Future Generation Computer Systems 164 (2025) 107578

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Review of deep learning-based pathological image classification: From task-specific models to foundation models

Haijing Luan [a,b], Kaixing Yang [b,c], Taiyuan Hu [a,b], Jifang Hu [a,b], Siyao Liu [d], Ruilin Li [a], Jiayin He [a], Rui Yan [e,f], Xiaobing Guo [g], Niansong Qian [h], Beifang Niu [a,b,*]

[a] Computer Network Information Center, Chinese Academy of Sciences, Beijing, 100083, China
[b] University of Chinese Academy of Sciences, Beijing, 100049, China
[c] Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, 100101, China
[d] Beijing ChosenMed Clinical Laboratory Co. Ltd, Beijing, 100176, China
[e] School of Biomedical Engineering, University of Science and Technology of China, Hefei, 230026, China
[f] Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, 215123, China
[g] Lenovo Research, Beijing, 100085, China
[h] Department of Thoracic Oncology, Senior Department of Respiratory and Critical CareMedicine, the Eighth Medical Center of PLA General Hospital, Beijing, 100091, China

ARTICLE INFO

Keywords:
Pathological images
Weakly supervised learning
Whole slide image
Foundation model

ABSTRACT

Pathological diagnosis is considered the gold standard in cancer diagnosis, playing a crucial role in guiding treatment decisions and prognosis assessment for patients. However, achieving accurate diagnosis of pathology images poses several challenges, including the scarcity of pathologists and the inherent subjective variability in their interpretations. The advancements in whole-slide imaging technology and deep learning methods provide new opportunities for digital pathology, especially in low-resource settings, by enabling effective pathological image classification. In this article, we begin by introducing the datasets, which include both unimodal and multimodal types, as essential resources for advancing pathological image classification. We then provide a comprehensive overview of deep learning-based pathological image classification models, covering task-specific models such as supervised, unsupervised, weakly supervised, and semi-supervised learning methods, as well as unimodal and multimodal foundation models. Next, we review tumor-related indicators that can be predicted from pathological images, focusing on two main categories: indicators that can be recognized by pathologists, such as tumor classification, grading, and region recognition; and those that cannot be recognized by pathologists, including molecular subtype prediction, tumor origin prediction, biomarker prediction, and survival prediction. Finally, we summarize the key challenges in digital pathology and propose potential future directions.

1. Introduction

Pathological diagnosis is considered the gold standard in cancer diagnosis, profoundly influencing therapeutic decisions and prognostic evaluations. Traditionally, pathologists have relied on high-power microscopes to examine stained specimens on glass slides. However, the landscape of diagnostic pathology has evolved with the widespread adoption of digital scanners, which integrate fragmented images to generate comprehensive and high-resolution whole slide image (WSI) of histopathological sections. The advent of WSI has addressed several limitations inherent in traditional glass slides, including fragility, retrieval challenges, and variability in diagnostic reproducibility, propelling pathology into a new stage of development driven by the transformative capabilities of digital pathology.

Currently, pathological diagnosis still faces huge challenges. Firstly, obtaining large-scale, finely annotated pathological image datasets is a complex and time-consuming task that requires experienced pathologists to manually annotate, thus imposing a substantial workload. Secondly, pathological diagnoses are influenced by the subjective biases of individual pathologists. Additionally, becoming a proficient pathologist requires the examination of at least 100,000 pathological images. However, as of 2018, China had only 10,000 licensed pathologists, indicating a significant shortfall of 90%. This severe shortage results in pathologists being overloaded with work and unable to meet the clinical demand for precise and efficient pathological diagnoses [1]. Consequently, there is an urgent need to develop automated and precise

* Corresponding author.
E-mail addresses: yanrui@ustc.edu.cn (R. Yan), guoxba@lenovo.com (X. Guo), qianniansong@301hospital.com.cn (N. Qian), bniu@sccas.cn (B. Niu).

https://doi.org/10.1016/j.future.2024.107578
Received 25 June 2024; Received in revised form 31 August 2024; Accepted 20 October 2024
Available online 28 October 2024

computational methods for analyzing pathological images to provide quantifiable auxiliary diagnoses.

Pathological image analysis tasks can be simply summarized as classification, registration, detection, segmentation, localization, and generation [2]. Image classification, the most critical and in-demand task in clinical pathological image analysis, serves as the foundation for subsequent tasks such as nucleus localization [3], mitosis detection [4], gland segmentation [5], pathological image retrieval [6], and pathological image registration. In the medical and natural image domains, deep learning-based image classification methods have achieved significant advancements [7]. However, given the extremely high resolution of WSI, directly applying deep learning methods to the entire WSI is often inefficient due to computational constraints and the risk of losing detailed image information during compression. Therefore, from task-specific models to foundation models, the computational pathology community has made significant efforts to address the challenges of WSI classification.

In this article, we review the advancements of deep learning-based hematoxylin and eosin (H&E)-stained pathological image classification methods. Specifically, we first introduce the unimodal and multimodal datasets for pathological image classification. Next, we outline core deep learning methods, covering task-specific approaches such as supervised, unsupervised, weakly supervised, and semi-supervised learning. We then shift our focus to foundation models, categorized into unimodal and multimodal types, and provide a detailed analysis of their applications and recent developments in computational pathology field. Additionally, we categorize pathological image classification tasks into two main types: *basic classification tasks*, which pathologists can identify in H&E-stained images with the naked eye, and *advanced classification tasks*, which cannot be identified by pathologists through visual inspection of H&E-stained images. We then provide a detailed introduction to both types of tasks. Finally, we summarize the existing challenges in pathological image analysis and discuss potential directions.

2. Datasset

2.1. Pathological images

Cancer diagnosis generally relies on the examination of tissue pathological sections. A white section is a slice of human tissue taken during surgery or biopsy and placed unstained on a glass slide, these white sections are usually processed into pathological sections through procedures including fixation, paraffin embedding, and cutting into thin slices, followed by staining techniques such as H&E staining, resulting in Formalin-Fixed Paraffin-Embedded (FFPE) slides that assist pathologists in morphological observation and diagnosis [8]. Fresh Frozen (FF) slides represent a rapid processing method conducted in a frozen laboratory during surgical procedures. This technique can often compromise tissue integrity, yet it is commonly employed for quick diagnostic assessments and decision-making, such as determining the nature of a lesion during surgery and guiding the extent of the operation. However, not all cases can be accurately diagnosed using H&E sections alone, necessitating the use of immunohistochemistry (IHC). IHC utilizes the chemical reaction of antigen antibodies to qualitatively and quantitatively identify antigens within cells. By binding to specific antibodies, IHC labels specific antigen molecules in the sections, rendering them to appear brownish-yellow. Fig. 1 illustrates examples of a FFPE tissue images, FF tissue images and an IHC images, respectively.

WSI is digital representations of histopathological sections. The process of creating WSI involves several steps, including sampling, fixation, dehydration, sectioning, staining, coverslipping, and scanning. WSI possesses a high resolution of $10^4 \times 10^4$ pixels, and are stored in a multi-resolution pyramids format [9]. This format enables pathologists to analyze the tissue at different levels of detail [10], ranging from the lowest to the highest magnification levels, including 5x, 10x, 20x, and 40x. A schematic diagram of WSI is presented in Fig. 2. Given the structural heterogeneity and cellular atypia of tumors, pathological images at various magnification levels enable pathologists to observe distinct details during analysis [10]. At low magnification, the overall morphological characteristics of the tissue (structural information) can be evaluated, whereas high magnification facilitates the detailed examination of critical features, such as the shape, arrangement, and size of cells (cellular information). It mirrors the actual diagnostic process in which pathologists diagnose diseases by analyzing cellular and structural details of pathological tissues through varying microscope magnifications.

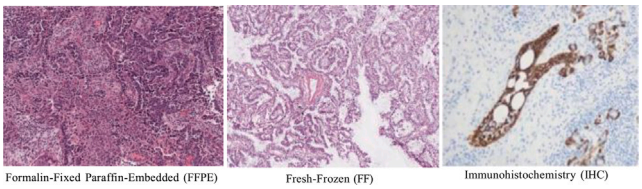


Fig. 1. Examples of pathological images with three different staining types.

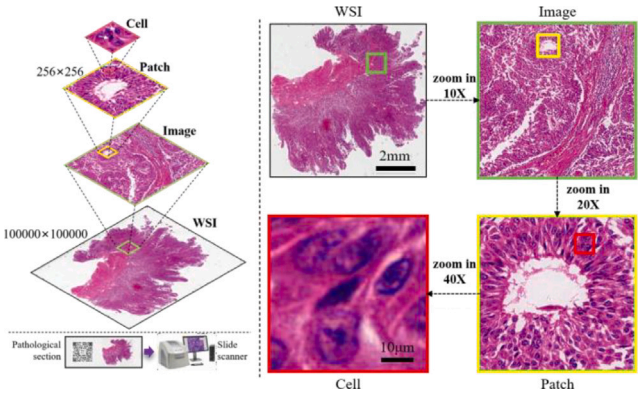


Fig. 2. Schematic diagram of whole slide image.

2.2. Multimodal

The successful development of pathology foundation models in digital pathology fundamentally relies on both the accumulation of multimodal data and the systematic construction and organization of large-scale multimodal datasets. Given the relative accessibility of pathology images and text data, most existing multimodal datasets predominantly focus on visual and language modalities, thereby limiting modal diversity. Here, we discuss various types of multimodal datasets, focusing on visual-language and visual-gene datasets, which are essential for advancing the development and research of pathology foundation models.

2.2.1. Pathology-language

The progress of pathology-language multimodal pathology foundation models largely depends on diverse medical text datasets, including literature, pathology reports, and social media.

Pathology reports. Pathology reports provide comprehensive diagnostic information on diseases, including anatomical locations and pathological findings, which are typically documented based on the sample processing and preparation procedures. The aggregation of diagnostic results from multiple WSI within pathology reports presents significant challenges in generating accurate image–text pairs for WSI. Ahmed et al. [11] addressed this issue by employing regular expressions

to process diagnostic texts from pathology reports, resulting in a de-identified dataset of over 350K WSIs and corresponding diagnostic text pairs, spanning a diverse range of diagnoses, procedure types, and tissue types.

Literature. Due to the limited privacy information within literature and the condensation of medical knowledge, large-scale literature datasets are publicly accessible. PubMed Central (PMC) [12] is a free digital archive of full-text biomedical and life sciences journal articles, providing comprehensive access to research papers, including figures, tables, and supplementary materials. For instance, Lu et al. [13] developed a dataset comprising 1.17 million image–text pairs for multimodal pathology foundation models by leveraging educational sources and data from PMC through image-caption matching techniques.

Social media. Social media platforms, including medical Twitter and YouTube, serve as significant sources of medical image and text data, with clinicians disseminating de-identified images, videos and medical insights, thereby providing substantial contributions to pathology research. Drawing on these resources, OpenPath [14] integrates over 200K pathology images from Twitter's medical knowledge-sharing platform and the Large-scale Artificial Intelligence Open Network, with each image accompanied by detailed descriptions from medical professionals. Similarly, QUILT [15] comprises 800K image–text pairs for histopathology, with images and text descriptions derived from YouTube videos, and has been further expanded with additional data from Twitter and research papers to create QUILT-1M.

2.2.2. Pathology-Gene

The advancements in high-throughput sequencing have laid a robust foundation for developing pathology-gene multimodal foundation models that integrate gene-related molecular data, such as genomics and transcriptomics, with pathology images to enable more accurate classification of tumor indicators.

Genomics and single-cell omics data. Genomics and single-cell omics data provide detailed insights into genetic information, gene expression patterns, and cellular functions. Genomics encompasses the study of an organism's complete genome, providing insights into genetic variations, gene functions, and interactions within the genetic landscape, while single-cell RNA sequencing (scRNA-seq) profiles gene expression at the individual cell level, offering detailed insights into cellular heterogeneity. For instance, Jin et al. [16] curated a dataset of 946 image-omic pairs to serve as input for GiMP, a model that integrates genomic data with pathological images.

Spatial transcriptomics. Spatial transcriptomics (ST) simultaneously measures gene expression while preserving spatial information within the tissue, providing detailed gene expression data for each spot image. STimage-1K4M [17] and HEST-1k [18] are currently regarded as representative datasets in the field of ST. HEST-1k consists of 1108 ST profiles, each associated with a WSI and corresponding metadata. In contrast, STimage-1K4M includes 1149 images derived from ST data, providing gene expression information at the resolution of individual spatial spots within pathology images.

3. Method

The development of deep learning-based pathological image classification has undergone a profound transformation, evolving from task-specific models designed for particular diagnostic tasks to more generalized foundation models that offer broader applicability across diverse pathological contexts. To elucidate this transition, this section will first present the methodologies and applications associated with task-specific models, followed by an examination of foundation models, which have increasingly come to predominate the field of computational pathology in recent years.

3.1. Task-specific model

Within the framework of task-specific models for tumor classification based on pathological images, deep learning methodologies are conventionally classified into supervised learning, unsupervised learning, weakly supervised learning, and semi-supervised learning. Fig. 3 illustrates the overall framework for tumor classification based on deep learning and pathological image. Firstly, the tissue regions in WSI are segmented using the CLAM processing pipeline [19], which is accessible on GitHub (https://github.com/mahmoodlab/CLAM). Next, WSI is tessellated into image patches of 256 × 256 pixels (without overlap). Subsequently, feature extraction and classification tasks are performed on each image patch, and attention heatmaps are used to provide model interpretability. The goal of feature extraction is to provide valuable information for subsequent pathological image classification.

3.1.1. Supervised learning

The goal of supervised learning is to construct a function that can effectively map input data to corresponding label. In pathological image analysis, these labels are associated with WSI or objects within WSIs. Supervised learning involves assigning labels at the patch level, necessitating pathologists to annotate regions of interest (ROIs) within WSIs based on their expertise. Common supervised learning algorithms include convolutional neural networks (CNNs), recurrent neural networks (RNNs), and graph neural networks (GNNs). CNNs were proposed to address the limitations of fully connected neural networks (FCNN) in image classification tasks, such as insufficient spatial structure representation and overfitting due to excessive parameters. CNNs improve performance by incorporating local receptive fields, weight sharing, and pooling operations. Compared to FCNN, CNNs are more robust to transformations like translation, rotation, scaling, and elastic deformation. Typical CNN architectures include LeNet [20], AlexNet [21], ResNet [22], GoogleNet [23], and DenseNet [24]. However, applying CNNs to process gigapixel WSIs is infeasible with current hardware capabilities. Therefore, a common approach is to segment WSIs into tissue regions and divide them into smaller patches (e.g., 256 × 256 pixels) to serve as inputs for CNN. However, the patches lose the neighboring and topological relationships present in the WSI. RNNs, designed for processing sequential data such as text and audio, can capture temporal dependencies and long-term relationships within sequences. By introducing time steps, RNNs enable the patches to retain contextual information and spatial relationships within WSIs, effectively improving the accuracy and reliability of pathological image analysis. In addition, the patches of WSIs exhibit different spatial topological structures, closely related to tumor heterogeneity. A common approach is to treat patches as vertices of a graph and construct edges based on the distances between patches, thus transforming the WSI into graph structure. GNNs can then extract features from graph structure for pathological image classification tasks.

Supervised learning is widely used in digital pathology, but it requires fine-grained labeled data, such as patch-level annotations for tumor classification tasks. However, The acquisition of a substantial amount of such detailed labeled data is both costly and challenging. It requires experienced pathologists to manually annotate specific regions of WSIs based on their prior knowledge of tumors. These factors limit the development of supervised learning in the field of medical image classification to some extent. Recent studies have sought to address the limitations posed by the requirement of extensive fine-grained labeling in supervised learning. The most representative approaches include weakly supervised learning based on weak annotations, semi-supervised learning based on limited annotations, and unsupervised learning based on unlabeled data.

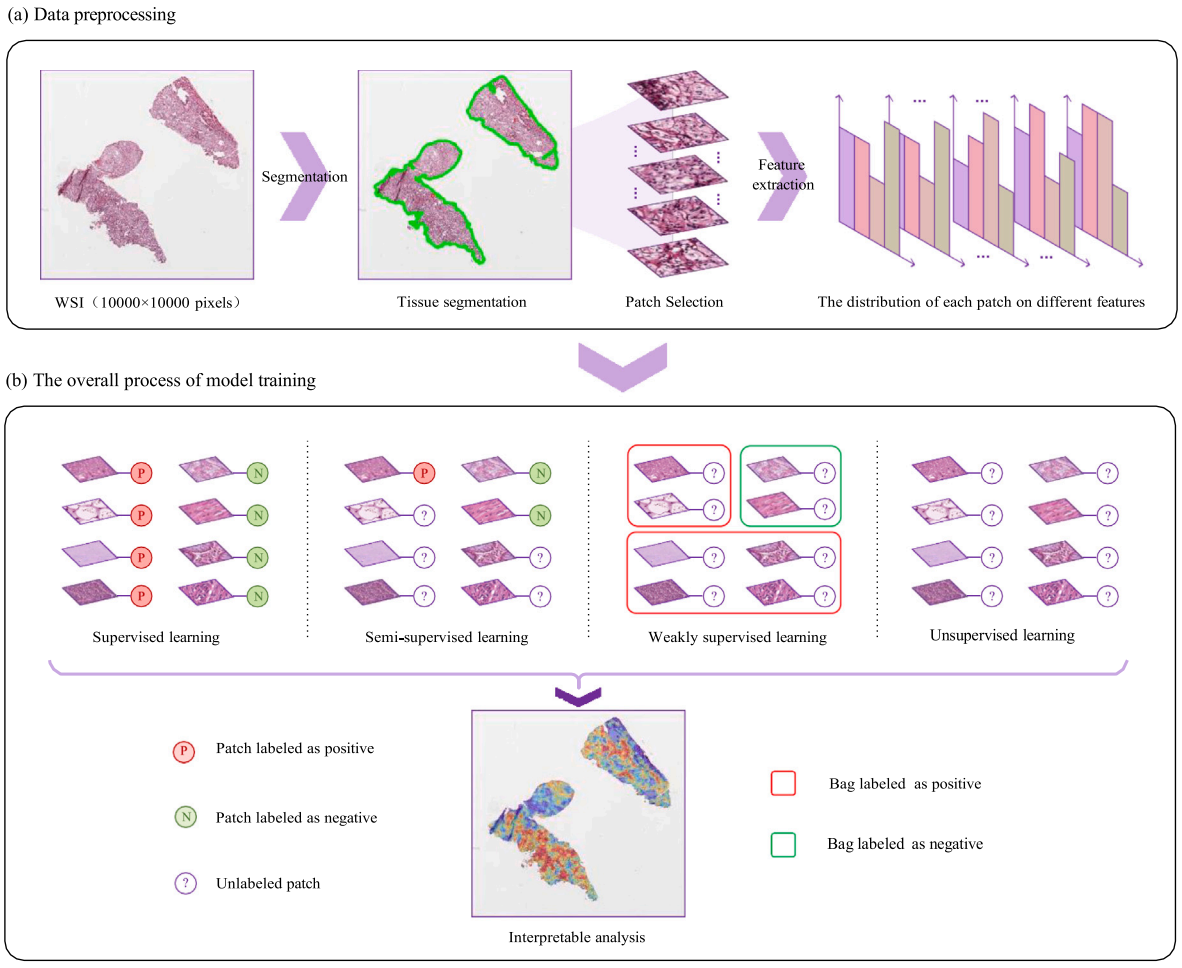


Fig. 3. Pathological image analysis framework.

3.1.2. Unsupervised learning

Unsupervised learning aims to identify latent relationships, patterns, and features in large amounts of unlabeled data without requiring manual labeling. The outcomes of unsupervised learning often inform subsequent classification tasks. This approach is particularly vital for tasks such as patch feature extraction, clustering, and image normalization in the analysis of pathological images.

Clustering and dimensionality reduction are common applications of unsupervised learning, designed to organize data into meaningful groups and reduce its dimensions by optimizing the probability distribution within specific constraints. K-Means is an unsupervised clustering methods, which is often used in digital pathology to ensure the representative of the selected patches within WSI [25]. Utilizing a self-supervised learning (SSL) pre-trained patch-level model as an encoder can substantially improve the performance of WSI classification tasks. CTransPath is an unsupervised feature extractor for histopathology images based on a Swin Transformer architecture and trained with semantically relevant contrastive learning (SRCL), a novel SSL technique derived from MoCo v3. Evaluations demonstrate that CTransPath exhibits robust performance across various tasks, including patch retrieval, classification, mitosis detection, and colorectal adenocarcinoma gland segmentation [26]. Additionally, zheng et al. [27] utilized the self-supervised representation learning framework BYOL [28] as a representation learner for CNNs, proposing a WSI classification method that performs exceptionally well in tumor subtype classification tasks. Recently, DINO [29] has emerged as an outstanding SSL framework, demonstrating exceptional performance and significant potential in WSI analysis [30,31].

Unsupervised learning offers several advantages but also presents significant challenges. A primary concern is the unverifiable nature of its output results, which complicates the assessment of their accuracy and reliability. Additionally, these outputs often result in difficulties with interpretation and understanding.

3.1.3. Semi-supervised learning

Obtaining fine-grained annotations, such as pixel-level or slice-level annotations, is often time-consuming and challenging. Consequently, relying solely on limited labeled data for model training can easily lead to overfitting, undermining the model's performance and generalization ability. In contrast, semi-supervised learning methods can effectively leverage a small number of labeled data points alongside a substantial amount of unlabeled data, enhancing the learning process and improving overall model robustness. Therefore, by employing semi-supervised learning, it is possible to harness the benefits of using a small labeled dataset to guide the learning process and leveraging a larger unlabeled dataset to enhance the generalizability of the model.

Peikari et al. [32] proposed a semi-supervised learning approach that first clusters and then labels high-density regions within the data space. This approach assists supervised support vector machines (SVM) in identifying decision boundaries effectively. Yu et al. [33] developed a mean teacher model for detecting malignant patches, they found that Semi-supervised learning significantly outperformed supervised learning with limited labels but observed no difference between Semi-supervised learning and supervised learning when using a fully labeled dataset of over 10,000 samples. Wenger et al. [34] employed a semi-supervised learning approach that utilizes consistency regularization

and self-ensembling to leverage the unlabeled portions of the dataset for detecting and grading bladder cancer samples. They reported an accuracy that was 19% higher than the baseline supervised learning model, achieved using only 3% of labeled data. Gao et al. [35] proposed a semi-supervised multi-task learning framework for cancer classification that leverages datasets annotated using a minimal point annotation strategy. Extensive experiments conducted across multiple datasets demonstrate the effectiveness of the proposed framework in terms of accuracy and generalization.

Semi-supervised learning demonstrates significant advantages over supervised learning when labeled data is limited and a large amount of unlabeled data is available, offering a viable solution to the challenges of obtaining labels in the medical field. However, it also faces issues related to interpretability. Additionally, a crucial prerequisite for the successful application of semi-supervised learning is that the labeled and unlabeled data must come from the same distribution [36].

3.1.4. Weakly supervised learning

Weakly supervised learning only requires WSI-level labels instead of fine-grained annotations. Multiple instance learning (MIL) is a typical method of weakly supervised learning. By treating each WSI (called bag) as a collection of smaller image regions (called instances) known as bags, it is possible to analyze the WSI without the need for manual extraction of ROIs. In the analysis of pathological images, it is essential to extract features from each instance within the WSI using deep neural networks (DNNs) and transform them into low-dimensional feature vectors. WSI classification is performed using various MIL pooling techniques and classifiers, which can be categorized into methods based on bag-level classifiers, instance-level classifiers, and combinations of bag-level and instance-level classifiers, as illustrated in Fig. 4.

Instance-level MIL methods involve training a instance-based classifier to predict instance scores, which are then aggregated using MIL pooling methods to obtain bag-level predictions. It provides robust interpretability, allowing for the identification of key instances that contribute to bag-level predictions. Bag-level MIL method uses DNNs to extract features from instances within a bag, aggregates these features through MIL pooling to form bag-level features, and trains a bag-level classifier to achieve prediction results. Embedding-based method follows the same principle as the MIL method based on a bag-level classifier but lacks the ability to locate key instances compared to instance-based methods, thus reducing interpretability. Attention mechanisms are employed to integrate information from all instances within WSIs [19,37,38]. Two-stage MIL with SSL uses self-supervised learning to obtain instance-level feature representations, which are subsequently aggregated using attention-based MIL for bag-level classification, enhancing performance and reducing overfitting [39,40]. The self-attention mechanism in Transformer models relationships between instances within WSIs, effectively capturing global and local features for WSI classification [40–45]. MIL methods that utilize both instance-level and bag-level classifiers leverage the strengths of each approach, enabling the prediction of bag-level labels while localizing key instances within the bags. Two-stage MIL method based on the predictions of the instance classifier, instances with high diagnostic relevance are selected for aggregation to train the bag-level classifier [46–48]. Hybrid MIL method simultaneously trains bag-level and instance-level classifiers in an end-to-end manner and subsequently integrates the results from both levels to achieve the overall classification of WSIs [40]. The End-to-End MIL method, in contrast to those relying solely on a bag-level classifier, incorporates an instance-level classifier, which introduces additional constraints and refines the feature space for bag-level classification [19,49,50].

In addition to MIL methods, several approaches leverage classical weakly supervised learning frameworks. These methods utilize CNN and vision transformer (ViT), operating under the assumption that all patches within a WSI inherit the WSI-level label for classification, despite the inherent weakness of these labels [51–54]. In these classical weakly supervised methods, each patch inherits the slide label, providing an effective form of data augmentation that is beneficial for specific tasks [55].

3.2. Foundation model

Currently, task-specific diagnostic methods in computational pathology, such as those used for the subtype classification of renal cell carcinoma and non-small cell lung cancer, as well as lymph node metastasis detection [19], exhibit significant advantages. Limited to specific subsets of pathology images, these methods are less likely to capture the full spectrum of variations in tissue morphology and laboratory conditions. In contrast, foundation models are trained on vast datasets–orders of magnitude larger than any previously employed in computational pathology–utilizing a suite of algorithms commonly referred to as SSL, which eliminates the need for meticulously curated labels. The embeddings generated by these models demonstrate robust generalizability across a broad range of predictive tasks, thereby enhancing the applicability and effectiveness of the models in diverse pathological contexts. Based on whether pathology images are combined with other modalities, such as biomedical text and genomic data, pathology foundation models can be categorized into unimodal foundation model (UFM) and multimodal foundation model (MFM). As shown in Fig. 5, a comparative analysis is provided of tumor pathological image classification methods utilizing various pathology foundation models across different modalities, including pathology, pathology-language, and pathology-gene.

3.2.1. Unimodal foundation model

The performance of foundation models is critically influenced by both the size of the dataset and the model, as evidenced by results from scaling laws [56]. In the field of natural images, modern foundation models are trained on millions of images, such as ImageNet [57] and JFT-300M [58], using models with hundreds of millions to billions of parameters, like ViTs. Despite the challenges in acquiring large-scale datasets in digital pathology, recent efforts [26,59,60] have utilized datasets ranging from 30,000 to 1.5 million WSIs to train foundation models from 28 million to 632 million parameters. These methods have demonstrated that image features generated through SSL on pathological images surpass those trained on natural images, with performance improving as the scale of the data increases. Based on the chronological order of research outcomes, a brief introduction to patch-based pathological foundation models will be provided, followed by an overview of the latest advancements in WSI-based pathological foundation models, as detailed in Table 1.

In patch-based pathology foundation models, a common approach involves selecting a limited number of tiles from each slide, treating each image tile as an independent sample for SSL pertaining. This process generates task-agnostic embeddings that can be utilized across various tasks, such as tumor classification, survival prediction, image search, and segmentation. Wang et al. [26] proposed the SRCL, a SSL method built on MoCo v3, along with CTransPath, a hybrid architecture that merges convolutional layers with the Swin Transformer, enabling it to simultaneously capture both local information and global contextual information. This is the first Transformer-based unsupervised feature extractor, which can be used for patch classification and WSI classification. In response to the limitations identified in retrieval capabilities, label efficiency, and potential biases related to H&E staining intensity in out-of-domain evaluation observed in CTransPath, Vorontsov et al. [59] introduced Virchow, the largest foundation model for computational pathology to date, which is a ViT-huge model trained on 381 million tiles using the DINOv2. These tiles were derived from nearly 1.5 million proprietary slides, covering 24 tissue types. The performance of downstream tasks was evaluated on tile-level and slide-level benchmarks, including tissue classification and biomarker prediction. Chen et al. [60] proposed a universal self-supervised framework, UNI, which also utilizes the DINOv2 backbone network and introduces data diversity through pre-training on the "Mass-100K" pathological slide dataset. This dataset, sourced from

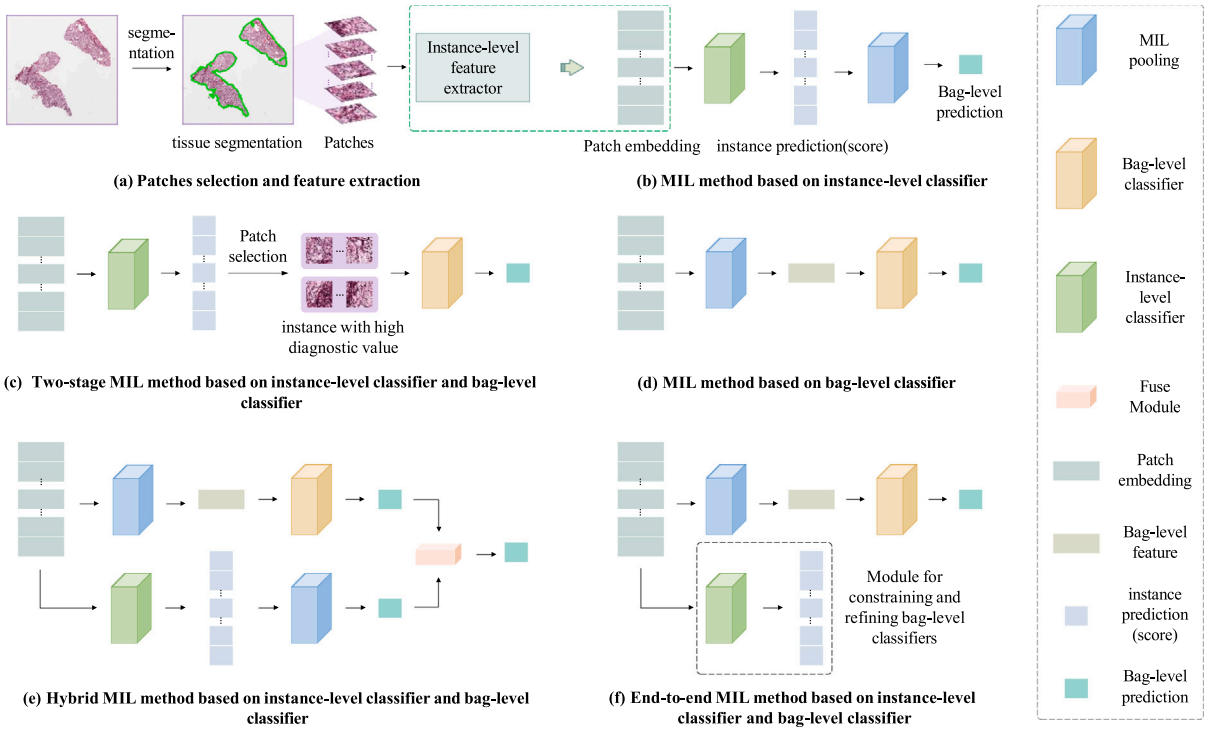


Fig. 4. Comparative analysis of tumor pathological image classification methods using multiple instance learning. Each method is color-coded to highlight the different components, such as instance-level classifiers, bag-level classifiers, and MIL pooling.

Massachusetts General Hospital (MGH), Brigham and Women's Hospital, and the Genotype-Tissue Expression (GTEx) consortium, includes a broad range of pathological samples. UNI is applicable to various tasks, including tissue classification, disease subtype classification, and cancer histological segmentation.

Recently, pathology foundation models have taken advantage of the multi-scale nature of WSIs. Juyal et al. [61] introduced PLUTO, a pathology foundation model based on a lightweight ViT, which also utilizes the DINOv2 backbone network by integrating various patch sizes and WSI tile resolutions to capture diverse biological contexts at the slide, tissue, and cellular & subcellular levels, across multiple resolutions. To enhance its performance, the DINOv2 loss function was modified by incorporating a masked autoencoder (MAE) objective term and a Fourier loss term, which collectively guide the model to capture both low-frequency and high-frequency features, thereby improving the representation of various biological contexts. PLUTO effectively emulates the process by which pathologists examine pathological slides at different magnification levels (0.25, 0.5, 1, and 2 μm per pixel), generating task-agnostic embeddings that perform exceptionally well across a wide range of downstream tasks, including WSI-level prediction, tile classification, and instance segmentation. BROW [62] used a transformer architecture with a self-distillation framework to extract superior feature representations from WSIs. It incorporates techniques like color augmentation, patch shuffling, and masked image modeling (MIM), while leveraging multi-scale pyramid of the WSI to achieve robust performance across various downstream tasks, such as patch-level classification and slide-level classification. To better leverage pathologist expertise, Dipple et al. [63] proposed RudolfV, a model that integrates medical expert knowledge into data curation and model training with a focus on data diversity. In this study, a dataset of 134k slides from 34k cases was curated, systematically grouped, and clustered to maximize diversity while ensuring balanced disease representation across the dataset. The RudolfV was then trained using an adapted DINOv2 framework with extended augmentations, making it applicable to various tasks in digital pathology.

Patch-level pathology foundation models face the challenge of capturing both local patterns within individual tiles and global patterns across entire slides. Existing models often treat each tile as an independent sample and use MIL for slide-level modeling [26,41,60], which limits their ability to capture complex global patterns in gigapixel WSIs. Additionally, given the limited sample sizes for MIL in many histology datasets, there is an increasing emphasis in digital pathology on transitioning from weakly-supervised to unsupervised methods for slide representation learning. To overcome these challenges, researchers have developed WSI-level pathology foundation models. Prov-GigaPath proposed by Xu et al. [64] is a slide-level pathology foudation model that leverages the GigaPath architecture, a ViT designed for gigapixel WSIs. The pretraining process involves two stages: first, tile-level SSL with DINOv2 and a standard ViT, followed by WSI-level SSL using a MAE with LongNet, which incorporates dilated self-attention to effectively capture both local and global patterns across the entire WSI. Song et al. [65] proposed PANTHER, a prototype-based aggregation framework that leverages morphological redundancy in tissue to create a task-agnostic, unsupervised slide representation by summarizing WSI patches into a set of morphological prototypes using a Gaussian mixture model (GMM), where each patch embedding is generated from the GMM and mapped to a morphological prototype using expectation–maximization (EM) to encapsulate morphological patterns. PANTHER outperforms or matches the performance of supervised MIL baselines across various tasks, providing new insights into model interpretability through per-prototype analysis.

3.2.2. *Multimodal foundation model*

Unlike unimodal foundation models, multimodal foundation models (MFMs) leverage the integration of various modalities, such as pathology images, text, and genetic data, to enhance diagnostic accuracy and interpretability. In this section, we will explore two prominent types of multimodal foundation models: Pathology-Language multimodal foundation models and Pathology-Gene multimodal foundation models, as detailed in Table 2.

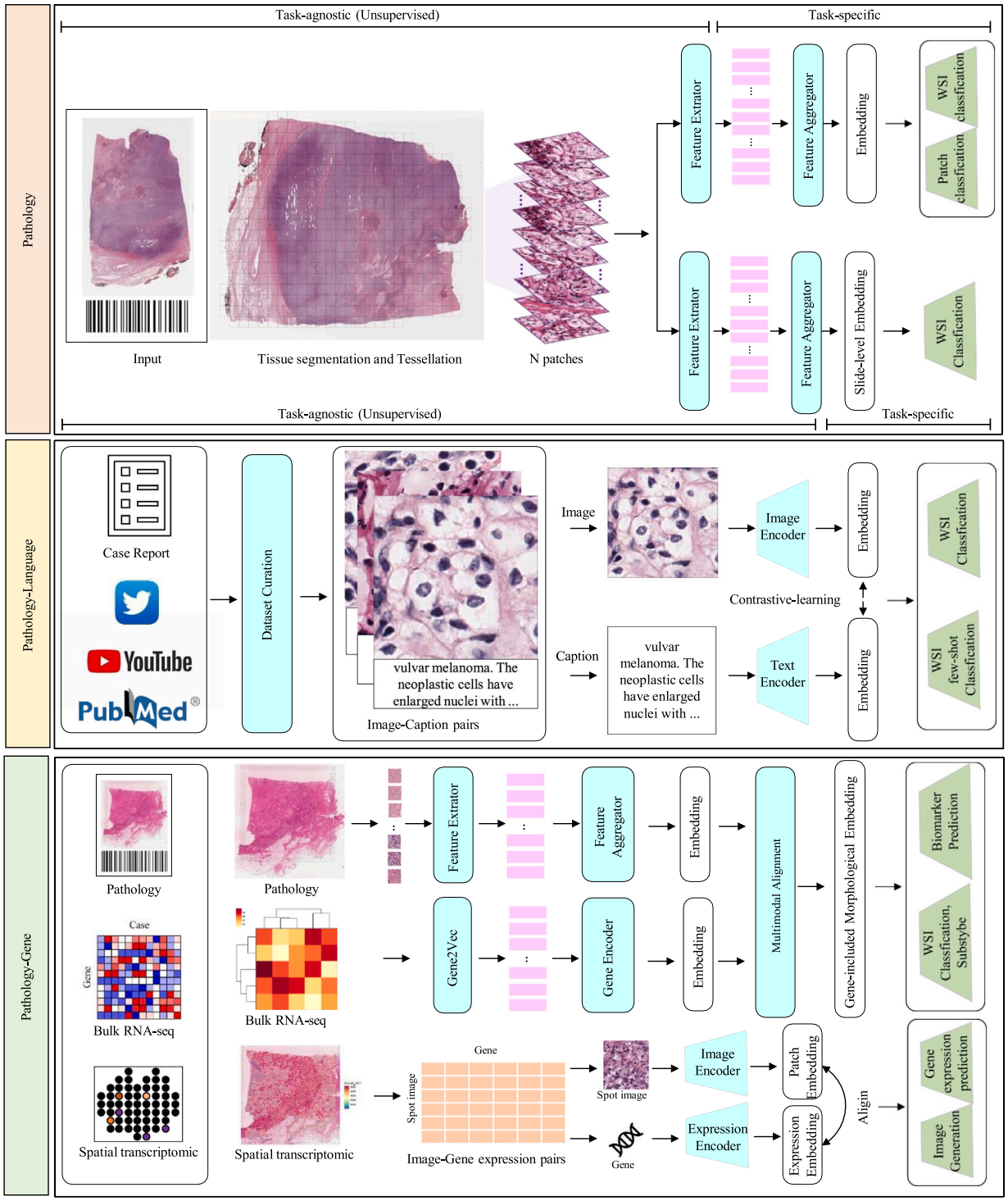


Fig. 5. Comparative analysis of tumor pathological image classification methods using pathology foundation models across different modalities, including Pathology, Pathology-Language, and Pathology-Gene. In the Pathology section, patch-based and WSI-based unimodal pathology foundation models are presented. The Pathology-Language section exclusively features patch-based multimodal pathology foundation models, while the Pathology-Gene section encompasses both patch-based and WSI-based multimodal pathology foundation models.

Pathology-Language multimodal foundation models. Multimodal data, particularly image–text pairs, has become increasingly significant and popular, largely due to the recent achievements of multimodal models such as Contrastive Language-Image Pre-Training (CLIP) and Contrastive Captioners (CoCa). In light of the current research landscape, where WSI-level visual-language pathological foundation models remain undeveloped, this section will concentrate on the introduction and advancements of patch-based visual-language pathological foundation models. Based on the CoCa, Lu et al. [13] introduced CONCH, a visual-language foundation model that underwent task-agnostic pretraining utilizing diverse sources, including histopathology images, biomedical text, and notably, over 1.17 million image–caption pairs from educational sources (EDU) and parts of the PMC dataset.

Table 1
Overview of Patch-based and WSI-based unimodal pathology foundation models.

Model	Type	Dataset source	WSIs	Patches	Architecture	Size	Objective
Virchow [59]	Patch	MSKCC (Proprietary)	1.5 M	2 B	ViT-H	632 M	DINOv2
UNI [60]	Patch	Mass-100K (Proprietary)	100 K	100 M	ViT-L	307 M	DINOv2
CTransPath [26]	Patch	TCGA+PAIP	32 K	15 M	Swin Trans.	28 M	MoCoV3
PLUTO [61]	Patch	TCGA+PathAI	158 K	195 M	FlexiViT-S	22 M	MAE+DINOv2
RudolfV [63]	Patch	TCGA+Proprietary	103 K	750 M	ViT-L	304 M	DINOv2
BROW [62]	Patch	TCGA+Camelyon+Proprietary	11 K	180 M	ViT-B	–	DINO
Prov-GigaPath [64]	WSI	Prov-Path (Proprietary)	171 K	1.3 B	ViT LongNet	23 M	DINOv2
PANTHER [65]	WSI	TCGA	11 K	–	ViT-L+GMM	32 K	DINOv2

Table 2
Overview of Pathology-Language and Pathology-Gene multimodal foundation models.

Model	Type	Modalities	Modality pairs	Backbone	Year
CONCH [13]	Patch	Pathology images, Text	1.2 M	ViT+Transformer	2024
MI-Zero [66]	Patch	Pathology images, Text	33 K	CTransPath+BERT	2023
PLIP [14]	Patch	Pathology images, Text	208 K	CLIP	2024
QUILTNet [15]	Patch	Pathology images, Text	1 M	CLIP	2024
PathChat [67]	Patch	Pathology images, Text	1.2 M	CONCH-ViT+LLaMA-2	2023
PathAsst [68]	Patch	Pathology images, Text	207 K	PLIP-ViT+Vicuna	2024
PathAlign [11]	Patch	Pathology images, Text	350 K	BLIP-2+LLM (PaLM-2 S)	2024
GiMP [16]	Patch	Pathology images, Genomic	946	ResNet+Transformer	2023
Med-PaLMM [69]	Patch	Multimodal images, Text, Genomic	–	PaLM-E	2023
TANGLE [70]	WSI	Pathology images, Transcriptomics	8629	ViT+iBOT+CTransPath	2024
mSTAR [71]	WSI	Pathology images, Transcriptomics	26 K	UNI+CL	2024

The model architecture, comprising an image encoder, a text encoder, and a multimodal fusion decoder, is trained through a combination of objectives: contrastive alignment for aligning image and text modalities within the model's representation space, and a captioning objective for generating the appropriate caption for each image. Lu et al. [66] proposed the MI-Zero, a zero-shot transfer framework that converts WSIs into multiple patches embedded in a visual-language latent space, utilizing set-based or graph-based aggregation strategies to classify pathological images.

Moreover, the growing availability of clinical image data and dissemination of medical knowledge by professionals on public platforms, such as medical Twitter and YouTube, presents an opportunity for further advancement of these models. Huang el al. [14] established OpenPath by curating 243,375 public pathology images identified from popular pathology Twitter hashtags, expanding the dataset with additional internet sources, and rigorously filtering it to yield 208,414 high-quality pathology image–text pairs. They introduced Pathology Language–Image Pretraining (PLIP), which trains on the OpenPath dataset using contrastive learning to optimize embedding vectors generated by text and image encoders, ensuring similarity for paired data and dissimilarity for non-paired data. Ikezogwo el al. [15] curated the QUILT dataset, comprising 800K image–text pairs sourced from YouTube, which was subsequently expanded with data from Twitter and others sources to create the comprehensive QUILT-1M dataset. Built on the CLIP framework, QUILTNET was then pretrained on the QUILT-1M dataset to develop a joint embedding space for image–text pairs through contrastive learning.

With the emergence of large language models (LLMs) [72,73] and the rapid advancements in multimodal large language models (MLLMs) and the broader field of generative AI [74], significant applications in the interpretation of natural images have been demonstrated. This is exemplified by models such as OpenAI's ChatGPT and GPT-4, which have excelled in human interaction through instruction tuning and human feedback. Although there have been efforts to investigate the performance of these models in addressing medical-related queries, their potential to assist professionals and researchers in the field of pathology remains relatively underexplored. Inspired by LLaVA, Lu el al. [67] introduced PathChat, a multimodal generative AI copilt that incorporates a vision encoder for transforming images into feature representations, a multimodal projector to map these features to the LLM's embedding space, and the LLM for processing the integrated visual and textual inputs, ultimately producing responses via auto-regressive next word prediction. Sun el al. [68] developed PathAsst, a multimodal generative foundational model tailored for pathology, by combining the vision encoder PathCLIP with the Vicuna-13b LLM through instruction-tuning. PathAsst integrates eight specialized pathological models, which can be seamlessly invoked within the system to enhance its diagnostic and classification capabilities across various pathology-related tasks. PathAlign, proposed by Ahmed et al. [11], enhances image–text alignment for gigapixel WSIs in digital histopathology using BLIP-2. It achieves vision-language alignment by associating WSIs with corresponding diagnostic texts from pathology reports, thereby facilitating tasks such as classification, cross-modal retrieval, and text generation. Additionally, it aligns the WSI encoder with the pre-trained LLM to generate accurate and contextually relevant text from WSIs.

Pathology-Gene multimodal foundation models. With the advancements in digital pathology and high-throughput sequencing, recent studies have increasingly integrated genomic information into WSIs to enhance the accuracy and specificity of pathological image classification. For instance, Jin et al. [16] proposed the Gene-induced Multimodal Pretraining (GiMP), an image-omic classification framework that integrates genomic data with pathological images. GiMP employs GroupMSA for the structured extraction of gene feature and applies Masked Patch Modeling (MPM) to optimize the learning process within WSIs. By incorporating triplet learning, GiMP effectively models modality relevance, thereby improving classification performance. Building on this concept, Med-PaLM M [69] further expands the multimodal paradigm by integrating genomic and pathological data with clinical language, thereby establishing a unified framework for interpreting a broader spectrum of biomedical information. Med-PaLM M is a comprehensive multimodal generative model that encodes and interprets diverse biomedical data, all utilizing a unified set of model weights.

Current research in multimodal pathology foundation models predominantly focuses on the patch-level approach, which can restrict the ability to capture comprehensive patterns across the entire WSI. This limitation emphasizes the necessity for models that more effectively integrate the full context of the slide. mSTAR [71] introduces a novel whole-slide pretraining paradigm by leveraging the largest multimodal dataset of H&E diagnostic WSIs, pathology reports, and RNA-Seq data. It employs a slide aggregator to integrate multimodal knowledge through slide-level contrastive learning and subsequently transfers this

knowledge to the patch extractor via self-taught training, ensuring alignment between patch-level and slide-level features. mSTAR excelled in both unimodal tasks, such as cancer classification, staging, and biomarker prediction (especially gene mutation prediction), and multimodal tasks like survival analysis and treatment response prediction. TANGLE [70] is a framework for transcriptomics-guided slide representation learning that integrates gene expression profiles and employs a multimodal contrastive learning strategy to effectively align slide and expression embeddings, addressing the challenges of generating slide embeddings from gigapixel WSIs. Trained on extensive datasets from human and rat tissue samples, the framework exhibits significant advantages in few-shot classification, prototype-based classification, and slide retrieval tasks when compared to conventional supervised and SSL methods.

While models like mSTAR have effectively integrated multimodal data and captured whole-slide level spatial information in pathology analysis, they still face challenges in addressing finer spatial relationships. Single-cell RNA sequencing enables analysis at the individual cell level, offering detailed insights into cellular heterogeneity. However, it lacks the spatial context within the tissue. To overcome this limitation, spatial transcriptomics (ST) technology has been developed in recent years. ST simultaneously measures gene expression while preserving spatial information within the tissue, providing detailed gene expression data for each sub-tile. STimage-1K4M [17] and HEST-1k [18] are recognized as representative datasets in the field of ST. To assess their effectiveness, contrastive learning fine-tuning has been conducted on various pre-trained multimodal image–text foundation models using these datasets. This process has enhanced the models' capability to integrate pathology images with corresponding gene expressions.

4. Classification of tumor-related indicators based on pathological images

Pathologists perform tumors diagnosis by observing pathological images, primarily relying on their accumulated medical knowledge and diagnostic experience to identify certain quantifiable or qualitative tumor-related indicators that are discernible through visual inspection. For example, the Nottingham grading system for breast cancer commonly uses three indicators: gland formation, mitotic count, and nuclear pleomorphism. We categorize pathological image classification tasks into two main types: *basic classification tasks* and *advanced classification tasks*. When deep learning was initially applied to pathological image analysis, research on whole-slide image (WSI) prediction primarily focused on basic classification tasks that could be recognized by the human eye, such as tumor classification, grading, and area recognition. As the field progressed, researchers began to explore advanced classification tasks, which involve predicting indicators that pathologists cannot identify with the naked eye, such as molecular subtype prediction, tumor origin prediction, tumor biomarker prediction, and survival prediction. This section will provide a detailed introduction to these two types of tasks, as illustrated in Fig. 6.

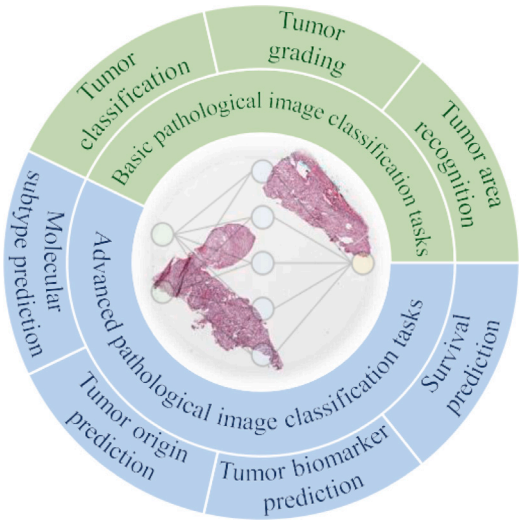


Fig. 6. Tumor-related indicators that can be predicted based-on deep learning and pathological images.

4.1. Basic pathological image classification tasks

4.1.1. Tumor classification

Automated classification of WSIs facilitates the development of detailed treatment plans and improves the precision of prognostic evaluations. Due to the high cost of pixel-level annotation for WSIs, numerous researchers have turned to weakly supervised learning approaches for WSI classification. In digital pathology, these method [19,42,43,47] facilitate automated pathological analysis and diagnostic assistance, thereby reducing subjectivity and individual variability in manual operations while providing consistent, standardized classification results. Pathology image classification methods based on weakly supervised learning have demonstrated robust performance across various cancer types, especially in prostate cancer [46], breast cancer [75], lung cancer [76,77], and renal cancer [19]. For instance, Coudray et al. [77] developed a deep learning method named DeepPATH for accurately predicting cancer subtypes, classifying WSIs into lung adenocarcinoma, lung squamous cell carcinoma, and normal tissue. Campanella et al. [46] trained an instance-based classifier using CNN and MIL to obtain the classification probabilities of all patches in the WSI, and aggregated patches with higher classification probabilities using RNN to train a bag-level classifier, achieving classification of breast cancer, prostate cancer, and basal cell carcinoma WSIs.

4.1.2. Tumor grading

Tumor grading is a fine-grained classification task due to the subtle differences in features observed pathology images of tumors at different grades. It entails the observation and analysis of histopathological images to assess and classify the severity, invasiveness, and prognosis of the tumor. Common grading systems in digital pathology include the Nottingham grading system (NGS) [78], Gleason grading system [79], and Fuhrman grading system [80]. For example, invasive breast cancer is stratified into high, medium, and low differentiation levels based on factors such as mitotic count, proportion of glandular formation, and nuclear morphology, as defined by the NGS. The Gleason grading system categorizes prostate cancer into different grades ranging from 1 to 5 based on the morphological features of the cell nucleus and variations in tissue structure. The Fuhrman grading system categorizes renal cell carcinoma into different grades based on the size, shape, and chromatin features of the renal cell nucleus. These grading systems are designed to assess and prognosticate various cancer types through objective metrics such as cell morphology, nucleolar morphology, mitotic count, and tissue structure. However, grading systems often necessitate pathologists to manually extract relevant attributes based on their experience and knowledge to ensure the accuracy and consistency of grading.

Applying deep learning to tumor grading tasks can alleviate the burden on experts by automating the extraction of relevant features. Specifically, weakly supervised learning can achieve tumor grading tasks based solely on WSIs without the need for pixel-level or patch-level fine-grained annotations. Behzadi et al. [81] proposed a Transformer-based MIL method for Gleason grading in prostate cancer pathology images. The approach involves extracting distinctive patches from WSIs and representing these patches as graphs. Subsequently, GNNs are employed to aggregate information from adjacent patches

within the WSI, thereby improving the accuracy of prostate cancer grading. Additionally, Raju et al. [82] introduced a graph attention MIL method combined with texture features, using GNN and attention-based methods to encode the spatial structure between patches, achieving colorectal cancer grading.

4.1.3. Tumor area recognition

The task of tumor region identification in WSI (also known as tumor region segmentation) is generally transformed into a patch classification problem. However, this approach has two main issues: (1) Inaccurate predictions (such as false positives) because a single patch contains only local information and cannot capture the overall context of the tissue microenvironment and tumor heterogeneity present in WSIs; (2) The boundaries of the predicted regions are coarse. To address these issues, the academic community has conducted extensive exploration.

Early studies on tumor region recognition based on deep learning began with the Camelyon challenge, organized by the International Symposium on Biomedical Imaging (ISBI). One of the tasks in the Camelyon challenge was to detect breast cancer metastases in sentinel lymph nodes from WSIs, which involved identifying and segmenting tumor regions. The Camelyon challenge dataset was the largest publicly available breast cancer WSI dataset at the time, containing 400 WSIs with corresponding manual annotations. Li et al. [83] incorporated a conditional random field algorithm with a CNN-based feature extractor to identify tumor metastasis regions in WSIs, effectively maintaining the contextual relationships between adjacent patches and achieving an average FROC score of 0.8 on the Camelyon16 test set. Wang et al. [84] first segmented the WSI into patches and then trained a CNN classifier on these patches. The patch prediction from the CNN classifier were aggregated to generate a tumor probability heatmap, ultimately achieving tumor identification.

4.2. Advanced pathological image classification tasks

4.2.1. Molecular subtype prediction

Molecular subtype prediction can be traced back to early studies on gene expression profiling. With advancements in high-throughput sequencing, researchers have increasingly utilized gene expression data to identify and classify molecular subtypes of tumors. However, traditional methods for analyzing gene expression data face several challenges, such as data noise, batch effects, and the complexities of high-dimensional data. Unlike molecular techniques, tissue sections strained with H&E are ubiquitously available. If molecular subtypes could be predicted from pathology images, it would facilitate screening, validation, and guidance for the detection of these subtypes. Liu et al. [85] proposed DPMIL, a framework that utilizes discriminative patch selection and MIL to predict four molecular subtypes of breast cancer (Luminal A, Luminal B, Her-2, and Basal-like) from WSIs. DPMIL utilizes attention mechanisms to identify key regions and dynamically adjusts weights based on the contribution of individual instances, thereby improving the prediction performance of breast cancer molecular subtypes. Experimental results demonstrate that DPMIL outperforms experienced pathologists in predicting molecular subtypes from breast cancer pathology images.

4.2.2. Tumor origin prediction

Cancers of unknown primary (CUPs) are metastatic tumors whose primary site cannot be determined; that is, when malignant cells are discovered in the body, but the primary site of the tumor cannot be determined [86,87]. CUPs are among the 10 most common cancers worldwide [88], accounting for 3%–5% of all cancers, and comprise the fourth most common cause of cancer-related deaths [89]. Patients with CUPs typically undergo standard examinations involving IHC, radiology, and endoscopy to determine the site of primary origin. If the primary site cannot be identified, patients with CUPs often undergo empirical treatment, with poor clinical prognosis; the median overall survival is only 6–12 months. Therefore, identifying the site of origin of CUPs can help determine the prognosis of patients, guide personalized clinical treatment, and provide guidance for targeted therapy of specific cancers. The advancement in deep learning and the growth of medical big data have facilitated the development of various software tools, including TOAD [37], CNA_origin [90], CUP-AI-Dx [91], CancerTypePrediction [92], Cancer_origin_prediction [93], CancerLocator [94], and DISMIR [95]. These methods employ deep learning combined with various types of molecular data like genomics, transcriptomics, or epigenomics, to predict the origin of CUPs. However, these methods did not integrate molecular data with histological images for origin prediction. TOAD, proposed by Lu et al. [37], was the first method to use WSIs for tumor origin prediction. It utilizes attention-based MIL, combined with a multi-branch network structure and multi-task learning, to determine whether the tumor is primary or metastatic and identify the primary site of the tumor. Subsequently, Shaban and Lu et al. [96] introduced a MIL-based multimodal deep learning approach that utilizes WSI, genomic data, and patient gender to achieve tumor origin prediction. Experimental results demonstrate that the multimodal fusion approach exhibits superior performance in tumor origin prediction compared to methods that rely solely on WSI or genomic data.

4.2.3. Tumor biomarker prediction

Tumor biomarkers are biological molecules that are closely associated with tumor occurrence, progression, prognosis, and treatment response. They provide critical insights into the biological characteristics of tumors and are widely used in the clinical diagnosis, treatment, and prognosis of cancer. Common tumor biomarkers include gene markers and protein markers, typically discovered through biological experiments. Compared to traditional biological experimental methods, computational approaches offer significant advantages in the prediction and discovery of biomarkers. These advantages include efficient processing and analysis of large and complex datasets, high-throughput capabilities, multidimensional data integration, complex pattern recognition, and high accuracy and sensitivity. Additionally, computational methods support personalized medicine and offer substantial time and cost savings. These benefits position computational techniques as essential tools for advancing the prediction and discovery of tumor biomarkers.

Gene mutation prediction. Gene mutation biomarkers are crucial for precise treatment for cancers. Consequently, clinical guidelines recommend genetic mutation testing for most cancer patients, which is typically performed using molecular biology methods. However, it is expensive and has a high turnaround time. In contrast, tissue sections strained with H&E are ubiquitously available. If genetic mutations could be predicted through pathological images, it would significantly improve the accessibility and efficiency of genetic mutation testing. The pioneering study by Coudray et al. [77] focused on predicting ten common gene mutations in non-small cell lung cancer using WSIs, demonstrating accurate predictions for six mutations: STK11, EGFR, FAT1, SETBP1, KRAS, and TP53. The area under the receiver operating characteristic curve (AUROC) for these predictions on the TCGA test set ranged from 0.733 to 0.856. Yan el al. [45] proposed a framework for WSI-based gene mutation prediction, which include three components: (1) cancerous area segmentation; (2) representative patch selection based on cancerous patch clustering; (3) hierarchical deep multi-instance learning (HDMIL) for gene mutation prediction. It was found that 5 of the 8 clinically relevant gene mutations in bladder cancer, including ATM, PIK3CA, ERBB2, FGFR3, and ERCC2, can be well predicted with an AUROC above 0.83.

Microsatellite status prediction. Microsatellite instability (MSI) is a genetic variation resulting from defects in the DNA mismatch repair (MMR) system. It was first identified in colorectal cancer in 1993, and subsequently observed in other cancer types such as endometrial

cancer, gastric cancer, and cholangiocarcinoma. Based on the degree of instability, MSI is typically classified into three categories: microsatellite instability-high (MSI-H), microsatellite instability-low (MSI-L), and microsatellite stable (MSS). Clinically, there is no significant difference in treatment and prognosis between MSI-L and MSS, resulting in their classification as MSS. The detection of MSI status in colorectal cancer can provide clinicians with crucial prognostic information and guide treatment strategies. However, traditional MSI detection methods are often constrained by high costs, inefficiency, and a limited scope, which present significant challenges for accurate MSI status prediction, particularly in resource-limited settings. In contrast, tissue sections strained with H&E are ubiquitously available. Currently, weakly supervised learning methods based on histopathological images have made significant advancements in predicting MSI status. Cao et al. [97] proposed an integrated multi-instance deep learning method based on histopathological images to predict the microsatellite status of colorectal cancer. The correlation analysis between the identified key histopathological image features and genomic/transcriptomic data provides great interpretation, establishing a foundation for the model's application in prospective clinical trials. Additionally, Hu et al. [98] developed an attention-based MIL method to predict MSI status from prostate cancer by leveraging a large dataset of real-world H&E WSIs and corresponding molecular testing results. Qiu et al. [99] utilized multimodal compact bilinear (MCB) pooling to integrate qualitative features from pathological images with quantitative information from corresponding clinical data to predict MSI status in colorectal cancer. This was the first method to combine pathological image and clinical features, demonstrating higher accuracy compared to existing unimodal prognostic methods.

Homologous recombination deficiency prediction. Homologous recombination deficiency (HRD) prediction is critical for guiding personalized cancer treatment, particularly in identifying patients who may benefit from poly-ADP ribose polymerase inhibitors (PARPi) or platinum-based chemotherapy. HRD occurs when the homologous recombination repair (HRR) pathway, responsible for repairing DNA double-strand breaks, is impaired, leading to increased genomic instability. HRD prediction typically relies on molecular biology assays, which have a high turnaround time, and cost. In contrast, H&E-stained tissue slides are routinely obtained in clinical practice, providing a feasible method for HRD detection. Therefore, recent studies have leveraged deep learning applied to pathology images for HRD prediction. Lazard et al. [100] developed a deep learning-based method for HRD prediction and identified related morphological phenotypes by applying it to a real dataset of untreated breast cancer WSIs. The method demonstrated high accuracy in HRD prediction and revealed morphological phenotypes, such as laminated fibrosis and clear tumor cells, which were associated with HRD, thus providing new insights into its phenotypic manifestations. Schirris et al. [101] introduced DeepSMILE, a novel contrastive SSL method designed to classify MSI and HRD in colorectal and breast cancers using WSIs. Experiments on colorectal and breast cancer datasets demonstrated that DeepSMIL significantly improved the accuracy of MSI and HRD classification through contrastive SSL, underscoring its potential as a valuable tool for early cancer diagnosis and personalized treatment in clinical practice. Although several studies have applied deep learning to pathology images for HRD prediction, research on multimodal approaches that include pathology images remains in its early stages. Currently, no published methods integrating pathology images with other data types for HRD prediction have been identified, highlighting the necessity for continued research and development in this domain.

Tumor mutation burden prediction. Tumor mutation burden (TMB), which refers to the number of non-synonymous mutations in tumor cells, has gained significant attention as a molecular marker with the rise of immunotherapy. The assessment of TMB has been formally integrated into the 2019 clinical practice guidelines for non-small cell lung cancer. However, the clinical application of tissue sequencing methods to determine TMB status is limited by factors such as time, cost, and tissue availability, and may also be compromised by spatial heterogeneity, leading to inconsistent results. Based on pathology images from 253 bladder cancer patients in TCGA, Xu et al. [102] developed a method that achieved 73% accuracy and an AUROC of 0.75 in distinguishing between patients with high and low TMB. The method consists of four modules: (1) tumor region detection, (2) selection of representative patches, (3) feature extraction from the selected patches using transfer learning, and (4) TMB classification of the image features using a SVM. Huang et al. [103] propose a multimodal deep learning framework based on ResNet and MCB pooling for predicting colorectal cancer TMB status directly from histopathological images and clinical information, achieving an AUROC of 0.817 and an accuracy of 0.85.

4.2.4. Survival prediction

Histopathological examination of tissue sections is essential for precise cancer diagnosis and the formulation of appropriate treatment strategies. In clinical practice, survival prediction predominantly relies on the manual assessment of histopathological features. Patients are categorized into different risk groups based on factors such as the extent of tumor invasion, necrosis, and proliferation to guide subsequent treatment. However, the subjective interpretation of histopathological features exhibits considerable variability, leading to low consistency among different observers and resulting in significant prognostic differences even when patients are assigned the same grade or stage. To address these limitations, Zhang et al. [104] proposed an unsupervised deep learning and contrastive clustering framework (DL-CC) for extracting and analyzing histomorphological features from H&E stained histopathological images of small cell lung cancer. They identified 50 histomorphological phenotype clusters and integrated two with significant prognostic value into a pathomics signature (PathoSig). PathoSig, rigorously validated across multicenter datasets, demonstrated effective risk stratification and improved prognostic accuracy.

Despite advancements with PathoSig, the heterogeneity of histopathology, genomics, and transcriptomics within the tumor microenvironment collectively influences responses to treatment and outcomes in cancer patients. Chen et al. [105] proposed a multimodal analysis platform PORPOISE, which integrates pathological image with copy number variations (CNV), gene mutation data, RNA-seq data, and other omics information to predict and interpret the relative risk of cancer mortality. PORPOISE also incorporates the spatial distribution of tumor, stromal, and immune cells within the tumor microenvironment, thereby accounting for the combined impact of such factors on patient risk and grading. Furthermore, the efficacy of the multimodal network was validated on TCGA gliomas and CCRCC, demonstrating its capability for patient stratification and the identification of prognostic features. By synthesizing multiple data sources and spatial distribution information, PORPOISE achieves refined stratification and diagnosis of cancer patients.

5. Challenges and future directions

5.1. Model interpretability

Deep learning-based pathology image analysis methods are often considered "black-box", where the decision-making process is difficult to interpret, posing a significant barrier to clinical application. However, these algorithms are extensively utilized for tasks such as detection and segmentation, where explaining the overall decision-making process is not required. Common approaches to enhancing the interpretability of deep learning include ablation studies [106], feature clustering [107], and class activation map [108]. In digital pathology, attention mechanisms are a prevalent method that visualizes key regions in WSIs to enhance the interpretability of pathology image classification methods. Fig. 7 illustrates the attention visualization for clear cell renal cell carcinoma based on the CLAM proposed by

Table 3
Summary of common color normalization methods in pathological images.

Method	Input transformation	Structure	Versatility	Complexity	Separation
Color Transfer [110]	Color mapping in Lab color space	Yes	Low	Low	Poor
SVD [111]	Converts RGB to optical density space	No	High	Low	Poor
Color Deconvolution [112]	Converts RGB to optical density space	Yes	High	High	Good
Sparse NMF [113]	Stain density transformation	Yes	Low	High	Poor

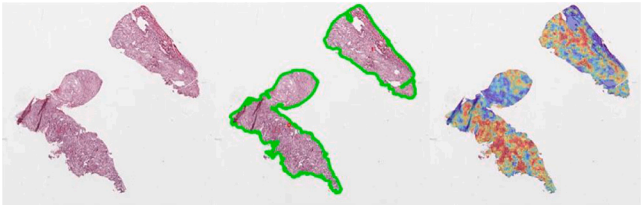


Fig. 7. Attention visualization of clear cell renal cell carcinoma based on CLAM.

Lu et al. [19]. The leftmost image represents the original WSI, the middle image shows the segmented tissue area from the WSI, and the rightmost image displays the attention heat maps based on the CLAM. The model's attention heat map is overlaid on the original WSI as a semi-transparent layer, with regions ranging from crimson (indicating high attention and high diagnostic relevance) to navy (indicating low attention and low diagnostic relevance).

5.2. Model generalization

Model generalization remains a significant bottleneck hindering the clinical deployment of deep learning in pathology image analysis. It is widely recognized that CNNs trained on small datasets often exhibit reduced predictive power, rendering them more susceptible to overfitting and unstable predictions. The application of transfer learning combined with cross validation and ensemble model strategies addresses these technical issues and enhances the generalizability of the results [109]. Variations in staining and acquisition protocols can lead to significant performance disparities when transferring a well-trained model to other datasets. Color normalization techniques, including color transfer, singular value decomposition (SVD), color deconvolution, and sparse non-negative matrix factorization (NMF), are frequently employed in pathology image analysis, alongside data augmentation. Common color normalization methods are shown in Table 3. Typically, gigapixel WSIs are converted to patch-level images for data augmentation or color normalization.

5.3. Shortcomings of labeled data

The widespread adoption of deep learning methods in natural image analysis is primarily due to availability of large-scale annotated datasets, such as ImageNet, Common Objects in Context (COCO), PASCAL Visual Object Classes (VOC), CIFAR-10, and CIFAR-100. However, in digital pathology, fine-grained annotations are sometimes required at the patch or pixel level, while most publicly available WSIs are labeled only at the WSI-level. Unlike natural image labels, which can be annotated through web search engines or crowdsourcing, the labeling of pathological images requires the specialized expertise of pathologists. Furthermore, fine-grained labeling in WSIs is an exceptionally labor-intensive task. To facilitate rapid digital pathology, we have summarized and organized common public available pathological image datasets, as detailed in Tables 4 and 5. Currently, public datasets such as TCGA and CPTAC provide a substantial amount of WSIs, which have fostered the application of weakly supervised learning in digital pathology. However, there is a notable scarcity of publicly available WSIs with fine-grained annotations. Therefore, we encourage individuals or organizations to release WSIs with fine-grained annotations to accelerate the advancement of digital pathology.

5.4. Challenges with artifacts in public datasets

WSIs, which are commonly utilized in computer-aided diagnosis, are generated through a process involving sectioning, mounting, staining, and scanning, each step of which may introduce undesirable artifacts into the data. During sectioning, artifacts may be introduced due to uneven tissue thickness, tissue tremors resulting from blade vibration, and tissue tearing caused by a dull blade. Improper mounting of slides may result in tissue warping, bubbling, and wrinkling. Furthermore, staining irregularities, frequently arising from inadequate sectioning and mounting, can introduce further artifacts. Additionally, the scanning of the slides can be adversely affected by dust and bacterial contamination, and incorrect focus calibration can result in blurred WSIs. In public WSI datasets, tumor regions are occasionally outlined using color markers, which are also considered artifacts. Detecting WSIs that contain artifacts is crucial, as these artifacts can influence the accuracy of subsequent WSI analyses. Table 6 summarizes widely used algorithms for the detection of artifacts, such as blurring, bubbles, and tissue warping and wrinkling, which are often employed as preprocessing steps for WSIs. Currently, typical tools for identifying and excluding image regions with these artifacts from model training and inference include HistoQC [122] and PathProfiler [123]. HistoQC is a digital pathology quality control tool that employ machine learning to automatically detect artifact and quantifies image characteristics through metrics such as contrast, brightness, and color balance, presenting the results visually. PathProfiler is a deep learning framework designed for quality assessment of colorectal cancer tissue slides at both patch-levels and WSI-levels. It implements image quality recognition and control through overall availability heatmaps and can be easily retrained for quality control across various tissue types.

5.5. Computational burden

The extremely high resolution of WSI poses a significant challenge for end-to-end WSI analysis, primarily because the computing facilities required for such analysis are often beyond the reach of general scientific research teams. Developing a computational framework tailored specifically for WSI could enable end-to-end WSI analysis, potentially leading to improved prediction results and the identification of more clinically relevant indicators. Achieving this goal necessitates concurrent research on deep learning methods for ultra-high-resolution images, focusing on areas such as network model optimization and high-performance computing.

6. Conclusion

In the paper, we first introduce the datasets, which include both unimodal and multimodal types, as foundational resources critical for advancing pathology image classification tasks. We then provide a thorough overview and analysis of deep learning-based pathological image classification, encompassing both task-specific models and foundation models, to clarify the current advancements of the field. Subsequently, we review tumor-related indicators that can be predicted from pathological images, addressing two primary aspects: indicators that can be read and recognized by pathologists, such as tumor classification, grading, and region recognition; and those that cannot be read and recognized by pathologists, including molecular subtype prediction, tumor origin prediction, tumor biomarker prediction, and survival prediction.

Table 4
Publicly available WSI-level datasets.

Dataset	Number	Types	Format	Magnification	Link
TCGA	18 394	Multiple tumors	svs	40×, 20×	✓
CPTAC	6106	Multiple tumors	svs	40×, 20×	✓
Camelyon17	1000	Breast cancer metastasis	tif	40×	✓
Camelyon16	400	Breast cancer lymph node metastasis	tif	40×	✓
Warwick QU	165	Colorectal cancer	bmp	20×	✓
MITOS-ATYPIA14	1302	Prostate cancer	tif	40×	✓
NADT-Prostate [114]	1401	Breast cancer	tif	20×	✓
PAIP [26]	2457	Multiple tumors	svs	40×	✓
NLST [115]	1225	Lung cancer	–	40×	✓
MCO [115]	1614	Colorectal cancer	–	40×	✓

Table 5
Manually annotated publicly available patch-level pathological image datasets.

Dataset	Size	Number	Types	Format	Magnification	Link
BACH [116]	2048 × 1536	400	Breast cancer	tif	40×	✓
BACH [117]	2048 × 1536	4020	Breast cancer	tif	20×, 40×	✓
BreakHis [118]	700 × 460	7909	Breast cancer	png	4×, 10×, 20×, and 40×	✓
PathoIDCG [119]	1000 × 1000	18 783	Prostate cancer	jpg	10X	✓
SICAPv2 [120]	512 × 512	3644	Breast cancer	tif	40×, 20×	✓
CRC [121]	4548 × 7520	139	Colorecta cancer	tif	20×	✓
Extended CRC [114]	5000 × 7300	300	Colorecta cancer	tif	20×	✓

Table 6
Artifact detection methods.

Artifact type	Method	Approach	References
Blurring	Feature-based	Texture features with machine learning classifiers	[124,125]
	CNN-based	CNNs for blurring detection	[126,127]
Tissue Warping	Color space transformation	Color variations detection	[128,129]
	CNN-based	CNNs for feature extraction	[130]
Bubbles	DNN-based	DNNs for artifact classification	
Bleeding	CNN-based	Deep learning techniques for tissue detection	[131,132]
	Feature-based	Color/texture features for identification	[133]

Finally, we highlighted the challenges faced in digital pathology and proposed possible solutions. Our review intends to serve as a useful reference for researchers and pathologists, offering clear insights to guide their work. Future research in digital pathology should aim to improve the generalizability and interpretability of deep learning methods, ensuring their efficacy across diverse pathological conditions and applicability in low-resource settings. Promising approaches include the integration of domain adaptation methods to better handle variations in data distributions across different medical institutions and geographic regions. Additionally, collaboration with and pathological experts is essential to ensure that research outcomes are effectively applied in clinical practice, providing support for tumor pathology diagnosis and treatment.

CRediT authorship contribution statement

Haijing Luan: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation. **Kaixing Yang:** Writing – review & editing, Writing – original draft. **Taiyuan Hu:** Writing – review & editing, Writing – original draft. **Jifang Hu:** Writing – review & editing, Visualization. **Siyao Liu:** Writing – review & editing. **Ruilin Li:** Writing – review & editing. **Jiayin He:** Writing – review & editing. **Rui Yan:** Supervision, Writing – original draft. **Xiaobing Guo:** Writing – review & editing, Writing – original draft, Supervision. **Niansong Qian:** Supervision, Methodology. **Beifang Niu:** Writing – review & editing, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant number 92259101) and the Strategic Priority Research Program of the Chinese Academy of Sciences (grant number XDB38040100).

Data availability

No data was used for the research described in the article.

References

[1] B. Xiuwu, P. Yifang, Challenges and opportunities facing the development of pathology in China, J. Third Mil. Med. Univ. (2019) 1815–1817.
[2] S. Deng, X. Zhang, W. Yan, E.I.-C. Chang, Y. Fan, M. Lai, Y. Xu, Deep learning in digital pathology image analysis: a survey, Front. Med. 14 (2020) 18.
[3] F. Xing, L. Yang, Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review, IEEE Rev. Biomed. Eng. 9 (2016) 234–263.
[4] H. Chen, Q. Dou, X. Wang, J. Qin, P. Heng, Mitosis detection in breast cancer histology images via deep cascaded networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 30, No. 1, 2016.
[5] H. Chen, X. Qi, L. Yu, P.-A. Heng, DCAN: Deep contour-aware networks for accurate gland segmentation, in: Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2016, pp. 2487–2496.
[6] Y. Zheng, Z. Jiang, F. Xie, J. Shi, H. Zhang, J. Huai, M. Cao, X. Yang, Diagnostic regions attention network (DRA-Net) for histopathology WSI recommendation and retrieval, IEEE Trans. Med. Imaging 40 (2020) 1090–1103.
[7] Z. Wang, Deep learning in medical ultrasound image segmentation: a review, 2020, http://dx.doi.org/10.48550/arXiv.2002.07703.
[8] S. Alheejawi, R. Berendt, N. Jha, S.P. Maity, M. Mandal, Detection of malignant melanoma in H&E-stained images using deep learning techniques, Tissue Cell 73 (2021) 101659.

[9] K. Jafari-Khouzani, H. Soltanian-Zadeh, Multiwavelet grading of pathological images of prostate, IEEE Trans. Biomed. Eng. 50 (2003) 697–704.

[10] K.-H. Yu, C. Zhang, G.J. Berry, R.B. Altman, C. Ré, D.L. Rubin, M. Snyder, Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features, Nature Commun. 7 (2016) 12474.

[11] F. Ahmed, A. Sellergren, L. Yang, S. Xu, B. Babenko, A. Ward, N. Olson, A. Mohtashamian, Y. Matias, G.S. Corrado, et al., PathAlign: A vision-language model for whole slide images in histopathology, 2024, arXiv preprint arXiv:2406.19578.

[12] R.J. Roberts, PubMed Central: The GenBank of the published literature, Proc. Natl. Acad. Sci. 98 (2) (2001) 381–382.

[13] M.Y. Lu, B. Chen, D.F. Williamson, R.J. Chen, I. Liang, T. Ding, G. Jaume, I. Odintsov, L.P. Le, G. Gerber, et al., A visual-language foundation model for computational pathology, Nat. Med. 30 (3) (2024) 863–874.

[14] Z. Huang, F. Bianchi, M. Yuksekgonul, T.J. Montine, J. Zou, A visual–language foundation model for pathology image analysis using medical twitter, Nat. Med. 29 (9) (2023) 2307–2316.

[15] W. Ikezogwo, S. Seyfioglu, F. Ghezloo, D. Geva, F. Sheikh Mohammed, P.K. Anand, R. Krishna, L. Shapiro, Quilt-1m: One million image-text pairs for histopathology, Adv. Neural Inf. Process. Syst. 36 (2024).

[16] T. Jin, X. Xie, R. Wan, Q. Li, Y. Wang, Gene-induced multimodal pre-training for image-omic classification, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 508–517.

[17] J. Chen, M. Zhou, W. Wu, J. Zhang, Y. Li, D. Li, STimage-1K4M: A histopathology image-gene expression dataset for spatial transcriptomics, 2024, arXiv preprint arXiv:2406.06393.

[18] G. Jaume, P. Doucet, A.H. Song, M.Y. Lu, C. Almagro-Pérez, S.J. Wagner, A.J. Vaidya, R.J. Chen, D.F. Williamson, A. Kim, et al., HEST-1k: A dataset for spatial transcriptomics and histology image analysis, 2024, arXiv preprint arXiv:2406.16192.

[19] M.Y. Lu, D.F. Williamson, T.Y. Chen, R.J. Chen, M. Barbieri, F. Mahmood, Data-efficient and weakly supervised computational pathology on whole-slide images, Nat. Biomed. Eng. 5 (2021) 555–570.

[20] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (1998) 2278–2324.

[21] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, vol. 25, 2012, pp. 1097–1105.

[22] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016.

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015.

[24] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016.

[25] Z. Wang, L. Yu, X. Ding, X. Liao, L. Wang, Lymph node metastasis prediction from whole slide images with transformer-guided multiinstance learning and knowledge transfer, IEEE Trans. Med. Imaging 41 (10) (2022) 2777–2787.

[26] X. Wang, S. Yang, J. Zhang, M. Wang, J. Zhang, W. Yang, J. Huang, X. Han, Transformer-based unsupervised contrastive learning for histopathological image classification, Med. Image Anal. 81 (2022) 102559.

[27] Y. Zheng, J. Li, J. Shi, F. Xie, J. Huai, M. Cao, Z. Jiang, Kernel attention transformer for histopathology whole slide image analysis and assistant cancer diagnosis, IEEE Trans. Med. Imaging 42 (9) (2023) 2726–2739.

[28] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, et al., Bootstrap your own latent-a new approach to self-supervised learning, Adv. Neural Inf. Process. Syst. 33 (2020) 21271–21284.

[29] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, A. Joulin, Emerging properties in self-supervised vision transformers, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 9650–9660.

[30] R.J. Chen, C. Chen, Y. Li, T.Y. Chen, A.D. Trister, R.G. Krishnan, F. Mahmood, Scaling vision transformers to gigapixel images via hierarchical self-supervised learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 16144–16155.

[31] K. Wu, Y. Zheng, J. Shi, F. Xie, Z. Jiang, Position-aware masked autoencoder for histopathology WSI representation learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 714–724.

[32] M. Peikari, S. Salama, S. Nofech-Mozes, A.L. Martel, A cluster-then-label semi-supervised learning approach for pathology image classification, Sci. Rep. 8 (1) (2018) 1–13.

[33] G. Yu, K. Sun, C. Xu, X.-H. Shi, C. Wu, T. Xie, R.-Q. Meng, X.-H. Meng, K.-S. Wang, H.-M. Xiao, et al., Accurate recognition of colorectal cancer with semi-supervised deep learning on pathological images, Nat. Commun. 12 (1) (2021) 6311.

[34] K. Wenger, K. Tirdad, A.D. Cruz, A. Mari, M. Basheer, C. Kuk, B.W. van Rhijn, A.R. Zlotta, T.H. van der Kwast, A. Sadeghian, A semi-supervised learning approach for bladder cancer grading, Mach. Learn. Appl. 9 (2022) 100347.

[35] Z. Gao, B. Hong, Y. Li, X. Zhang, J. Wu, C. Wang, X. Zhang, T. Gong, Y. Zheng, D. Meng, et al., A semi-supervised multi-task learning framework for cancer classification with weak annotation in whole-slide images, Med. Image Anal. 83 (2023) 102652.

[36] J.-N. Eckardt, M. Bornhäuser, K. Wendt, J.M. Middeke, Semi-supervised learning in cancer diagnostics, Front. Oncol. 12 (2022) 960984.

[37] M.Y. Lu, T.Y. Chen, D.F. Williamson, M. Zhao, M. Shady, J. Lipkova, F. Mahmood, AI-based pathology predicts origins for cancers of unknown primary, Nature 594 (2021) 106–110.

[38] J. Li, W. Li, A. Sisk, H. Ye, W.D. Wallace, W. Speier, C.W. Arnold, A multi-resolution model for histopathology image classification and localization with multiple instance learning, Comput. Biol. Med. 131 (2021) 104253.

[39] M.Y. Lu, R.J. Chen, J. Wang, D. Dillon, F. Mahmood, Semi-supervised histology classification using deep multiple instance learning and contrastive predictive coding, 2019, arXiv preprint arXiv:arXiv:1910.10825.

[40] B. Li, Y. Li, K.W. Eliceiri, Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning, in: Computer Vision and Pattern Recognition, IEEE, 2021, pp. 14318–14328.

[41] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, 2017, arXiv e-prints arXiv:arXiv:1706.03762.

[42] Z. Shao, H. Bian, Y. Chen, Y. Wang, J. Zhang, X. Ji, et al., TransMIL: Transformer based correlated multiple instance learning for whole slide image classification, in: Advances in Neural Information Processing Systems, 2021, pp. 2136–2147.

[43] X. Wang, S. Yang, J. Zhang, M. Wang, J. Zhang, J. Huang, W. Yang, X. Han, TransPath: Transformer-based self-supervised learning for histopathological image classification, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 186–195.

[44] H. Chen, C. Li, G. Wang, X. Li, M.M. Rahaman, H. Sun, W. Hu, Y. Li, W. Liu, C. Sun, et al., GasHis-transformer: A multi-scale visual transformer approach for gastric histopathology image classification, in: Pattern Recognition: The Journal of the Pattern Recognition Society, Vol. 130, 2021, 108827.

[45] R. Yan, Y. Shen, X. Zhang, P. Xu, J. Wang, J. Li, F. Ren, D. Ye, S.K. Zhou, Histopathological bladder cancer gene mutation prediction with hierarchical deep multiple-instance learning, Med. Image Anal. 87 (2023) 102824.

[46] G. Campanella, M.G. Hanna, L. Geneslaw, A. Miraflor, V. Werneck Krauss Silva, K.J. Busam, E. Brogi, V.E. Reuter, D.S. Klimstra, T.J. Fuchs, Clinical-grade computational pathology using weakly supervised deep learning on whole slide images, Nat. Med. 25 (2019) 1301–1309.

[47] X. Wang, H. Chen, C. Gan, H. Lin, Q. Dou, E. Tsougenis, Q. Huang, M. Cai, P.-A. Heng, Weakly supervised deep learning for whole slide lung cancer image analysis, IEEE Trans. Cybern. (2019) 1–13.

[48] S. Wang, Y. Zhu, L. Yu, H. Chen, H. Lin, X. Wan, X. Fan, P.-A. Heng, RMDL: Recalibrated multi-instance deep learning for whole slide gastric image classification, Med. Image Anal. 58 (2019) 101549.

[49] Y. Sharma, A. Shrivastava, L. Ehsan, C.A. Moskaluk, S. Syed, D. Brown, Cluster-to-conquer: A framework for end-to-end multi-instance learning for whole slide image classification, in: Proceedings of the Fourth Conference on Medical Imaging with Deep Learning, in: Proceedings of Machine Learning Research, PMLR, 2021, pp. 682–698.

[50] P. Chikontwe, M. Kim, S.J. Nam, H. Go, S.H. Park, Multiple instance learning with center embeddings for histopathology classification, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer International Publishing, 2020, pp. 519–528.

[51] H.-J. Jang, I.-H. Song, S.-H. Lee, Deep learning for automatic subclassification of gastric carcinoma using whole-slide histopathology images, Cancers 13 (15) (2021) 3811.

[52] F. Kanavati, G. Toyokawa, S. Momosaki, M. Rambeau, Y. Kozuma, F. Shoji, K. Yamazaki, S. Takeo, O. Iizuka, M. Tsuneki, Weakly-supervised learning for lung carcinoma classification using deep learning, Sci. Rep. 10 (1) (2020) 9297.

[53] P.L. Schrammen, N. Ghaffari Laleh, A. Echle, D. Truhn, V. Schulz, T.J. Brinker, H. Brenner, J. Chang-Claude, E. Alwers, A. Brobeil, et al., Weakly supervised annotation-free cancer detection and prediction of genotype in routine histopathology, J. Pathol. 256 (1) (2022) 50–60.

[54] X. Wang, C. Zou, Y. Zhang, X. Li, C. Wang, F. Ke, J. Chen, W. Wang, D. Wang, X. Xu, et al., Prediction of BRCA gene mutation in breast cancer based on deep learning and histopathology images, Front. Genet. 12 (2021) 661109.

[55] N.G. Laleh, H.S. Muti, C.M.L. Loeffler, A. Echle, O.L. Saldanha, F. Mahmood, M.Y. Lu, C. Trautwein, R. Langer, B. Dislich, et al., Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology, Med. Image Anal. 79 (2022) 102474.

[56] J. Kaplan, S. McCandlish, T. Henighan, T.B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, D. Amodei, Scaling laws for neural language models, 2020, arXiv preprint arXiv:2001.08361.

[57] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248–255.

[58] C. Sun, A. Shrivastava, S. Singh, A. Gupta, Revisiting unreasonable effectiveness of data in deep learning era, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 843–852.

[59] E. Vorontsov, A. Bozkurt, A. Casson, G. Shaikovski, M. Zelechowski, K. Severson, E. Zimmermann, J. Hall, N. Tenenholtz, N. Fusi, et al., A foundation model for clinical-grade computational pathology and rare cancers detection, Nat. Med. (2024) 1–12.

[60] R.J. Chen, T. Ding, M.Y. Lu, D.F. Williamson, G. Jaume, A.H. Song, B. Chen, A. Zhang, D. Shao, M. Shaban, et al., Towards a general-purpose foundation model for computational pathology, Nat. Med. 30 (3) (2024) 850–862.

[61] D. Juyal, H. Padigela, C. Shah, D. Shenker, N. Harguindeguy, Y. Liu, B. Martin, Y. Zhang, M. Nercessian, M. Markey, et al., PLUTO: Pathology-Universal Transformer, 2024, arXiv preprint arXiv:2405.07905.

[62] Y. Wu, S. Li, Z. Du, W. Zhu, BROW: Better features for whole slide image based on self-distillation, 2023, arXiv preprint arXiv:2309.08259.

[63] J. Dippel, B. Feulner, T. Winterhoff, S. Schallenberg, G. Dernbach, A. Kunft, S. Tietz, P. Jurmeister, D. Horst, L. Ruff, et al., RudolfV: a foundation model by pathologists for pathologists, 2024, arXiv preprint arXiv:2401.04079.

[64] H. Xu, N. Usuyama, J. Bagga, S. Zhang, R. Rao, T. Naumann, C. Wong, Z. Gero, J. González, Y. Gu, et al., A whole-slide foundation model for digital pathology from real-world data, Nature (2024) 1–8.

[65] A.H. Song, R.J. Chen, T. Ding, D.F. Williamson, G. Jaume, F. Mahmood, Morphological prototyping for unsupervised slide representation learning in computational pathology, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 11566–11578.

[66] M.Y. Lu, B. Chen, A. Zhang, D.F. Williamson, R.J. Chen, T. Ding, L.P. Le, Y.-S. Chuang, F. Mahmood, Visual language pretrained multiple instance zero-shot transfer for histopathology images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 19764–19775.

[67] M.Y. Lu, B. Chen, D.F. Williamson, R.J. Chen, M. Zhao, A.K. Chow, K. Ikemura, A. Kim, D. Pouli, A. Patel, et al., A multimodal generative AI copilot for human pathology, Nature (2024) 1–3.

[68] Y. Sun, C. Zhu, S. Zheng, K. Zhang, L. Sun, Z. Shui, Y. Zhang, H. Li, L. Yang, Pathasst: A generative foundation ai assistant towards artificial general intelligence of pathology, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 38, No. 5, 2024, pp. 5034–5042.

[69] T. Tu, S. Azizi, D. Driess, M. Schaekermann, M. Amin, P.-C. Chang, A. Carroll, C. Lau, R. Tanno, I. Ktena, et al., Towards generalist biomedical AI, NEJM AI 1 (3) (2024) AIoa2300138.

[70] G. Jaume, L. Oldenburg, A. Vaidya, R.J. Chen, D.F. Williamson, T. Peeters, A.H. Song, F. Mahmood, Transcriptomics-guided slide representation learning in computational pathology, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 9632–9644.

[71] Y. Xu, Y. Wang, F. Zhou, J. Ma, S. Yang, H. Lin, X. Wang, J. Wang, L. Liang, A. Han, et al., A multimodal knowledge-enhanced whole-slide pathology foundation model, 2024, arXiv preprint arXiv:2407.15362.

[72] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al., Training language models to follow instructions with human feedback, Adv. Neural Inf. Process. Syst. 35 (2022) 27730–27744.

[73] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al., Llama 2: Open foundation and fine-tuned chat models, 2023, arXiv preprint arXiv:2307.09288.

[74] M. Moor, O. Banerjee, Z.S.H. Abad, H.M. Krumholz, J. Leskovec, E.J. Topol, P. Rajpurkar, Foundation models for generalist medical artificial intelligence, Nature 616 (7956) (2023) 259–265.

[75] P. Sudharshan, C. Petitjean, F. Spanhol, L.E. Oliveira, L. Heutte, P. Honeine, Multiple instance learning for histopathological breast cancer image classification, Expert Syst. Appl. 117 (2019) 103–111.

[76] M. Adnan, S. Kalra, H.R. Tizhoosh, Representation learning of histopathology images using graph neural networks, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, 2020.

[77] N. Coudray, P.S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A.L. Moreira, N. Razavian, A. Tsirigos, Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning, Nat. Med. 24 (2018) 1559–1567.

[78] M. Fang, L. Jianying, X. Weicheng, Histological grading of invasive breast cancer: Nottingham histological grading system, Chinese J. Pathol. 48 (2019) 6.

[79] S. Otálora, M. Atzori, A. Khan, O. Jimenez-del Toro, V. Andrearczyk, H. Müller, Systematic comparison of deep learning strategies for weakly supervised gleason grading, in: SPIE Medical Imaging, 2020.

[80] S. Menon, G. Bahirwade, G. Bakshi, G. Prakash, A. Joshi, S. Desai, Comparison of fuhrman nuclear grading system with a novel tumour grading system for chromophobe renal cell carcinoma and its correlation with clinical outcome, Eur. Urol. Suppl. 17 (2018) e2921.

[81] M.M. Behzadi, M. Madani, H. Wang, J. Bai, A. Bhardwaj, A. Tarakanova, H. Yamase, G.H. Nam, S. Nabavi, Weakly-supervised deep learning model for prostate cancer diagnosis and gleason grading of histopathology images, 2022, arXiv preprint, arXiv:arXiv:.12844.

[82] A. Raju, J. Yao, M.M. Haq, J. Jonnagaddala, J. Huang, Graph attention multi-instance learning for accurate colorectal cancer staging, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V, Springer-Verlag, 2020, pp. 529–539.

[83] Y. Li, W. Ping, Cancer metastasis detection with neural conditional random field, 2018, arXiv preprint arXiv:1806.07064.

[84] D. Wang, A. Khosla, R. Gargeya, H. Irshad, A.H. Beck, Deep learning for identifying metastatic breast cancer, 2016, arXiv preprint arXiv:1606.05718.

[85] H. Liu, W.-D. Xu, Z.-H. Shang, X.-D. Wang, H.-Y. Zhou, K.-W. Ma, H. Zhou, J.-L. Qi, J.-R. Jiang, L.-L. Tan, et al., Breast cancer molecular subtype prediction on pathological images with discriminative patch selection and multi-instance learning, Front. Oncol. 12 (2022) 858453.

[86] N. Pavlidis, G. Pentheroudakis, Cancer of unknown primary site, Lancet 379 (2012) 1428–1435.

[87] D. Petrakis, G. Pentheroudakis, E. Voulgaris, N. Pavlidis, Prognostication in cancer of unknown primary (CUP): development of a prognostic algorithm in 311 cases and review of the literature, Cancer Treat. Rev. 39 (2013) 701–708.

[88] N. Pavlidis, K. Fizazi, Cancer of unknown primary (CUP), Crit. Rev. Oncol. Hematol. 54 (2005) 243–250.

[89] E.T. Krementz, E.J. Cerise, D.S. Foster, L.R. Morgan Jr., Metastases of undetermined source, Curr. Probl. Cancer 4 (1979) 1–37.

[90] Y. Liang, H. Wang, J. Yang, X. Li, C. Dai, P. Shao, G. Tian, B. Wang, Y. Wang, A deep learning framework to predict tumor tissue-of-origin based on copy number alteration, Front. Bioeng. Biotechnol. 8 (2020) 701.

[91] Y. Zhao, Z. Pan, S. Namburi, A. Pattison, A. Posner, S. Balachander, C.A. Paisie, H.V. Reddi, J. Rueter, A.J. Gill, et al., CUP-AI-Dx: a tool for inferring cancer tissue of origin and molecular subtype using RNA gene-expression data and artificial intelligence, EBioMedicine 61 (2020) 103030.

[92] M. Mostavi, Y.-C. Chiu, Y. Huang, Y. Chen, Convolutional neural network models for cancer type prediction based on gene expression, BMC Med. Genom. 13 (2020) 1–13.

[93] C. Zheng, R. Xu, Predicting cancer origins with a DNA methylation-based deep neural network model, PLoS One 15 (2020) e0226461.

[94] S. Kang, Q. Li, Q. Chen, Y. Zhou, S. Park, G. Lee, B. Grimes, K. Krysan, M. Yu, W. Wang, et al., CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA, Genome Biol. 18 (2017) 1–12.

[95] J. Li, L. Wei, X. Zhang, W. Zhang, H. Wang, B. Zhong, Z. Xie, H. Lv, X. Wang, DISMIR: Deep learning-based noninvasive cancer detection by integrating DNA sequence and methylation information of individual cell-free DNA reads, Brief. Bioinform. 22 (2021) bbab250.

[96] M. Shaban, M.Y. Lu, D.F. Williamson, R.J. Chen, J. Lipkova, T.Y. Chen, F. Mahmood, Abstract PR005: Deep learning-based multimodal integration of histology and genomics improves cancer origin prediction, Cancer Res. 83 (2023) PR005.

[97] R. Cao, F. Yang, S.-C. Ma, L. Liu, Y. Zhao, Y. Li, D.-H. Wu, T. Wang, W.-J. Lu, W.-J. Cai, et al., Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in colorectal cancer, Theranostics (2020) 11080–11091.

[98] Q. Hu, A.A. Rizvi, G. Schau, K. Ingale, Y. Muller, R. Baits, S. Pretzer, A. BenTaieb, A. Gordhamer, R. Nussenzveig, et al., Development and validation of a deep learning-based microsatellite instability predictor from prostate cancer whole-slide images, NPJ Precis. Oncol. 8 (1) (2024) 88.

[99] W. Qiu, J. Yang, B. Wang, J. Yang, G. Tian, P. Wang, J. Yang, Predicting microsatellite instability in colorectal cancer based on a novel multimodal fusion deep learning model integrating both histopathological images and clinical information, Cell Rep. Methods (2022).

[100] T. Lazard, G. Bataillon, P. Naylor, T. Popova, F.-C. Bidard, D. Stoppa-Lyonnet, M.-H. Stern, E. Decencière, T. Walter, A. Vincent-Salomon, Deep learning identifies morphological patterns of homologous recombination deficiency in luminal breast cancers from whole slide images, Cell Rep. Med. 3 (12) (2022).

[101] Y. Schirris, E. Gavves, I. Nederlof, H.M. Horlings, J. Teuwen, DeepSMILE: Contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer, Med. Image Anal. 79 (2022) 102464.

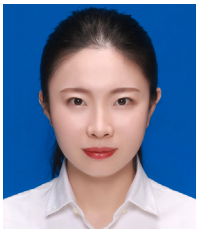
[102] H. Xu, S. Park, S.H. Lee, T.H. Hwang, Using transfer learning on whole slide images to predict tumor mutational burden in bladder cancer patients, 2019, 554527, BioRxiv.

[103] K. Huang, B. Lin, J. Liu, Y. Liu, J. Li, G. Tian, J. Yang, Predicting colorectal cancer tumor mutational burden from histopathological images and clinical information using multi-modal deep learning, Bioinformatics 38 (22) (2022) 5108–5115.

[104] Y. Zhang, Z. Yang, R. Chen, Y. Zhu, L. Liu, J. Dong, Z. Zhang, X. Sun, J. Ying, D. Lin, et al., Histopathology images-based deep learning prediction of prognosis and therapeutic response in small cell lung cancer, NPJ Digit. Med. 7 (1) (2024) 15.

[105] R.J. Chen, M.Y. Lu, D.F. Williamson, T.Y. Chen, J. Lipkova, M. Shaban, M. Shady, M. Williams, B. Joo, Z. Noor, et al., Pan-cancer integrative histology-genomic analysis via interpretable multimodal deep learning, Cancer Cell 40 (2021) 865–878.

[106] T.S. Sheikh, Y. Lee, M. Cho, Histopathological classification of breast cancer images using a multi-scale input and multi-feature network, Cancers 12 (2020) 2031.
[107] P. Ström, K. Kartasalo, H. Olsson, L. Solorzano, B. Delahunt, D.M. Berney, D.G. Bostwick, A.J. Evans, D.J. Grignon, P.A. Humphrey, et al., Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study, Lancet Oncol. 21 (2020) 222–232.
[108] A.-C. Woerl, M. Eckstein, J. Geiger, D.C. Wagner, T. Daher, P. Stenzel, A. Fernandez, A. Hartmann, M. Wand, W. Roth, et al., Deep learning predicts molecular subtype of muscle-invasive bladder cancer from conventional histopathological slides, Eur. Urol. 78 (2020) 256–264.
[109] M. Khened, A. Kori, H. Rajkumar, G. Krishnamurthi, B. Srinivasan, A generalized deep learning framework for whole-slide image segmentation and analysis, Sci. Rep. 11 (1) (2021) 11579.
[110] E. Reinhard, M. Adhikhmin, B. Gooch, P. Shirley, Color transfer between images, IEEE Comput. Graph. Appl. 21 (2001) 34–41.
[111] M. Macenko, M. Niethammer, J.S. Marron, D. Borland, J.T. Woosley, X. Guan, C. Schmitt, N.E. Thomas, A method for normalizing histology slides for quantitative analysis, in: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009, pp. 1107–1110.
[112] A.M. Khan, N. Rajpoot, D. Treanor, D. Magee, A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution, IEEE Trans. Biomed. Eng. 61 (2014) 1729–1738.
[113] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A.M. Schlitter, I. Esposito, N. Navab, Structure-preserving color normalization and sparse stain separation for histological images, IEEE Trans. Med. Imaging 35 (2016) 1962–1971.
[114] M. Shaban, R. Awan, M.M. Fraz, A. Azam, Y.-W. Tsang, D. Snead, N.M. Rajpoot, Context-aware convolutional neural network for grading of colorectal cancer histology images, IEEE Trans. Med. Imaging 39 (7) (2020) 2395–2405.
[115] J. Yao, X. Zhu, J. Jonnagaddala, N. Hawkins, J. Huang, Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks, Med. Image Anal. 65 (2020) 101789, http://dx.doi.org/10.1016/j.media.2020.101789.
[116] G. Aresta, T. Araújo, S. Kwok, S.S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan, et al., BACH: Grand challenge on breast cancer histology images, Med. Image Anal. 56 (2019) 122–139.
[117] R. Yan, F. Ren, Z. Wang, L. Wang, T. Zhang, Y. Liu, X. Rao, C. Zheng, F. Zhang, Breast cancer histopathological image classification using a hybrid deep neural network, Methods 173 (2020) 52–60.
[118] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, IEEE Trans. Biomed. Eng. 63 (7) (2015) 1455–1462.
[119] R. Yan, F. Ren, J. Li, X. Rao, Z. Lv, C. Zheng, F. Zhang, Nuclei-guided network for breast cancer grading in HE-stained pathological images, Sensors 22 (11) (2022) 4061.
[120] J. Silva-Rodríguez, A. Colomer, M.A. Sales, R. Molina, V. Naranjo, Going deeper through the Gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection, Comput. Methods Programs Biomed. 195 (2020) 105637.
[121] R. Awan, K. Sirinukunwattana, D. Epstein, S. Jefferyes, U. Qidwai, Z. Aftab, I. Mujeeb, D. Snead, N. Rajpoot, Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images, Sci. Rep. 7 (1) (2017) 16852.
[122] A. Janowczyk, R. Zuo, H. Gilmore, M. Feldman, A. Madabhushi, HistoQC: An open-source quality control tool for digital pathology slides, JCO Clin. Cancer Inform. 3 (2019) 1–7.
[123] M. Haghighat, L. Browning, K. Sirinukunwattana, S. Malacrino, N. Khalid Alham, R. Colling, Y. Cui, E. Rakha, F.C. Hamdy, C. Verrill, et al., Automated quality assessment of large digitised histology cohorts by artificial intelligence, Sci. Rep. 12 (2022) 5002.
[124] H. Wu, J.H. Phan, A.K. Bhatia, C.A. Cundiff, B.M. Shehata, M.D. Wang, Detection of blur artifacts in histopathological whole-slide images of endomyocardial biopsies, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2015, pp. 727–730.
[125] G. Campanella, A.R. Rajanna, L. Corsale, P.J. Schüffler, Y. Yagi, T.J. Fuchs, Towards machine learned quality control: A benchmark for sharpness quantification in digital pathology, Comput. Med. Imaging Graph. 65 (2018) 142–151.
[126] T. Albuquerque, A. Moreira, J.S. Cardoso, Deep ordinal focus assessment for whole slide images, in: 2021 IEEE/CVF International Conference on Computer Vision Workshops, ICCVW, 2021, pp. 657–663.
[127] T. Kohlberger, Y. Liu, M. Moran, P.-H.C. Chen, T. Brown, J.D. Hipp, C.H. Mermel, M.C. Stumpe, Whole-slide image focus quality: Automatic assessment and impact on AI cancer detection, J. Pathol. Inform. 10 (2019) 39.
[128] P.A. Bautista, Y. Yagi, Detection of tissue folds in whole slide images, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, 2009, pp. 3669–3672.
[129] S. Palokangas, J. Selinummi, O. Yli-Harja, Segmentation of folds in tissue section images, in: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, pp. 5641–5644.
[130] H.M. Shakhawat, T. Nakamura, F. Kimura, Y. Yagi, M. Yamaguchi, Automatic quality evaluation of whole slide images for the practical use of whole slide imaging scanner, ITE Trans. Media Technol. 8 (2020) 252–268.
[131] R. Wetteland, K. Engan, T. Eftestøl, V. Kvikstad, E.A. Janssen, Multiclass tissue classification of whole-slide histological images using convolutional neural networks, in: ICPRAM 2019, Vol. 1, 2019, pp. 320–327.
[132] Z. Swiderska-Chadaj, T. Markiewicz, J. Gallego, G. Bueno, B. Grala, M. Lorent, Deep learning for damaged tissue detection and segmentation in Ki-67 brain tumor specimens based on the U-net model, Bull. Pol. Acad. Sci. Tech. Sci. 66 (2018) 849–856.
[133] E. Mercan, S. Aksoy, L.G. Shapiro, D.L. Weaver, T.T. Brunyé, J.G. Elmore, Localization of diagnostically relevant regions of interest in whole slide images, in: 2014 22nd International Conference on Pattern Recognition, IEEE, 2014, pp. 1179–1184.



Haijing Luan is currently pursuing the Ph.D. degree in the Institute of the Computer Network Information Center at Chinese Academy of Sciences, Beijing, China. She is mainly engaged in high-performance computing, digital pathology, deep learning, biomarker prediction and cancer genomics.

Kaixing Yang received the B.S. degree from the Department of Computer Science and Technology, Yanshan University, in 2023. She is currently pursuing the M.S. degree in Beijing Institute of Genomics Chinese Academy of Sciences, Beijing, China. She is mainly engaged in high-performance computing, and deep learning.

Taiyuan Hu is currently pursuing the Ph.D. degree in computer science at the Computer Network Information Center, Chinese Academy of Sciences, Beijing, China. His research focuses on high-performance computing, deep learning, medical image analysis, and biomarker prediction.

Jifang Hu received the B.S. degree in Computer Science and Technology from Shandong Normal University, in 2020. He is currently pursuing the Ph.D. degree in Computer Network Information Center at Chinese Academy of Sciences, Beijing, China. He is mainly engaged in high-performance computing, deep learning and digital pathology.

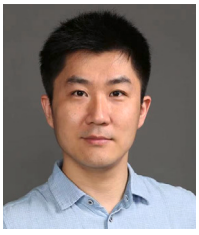
Siyao Liu is currently employed by the Beijing ChosenMed Clinical Laboratory Co. Ltd. She adept at applying biostatistics methods in biomedical research. Additionally, she is skilled in the establishment of computational models through various machine learning or bioinformatics mining algorithms, and excels in the screening of tumor-related biomarkers.

Ruilin Li currently works at the Computer Network Information Center, Chinese Academy of Sciences as an assistant research professor. She received her Ph.D. degree in computer software and theory from the University of the Chinese Academy of Sciences. She has long been engaged in research in the interdisciplinary field of high-performance computing and bioinformatics. Her main research interests include cancer genomics, data mining, large-scale model development, and artificial intelligence.

Jiayin He received M.S. degree in Statistics from the George Washington University. She is currently conducting research work at the Computer Network Information Center, Chinese Academy of Sciences. Her research interests include biostatistics and computational statistics.

Rui Yan received the Ph.D. degree in computer science from University of Chinese Academy of Sciences, China in 2023. He is currently a postdoctoral researcher at School of Biomedical Engineering, University of Science and Technology of China. His research interests include deep learning, bioinformatics, and medical image analysis.

Xiaobing Guo received the B.S. degree and M.S. degree from the Department of Computer Science, Xi'an Jiaotong University, Xian, Shanxi, in 1999 and 2002, respectively. He received the Ph.D. degree in the Institute of Computing Technology at Chinese Academy of Sciences, Beijing, China, in 2013. He is currently a distinguished researcher within Lenovo Research. His current research interests focus on privacy enhanced computation and AI applications.



Niansong Qian is the Deputy Chief Physician of the Department of Thoracic Oncology, Senior Department of Respiratory and Critical CareMedicine, the Eighth Medical Center of PLA General Hospital. He is experienced in the diagnosis of common respiratory diseases, and has long been engaged in the precise and comprehensive treatments of chemotherapy, targeting, and immunotherapy for lung cancer and other cancers.

Beifang Niu is a Professor with the Computer Network Information Center, Chinese Academy of Sciences, and the University of Chinese Academy of Sciences, China. He received Ph.D. degree from the Computer Network Information Center, Chinese Academy of Sciences in 2009. His research interests include the research of high-performance computing technology for biological and medical big data.