Liu X, Lu Q, He J, et al. TOOme: A Novel Computational Framework to Infer Cancer Tissue-of-Origin by Integrating Both Gene Mutation and Expression. *Front Bioeng Biotechnol* (2020) 8:394. doi: 10.3389/fbioe.2020.00394
4. Zhang Y, Huang H, Zhang D, Qiu J, Zhang J, Wang K, et al. A Review on Recent Computational Methods for Predicting Noncoding RNAs. *BioMed Res Int* (2017) 2017:9139504. doi: 10.1155/2017/9139504
5. Yang J, Hui Y, Zhang Y, Zhang M, Ji B, Tian G, et al. Application of Circulating Tumor DNA as a Biomarker for Non-Small Cell Lung Cancer. *Front Oncol* (2021) 11:725938. doi: 10.3389/fonc.2021.725938
6. Yang J, Peng S, Zhang B, Houten S, Schadt E, Zhu J, et al. Human Geroprotector Discovery by Targeting the Converging Subnetworks of Aging and Age-Related Diseases. *Geroscience* (2020) 42(1):353–72. doi: 10.1007/s11357-019-00106-x
7. Ma X, Xi B, Zhang Y, Zhu L, Sui X, Tian G, et al. A Machine Learning-Based Diagnosis of Thyroid Cancer Using Thyroid Nodules Ultrasound Images. *Curr Bioinf* (2020) 15(4):349–58. doi: 10.2174/1574893614666191017091959
8. Stålhammar G, Fuentes Martinez N, Lippert M, Tobin NP, Mølholm I, Kis L, et al. Digital Image Analysis Outperforms Manual Biomarker Assessment in Breast Cancer. *Modern Pathol* (2016) 29(4):318–29. doi: 10.1038/modpathol.2016.34
9. Acs B, Rantalainen M, Hartman J. Artificial Intelligence as the Next Step Towards Precision Pathology. *J Intern Med* (2020) 288(1):62–81. doi: 10.1111/joim.13030
10. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting Cancer Outcomes From Histology and Genomics Using Convolutional Networks. *Proc Natl Acad Sci USA* (2018) 115 (13):E2970–9. doi: 10.1073/pnas.1717139115
11. Xu J, Janowczyk A, Chandran S, Madabhushi A. A High-Throughput Active Contour Scheme for Segmentation of Histopathological Imagery. *Medical Image Analysis* (2011) 15(6):851–62. doi: 10.1016/j.media.2011.04.002
12. Yu KH, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-Small Cell Lung Cancer Prognosis by Fully Automated Microscopic Pathology Image Features. *Nat Commun* (2016) 7:12474. doi: 10.1038/ncomms12474
13. Vaidya P, Wang X, Bera K, Khunger A, Choi H, Patil P, et al. RaPtomics: Integrating Radiomic and Pathomic Features for Predicting Recurrence in Early Stage Lung Cancer. *Digital Pathol* (2018). doi: 10.1117/12.2296646
14. Wu Z, Wang L, Li C, Cai Y, Liang Y, Mo X, et al. DeepLRHE: A Deep Convolutional Neural Network Framework to Evaluate the Risk of Lung Cancer Recurrence and Metastasis From Histopathology Images. *Front Genet* (2020) 11:768. doi: 10.3389/fgene.2020.00768
15. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors With Mismatch-Repair Deficiency. *N Engl J Med* (2015) 372(26):2509–20. doi: 10.1056/NEJMoa1500596
16. Kather JN, Halama N, Jaeger D. Genomics and Emerging Biomarkers for Immunotherapy of Colorectal Cancer. *Semin Cancer Biol* (2018) 52(Pt 2):189–97. doi: 10.1016/j.semcancer.2018.02.010
17. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep Learning can Predict Microsatellite Instability Directly From Histology in Gastrointestinal Cancer. *Nat Med* (2019) 25(7):1054–6. doi: 10.1038/s41591-019-0462-y
18. Khan AM, Rajpoot N, Treanor D, Magee D. A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution. *IEEE Trans Biomed Eng* (2014) 61(6):1729–38. doi: 10.1109/TBME.2014.2303294
19. Chi J, Eramian M. Enhancement of Textural Differences Based on Morphological Component Analysis. *IEEE Trans Image Process* (2015) 24 (9):2671–84. doi: 10.1109/TIP.2015.2427514
20. Erik Reinhard MA, Gooch B, Shirley P. Color Transfer Between Images. *IEEE Comput Graphics Appl* (2001) 21(5):34–41. doi: 10.1109/38.946629

21. Magee D, Treanor D, Crellin D, Shires M, Smith K, Mohee K, et al. Color Normalization in Digital Histopathology Images. *Comput Sci* (2009) p:100– 111.
22. Macenko M, Niethammer M, Marron J, Borland D, Woosley JT, Guan X, et al. A Method for Normalizing Histology Slides for Quantitative Analysis. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. (2009). doi: 10.1109/ISBI.2009.5193250
23. Niethammer M, Borland D, Marron JS, Woosley JT, Thomas NE. Appearance Normalization of Histology Slides. (2010). doi: 10.1007/978-3-642-15948-0_8
24. Vahadane A, Sethi A. Towards Generalized Nuclear Segmentation in Histological Images. (2014). doi: 10.1109/BIBE.2013.6701556
25. Ramakrishnan G, Anand D, Sethi A. Fast GPU-Enabled Color Normalization for Digital Pathology. In: *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE (2019).
26. Wu H-S, Xu R, Harpaz N, Burstein D, Gil J. Segmentation of Microscopic Images of Small Intestinal Glands With Directional 2-D Filters. *Analytical and quantitative cytology and histology* (2005) 27(5):291–300.
27. Farjam R, Soltanian-Zadeh H, Jafari-Khouzani K, Zoroofi RA. An Image Analysis Approach for Automatic Malignancy Determination of Prostate Pathological Images. *Cytomet Part B Clin Cytomet* (2007) 72B(4):227–40. doi: 10.1002/cyto.b.20162
28. Naik S, Doyle S, Agner S, Madabhushi A, Tomaszewski J. Automated Gland and Nuclei Segmentation for Grading Prostate and Breast Cancer Histopathology. In: *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. (2008).
29. Nguyen K, Sarkar A, Jain AK. Structure and Context in Prostatic Gland Segmentation and Classification. In: Ayache N, Delingette H, Golland P, Mori K, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012. MICCAI 2012. Lecture Notes in Computer Science, vol 7510. Berlin, Heidelberg: Springer (2012). doi: 10.1007/978-3-642-33415-3_15
30. Gunduz-Demir C, Kandemir M, Tosun AB, Sokmensuer C. Automatic Segmentation of Colon Glands Using Object-Graphs. *Medical Image Analysis* (2010) 14(1):1–12. doi: 10.1016/j.media.2009.09.001
31. Nosrati MS, Hamarneh G. Local Optimization Based Segmentation of Spatially-Recurring, Multi-Region Objects With Part Configuration Constraints. *IEEE Trans Med Imaging* (2014) 33(9). doi: 10.1109/TMI.2014.2323074
32. Cohen A RE, Shimshoni I, Sabo E. Memory Based Active Contour Algorithm Using Pixel-Level Classified Images for Colon Crypt Segmentation. *Comput Med Imaging Graphics* (2015) 2015:43:150–64. doi: 10.1016/j.compmedimag.2014.12.006
33. Sirinukunwattana K, Snead D, Rajpoot N. A Stochastic Polygons Model for Glandular Structures in Colon Histology Images. *IEEE Trans Med Imaging* (2015) 34(11):1–1. doi: 10.1109/TMI.2015.2433900
34. Paramanandam M, O'Byrne M, Ghosh B, Mammen JJ, Manipadam MT, Thamburaj R, et al. Automated Segmentation of Nuclei in Breast Cancer Histopathology Images. *PloS One* (2016) 11(9):e0162053. doi: 10.1371/journal.pone.0162053
35. Xu J, Lei X, Liu Q, Gilmore H, Wu J, Tang J, et al. Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images. (2014). doi: 10.1109/ISBI.2014.6868041
36. Irshad H, Veillard A, Roux L, Racoceanu D. Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review—Current Status and Future Potential. *IEEE Rev Biomed Eng* (2017) 7:97–114. doi: 10.1109/RBME.2013.2295804
37. Qi X, Xing F, Foran DJ, Yang L. Robust Segmentation of Overlapping Cells in Histopathology Specimens Using Parallel Seed Detection and Repulsive Level Set. *Biomed Eng IEEE Trans* (2011) 59(3):754–65. doi: 10.1109/TBME.2011.2179298
38. Lu C, Xu H, Xu J, Gilmore H, Mandal M, Madabhushi A. Multi-Pass Adaptive Voting for Nuclei Detection in Histopathological Images. *Sci Rep* (2016) 6 (1):33985. doi: 10.1038/srep33985
39. Parvin B, Yang Q, Han J, Chang H, Rydberg M, Barcellos-Hoff MH. Iterative Voting for Inference of Structural Saliency and Characterization of Subcellular Events. *IEEE Trans Image Process* (2007) 16(3):615–23. doi: 10.1109/TIP.2007.891154
40. Arteta C, Lempitsky V, Noble JA, Zisserman A. Learning to Detect Cells Using Non-Overlapping Extremal Regions. In: *Proceedings of the 15th International Conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part I* (2012).
41. Cheng J, Rajapakse J. Segmentation of Clustered Nuclei With Shape Markers and Marking Function. *IEEE Transactions on Biomedical Engineering* (2009) 56(3):741–8.
42. Veta M, Van Diest PJ, Kornegoor R, Huisman A, Viergever MA, Pluim JP. Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images. *PloS One* (2013) 8(7):e70221. doi: 10.1371/journal.pone.0070221
43. Vink JP, Van Leeuwen M, Van Deurzen C, de Haan G. Efficient Nucleus Detector in Histopathology Images. *J Microscopy* (2013) 249(2):124–35. doi: 10.1111/jmi.12001
44. Ali S, Madabhushi A. An Integrated Region-, Boundary-, Shape-Based Active Contour for Multiple Object Overlap Resolution in Histological Imagery. *IEEE Trans Med Imaging* (2012) 31:1448–60. doi: 10.1109/TMI.2012.2190089
45. Veta M, Pluim JP, Van Diest PJ, Viergever MA. Breast Cancer Histopathology Image Analysis: A Review. *IEEE Trans Biomed Eng* (2014) 61(5):1400–11. doi: 10.1109/TBME.2014.2303852
46. Xing F, Yang L. Robust Nucleus/Cell Detection and Segmentation in Digital Pathology and Microscopy Images: A Comprehensive Review. *IEEE Rev BioMed Eng* (2016) 9:234–63.
47. Xing F, Xie Y, Yang L. An Automatic Learning-Based Framework for Robust Nucleus Segmentation. *IEEE Trans Med Imaging* (2016) 35(2):550–66. doi: 10.1109/TMI.2015.2481436
48. Yang X, Li H, Zhou X. Nuclei Segmentation Using Marker-Controlled Watershed, Tracking Using Mean-Shift, and Kalman Filter in Time-Lapse Microscopy. *IEEE Transactions on Circuits and Systems I: Regular Papers* (2006) 53(11):2405–14. doi: 10.1109/TCSI.2006.884469
49. Xue JH, Titterington DM. T-Tests, F-Tests and Otsu's Methods for Image Thresholding. *IEEE Trans Image Process A Publ IEEE Signal Process Soc* (2011) 20(8):2392–6.
50. Zhang C, Sun C, Pham TD. Segmentation of Clustered Nuclei Based on Concave Curve Expansion. *J Microscopy* (2013) 251(1):57–67. doi: 10.1111/jmi.12043
51. Sirinukunwattana K, Raza SEA, Tsang Y-W, Snead DR, Cree IA, Rajpoot NM. Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. *IEEE Trans Med Imaging* (2016) 35:1196–206.
52. Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A, Sethi A. A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology. *IEEE Trans Med Imaging* (2017) 36:1550–60.
53. Ciresan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis Detection in Breast Cancer Histology Images With Deep Neural Networks. *Med Image Comput Comput Assist Interv* (2013) 16(Pt 2):411–8. doi: 10.1007/978-3-642-40763-5_51
54. Liu F, Lin Y. A Novel Cell Detection Method Using Deep Convolutional Neural Network and Maximum-Weight Independent Set. In: *Deep Learning and Convolutional Neural Networks for Medical Image Computing*. Springer International Publishing (2017).
55. Song Y, Zhang L, Chen S, Ni D, Lei B, Wang T. Accurate Segmentation of Cervical Cytoplasm and Nuclei Based on Multiscale Convolutional Network and Graph Partitioning. *IEEE Trans Biomed Eng* 62(10):2421–33. doi: 10.1109/TBME.2015.2430895
56. Chang JM. Back to Basics: Traditional Nottingham Grade Mitotic Counts Alone Are Significant in Predicting Survival in Invasive Breast Carcinoma. *Ann Surg Oncol* (2015) 22:509–15. doi: 10.1245/s10434-015-4616-y
57. Belien J, Baak J, Van Diest P, Van Ginkel A. Counting Mitoses by Image Processing in Feulgen Stained Breast Cancer Sections: The Influence of Resolution. *Cytometry* (1997) 28(2):135–40. doi: 10.1002/(SICI)1097-0320(19970601)28:2<135::AID-CYTO6>3.0.CO;2-E
58. Roux L, Racoceanu D, Loménie N, Kulikova M, Irshad H, Klossa J, et al. Mitosis Detection in Breast Cancer Histological Images An ICPR 2012 Contest. *J Pathol Inf* (2013) 4(1):8. doi: 10.4103/2153-3539.112693
59. Zhang X, Dou H, Ju T, Xu J, Zhang S. Fusing Heterogeneous Features From Stacked Sparse Autoencoder for Histopathological Image Analysis. *IEEE J Biomed Health Inf* (2017) 20(5):1377–83.
60. Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, Van De Vijver MJ, et al. Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated With Survival. *Sci Trans Med* (2011) 3(108):108ra113–108ra113. doi: 10.1126/scitranslmed.3002564
61. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PloS Med* (2018) 10(2):e1001381. doi: 10.1371/journal.pmed.1001381

62. Tizhoosh HR PL. Artificial Intelligence and Digital Pathology: Challenges and Opportunities. *J Pathol Inform* (2018) 9:38.
63. Tutac AE, Racoceanu D, Putti T, Xiong W, Leow W-K, Cretu V, et al. Knowledge-Guided Semantic Indexing of Breast Cancer Histopathology Images. In: *2008 International Conference on Biomedical Engineering and Informatics*. IEEE (2008).
64. Dalle JR, Leow WK, Racoceanu D, Tutac AE, Putti TC. Automatic Breast Cancer Grading of Histopathological Images. *Conf Proc IEEE Eng Med Biol Soc* (2008) 2008:3052–5. doi: 10.1109/IEMBS.2008.4649847
65. Wang C, Pécot T, Zynger DL, Machiraju R, Shapiro CL, Huang K. Research and Applications: Identifying Survival Associated Morphological Features of Triple Negative Breast Cancer Using Multiple Datasets. *J Am Med Inform Assoc* (2013) 20:680–7.
66. Basavanhally A, Ganesan S, Feldman M, Shih N, Madabhushi A. Multi-Field-Of-View Framework for Distinguishing Tumor Grade in ER+ Breast Cancer From Entire Histopathology Slides. *IEEE Trans Biomed Eng* (2013) 60 (8):2089–99. doi: 10.1109/TBME.2013.2245129
67. Yuan Y, Failmezger H, Rueda OM, Ali HR, Gräf S, Chin S-F, et al. Quantitative Image Analysis of Cellular Heterogeneity in Breast Tumors Complements Genomic Profiling. *Sci Trans Med* (2020) 4(157):157ra143–157ra143.
68. Chen RJ, Lu MY, Wang J, Williamson DF, Rodig SJ, Lindeman NI, et al. Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis. *IEEE Trans Med Imaging* (2020). doi: 10.1109/TMI.2020.3021387
69. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting Cancer Outcomes From Histology and Genomics Using Convolutional Networks. *Proc Natl Acad Sci USA* (2018) 115 (13):E2970–9. doi: 10.1073/pnas.1717139115
70. Serag A I-MA, Qureshi H. Translational AI and Deep Learning in Diagnostic Pathology. *Front Med* (2019) 6:185. doi: 10.3389/fmed.2019.00185
71. Chang HY JC, Woo JI. Artificial Intelligence in Pathology. *J Pathol Transl Med* (2019) 53:1–12. doi: 10.4132/jptm.2018.12.16
72. Daneshjou R, He B, Ouyang D, Zou JY. How to Evaluate Deep Learning for Cancer Diagnostics - Factors and Recommendations. *Biochim Biophys Acta Rev Cancer* (2021) 1875(2):188515.
73. Pope WB, Sayre J, Perlina A, Villablanca JP, Mischel PS, Cloughesy TF. MR Imaging Correlates of Survival in Patients With High-Grade Gliomas. *AJNR Am J Neuroradiol* (2005) 26(10):2466–74.
74. Cao R, Yang F, Ma SC, Liu L, Zhao Y, Li Y, et al. Development and Interpretation of a Pathomics-Based Model for the Prediction of Microsatellite Instability in Colorectal Cancer. *Theranostics* (2020) 10 (24):11080–91. doi: 10.7150/thno.49864
75. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep Learning in Cancer Pathology: A New Generation of Clinical Biomarkers. *Br J Cancer* (2021) 124(4):686–96. doi: 10.1038/s41416-020-01122-x
76. Colling R, Pitman H, Oien K, Rajpoot N, Macklin P, Snead D, et al. Artificial Intelligence in Digital Pathology: A Roadmap to Routine Use in Clinical Practice. *J Pathol* (2019) 249(2):143–50. doi: 10.1002/path.5310

Conflict of Interest: Author XShi, GT, YL, JY and YW were employed by the company Geneis Beijing Co., Ltd. Author LL was employed by the company Beijing Shanghe Jiye Biotech Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.