Towards label-efficient automatic diagnosis and analysis: a comprehensive survey of advanced deep learning-based weakly-supervised, semi-supervised and self-supervised techniques in histopathological image analysis

Linhao Qu, Siyu Liu, Xiaoyu Liu, Manning Wang and Zhijian Song

Digital Medical Research Center, School of Basic Medical Science, Fudan University, Shanghai Key Lab of Medical Image Computing and Computer Assisted Intervention, Shanghai 200032, People’s Republic of China

* Authors to whom any correspondence should be addressed.

E-mail: lhqu20@fudan.edu.cn, mnwang@fudan.edu.cn and zjsong@fudan.edu.cn

Keywords: histopathological images, automatic analysis, deep learning

Abstract

Histopathological images contain abundant phenotypic information and pathological patterns, which are the gold standards for disease diagnosis and essential for the prediction of patient prognosis and treatment outcome. In recent years, computer-automated analysis techniques for histopathological images have been urgently required in clinical practice, and deep learning methods represented by convolutional neural networks have gradually become the mainstream in the field of digital pathology. However, obtaining large numbers of fine-grained annotated data in this field is a very expensive and difficult task, which hinders the further development of traditional supervised algorithms based on large numbers of annotated data. More recent studies have started to liberate from the traditional supervised paradigm, and the most representative ones are the studies on weakly supervised learning paradigm based on weak annotation, semi-supervised learning paradigm based on limited annotation, and self-supervised learning paradigm based on pathological image representation learning. These new methods have led a new wave of automatic pathological image diagnosis and analysis targeted at annotation efficiency. With a survey of over 130 papers, we present a comprehensive and systematic review of the latest studies on weakly supervised learning, semi-supervised learning, and self-supervised learning in the field of computational pathology from both technical and methodological perspectives. Finally, we present the key challenges and future trends for these techniques.

1. Introduction

Histopathological images contain abundant phenotypic information and pathological patterns, which are the gold standards for disease diagnosis and essential for the prediction of patient prognosis and treatment outcome (Wang et al 2019, Myronenko et al 2021, Srinidhi et al 2021). For clinical diagnosis, experienced pathologists usually require exhaustive examination and interpretation of hematoxylin-eosin-stained (H&E) tissue slides under a high magnification microscope, including differentiation of tumor areas from large areas of normal tissues, elaborate grading of tumors, and detailed assessment of tumor progression and invasion (e.g. presence of invasive carcinoma or proliferative changes, etc.). This is a highly time-consuming and labor-intensive task, and for example, it usually takes an experienced histopathologist 15–30 min to examine a complete slide (Wang et al 2019). Moreover, even an experienced pathologist may not be able to accurately determine the deep features hidden in the pathological images, such as predicting lymph node metastasis and prognosis from the primary lesion. Therefore, computer-assisted automatic analysis techniques for histopathological images are in urgent need in clinical practice.
With the advent and development of digital slide scanners in the past two decades, tissues on biopsies can be converted into digital whole slide images (WSIs) that fully preserve the original tissue structure, laying the foundation for automatic pathological image analysis. Early studies in the field of digital pathology diagnosis primarily focused on extracting hand-crafted features from manually selected regions of interest (ROI) by pathologists (Jafar-Khouzani and Soltanian-Zadeh 2003, Basavanahally et al 2013, Yu et al 2016, Quaiser et al 2016, Luo et al 2017, Mercan et al 2017) and using machine learning methods (Rajpoot and Rajpoot 2004, Doyle et al 2006, 2007, Qureshi et al 2008) for automatic analysis and diagnosis. In this regard, Gurcan et al (2009) and Madabhushi and Lee (2016) have presented an elaborate review.

In recent years, thanks to the powerful and automatic feature extraction capability, deep learning methods represented by convolutional neural network (CNN) have gradually become the mainstream in the field of digital pathology. However, a major challenge is the huge size of WSIs, typically reaching 100 000 × 100 000 pixels at the highest resolution, which prevents the direct use of the entire WSIs as the input to deep learning models. Therefore, when using CNNs to process pathological images, WSIs are usually tiled into many small patches to reduce the computational burden. Earlier studies usually adopted a strongly supervised approach based on these patches to train the network and perform the corresponding classification (Cruz-Roa et al 2014, 2017, Wei et al 2019, Ehteshami Bejnordi et al 2018, Nagpal et al 2019, Shaban et al 2019, Halicek et al 2019) and segmentation tasks (Chen et al 2017, Gu et al 2018, Swiderska-Chadaj et al 2019). In these works, detailed patch-level annotation is essential, e.g. supervised classification problems require pathologists to give detailed class labels for each patch, and segmentation problems require pathologists to give more detailed pixel-level annotation for each patch.

Although supervised deep learning methods have achieved unprecedented success in digital pathology, they share a common drawback: they all require large amounts of high-quality fine-grained labeled data (patch-level labeled data for classification problems or pixel-level labeled data for segmentation problems) for training. Unfortunately, in the field of digital pathology, obtaining a large amount of data with fine-grained annotation is a very expensive and challenging task, mainly because (1) only experienced pathologists can perform the annotation, and these pathologists are scarce; (2) histopathological images often contain complex and diverse instances of objects, resulting in a large amount of time-consuming and laborious manual annotation effort (Yang et al 2017, Tajbakhsh et al 2020, Srinidhi et al 2021). Arguably, the lack of a large amount of annotated data limits the application of deep learning techniques in computational pathology. For this reason, some new studies have recently attempted to liberate from the traditional strongly supervised paradigms, the most representative of which are the weakly supervised learning paradigm based on weak annotations, the semi-supervised learning paradigm based on limited annotations, and the self-supervised paradigm based on the representation learning of pathological images.

The weakly supervised learning paradigm no longer requires pathologists to give annotations of all pixels or regions on the entire WSI, but only class labels or sparse region annotations on the entire WSI; the semi-supervised learning paradigm no longer requires pathologists to give fine-grained annotations of a large amount of data, but only a small fraction of fine-grained labeled data and a large amount of unlabeled data; while the self-supervised learning paradigm can create supervised information through a large amount of unlabeled data for self-supervised training to learn an accurate feature representation of the data. In the process of training with limited labeled data, using the features trained by self-supervised learning to determine the initial model weights can significantly improve the performance of the model. Therefore, weakly supervised learning, semi-supervised learning and self-supervised learning are leading a new study direction of the automatic diagnosis and analysis for pathological images.

However, there are very few related reviews. Srinidhi et al (2021) reviewed representative supervised learning, weakly supervised learning, unsupervised learning, and transfer learning studies in the field of computational pathology until December 2019. Romy et al (2019) reviewed representative weakly supervised learning studies until 2020. Nevertheless, in recent years, deep learning techniques have been developing rapidly and the new techniques continue to emerge. Therefore, a review regarding the applications of these techniques in the automatic diagnosis of pathological images has important theoretical value and clinical significance.

In this review, we summarize more than 130 recent technical studies systematically on weakly supervised learning, semi-supervised learning, and self-supervised learning in the field of computational pathology. We performed this extensive review by searching Google Scholar, PubMed, and arXiv for papers including keywords such as (‘deep learning’ or ‘weakly supervised learning’ or ‘semi-supervised learning’ or ‘self-supervised learning’) and (‘digital pathology’ or ‘histopathology’ or ‘computational pathology’). Notably, on the one hand, we focus on papers presenting novel techniques and theories with high impact (h-index, citations and impact factors of journals), thus we concentrate more on studies published in top conferences (including CVPR, NeurIPS, MICCAI, ISBI, MIDL, IPMI, AAAI, ICCV, ECCV, etc) and top journals (including TPAMI, TMI, MIA, etc) on weakly supervised, semi-supervised, and self-supervised learning in the field of computational pathology. On the other hand, since technical research in this area is growing rapidly and more new techniques...
have been proposed, we mainly cover papers published in 2019–2021. On the other hand, we also present a meticulous summary of the disease types, tasks, datasets, and performance covered by these papers. In total, this review contains more than 200 relevant references.

The rest of the paper is organized as follows: section 2 expounds a general overview of the weakly supervised, semi-supervised, and self-supervised learning paradigms in the context of computational pathology; section 3 includes a detailed review of the weakly supervised (section 3.1), semi-supervised (section 3.2), and self-supervised (section 3.3) learning paradigms; We discuss the three learning paradigms and their future trends in section 4, and conclude the whole paper in section 5. The list of all the acronyms used in this review is shown in table 1.

2. Overview of learning paradigms and problem formulation

In this section, we provide a general overview and problem formulation of the three learning paradigms reviewed in this paper, and compare them with the traditional strongly supervised paradigm. To make the description more specific and vivid, we present an example of accurately classifying normal and cancerous tissues in a WSI, as shown in figure 1. The raw data for this example WSI comes from a study on predicting lymph node metastasis in breast cancer using deep learning (Bejnordi et al 2017). We also intuitively compare and summarize these paradigms in table 2.

For the dataset \( W = \{ W_i \}_{i=1}^{n} \) consisting of NWSIs, each WSI \( W_i \) is now cut into patches \( \{ p_{i,j} \}_{j=1}^{n_i} \), and \( n_i \) is the number of patches cut out of \( W_i \). In the supervised learning paradigm, a large number of patches with fine-grained labels are available for training, so each patch is given a label \( y_{ij} \in \mathbb{R}^C \), and \( C \) denotes the possible class. For example, in the binary classification task, \( C = 2 \) and the label takes the scalar form \( \{0, 1\} \) while in the regression task, \( C \) takes the form of a continuous set of real numbers \( \mathbb{R} \). The goal of the supervised learning paradigm is to train a model \( f_{\theta}: x \rightarrow y \) to optimally predict the labels \( y_{ij} \) of the unknown patches \( p_{i,j} \) in the test WSI based on the loss function \( L \). Figure 1(a) illustrates the main process of this paradigm. During training, the model is trained in a supervised manner using patches cut out of the training WSIs and their labels (green for negative and red for positive) by pathologists; during testing, the trained model is used to predict the labels of the patches cut out of the unseen test WSIs.

In the weakly supervised learning paradigm, the label \( y_{ij} \) of each patch is typically unknown, while only the label of each WSI is available, and thus the traditional strongly supervised learning paradigm cannot work. In this review, we focus on the most dominant weakly supervised paradigm currently used in computational pathology, the deep multiple instance learning (MIL) approach. In MIL, each WSI is considered as a bag containing many patches (also called instances). If a WSI (bag) is labeled as disease-positive, then at least one patch (instance) in that WSI is disease-positive; if a WSI is disease-negative, then all patches in that WSI are negative. The relationship between a WSI (bag) and its patches (instances) can be expressed mathematically as follows.

### Table 1. List of all the acronyms in this review.

| Full name | Acronyms | Full name | Acronyms |
|-----------|----------|-----------|----------|
| Area Under ROC Curve | AUC | Graph Neural Network | GNN |
| Auxiliary Classifier Generative Adversarial Networks | AC-GAN | Hematoxylin-Eosin-Stained | H&E |
| Average Hausdorff Distance | AHD | Magnification Prior Contrastive Similarity | MPCS |
| Average Jaccard Index | AJI | Mean Average Precision | MAP |
| Calinski-Harabaz Index | CHI | Mean Teachers | MT |
| Contrastive Predictive Coding | CPC | Microsatellite Instability | MSI |
| Convolutional Autoencoder | CAE | Multiple Instance Fully Convolutional Network | MI-FCN |
| Convolutional Neural Network | CNN | Multiple Instance Learning | MIL |
| Deep Learning Hashing | DLH | Noise Contrastive Estimation | NCE |
| Deformation Representation Learning | DRL | Percentage Of Tumor Cellularity | TC |
| Diffusion-Convolutional Neural Networks | DCNNs | Recurrent Neural Network | RNN |
| Dual-Stream Multiple Instance Learning | DSMIL | Regions Of Interest | ROI |
| Expectation-Maximization | EM | Resolution Sequence Prediction | RSP |
| Exponential Moving Average | EMA | Silhouette Index | SI |
| Focal-Aware Module | FAM | Support Vector Machines | SVM |
| Frechet Inception Distance | FID | Temporal Ensembling | TE |
| Generative Adversarial Networks | GAN | The Cancer Genome Atlas Program | TCGA |
| Graph Convolutional Neural Network | GCN | Whole Slide Images | WSI |
Given a dataset $W = \{W_i\}_{i=1}^N$ consisting of $N$ WSIs, each image $W_i$ has a corresponding label $Y_i \in \{0, 1\}$, $i = \{1, 2, \ldots, N\}$. Now each WSI $W_i$ is cut into small patches $\{p_{i,j} \mid j = 1, 2, \ldots, n_i\}$ without overlapping each other, and $n_i$ is the number of patches. All patches $\{p_{i,j} \mid j = 1, 2, \ldots, n_i\}$ in $W_i$ form a bag, the bag-level label is the label $Y_i$ of $W_i$, and each small patch is called an instance of this bag, while the instance-level label $y_{i,j}$ and its corresponding bag-level label $Y_i$ have the following relationship:

$$
Y_i = \begin{cases} 
0, & \text{if } \sum_j y_{i,j} = 0 \\
1, & \text{else} 
\end{cases}
$$

(1)

It means that the labels of all instances in the negative bag are negative, while at least one positive instance exists in the positive bag and the labels of instances $y_{i,j}$ are unknown.

As shown in figure 1(b), generally, there are two main goals of deep learning-based WSI analysis, one is global slide classification, i.e. to accurately classify each WSI, and the other is positive patch localization, i.e. to accurately classify each instance in positive bags. A review of the current state-of-the-art weakly supervised learning methods is presented in section 3.1.
Table 2. Intuitive summary and comparison of the four paradigms.

| Methods                     | Input                                                                 | Suitable tasks                                                                 | Technical paradigms                                                                 | Strengths                                      | Weaknesses                                      |
|-----------------------------|-----------------------------------------------------------------------|--------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------|------------------------------------------------|
| Supervised learning paradigm | A large number of small patches (tiled from WSIs) with fine-grained labels | WSI-level and patch-level classification/segmentation/regression                | —                                                                                | Broad application, effective and simple training | Require large amount of fine-grained labeled data |
| Weakly supervised learning paradigm | Entire WSIs with overall labels or sparse labels | WSI-level classification/segmentation/regression, Patch-level coarse-grained localization | Instance-based approach, Bag-based approach, hybrid approach | No need for fine-grained annotation, effectively reduce the burden of data annotation | Achieve limited performance for fine-grained tasks |
| Semi-supervised learning paradigm | A limited number of small patches (tiled from WSIs) with fine-grained labels | WSI-level and patch-level classification/segmentation/regression                  | Pseudo-labelling-based approach, Consistency-based approach, Graph-based approach, Unsupervised-preprocessing-based approach, GAN-based approach and others | Require only a small amount of fine-grained annotation, effectively reduce the burden of data annotation | Need to satisfy various semi-supervised assumptions |
| Self-supervised learning paradigm | A large number of small patches (tiled from WSIs) without labels | Patch-level feature representations, Multiple related down-stream tasks          | Predictive approach, Generative approach, Contrastive approach, Hybrid approach | Efficiently extract image features from a large amount of unsupervised data, effectively reduce the data annotation burden | May result in information loss when the extracted features are not applicable to downstream tasks |
In the semi-supervised learning paradigm, we only have a very small number of patches with labels, in addition to a large number of unlabeled patches that can also be used for training. Therefore, the main goal of the semi-supervised learning paradigm is how to use the unlabeled data to improve the performance of the models trained with limited labeled data. As shown in figure 1(c), in contrast to the supervised learning paradigm, the semi-supervised learning paradigm makes use of a large amount of unlabeled data while training with the labeled data. During testing, the trained model is used to predict the labels of the patches in test WSIs. See section 3.2 for a detailed review of the semi-supervised learning methods.

Self-supervised learning is a hybrid learning approach, which combines unsupervised and supervised learning paradigms in a pre-training and fine-tuning manner. The aim is to get better results of supervised training through generating supervised information from a large amount of unlabeled data, which can learn better feature representations, and can reduce manual annotation in the subsequent tasks. Due to the small amount of annotated data, it is not sufficient to use these data directly to train the model. Therefore, the self-supervised learning paradigm first learns a primary feature representation from a large amount of unlabeled data, which is called the pre-training process. The feature representations learned in the self-supervised auxiliary tasks are then transferred for further training in downstream tasks using limited labeled data, which is called the fine-tuning process. In this way, the primary feature representations can effectively help the network to achieve an effective training result with less labeled data.

As shown in figure 1(d), the pre-training process of the self-supervised learning paradigm is typically performed through self-supervised auxiliary tasks. In the self-supervised auxiliary tasks, certain inherent properties of the unlabeled data are first utilized to generate supervised information, and then the network is trained by the self-supervised information, such as self-reconstruction, random rotation followed by angle prediction, color information discarding followed by colorization, and patch position disruption followed by restoration. Once accomplishing these self-supervised auxiliary tasks, the effective feature representations can be extracted. The fine-tuning process of self-supervised learning is done in the downstream tasks. During the fine-tuning process, a small amount of labeled data is used to perform the supervised training, and the model is not trained from scratch, but is further trained using the feature representations learned in the auxiliary tasks as the initial weights of the network. Finally, the trained network is used for testing. A review of the state-of-the-art self-supervised learning methods is presented in section 3.3.

3. Paradigms

3.1. Weakly supervised learning paradigm

In this section, we provide a comprehensive review of the primary deep multiple instance learning (MIL) methods currently used in the weakly supervised learning paradigm for computational pathology. In MIL, each WSI is considered as a bag containing many patches (also called instances). If a WSI (bag) is labeled disease-positive, then at least one patch (instance) in that WSI is disease-positive; if a WSI is disease-negative, then all patches in that WSI are negative.

We categorize the current deep MIL methods for WSI analysis into instance-based methods, bag-based methods, and hybrid methods. Our categorization is mainly based on whether the methods contain an instance classifier or a bag classifier, i.e. instance-based methods contain only an instance classifier; bag-based methods contain only a bag classifier; while hybrid methods contain both an instance classifier and a bag classifier. In this way, the categories clearly cover almost current deep MIL methods for WSI analysis. A diagram of the three methods above is shown in figure 2. The detailed literatures in this section are summarized in table 3.

3.1.1. Instance-based approach

The main idea of the instance-based approach is to train a good instance classifier to accurately predict the potential labels of instances in each bag, and then use MIL-pooling to aggregate the predictions of all instances in each bag to obtain the prediction of the bag. The details are shown in figure 2(a). Since the true labels of each instance are unknown, these approaches usually first assign the labels of each instance with their corresponding bags as the pseudo labels (i.e. all instances in a positive bag are given positive labels, and all instances in a negative bag are given negative labels), and then train the instance classifier using a supervised way until it converges. The loss function is usually the cross-entropy function defined between the predictions of the instance classifier and the pseudo labels. After training, the instance classifier is used to make predictions for all instances in the test bag, and then the predictions of each instance are aggregated to obtain the prediction of the bag, and this aggregation process is called MIL-pooling. Commonly used MIL pooling methods include Mean-pooling (Wang et al 2018), Max-pooling (Wu et al 2015, Feng and Zhou 2017, Wang et al 2018), Voting (Cruz-Roa et al 2014), log-sum-exp-pooling (Ramon and De Raedt 2000), Noisy-or-pooling (Maron and Lozano-Pérez 1997), Noisy-and-pooling (Kraus et al 2016), and Dynamic pooling (Yan et al 2018) among others.
Instance-based approach is more common in early studies, and its main advantage lies in the direct prediction of each instance so that the localization task can be performed conveniently. However, it has two major drawbacks. First, since the true labels of each instance in the positive bags are not necessarily all positive, the pseudo labels assigned to the instances in the positive bags are noisy, which will lead to inaccurate training of the instance classifier; Second, the MIL-pooling method, which aggregates the predictions of instances in each bag, is manually designed and non-trainable, making it less flexible and robust. Therefore, the performance of these methods is usually limited.

3.1.2. Bag-based approach

The main idea of the bag-based approaches is to first extract the features of each instance in a bag using shared instance-level feature extractors, then use MIL-pooling to aggregate the instance-level features to obtain the bag-level features, and then train the bag classifier in a supervised manner until it converges. The specific diagram is shown in figure 2(b). The loss function is usually defined as the cross-entropy loss between the predictions of the bag classifier and the true bag labels.

MIL-pooling also exists in bag-based methods, but unlike instance-based methods, MIL-pooling here aggregates not the predictions of instances, but the features of instances. Mean-pooling, Max-pooling and other aggregation methods can also be used as aggregation methods for instance features, but their drawbacks remain, i.e. they cannot be trained and adjusted adaptively, so they are often not flexible enough.

The key of the bag-based methods is the training of the bag classifier. Since the true labels of the bags are available, there is no noise in their training process, so these methods tend to be more accurate than instance-based methods in bag classification. However, a serious problem of the bag-based approaches is that they cannot
| Reference        | Approach     | Disease Type                                      | Staining | Task                          | Dataset                                      | Dataset Scale | Dataset Link                  | Performance            |
|------------------|--------------|--------------------------------------------------|----------|------------------------------|----------------------------------------------|---------------|-------------------------------|------------------------|
| Yan et al (2018) | Instance-based | Breast Cancer Diabetes (from eye fundus images) | H&E      | Benign and malignant classification | UCSB breast dataset                          | 58 cases      | Kandemir et al (2014)       | Accuracy: 0.927        |
|                  |              |                                                  |          |                              | Messidor dataset                             | 1200 cases    | Decencièr et al (2014)      | Accuracy: 0.740        |
| Kraus et al (2016) | Instance-based | Breast Cancer                                    | H&E      | Classification of 12 distinct categories | Broad Bioimage Benchmark Collection (BBBC021v1) dataset | 340 cases      | Ljosa et al (2012)         | Accuracy: 0.958 for fullimage, 0.971 for treatment |
| Cruz-Roa et al (2014) | Instance-based | Breast Cancer                                    | H&E      | Automatic detection of invasive ductal carcinoma tissue regions | Clinical histopathology dataset collected from multiple hospitals | 162 cases      | inhouse                      | Accuracy: 0.842        |
| Ilse et al (2018) | Bag-based    | Breast Cancer                                    | H&E      | Automatic detection of cancerous regions | Breast cancer dataset                         | 58 cases      | Gelasca et al (2008)       | Accuracy: 0.755        |
|                  |              | Colon Cancer                                     | H&E      |                              | Colon cancer dataset                         | 100 cases     | Sirimukunwattana et al (2016) | Accuracy: 0.904        |
| Tu et al (2019)  | Bag-based    | Diabetes (from eye fundus images)                | H&E      | Diagnosing diabetes from weakly labeled retinal images | Messidor dataset                             | 1200 cases    | Decencièr et al (2014)      | Accuracy: 0.742        |
| Hashimoto et al (2020) | Bag-based    | Malignant Lymphoma                               | H&E      | Classification of malignant lymphoma sub-types | Clinical histopathology dataset collected from multiple hospitals | 196 cases      | inhouse                      | Accuracy: 0.871        |
|                  |              | Lung Cancer                                      | H&E      |                              | National Lung Screening Trial (NLST) dataset | 387 cases     | Team et al (2011)          | AUC: 0.652              |
| Yao et al (2020) | Bag-based    | Colorectal Cancer                                | H&E      | Cancer survival prediction   | Molecular and Cellular Oncology (MCO) dataset | 1146 cases    | Ward and Hawkins (2015)     | AUC: 0.7143             |
| Lu et al (2019)  | Bag-based    | Breast Cancer                                    | H&E      | Classification of normal or benign | BACH dataset                                 | 400 cases     | Aresta et al (2019)        | Accuracy: 0.95         |
| Zhao et al (2020) | Bag-based    | Colon Adenocarcinoma Breast Cancer               | H&E      | Prediction of lymph node metastasis | The Cancer Genome Atlas (TCGA) dataset        | 425 cases     | Kandoth et al (2013)       | Accuracy: 0.6761        |
|                  |              |                                                  |          |                              | Camelyon16 dataset                           | 400 cases     | Bejnordi et al (2017)      | Accuracy: 0.8992        |
| Li et al (2021)  | Bag-based    | Lung Cancer                                      | H&E      | Diagnosis of lung cancer subtypes | The Cancer Genome Atlas (TCGA) lung cancer dataset | 1054 cases   | https://portal.gdc.cancer.gov/ | Accuracy: 0.9571        |
|                  |              |                                                  |          |                              | Camelyon16 dataset                           | 400 cases     | Bejnordi et al (2017)      | Accuracy: 0.8837        |
| Reference          | Approach    | Disease Type       | Staining | Task                                      | Dataset                      | Dataset Scale | Dataset Link                                      | Performance |
|--------------------|-------------|--------------------|----------|------------------------------------------|------------------------------|---------------|--------------------------------------------------|-------------|
| Shao et al (2021)  | Bag-based   | Lung Cancer        | H&E      | Diagnosis of cancer subtypes             | TCGA-NSCLC dataset          | 993 cases     | https://portal.gdc.cancer.gov/                    | Accuracy: 0.8835 |
|                    |             | Kidney Cancer      | H&E      | Diagnosis of cancer subtypes             | TCGA-RCC dataset            | 884 cases     | https://portal.gdc.cancer.gov/                    | Accuracy: 0.9466 |
|                    |             | Breast Cancer      | H&E      | Detection of lymph node metastases       | BREAST-LNM dataset          | 3957 cases    | inhouse                                          | AUC: 0.7288  |
| Li et al (2021)    | Bag-based   | Lung Cancer        | H&E      | Diagnosis of lung cancer subtypes        | CPTAC-LUAD dataset          | 1065 cases    | https://portal.gdc.cancer.gov/                    | Accuracy: 0.9906 |
|                    |             | Glioma             |          | Classification of glioma                 |                             | 209 cases     | https://portal.gdc.cancer.gov/                    | Accuracy: 0.771 |
| Hou et al (2016)   | Hybrid      | Lung Cancer        | H&E      | Diagnosis of non-small-cell lung carcinoma subtypes | The Cancer Genome Atlas (TCGA dataset) | 316 cases     | https://portal.gdc.cancer.gov/                    | Accuracy: 0.798 |
|                    |             | Prostate Cancer    | H&E      | Benign and malignant classification      | Prostate core biopsy dataset | 24859 cases   | inhouse                                          | AUC: 0.986   |
|                    |             | Skin Cancer        |          | Benign and malignant classification      | Skin dataset                | 9,962 cases   | inhouse                                          | AUC: 0.986   |
| Campanella et al (2019) | Hybrid  | Breast Cancer      | H&E      | Detection of lymph node metastases       | Breast dataset              | 9894 cases    | https://cancerimagingarchive.net/datascope/cptac/ | Accuracy: 0.965 |
| Wang et al (2019)  | Hybrid      | Lung Cancer        | H&E      | Diagnosis of lung cancer subtypes        | Lung cancer dataset         | 939 cases     | http://thomasfuchslab.org/data/                   | Accuracy: 0.973 |
| Chen et al (2021)  | Hybrid      | Breast Cancer      | IHC      | HER2 scoring (negative (0/1+), equivocal (2+) and positive (3+)) | HER2 scoring dataset       | 1105 cases    | inhouse                                          | Accuracy: 0.8970 |
| Chikontwe et al (2020) | Hybrid | Colectoral Cancer  | H&E      | Prediction of normal and malignant tissues | CRC WSI Dataset I           | 173 cases     | inhouse                                          | Accuracy: 0.9231 |
|                    |             | Gastrointestinal  |          | Prediction of patients with celiac disease or being healthy | CRC WSI Dataset II         | 193 cases     | inhouse                                          | Accuracy: 0.9872 |
|                    |             | Celiac Disease     |          |                                           | Gastrointestinal dataset    | 413 cases     | inhouse                                          | Accuracy: 0.862 |
| Sharma et al (2021)| Hybrid      | Breast Cancer      | H&E      | Detection of lymph node metastases       | Camelyon16 dataset         | 400 cases     | Bejnordi et al (2017)                             | AUC: 0.9112  |
| Lu et al (2021)    | Hybrid      | Renal Cell Carcinoma | H&E      | subtyping and the detection of lymph node metastasis | RCC dataset                | 884 cases     | https://portal.gdc.cancer.gov/                    | Accuracy: 0.991 |
|                    |             | Non-small-cell Lung Cancer | H&E      |                                             | NSCLC dataset               | 993 cases     | https://cancerimagingarchive.net/datascope/cptac/ | Accuracy: 0.956 |
|                    |             | Breast Cancer      |          |                                             | CAMELYON16 and CAMELYON17 dataset | 899 cases | https://camelyon17.grand-challenge.org/Data/ | AUC: 0.936 |
| Reference | Approach | Disease Type | Staining | Task | Dataset | Dataset Scale | Dataset Link | Performance |
|-----------|----------|--------------|----------|------|---------|--------------|--------------|-------------|
| Myronenko et al (2021) | Hybrid | Prostate Cancer | H&E | Classifying cancer tissue into Gleason patterns | Prostate cancer grading challenge dataset | 11,000 cases | https://panda.grand challenge. org/home/ | Accuracy: 0.805 |
| Naik et al (2020) | Clinical studies | Breast Cancer | H&E | Determination of hormonal receptor status | Australian Breast Cancer Tissue Bank (ABCTB) dataset | 2535 cases | https://abctb.org.au/abctbNew2/ACCESSPOLICY.pdf | AUC: 0.92 |
| Tomita et al (2019) | Clinical studies | Esophagus Cancer | H&E | Detection of cancerous and precancerous esophagus tissue | Esophagus cancer dataset | 180 cases | inhouse | Accuracy: 0.83 |
| Skrede et al (2020) | Clinical studies | Colorectal Cancer | H&E | Prediction of colorectal cancer outcome | Colorectal cancer dataset | 2473 cases | inhouse | Ratio for poor versus good prognosis: 3.84 AUC: 0.81 |
| Kather et al (Kather et al (2019)) | Clinical studies | Gastrointestinal Cancer | H&E | Prediction of microsatellite instability | TCGA-STAD dataset | 315 cases | https://portal.gdc.cancer.gov/ | AUC: 0.84 |
| Coudray et al (2018) | Clinical studies | Lung Cancer | H&E | Classification of subtypes | The Cancer Genome Atlas (TCGA) dataset | 1634 cases | https://portal.gdc.cancer.gov/ | AUC: 0.97 |
| Bejnordi et al (2017) | Clinical studies | Breast Cancer | H&E | Detection of lymph node metastases | CAMELYON16 dataset | 400 cases | https://came lyon16.grandchallenge.org/ | AUC of six of commonly mutated genes from 0.733 to 0.856 |
| Wessels et al (2021) | Clinical studies | Prostate Cancer | H&E | Prediction of lymph node metastasis | Prostate cancer dataset | 218 cases | inhouse | AUC: 0.994 |
| Anand et al (2021) | Clinical studies | Thyroid Cancer | H&E | Prediction of BRAF mutation | ISBI 2017 Thyroid Tissue Microarray (TH-TMA17) dataset | 85 cases | Wang et al (2018) | AUC: 0.68 |
| Yang et al (2022) | Clinical Studies | Breast Cancer | H&E | Prediction of HER2-positive breast cancer recurrence and metastasis risk | TCGA-THCA dataset | 444 cases | https://github.com/bensteven2/HE_breast_recurrence/ | AUC: 0.98 |
| | | | | | HER2-positive breast cancer dataset | 127 cases | https://portal.gdc.cancer.gov/ | AUC: 0.76 |
| | | | | | The Cancer Genome Atlas (TCGA) dataset | 123 cases | | AUC: 0.72 |
| Reference          | Approach            | Disease Type       | Staining | Task                                                                 | Dataset                                      | Dataset Scale | Dataset Link                                                                 | Performance |
|--------------------|---------------------|--------------------|----------|----------------------------------------------------------------------|----------------------------------------------|---------------|-----------------------------------------------------------------------------|-------------|
| Li et al (2021)    | Clinical studies    | Breast Cancer      | H&E      | Predicting biomarker of pathological complete response to neoadjuvant chemotherapy | Breast cancer dataset                        | 540 cases     | inhouse                                                                     | AUC: 0.847   |
| Saillard et al (2020) | Clinical studies | Hepatocellular Carcinoma | H&E | Predicting survival after hepatocellular carcinoma resection | Discovery set                                 | 194 cases     | inhouse                                                                     | C-Indices: 0.78 |
| Velmahos et al (2021) | Clinical studies | Bladder Cancer     | H&E      | Identifying FGFR-activating mutations                                | The Cancer Genome Atlas (TCGA) dataset       | 328 cases     | https://portal.gdc.cancer.gov/                                             | C-Indices: 0.70 |
| Woerl et al (2020) | Clinical studies    | Bladder Cancer     | H&E      | Prediction of molecular subtypes                                     | The Cancer Genome Atlas (TCGA) Urothelial Bladder Carcinoma Dataset | 407 cases     | https://portal.gdc.cancer.gov/                                             | AUC=0.89     |
|                    |                     |                    |          |                                                                      | CCC-EMN cohort                               | 16 cases       | inhouse                                                                     | AUC=0.85     |
perform the localization task easily. Furthermore, the aggregation functions for instance features are not flexible enough to show the contribution of different instances to bag classification.

Attention-based Approach. Ilse et al. (2018) have alleviated these dilemmas. They first proposed to use the trainable attention mechanism to aggregate instance features, and started a wave of study on attention-based aggregation methods by subsequent bag-based methods. They trained both the instance-level feature extractor and a bag-level classifier using an end-to-end manner, and used the attention mechanism to aggregate the features and measure the significance of each instance. Tu et al. (2019) proposed a new end-to-end graph neural network (GNN) for instance aggregation. This work is the first GNN-based MIL work. Hashimoto et al. (2020) proposed a novel end-to-end method for cancer subtype classification by combining MIL, domain adversarial and multiscale learning frameworks. Zhu et al. (2017), Yao et al. (2020) proposed a deep attention guided MIL framework for cancer survival analysis. They first used a pre-trained model from ImageNet (Deng et al. 2009) to extract the features of instances in each bag, and then used K-means algorithm to cluster the instances in each bag to obtain the phenotypic patterns, and finally applied attention mechanism to aggregate the features of these patterns and performed prediction.

Self-supervised pre-training-based approach. Due to the extremely large size of WSIs and the large number of instances cut out, direct end-to-end training of all instances is easily limited by computational resources. Therefore, some studies first use advanced self-supervised pre-training methods to characterize each instance and then perform subsequent training. Lu et al. (2019) first proposed to obtain instance-level feature representations by self-supervised contrastive predictive coding (CPC), and then used the attention-based MIL method for instance aggregation to perform bag-level classification. This is the first MIL study using self-supervised contrastive learning. Zhao et al. (2020) used a pre-trained VAE-GAN (Larsen et al. 2016) to extract instance-level features, and then used GNN to aggregate instance features and perform bag-level classification. Li et al. (2021) proposed DSMIL, where they used contrastive pre-training (Chen et al. 2020) to obtain the instance features, and then proposed the masked non-local operation-based dual-stream aggregator to perform both instance-level classification and bag-level classification.

Transformer based approach. In MIL-based WSI analysis, not only the contribution of different instances to bag classification should be considered, the relationships among different instances should also be fully explored, because different instances in a WSI are not isolated from each other, but have strong correlation. To address this issue, Shao et al. (2021) and Li et al. (2021) et al used transformer-based architectures to aggregate instances and both achieved promising results. The former designed a transformer-based related proofs. The latter presented a MIL framework based on the deformable transformer and convolutional layers.

3.1.3. Hybrid approach

The hybrid approach combines the advantages of the above two approaches. It trains both the instance-level classifier and the bag-level classifier, and uses the former to predict the instance-level results while the latter for bag-level results. Overall, there are two types of the hybrid approaches. One is the two-stage approach and the other is the end-to-end approach.

Two-stage hybrid approach. The two-stage hybrid approach generally trains the instance classifier by assigning each instance in each bag with their corresponding bag labels as pseudo labels, and then trains the bag classifier to complete the bag classification based on the predictions of the instance classifier. Some studies have also attempted to select the key instances in each bag based on the predictions of the instance classifier, and then train the bag classifier based on these key instances. The specific diagram is shown in figure 2(c). Hou et al. (2016) proposed a new Expectation-Maximization (EM) based model. They selected discriminative instances based on spatial relationship to train the instance classifier and fed the histogram of instance predictions into the multiclass logistic regression model and the SVM model (Chang and Lin 2011) for bag prediction. Campanella et al. (2019) first selected key instances with the maximum prediction probability of the instance classifier in the current iteration and assigned pseudo labels of the corresponding bag labels to them. Then they fed the features of these key instances into the recurrent neural network (RNN) to perform the aggregation and prediction of the bags. Wang et al. (2019) selected key instances based on the predictions of positive instance probability and fed their features into the global feature descriptor and used the random forest algorithm to classify the bags. Chen et al. (2021) proposed a focal-aware module (FAM) and used thumbnails of WSI to automatically estimate the key regions associated with the diagnosis. Then, the instance features at different scales were extracted based on these key regions and aggregated using GNN to perform the bag classification.

End-to-end hybrid approach. The end-to-end hybrid approach generally trains the instance-level classifier and the bag-level classifier at the same time. A common approach is to train the two classifiers simultaneously by assigning each instance the corresponding bag labels as pseudo labels on top of the bag classifier. Some studies also train the instance classifier to select the key instances in an epoch first, and then train the bag classifier after 12
aggregating the instance features. The specific diagram is shown in figure 2(d). Shi et al (2020) proposed loss-based attention MIL. They added an instance-level loss function weighted by the instance attention scores based on AB-MIL (Ilse et al 2018) as a regularization term to improve the recall of instances and used consistency constraints to smooth the training process to improve the generalization ability. Chikontwe et al (2020) combined top-k instance selection, instance-level representation learning, and bag-level representation in an end-to-end framework. Sharma et al (2021) also combined instance selection, instance-level representation learning and bag-level representation in an end-to-end framework. Unlike Chikontwe et al (2020), they proposed to use a clustering-based sampling method to select key instances. Lu et al (2021) also proposed a MIL framework based on clustering and attention mechanisms. They selected the instances with the largest and smallest attention scores in the current bag for clustering to enhance the learning of feature space. Myronenko et al (2021) proposed a MIL framework combining the transformer and CNN architectures to compute the interrelationships between instances and aggregate the instances features to accomplish the bag classification. They added the instance loss to assist the optimization process.

3.1.4. Representative clinical studies
A large number of outstanding studies have been dedicated to address significant clinical problems using weakly supervised methods. For example, Coudray et al (2018) et al developed deep learning models for accurate prediction of cancer subtypes and genetic mutations and sparked the whole field of weakly supervised computational pathology. Naik et al (2020) et al presented an attention-based deep MIL framework to predict directly estrogen receptor status from H&E slices. Another typical clinical work comes from Tomita et al (2019), who proposed a grid-based attention network to perform 4-class classification of high-resolution endoscopic esophagus and gastroesophageal junction mucosal biopsy images from 379 patients. Skrede et al (2020) developed a multi-scale deep MIL-based model to analyze conventional HE-stained slides and developed a model that can effectively predict the diagnosis of patients after colorectal cancer surgery. Another gastrointestinal tract oncology study (Kather et al 2019) predicted microsatellite instability (MSI) based on a deep MIL model directly on HE-stained slides. Currently, weakly supervised deep-learning models for digital pathological analysis has been applied in a wide range of cancer types including breast, colorectal, lung, liver, cervical, thyroid, and bladder cancers (Coudray et al 2018, Chaudhary et al 2018, Campanella et al 2019, Saillard et al 2020, Woerl et al 2020, Anand et al 2021, Li et al 2021, Velmabos et al 2021, Wessels et al 2021, Yang et al 2022).

3.2. Semi-supervised learning paradigm
Semi-supervised learning is a branch of machine learning that combines both supervised and unsupervised learning tasks and improves model performance by exploiting the information associated between tasks (Zhu 2005, Van Engelen and Hoos 2020). In semi-supervised learning, only a small amount of labeled data is generally available, and besides that, a large amount of unlabeled data can be utilized for network training. Consequently, the main goal of semi-supervised learning is how to use these unlabeled data to improve the performance of the model trained with limited labeled data. Scenarios of the semi-supervised learning paradigm are very common in the field of pathological image analysis, both in diagnostic tasks and in segmentation tasks. Due to the expensive and time-consuming fine-grained annotation, pathologists often can only provide a small number of precise annotations for supervised training of the models, while a large amount of unannotated data cannot be used. Training deep models with only these limited labeled data can easily lead to over-fitting, thus significantly harming the performance and generalization of the models. In the semi-supervised learning paradigm, a large number of unlabeled images can be used to assist in training and thus further improve the performance, generalization, and robustness of the models.

In the past two decades, numerous semi-supervised learning algorithms have been proposed and widely used in the fields of natural image processing and pathological image analysis. The representative approaches in the field of semi-supervised learning are divided into five categories, namely pseudo-labelling-based approach (section 3.2.1), consistency-based approach (section 3.2.2), graph-based approach (section 3.2.3), unsupervised-preprocessing approach (section 3.2.4), and other approaches (section 3.2.5). We introduce these methods below, respectively. For each category, we first describe their fundamental principles and then elaborate on their representative studies in the field of pathological image analysis. For a systematic review of the assumptions, concepts and representative methods of semi-supervised learning in the field of natural images, we recommend the review by Van Engelen and Hoos (2020). Table 4 summarizes the detailed list of literatures in this section.

3.2.1. Pseudo-labelling-based approach
Fundamental principles. The pseudo-labeling-based approach is a classical and well-known semi-supervised method (Zhu 2005), which mainly consists of two alternating processes, training and pseudo-labeling. Taking
Table 4. List of literatures in the semi-supervised learning section.

| Reference     | Approach          | Disease type | Staining         | Task                                         | Dataset                      | Dataset scale                  | Dataset link                          | Performance                           |
|---------------|-------------------|--------------|------------------|----------------------------------------------|------------------------------|---------------------------------|---------------------------------------|---------------------------------------|
| Singh et al (2011) | Pseudo-labelling-based | Breast Cancer | 3D fluorescence microscopy | Identifying nuclear phenotypes | Nuclei image dataset         | 984 images                     | Inhouse                              | Mean accuracy: 0.8                     |
| Bulten et al (2020) | Pseudo-labelling-based | Prostate Cancer | H&E               | Gleason grading                            | Inhouse dataset              | 5739 biopsies from 1243 patients | Inhouse                              | AUC = 0.99                            |
| Tolkach et al (2020) | Pseudo-labelling-based | Prostate Cancer | H&E               | Detection of prostate cancer tissue         | The Cancer Genome Atlas Program (TCGA) dataset | 1.67 million patches            | http://portal.gdc.cancer.gov |                                               |
| Jaiswal et al (2019) | Pseudo-labelling-based | Breast Cancer | H&E               | Detection of lymph node metastases          | PatchCamelyon dataset        | 327 680 patches                 | https://zenodo.org/ deposit/3823933 | Accuracy = 0.98                      |
| Shaw et al (2020) | Pseudo-labelling-based | Colorectal Cancer | H&E               | Classification of 9 categories of pathology patterns | Public dataset              | 100 000 patches                 | https://zenodo.org/ record/1214456#.YvyX3ZByw4 | Mean accuracy = 0.943                 |
| Marini et al (2021) | Pseudo-labelling-based | Prostate Cancer | H&E               | Gleason grading                            | Tissue MicroArray dataset, Zurich dataset | 886 cases                      | Inhouse                              | κ-score: 0.7645                       |
| TGCA (2019) |                    | Breast Cancer |                  |                                              | TCGA-PRAD dataset            | 449 cases                      | http://portal.gdc.cancer.gov | κ-score: 0.4529                       |
| Cheng et al (2020) | Pseudo-labelling-based | Prostate Cancer | H&E               | Automated segmentation of cancerous regions | TVGH TURP dataset            | 71 cases                       | Inhouse                              | Dice: 93.76                           |
| Zhou et al (2020) | Consistency-based  | —            | Liquid-based pap test specimen | Cervical cell instance segmentation | liquid-based Pap test specimen dataset | 4439 cytoplasm                  | Inhouse                              | AJI: 73.45, MAP: 46.01               |
| Su et al (2019)  | Consistency-based  | —            | H&E               | Nuclei classification                       | MoNuSeg dataset              | 22462 nuclei                    | Sirinukunwattana et al (2016)       | F1 score: 75.02 (5% labels)            |
|                 |                   |              |                  | Detection of tumor metastasis               | Ki-67 nucleus dataset        | 17 516 nuclei                   | Inhouse                              | F1 score: 79.32 (5% labels)           |
|                 |                   |              |                  |                                               | BreastPathQ dataset          | 2579 patches                    | Martel et al (2019)                 | TC: 0.876 (10% labels)               |
| Reference                  | Approach                        | Disease type                               | Staining | Task                                    | Dataset         | Dataset scale | Dataset link                                                                 | Performance       |
|----------------------------|---------------------------------|--------------------------------------------|----------|-----------------------------------------|-----------------|---------------|-------------------------------------------------------------------------------|--------------------|
| Srinidhi et al (2022)      | Consistency-based               | Breast Cancer, Colorectal Cancer           | H&E      | Classification of tissue types          | Camelyon16 dataset | 399 WSIs      | https://camelyon16.grand-challenge.org/Data/                                  | AUC: 0.855 (10% labels) |
|                           |                                 |                                            |          | Quantification of tumor cellularity     | Kather multiclass dataset | 100K patches  | Kather et al (2019)                                                           | Accuracy: 0.982 (10% labels) |
| Xu et al (2016)            | Graph-based                     | —                                          | Microscopy images | Neuron segmentation                 | Neural morphology image dataset | 2000 neuron regions with with annotations | Inhouse                                                                     | F1 score: 0.96 (40% labels) |
| Su et al (2015)            | Graph-based                     | —                                          | Microscopy images | Cell segmentation                    | Phase contrast microscopy image dataset | Multiple sequences of total 1404 frames 2904 patches                         | http://www.celltracking.rri.cmu.edu/downloads.html | TC: 0.9813 |
| Shi et al (2020)           | Graph-based                     | Lung Cancer                                | H&E      | Predictions of subtypes                 | The Cancer Genome Atlas (TCGA) | 1763 patches   | http://portal.gdc.cancer.gov                                                   | Accuracy: 0.905 (20% labels) |
|                           |                                 | Breast Cancer                              |          |                                        | Pathology triaging image dataset | 4402 patches    | Inhouse                                                                     | Accuracy: 0.895 (20% labels) |
| Peikari et al (2018)       | Unsupervised-preprocessing-based| Breast Cancer                              | H&E      | Identifying different breast tissue regions | Nuclei figure classification dataset | 30 000 figures | Inhouse                                                                     | AUC: 0.86          |
| Lu et al (2019)            | Unsupervised-preprocessing-based| Breast Cancer                              | H&E      | Benign and malignant classification     | BACH dataset     | 400 cases      | BACH: Grand challenge on Breast Cancer histology images                      | Accuracy: 0.95 |
| Koohbanani et al (2021)    | Unsupervised-preprocessing-based| Breast Cancer                              | H&E      | Detection of tumor regions              | Camelyon16 dataset | 399 slides     | https://camelyon16.grand-challenge.org/Data/                                  | AUC: 0.817 (1% labeled) |
|                           |                                 | Oral Squamous Cell Carcinoma               |          | Prediction of metastases in the cervical lymph nodes | LNM-OSCC dataset | 217 slides     | Inhouse                                                                     | AUC: 0.806 (1% labeled) |
|                           |                                 | Colorectal Cancer                          |          | Classification of tissue types          | Kather multiclass dataset | 100K patches   | Kather et al (2019)                                                           | AUC: 0.903 (1% labeled) |
| Reference                        | Approach                                   | Disease type                          | Staining | Task                                      | Dataset                        | Dataset scale | Dataset link                                         | Performance                  |
|---------------------------------|--------------------------------------------|---------------------------------------|----------|-------------------------------------------|-------------------------------|----------------|-----------------------------------------------------|-----------------------------|
| Srinidhi et al (2022)           | Unsupervised-pre-processing-based          | Breast Cancer, Colorectal Cancer      | H&E      | Detection of tumor metastasis            | BreastPathQ dataset          | 2579 patches  | Martel et al (2019)                                  | TC: 0.876 (10% labels)       |
|                                 |                                            |                                       |          | Classification of tissue type             | Camelyon16 dataset           | 399 WSIs      | https://camelyon16.grand-challenge.org/Data/        | AUC: 0.855 (10% labels)      |
|                                 |                                            |                                       |          | Quantification of tumor cellularity      | Kather multiclass dataset    | 100K patches  | Kather et al (2019)                                  | ACC: 0.982 (10% labels)      |
| Kapil et al (2018)              | GAN-based                                  | Lung Cancer                           | Ventana PD-L1 (SP263) assay             | Automated tumor proportion scoring | NSCLC needle biopsy dataset  | 270 slides    | Inhouse                                             | Ratio of the number of tumor positive cell pixels to the total number of tumor cell pixels: 0.94 |
| Cong et al (2021)               | GAN-based                                  | Brain Cancer                          | H&E      | Stain normalisation                       | TCGA1 glioma cohort          | 22 229 images | Liu et al (2020)                                    | F1 score: 0.937              |
| Sparks and Madabhushi (2016)    | Manifold-learning-based                    | Breast Cancer                         | H&E      | Image retrieval                           | BreakHis database            | 7909 images   | Spanhol et al (2015)                                | F1 score: 0.980              |
| Li et al (2018)                 | Expectation-Maximization-based             | Prostate Cancer                       | H&E      | Semantic segmentation                     | Prostate histpathology dataset | 58 patients   | Inhouse                                             | SE: 0.14                    |
| Su et al (2021)                 | Association-learning-based                 | Breast Cancer                         | H&E      | Classification of cancerous and non-cancerous slides | Bioimaging 2015 challenge dataset | 135 fully annotated and 1800 weakly annotated tiles | Gertych et al (2015) | AJI: 0.495 |
|                                 |                                            |                                       |          |                                           | Bioimaging 2015 challenge dataset | 285 images     | Aratuo et al (2017)                                 | F1 score: 0.75               |
|                                 |                                            |                                       |          |                                           | BACH dataset                 | 400 images     | Aresta et al (2019)                                 | F1 score: 0.77               |
the classification problem as an example, in the training process, one or more classifiers are first trained in a supervised manner on the labeled data. The labeled data may be derived from the initial accurately labeled data or from the pseudo-labeled data from the previous iterations. In the pseudo-labeling process, all the unlabeled data are first predicted using the classifier trained in the previous process, and then the most confidently predicted portion of the data are selected for pseudo-labeling. Finally, these pseudo-labeled data are added to the labeled data for the next iteration. This process is repeated until no data with high confidence are found or all data are labeled.

The pseudo-labeling-based methods are firstly applied to the field of natural image processing and typically contain self-training methods (Lee et al 2013) and co-training methods (Blum and Mitchell 1998, Zhou and Li 2005).

Study in pathological image analysis. In pathological image analysis, Singh et al (2011) proposed a semi-supervised method of learning distance metrics from labeled data and performing label propagation for identifying the subtypes of nuclei, which was locally adaptive and could fully consider the heterogeneity of the data. Bulten et al (2020) developed a deep learning system for Gleason scoring of prostate biopsies based on semi-supervised learning. They first trained the network on a small training dataset with pure Gleason scores, and then applied the trained network to other internal training datasets to set reference standards. Then, the labels were corrected and relabeled using reports from pathologists. Tolkach et al (2020) used a pseudo-labeling-based semi-supervised strategy to train the CNN network to accomplish Gleason pattern classification. Jaiswal et al (2019) proposed a semi-supervised method based on pseudo-labeling and entropy regularization for breast cancer pathological image classification. Shaw et al (2020) extended the study of Yalniz et al (2019) by proposing a semi-supervised teacher-student distillation method for the classification of colorectal cancer pathological images. Marini et al (2021) proposed a deep pseudo-labeling-based semi-supervised learning method for strongly heterogeneous pathology data containing only a small number of local annotations. Their method consists of a high-volume teacher model and a small-volume student model, where the teacher model is automatically labeled with pseudo labels for the training of the student model. Cheng et al (2020) proposed a semi-supervised learning framework based on a teacher-student model with similarity learning for the segmentation of breast cancer lesions containing a small number of annotations and noisy annotations.

3.2.2. Consistency-based approach

Fundamental principles. The consistency-based semi-supervised learning approach is mainly based on the smoothing assumption. In the smoothing assumption, the prediction model should be robust to local perturbations within its input. This means that when we perturb the data points with a small amount of noise, the network’s predictions for the perturbed data points and the clean original data points should be similar. In the implementation of deep neural networks, the consistency-based approach can be easily extended to a semi-supervised learning setup by directly adding unsupervised consistency loss functions to the original supervised loss functions. In the field of natural image processing, typical methods include ρ-model (Laine and Aila 2016), Temporal Ensembling model (Laine and Aila 2016), Mean Teachers (Tarvainen and Valpola 2017) and UDA (Xie et al 2020).

Study in pathological image analysis. In pathological image analysis, Zhou et al (2020) proposed a new mean-teacher (MT) framework based on template-guided perturbation-sensitive sample mining. This framework consists of a teacher network and a student network. The teacher network is an integrated prediction network from K-times randomly augmented data, which is used to guide the student network to remain invariant to small perturbations at both feature and semantic levels. Su et al (2019) proposed a novel global and local consistency loss and performed the nuclei classification task based on the Mean-Teacher framework.

3.2.3. Graph-based approach

Fundamental principles. Methods of graph-based semi-supervised learning typically construct graphs to preserve the relationships of neighboring nodes, and use the graph transformations to simultaneously exploit information from labeled data and explore the underlying structure of unlabeled data. The key step of the graph-based semi-supervised learning methods is to construct a better graph to represent the original data structure. They usually define a graph on all data points (both labeled and unlabeled data points) and use weights to encode the similarity between pairs of the data points. In this way, the labeled information can be propagated through the graph to the unlabeled data points. For labeled data points, the predicted labels should match the true labels; similar data points defined by a similarity graph should have the same predictions. Graph-based semi-supervised methods are a relatively complex and long-developed field, and we recommend Van Engelen and Hoos (2020), Chong et al (2020) for a more thorough understanding.

Study in pathological image analysis. In pathological image analysis, Xu et al (2016) proposed a new framework that combines a CNN with a semi-supervised regularization term. They first generated a hypothetical label for each unlabeled sample, then proposed a graph-based smoothing term for regularization.
Su et al (2015) proposed an active learning and graph-based semi-supervised learning method for interactive cell segmentation. Inspired by the temporal ensembling model (Laine and Aila 2016), Shi et al (2020) proposed a graph-based temporal ensembling model GTE. This method creates ensemble targets for both features and label predictions for each training sample, and encourages the model to form consistent predictions under different perturbations to exploit the semantic information of unlabeled data and improve the robustness of the model to noisy labels.

3.2.4. Unsupervised-preprocessing-based approach

Fundamental principles. Unlike the previous approaches, unsupervised preprocessing-based approaches are typically dedicated to the unsupervised feature extraction, clustering (cluster-then-label), or initialization of the parameters of the subsequent supervised learning process (pre-training) from a large amount of unlabeled data. The most popular methods include autoencoders and their variants (Vincent et al 2008, Rifai et al 2011). Clustering is another method that enables adequate learning of the overall data distribution, thus many semi-supervised learning algorithms (Demiriz et al 1999, Dara et al 2002, Goldberg et al 2009) guide the subsequent classification process through clustering. The idea of the pre-training is to first pre-train a model using unsupervised methods with unlabeled data, and then use the parameters of this model as the initial parameters of the subsequent supervised training model, i.e. the subsequent supervised learning is fine-tuned on the basis of these initial parameters. On this basis, the large number of unlabeled data can fully guide the subsequent classification models with limited labeled data thus improving the performance of semi-supervised learning (Erhan et al 2010).

Study in pathological image analysis. In pathological image analysis, Peikari et al (2018) proposed a cluster-then-label semi-supervised learning method for identifying high-density regions in the data space and then utilized these regions to help support vector machines find decision boundaries. Lu et al (2019) proposed a semi-supervised method based on feature extraction and pre-training for the WSI-level breast cancer classification task, which is the first work that relies on self-supervised feature learning using contrastive predictive coding for weakly supervised histopathological image classification. Koohbanani et al (2021) proposed a joint framework of self-supervised learning and semi-supervised learning for pathological images. They proposed three pathology-specific self-supervised tasks, magnification prediction, magnification jigsaw prediction and hematoxylin channel prediction, to learn high-level semantic information and domain invariant information in pathological images. Srinidhi et al (2022) also proposed a framework that combines self-supervised learning with semi-supervised learning. They first proposed the resolution sequence prediction (RSP) self-supervised auxiliary task to pre-train the model through unlabeled data, and then they performed fine-tuning of the model on the labeled data. After that they used the trained model from the above two steps as the initial weights of the model for further semi-supervised training based on the teacher-student consistency framework.

3.2.5. Other approaches

Among semi-supervised adversarial networks, there are many other approaches, such as the methods based on generative adversarial networks (GAN) (Goodfellow et al 2014, Goodfellow 2016, Salimans et al 2016, Odena 2016, Dai et al 2017), Manifold-based methods (Belkin et al 2005, 2006, Rifai et al 2011, 2011, Weston et al 2012) and association learning based methods (Haeusser et al 2017).

In pathological image analysis, Kapil et al (2018) first used auxiliary classifier generative adversarial networks (AC-GAN) for the pathological image semi-supervised classification task and achieved favorable results. Cong et al (2021) proposed to use a GAN-based semi-supervised learning method to accomplish the stain normalization problem for pathological images. Sparks and Madabhushi (2016) proposed a semi-supervised method based on epidemic learning to accomplish a content-based histopathological image retrieval task. Li et al (2018) developed an Expectation-Maximization (EM)-based semi-supervised method for the semantic segmentation task of radical prostatectomy histopathological images. Su et al (2021) proposed a new semi-supervised method based on association learning for pathological image classification task inspired by Haeusser et al (2017). Some studies (Foucart et al 2019) have also attempted to analyze the weaknesses and effectiveness of semi-supervised, noisy learning and weak label learning based on deep learning for pathological image analysis.

3.3. Self-supervised learning paradigm

Unlike the former two paradigms, the self-supervised learning paradigm does not perform the classification or segmentation of pathological images directly, but in a two-stage ‘pre-training and fine-tuning’ approach. Due to the small number of annotated pathological images, it is not enough to use these data to directly train the model. Therefore, the self-supervised learning paradigm aims to first learn effective feature representations from a large amount of unlabeled data, which is called the pre-training process. Afterwards, the feature representations learned in the self-supervised auxiliary tasks are used to be transferred to train the downstream tasks using...
limited labeled data, which is called the fine-tuning process. In this way, good feature representations can effectively help the model to achieve good results even if it is trained with only a small amount of labeled data.

The process of pre-training, i.e. the learning process of good feature representations, is the key to self-supervised learning. Typically, self-supervised learning learns good feature representations by performing self-supervised auxiliary tasks. In a self-supervised auxiliary task, certain inherent properties of the unlabeled data are first used to generate supervised signals, and then the network is trained by these self-supervised signals. Different studies usually focus on designing different self-supervised auxiliary tasks to perform feature representation learning efficiently. According to the properties of the auxiliary tasks, existing self-supervised learning paradigms can be mainly classified into predictive self-supervised learning, generative self-supervised learning, and contrastive self-supervised learning. Predictive self-supervised learning learns good feature representations by constructing the auxiliary tasks as classification problems with unlabeled data; generative self-supervised learning learns good feature representations by reconstructing the input images; and contrastive self-supervised learning learns good feature representations by learning to distinguish between similar samples (positive samples) and dissimilar samples (negative samples). For a systematic review of self-supervised methods in the natural image domain and medical image domain, we recommend the reviews by Liu et al (2021) and Shurrab and Duwairi (2021).

In this section, we provide a detailed review of the studies on self-supervised learning for pathological image analysis. Currently, some studies focus on proposing innovative self-supervised frameworks for pathological images (we call them study on novel self-supervised frameworks), while others attempt to apply existing self-supervised learning methods to pathological image analysis (we call them study on application of self-supervised frameworks). We introduce studies on novel self-supervised frameworks in section 3.3.1, where we focus on predictive self-supervised learning, generative self-supervised learning, contrastive self-supervised learning, and hybrid self-supervised learning and their state-of-the-art approaches in pathological image analysis. We introduce the study on application of self-supervised frameworks in section 3.3.2. Table 5 summarizes a detailed list of literatures in this section.

3.3.1. Study on novel self-supervised frameworks
Predictive self-supervised learning approach

Fundamental principles. Predictive self-supervised learning learns good feature representations by constructing the auxiliary tasks as classification problems with unlabeled data, and the class labels for classification are constructed from the unlabeled data itself. Currently, predictive self-supervised auxiliary tasks widely applied in natural image processing are relative position prediction (Doersch et al 2015), solving jigsaw puzzles (Noroozi and Favaro 2016), and rotation angle prediction (Gidaris et al 2018), etc.

Study in pathological image analysis. In the field of pathological image processing, Sahasrabudhe et al (2020) proposed the auxiliary task of predicting patch magnification for cell nuclei segmentation. Their main idea is that given WSIs of different magnification classes (e.g. 5 ×, 10 ×, 20 ×), they first obtained patches of different magnifications from them and then predicted the magnification class of those patches by examining the size and texture of the cell nuclei in the patches. Srinidhi et al (2022) proposed the resolution sequence prediction (RSP) auxiliary task. First they used patches with different magnifications to construct different combinations of resolution sequences, and then trained the network to predict the order of the resolution sequences. Koohbanani et al (2021) proposed magnification prediction and solving magnification puzzles auxiliary tasks for pathological images. They first trained the network to accurately predict the magnification category, and then trained the network to predict the order of the patches with different magnifications.

Generative self-supervised learning approach

Fundamental principles. Generative self-supervised learning learns good feature representations by reconstructing the input images. They argue that the image itself is a useful self-supervised information and that the network can learn the potential feature representations of the generated image during the image reconstruction process. In natural image processing, autoencoders (Goodfellow et al 2016) are representative of early work on generative self-supervised feature representation learning. Later, denoising autoencoders (Vincent et al 2008) enhanced the feature representation capability of the model by introducing noise. Subsequently, researchers proposed a series of reconstructive self-supervised auxiliary tasks, including inpainting (Pathak et al 2016), colorization (Zhang et al 2016), patch shuffling and restoration (Chen et al 2019, Zhou et al 2021) to further enhance the feature representation capability of the network and achieved promising results. On the other hand, a series of GAN-based models (e.g. DCGAN 2015, BiGAN 2016) have also been used to perform self-supervised representation learning. In the latest self-supervised studies on natural images, a series (e.g. Bao et al 2022, MAE 2021, PeCo 2021, etc.) of self-supervised studies based on masked image blocks and reconstruction using Transformer achieved the highest performance, which is expected to start a new wave of research on reconstruction-based self-supervised representation learning.
Table 5. List of literatures in the self-supervised learning section.

| Reference                  | Approach       | Disease type               | Staining | Dataset                      | Dataset scale | Dataset link                                      | Self-supervised method                                                                 | Downstream task                                                                 | Downstream performance                  |
|----------------------------|----------------|-----------------------------|----------|------------------------------|---------------|---------------------------------------------------|----------------------------------------------------------------------------------------|------------------------------------------|------------------------------------------|
| Sahasrabudhe et al (2020)  | Predictive     | —                           | H&E      | MoNuSeg database             | 1125 737 tiles | Kumar et al (2017)                                | Identification of the magnification levels for tiles                                 | Nuclei segmentation                     | AUC: 0.5354, AHD: 7.7502, Dice: 0.7477  |
| Srinidhi et al (2022)      | Predictive     | Breast Cancer, Colorectal Cancer | H&E      | Camelyon16 dataset          | 2579 patches  | Martel et al (2019)                               | Predicting the resolution sequences                                                  | Detection of tumor metastasis             | AUC: 0.876 (10% labels)                  |
|                            |                |                              |          | Kather multiclass dataset   | 100K patches   | Kather et al (2019)                               | Classification of tissue types                                                       | Quantification of tumor cellularity      | Accuracy: 0.982 (10% labels)             |
|                            |                |                              |          | Breast Cancer                | 399 slides     | https://camelyon16. grand-challenge. org/Data/     |                                                                                      | Detection of tumor regions               | AUC: 0.817 (1% labeled)                 |
| Koohbanani et al (2021)    | Predictive     | oral Squamous Cell Carcinoma | H&E      | LNM-OSSC dataset             | 217 slides     | Inhouse                                           | Magnification prediction and solving magnification puzzles                          | Prediction of metastases in the cervical lymph nodes                               | AUC: 0.806 (1% labeled)                 |
|                            |                |                              |          | Kather multiclass dataset   | 100K patches   | Kather et al (2019)                               |                                                                                      |                                                                                      |                                          |
| Muhammad et al (2019)      | Generative     | Cholangiocarcinoma           | H&E      | Intrahepatic cholangiocarcinoma (ICC) dataset | 246 patients | Inhouse                                           | Deep clustering convolutional autoencoder                                           | Classification of tissue types           | AUC: 0.903 (1% labeled)                 |
|                            |                |                              |          | CAMELYON16 dataset           | 100 000 patches | https://camelyon16. grand-challenge. org/Data/     |                                                                                      | Subtyping of cholangiocarcinoma         | CHL: 3863 (5 clusters) and 4314 (clustering weight = 0.2) |
| Mahapatra et al (2020)     | Generative     | Breast Cancer                | H&E      | CAMELYON17 dataset           | 100 000 patches | https://camelyon16. grand-challenge.org/Data/, inhouse |                                                                                      | Stain normalization                     | Average AUC: 0.9320                     |
|                            |                |                              |          | National Center for Tumor diseases (NCT) dataset | 86 slides      | https://zenodo.org/record/1214456#.Yvzd-nZBshE    |                                                                                      |                                                                                      | FID: 16.65                              |
| Reference                  | Approach       | Disease type          | Staining | Dataset                                                                 | Dataset scale                          | Dataset link                                                                 | Self-supervised method                                                                 | Downstream task                                                                 | Downstream performance |
|---------------------------|----------------|-----------------------|----------|--------------------------------------------------------------------------|-----------------------------------------|--------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------|----------------------|
| Quiros et al (2019)       | Generative     | Breast Cancer         | H&E      | Netherlands Cancer Institute (NKI) dataset and Vancouver General Hospital (VGH) dataset | 576 tissue microarrays (TMAs)           | Beck et al (2011)                                                             | Using generative adversarial networks (GANs) to capture key tissue features and structure information | Count of cancer, lymphocytes, or stromal cells                                      | FID: 32.05          |
| Breast Cancer             |                |                       |          |                                                                          |                                         | Beck et al (2011)                                                             |                                                                                                                                     |                                                               |
| Quiros et al (2021)       | Generative     | Colon cancer          | H&E      | National Center for Tumor diseases (NCT, Germany) dataset                | 100K tissue tiles                      | https://zenodo.org/record/1214456#.YvdnZBshE                                 | Presenting an adversarial learning model to extract feature representations of cancer tissue | Classifying tissue types and predicting the presence of tumor in Whole Slide Images (WSIs) using multiple instance learning (MIL) | AUC: 0.97 and Accuracy: 0.85; AUC: 0.98 and Accuracy: 0.94 |
| Lung Cancer               |                |                       |          |                                                                          |                                         | http://portal.gdc.cancer.gov                                               |                                                                                                                                     |                                                               |
| Breast Cancer             |                |                       |          |                                                                          |                                         | https://camelyon16.grand-challenge.org/Data/                                 |                                                                                                                                     |                                                               |
| Boyd et al (2021)         | Generative     | Colorectal Cancer     | H&E      | CRC benchmark dataset                                                   | 100K image tiles                       | https://doi.org/10.5281/zenodo.1214456                                      | Visual field expansion                                                                 | Classification of tiles into metastatic and non-metastatic classes               | Accuracy: 0.8569                                                                 |
| Breast Cancer             |                |                       |          |                                                                          |                                         | https://camelyon16.grand-challenge.org/Data/                                 |                                                                                                                                     |                                                               |
| Koohbanani et al (2021)   | Generative     | Oral Squamous Cell Carcinoma | H&E | LNM-OSCC dataset                                                        | 217 slides                             | Inhouse                                                                       | Hematoxylin channel prediction auxiliary task                                                                                      | Prediction of metastases in the cervical lymph nodes                             | AUC: 0.806 (1% labeled)                                                    |
| Colorectal Cancer         |                |                       |          |                                                                          |                                         | Kather et al (2019)                                                          |                                                                                                                                     |                                                               |
|                           |                |                       |          |                                                                          |                                         | Inhouse                                                                       | Classification of tissue types                                                                                                  | Classification of tissue types                                                                                               | AUC: 0.903 (1% labeled) Nucleus Classification: Lymphocyte Classification AUC 0.7856 |
|                           |                |                       |          |                                                                          |                                         | Kather et al (2019)                                                          |                                                                                                                                     |                                                               |
| Reference | Approach | Disease type | Staining | Dataset | Dataset scale | Dataset link | Self-supervised method | Downstream task | Downstream performance |
|-----------|----------|--------------|----------|---------|---------------|--------------|------------------------|-----------------|-----------------------|
| Hou et al (2019) | Generative | — | H&E | CRCHistoPhenotypes nuclear detection dataset | 100 images | Sirinsukunwattana et al (2016) | Sparse Convolutional auto-encoder (CAE) | Nucleus detection | F-measure: 0.8345 |
| | | | | MICCAI 2015 nucleus segmentation challenge dataset | 763 images | https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=20644646 | | | |
| Xie et al (2020) | Contrastive | — | H&E | MoNuSeg dataset | 44 images | Naylort et al (2018) | Scale-wise triplet learning and counting ranking | Nuclei segmentation | |
| Chhipa et al (2022) | Contrastive | Breast Cancer | H&E | BreakHis dataset | 7909 images | Spanhol et al (2015) | Magnification prior contrastive similarity | Classifying histopathological images | |
| | | Breast Cancer | | MICCAI 2015 Gland Segmentation Challenge (GLaS) dataset | 165 images | Sirinsukunwattana et al (2017) | Gland segmentation | |
| Xu et al (2020) | Contrastive | Colon Cancer | H&E | PatchCamelyon (PCam) image classification dataset | 327 680 patches | Veeling et al (2018) | Deformation representation learning | Semi-supervised classification | F1-score 0.900, Accuracy 0.8548 (10% labeled) |
| Wang et al (2021) | Contrastive | Liver, Renal, Colorectal, Prostatic, Pancreatic, and Cholangio Breast Cancers | H&E | Multiple histopathological image datasets including MHIST, NCT-CRC-HE, PatchCamelyon dataset | 2.7 million images | https://github.com/Xiyue-Wang/TransPath | Contrastive learning like BYOL, (Bootstrap your own latent: a new approach to self-supervised learning) | Histopathological image classification tasks | F1-score 0.8993, 0.9582, 0.8983 on MHIST, NCT-CRC-HE, PatchCamelyon dataset |
| Abbet et al (2020) | Generative + Contrastive | Colorectal Cancer | H&E | Clinicopathological dataset | 660 WSIs | Inhouse | Colorization, Image reconstruction and Contrastive learning | Survival analysis | C-Index: 0.6943 |
| Yang et al (2021) | Generative + Contrastive | Colorectal Cancer | H&E | NCTCRC-HE-100K dataset | 100K images | https://zenodo.org/record/1214456#.YvzdZ8kE | Cross-stain prediction, Contrastive training | Nine-class classification of histopathological images | Accuracy of eight-class classification with only 1,000 labeled data: 0.915 C-Index: 0.826 |
| Chen et al (2020) | Application | Glioma and Cell Carcinoma | H&E | The Cancer Genome Atlas (TCGA) dataset | 1505 images | http://portal.gdc.cancer.gov | Contrastive predictive coding (CPC) | Survival prognosis prediction | |
| Reference          | Approach   | Disease type | Staining | Dataset                  | Dataset scale                                      | Dataset link                                                                 | Self-supervised method | Downstream task                      | Downstream performance |
|--------------------|------------|--------------|----------|--------------------------|---------------------------------------------------|--------------------------------------------------------------------------------|------------------------|--------------------------------------|----------------------|
| Ciga et al (2022)  | Application| Multiple Types| H&E      | Out of the total 57 datasets from various institutions | A large number of images                         | https://github.com/ozanciga/self-supervised-histopathology                  | Contrastive learning  | Classification, Regression, and Segmentation | Multiple results       |
| Tellez et al (2019)| Application| Breast Cancer | H&E      | Camelyon16 dataset       | 400 WSIs                                          | https://camelyon16.grand-challenge.org/Data/Veta et al (2019)                 | Variational auto-encoder, Contrastive learning and BiGAN | Predicting the presence of metastasis | AUC: 0.725          |
| Stacke et al (2021)| Application| Multiple Types| H&E      | Camelyon16 dataset       | 400 slides                                        | https://github.com/k-stacke/ssl-pathology                                   | Contrastive learning  | Predicting tumor proliferation speed   | Spearman correlation: 0.522 | Multiple results |
|                    |            |              |          | AIDA-LNSK dataset        | 492 WSIs                                          | Veta et al (2019)                                                            |                        |                       |                  |
|                    |            |              |          | Multidata (samples from 60 publicly available datasets)| 96 slides                                         | Aresta et al (2019)                                                          |                        |                       |                  |
|                    |            |              |          | CRC-100K dataset         | 100K images                                       | Kather et al (2016)                                                           | Weakly-supervised cancer subtyping |                       | AUC: 0.886          |
|                    |            |              |          | BreastPathQ dataset      | 2766 patches                                      | Petrick et al (2021)                                                         | Contrastive learning  | Patch-level tissue phenotyping        | AUC: 0.987           |
|                    |            |              |          | TCGA-CRC dataset         | 355 patients                                      |                                                                           |                        |                       |                  |
|                    |            |              |          | TCGA-Gastric dataset     | 375 patients                                      |                                                                           |                        |                       |                  |
|                    |            |              |          | Camelyon16 dataset       | 400 slides                                        |                                                                           |                        |                       |                  |
|                    |            |              |          | TCGA-COAD dataset        | 461 slides                                        |                                                                           |                        |                       |                  |
| Lu et al (2019)    | Application| Breast Cancer | H&E      | BACH dataset             | 400 cases                                         | Aresta et al (2019)                                                          | Contrastive predictive coding (CPC) | Classification and localization of clinically relevant histopathological classes | Accuracy: 0.95       |

Colorectal Cancer

Gastric Cancer

Breast Cancer

TCGA-COAD dataset 461 slides Guinney et al (2015) | Contrastive learning Colorectal Cancer subtyping AUC: 0.882 (CMS1) and AUC: 0.829 (CMS3) | Multiple results

Aresta et al (2019) | Contrastive predictive coding (CPC) classification and localization of clinically relevant histopathological classes | Accuracy: 0.95
| Reference          | Approach                | Disease type                  | Staining | Dataset                                      | Dataset scale | Dataset link                                                      | Self-supervised method                                                                 | Downstream task                                                                 | Downstream performance      |
|--------------------|-------------------------|-------------------------------|----------|----------------------------------------------|---------------|------------------------------------------------------------------|---------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|-----------------------------|
| Zhao et al (2020)  | Application             | Colon Adenocarcinoma          | H&E      | The Cancer Genome Atlas (TCGA) dataset       | 425 patients  | http://portal.gdc.cancer.gov                                      | Variational Auto Encoder and Generative Adversial Network (VAE-GAN)                  | Predicting lymph node metastasis                                               | Accuracy: 0.6761              |
|                    |                         | Breast Cancer                 |          | Camelyon16 dataset                           | 400 cases     | https://camelyon16.grand-challenge.org                            |                                                                                      | Detection of lymph node metastases                                               | Accuracy: 0.8992              |
| Li et al (2021)    | Application             | Lung Cancer                   | H&E      | TCGA lung cancer dataset                     | 1054 cases    | https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/https://camelyon16.grand-challenge.org | Contrastive learning                                                                | Diagnosis of lung cancer subtypes                                               | Accuracy: 0.9571              |
|                    |                         |                                |          | Camelyon16 dataset                           | 399 slides    | Inhouse                                                          |                                                                                      | Detection of tumor regions                                                      | AUC: 0.817 (1% labeled)        |
| Koohbanani et al (2021) | Application            | Oral Squamous Cell Carcinoma | H&E      | LNM-OSCC dataset                             | 217 slides    |                                                                  | Magnification prediction, JigMag prediction and Hematoxylin channel prediction     | Prediction of metastases in the cervical lymph nodes                             | AUC: 0.806 (1% labeled)        |
|                    |                         | Colorectal Cancer             |          | Kather multiclass dataset                    | 100K patches  | Kather et al (2019)                                              | Classification of tissue types                                                     | AUC: 0.903 (1% labeled)          |                                                                            |
|                    |                         |                                |          | BreastPathQ dataset                          | 2579 patches  | Martel et al (2019)                                             | Detection of tumor metastasis                                                     | TC: 0.876 (10% labeled)           |                                                                            |
| Srinidhi et al (2022) | Application            | Breast Cancer, Colorectal Cancer | H&E      | Camelyon16 dataset                           | 399 WSIs      | https://camelyon16.grand-challenge.org/kather                      | Resolution sequence prediction                                                     | Classification of tissue types                                                  | AUC: 0.855 (10% labeled)       |
|                    |                         |                                |          | Kather multiclass dataset                    | 100K patches  | Kather et al (2019)                                              |                                                                                      | Quantification of tumor cellularity                                               | ACC: 0.982 (10% labeled)        |
|                    |                         |                                |          | MICCAI 2015 Gland Segmentation Challenge (GlaS) dataset | 165 images    | Sirinukunwattana et al (2017)                                    |                                                                                      |                                                                                |
| Zheng et al (2019) | Application             | Colon Cancer                  | H&E      | Fungus dataset                               | 84 images     | Zhang et al (2017)                                               | Variational auto encoder (VAE)                                                     | Active learning in biomedical image segmentation                              | F1 score: 0.909, 0.9252 (30% labeled) |
Study in pathological image analysis. In pathological image analysis, Muhammad et al (2019) proposed a new deep convolutional autoencoder-based clustering model to learn the feature representations of pathological images. Mahapatra et al (2020) incorporated semantic information into a GAN-based generative model for self-supervised feature representation learning and used it for the stain normalization task of pathological images. Quiros et al (2019, 2021) designed GANs for pathological images to extract key feature representations of tissues. Boyd et al (2021) proposed a new generative auxiliary task which performs representation learning by extending the view of image patches. Hou et al (2019) proposed a sparse convolutional autoencoder (CAE) for simultaneous nuclei detection and feature extraction in histopathological images. Koohbanani et al (2021) proposed the hematoxylin channel prediction auxiliary task, where they used hematoxylin and eosin (H&E) stained images to predict the hematoxylin channel pixel by pixel.

Contrastive self-supervised learning approach

Fundamental principles. The contrastive self-supervised approach is one of the most popular self-supervised paradigms, which focuses on learning good feature representations by encouraging the model to learn to distinguish between similar samples (positive samples) and dissimilar samples (negative samples).

Contrastive predictive coding (CPC) (Van den Oord et al 2018) is an early contrastive self-supervised method applied to natural image processing whose goal is to maximize the mutual information between patches (positive samples) from the same image and minimize the mutual information between patches (negative samples) from different images within a mini-batch. Typical subsequent studies have been devoted to constructing negative samples. MoCo (He et al 2020) is a momentum-based contrastive self-supervised framework, which is mainly based on the ideas of dynamic dictionary-lookup and queues. SimCLR (Chen et al 2020) is a simple contrastive learning framework that aims to maximize the cosine similarity between two augmented views of the same image (positive samples) and minimize the similarity between different images in a minibatch (negative samples).

These methods rely heavily on a large number of negative samples since only positive samples will easily lead to model degeneration, i.e. mapping the features of all samples to an identical vector. However, recent studies have shown that negative samples are not necessary. Caron et al (2020) introduced clustering into contrastive learning, thus eliminating the need for negative samples. Chen and He (2021) explored stop-gradient operation applied to siamese networks without the need for a large number of negative samples. Grill et al (2020), Caron et al (2021) proposed a self-supervised learning model based on a teacher-student knowledge distillation framework that achieves state-of-the-art performance without any negative samples.

Study in pathological image analysis. In pathological image analysis, Xie et al (2020) employed patches from different magnifications as positive samples and patches from different magnifications as negative samples and constructed scale-wise triplet loss to perform contrastive learning for the nuclei segmentation. Chhipa et al (2022) proposed Magnification Prior Contrastive Similarity (MPCS) to construct contrastive loss. Xu et al (2020) proposed a self-supervised Deformation Representation Learning (DRL) framework to learn semantic features from unlabeled pathological images. They used mutual information to train the network to distinguish original histopathological images from those deformed in local structure, while consistent global contextual information was maintained using noise contrastive estimation (NCE). Wang et al (2021) proposed Transpath based on the BYOL framework 2020. They first collected the current largest histopathological image dataset for self-supervised pre-training, which includes about 2.7 million images from 32,529 WSIs. Then they proposed a hybrid framework combining CNN and Transformer to extract both local structural features and global contextual features, and proposed a TAE module to further enhance the feature extraction capability.

Hybrid self-supervised learning approach. Many studies have also presented hybrid self-supervised methods for pathological images. Abbet et al (2020) proposed a combination of generative and contrastive self-supervised representation learning method for pathological images. They first applied colorization as a generative auxiliary task. Then, they constructed the contrastive loss using spatially neighboring patches as positive samples and distant patches as negative samples. Yang et al (2021) also proposed a self-supervised representation method combining generative and contrastive approaches for pathological images. They first proposed a generative-based self-supervised task called cross-stain prediction, in which they defined two encoders and decoders to predict the E-channel and H-channel, respectively, and then they used the encoders trained in the previous task to perform further contrastive training.

3.3.2. Study on applications of self-supervised frameworks

In addition to studies that aim to propose innovative self-supervised frameworks for pathological images, more studies have attempted to apply existing self-supervised learning methods to various pathological image analysis tasks. Chen et al (2020) proposed an end-to-end multimodal fusion framework for histopathological images and genomic data for survival prognosis prediction, in which they used contrastive predictive coding (CPC) pre-trained self-supervised features for initialization of the network model. Ciga et al (2022) showed through extensive experiments that using self-supervised pre-training methods can yield better features to improve
performance on several downstream tasks. They found that the success of contrastive self-supervised pre-training methods depended heavily on the diversity of the unlabeled training set rather than the number of images. On the other hand, positive and negative samples that are visually significantly different facilitate contrastive self-supervised learning, while positive and negative sample that contain only minor differences but are generally similar (e.g. normal patches versus patches containing only a small percentage of tumor regions) are not conducive to contrastive learning. However, this is uncommon in natural images, so it is particularly important to design targeted self-supervised tasks for the characteristics of pathological images. Tellez et al. (2019) used the variational autoencoder 2013, contrastive learning 2016 and BiGAN 2016 for the compression of gigapixel pathological images and evaluated the performance on a synthetic dataset and two public histopathology datasets, respectively, achieving promising results. Stacke et al. (2021) investigated how SimCLR 2020 could be extended for pathological images to learn useful feature representations. They systematically compared the differences between ImageNet data and histopathology data and how this affected the goals of self-supervised learning, and pointed out the impact that designing for different self-supervised goals would have on the results. Chen and Krishnan (2022) comprehensively compared the performance of ImageNet pre-trained features, SimCLR pre-trained features, and DINO 2021 pre-trained features in weakly supervised classification and fully supervised classification tasks for histopathological images. They found that the DINO-based knowledge distillation framework could better learn effective and interpretable features in pathological images. Saillard et al. (2021) and Dehaene et al. (2020) used the MoCo V2 2020 self-supervised learning method to train pathological images and the experimental results showed that the results using the self-supervised pre-trained features were consistently better than those using features pre-trained on ImageNet under the same conditions. Lu et al. (2019), Zhao et al. (2020), and Li et al. (2021) used contrastive predictive coding (CPC) 2018, VAE-GAN 2016, and SimCLR 2020 self-supervised pre-trained features for weakly supervised WSI classification, respectively, and achieved the current state-of-the-art performance. Kooebanani et al. (2021) developed a semi-supervised learning framework facilitated by self-supervised learning with a multi-task learning approach for training, i.e. training with a small amount of labeled data as the main task and self-supervised tasks as auxiliary tasks. In their study, they also compared the effectiveness of various commonly used pathology-agnostic self-supervised auxiliary tasks (including rotation, flipping, auto-encoder, real/fake prediction, domain prediction, etc.) to facilitate semi-supervised learning. Srinidhi et al. (2022) also attempted to use self-supervised pre-trained features to enhance semi-supervised learning. They first proposed the resolution sequence prediction (RSP) self-supervised auxiliary task to pre-train the model through unlabeled data, and then they fine-tuned the model on the labeled data. After that, they used the trained model from the above two steps as the initial weights of the model for further semi-supervised training based on the teacher-student consistency framework.

In addition, self-supervised learning has been used for a variety of other pathological tasks, such as pathological image retrieval (Shi et al. 2018, Yang et al. 2020), active learning (Zheng et al. 2019), and molecular signature prediction (Ding et al. 2020, Fu et al. 2020, Kather et al. 2020), etc.

4. Discussion and future trends

4.1. For weakly supervised learning paradigm

The two main goals of WSI analysis using the weakly supervised learning paradigm are global slide classification, which aims to accurately predict the labels of each WSI, and positive patch localization, which aims to accurately predict the labels of each positive patch in the positive bags. Among above two tasks, the former can be used for rapid automatic diagnosis of clinical pathology slides, such as early clinical screening, and the latter can be used for precise localization of tumor cells, as well as interpretable analysis of clinical diagnosis by deep learning networks. Based on the diagnostic results obtained from the whole slides, pathologists are often more interested in the precise location of tumor cells, the cell morphology and other microstructures for further analysis and corroborations. On the other hand, pathologists also expect new knowledge from the diagnosis of the deep neural networks, such as discovering new pathological patterns and structures, etc. A few current algorithms can perform the task of global slide classification well, but the task of positive patch localization is another challenge for most algorithms. A primary reason is that the loss functions of most bag-based deep MIL algorithms are defined only at the bag-level, and although mechanisms such as attention (Ilse et al. 2018) can be used to measure the contribution of each instance to the bag-level classification, the network does not have enough motivation to classify all instances accurately (Shi et al. 2020, Qu et al. 2022). On the other hand, instance-based methods and hybrid methods, although defining instance-level classifiers, usually face a high risk of errors in pseudo-labeling or key instance selection. Therefore, it is a new challenge for the weakly supervised learning paradigm to further improve the ability to classify instances while obtaining a better slide-level diagnosis.
Further, with the emergence of the methods of the weakly supervised segmentation in the natural image processing field (Lee et al. 2021, Ru et al. 2022, Xu et al. 2022, Pan et al. 2022, Chen et al. 2022), a new challenging direction for WSI analysis is to perform pixel-level semantic segmentation of the entire WSI based on weak or sparse labels. The task of the positive patch localization, which described in the previous section is still based on the classification of patches, and it is a more challenging task to further obtain pixel-level segmentation results based on the weak labels. A few current studies (Xu et al. 2019, Qu et al. 2020, Lerousseau et al. 2020, Belharbi et al. 2021) have made attempts in this new direction, but they still face many problems such as lack of details and precision on the segmentation results. Overall, for the weakly supervised learning paradigm, how to obtain the most detailed segmentation results as possible with weak labels is another promising study direction.

Another urgent need is the publicly available WSI datasets with fine-grained annotations at the patch level. As we all know, the scarcity of the publicly available pathological image datasets is an important factor hindering the development of the field. In recent years, we are grateful for the support of large public pathology datasets such as TCGA (2019), but public pathology datasets with fine-grained annotations are still in short supply for deeper research. To our knowledge, the large public WSI dataset with detailed annotation at the patch level is merely CAMELYON (Bejnordi et al. 2017). We should encourage an individual or organization to provide more public WSI datasets with detailed patch-level annotations to promote the development of this study field.

4.2. For semi-supervised learning paradigm
For semi-supervised learning paradigm, a new study direction is the combination with active learning, the purpose of which is to use the most effective labeled data to obtain the highest performance. Active learning aims to find the most valuable samples in the unlabeled dataset to be annotated through iterative interactions with experts, which allows to further exploit the effects of semi-supervised learning. There are already a lot of studies on pathological image analysis with the help of active learning (Yang et al. 2017, Zheng et al. 2019) or combination with semi-supervised learning and active learning (Parag et al. 2014, Su et al. 2015).

Another challenge is the effect that noisy data and domain variation have on the performance of semi-supervised learning algorithms. In the field of computational pathology, noisy annotations are very common, because the instance features of pathological images are very complex and variable, and their sizes are so huge that doctors are likely to suffer from missing and mislabeling during annotation. When performing multicenter validation, significant staining variation between the slides from different centers is also very common as there is no uniform standard for staining pathological images among different centers. Both the noisy labels and the domain variation are powerful factors that affect the performance of semi-supervised learning in real-world scenarios. Recent studies (Foucart et al. 2019, Cheng et al. 2020, Shi et al. 2020, Koohbanani et al. 2021, Marini et al. 2021) have made efforts on these two problems, and more studies in this field are expected.

4.3. For Self-supervised learning paradigm
For self-supervised learning paradigm, although current relevant studies in the field of natural images are developing rapidly, the direct applications of these methods to pathological images will be hindered by the strong domain discrepancy (Koohbanani et al. 2021, Ciga et al. 2022). Therefore, how to design more effective self-supervised auxiliary tasks for pathological images is a promising direction.

On the other hand, self-supervised learning has been promoting the development of weakly supervised learning and semi-supervised learning in pathological image analysis. As we all know, it is difficult for a network to learn effective feature representations with very limited annotations. In contrast, self-supervised learning is very suitable for learning effective feature representations from a lot of unlabeled data. Therefore, it will be a popular way to combine the features extracted by self-supervised pre-training with the weakly supervised or semi-supervised downstream tasks in the future. On the one hand, the efficient feature representations obtained from self-supervised pre-training will greatly improve the efficiency of weakly supervised learning and semi-supervised learning, and on the other hand, weakly supervised learning or semi-supervised learning will fully release the new potential of self-supervised learning in the field of computational pathology.

4.4. Limitations
This review also has several limitations. First, due to space limitations, this review does not include more clinical studies. We focus more on top technical conferences and journals and do not include more excellent papers published in clinical journals. For more systematic reviews of clinical studies, see Cifci et al. (2022) and Kleppe et al. (2021) for details. In addition, since there are so many technical studies on artificial intelligence applied to computational pathology, it is difficult to summarize them all, and due to space limitations, we have tried to include as many recent articles as possible, while some of them have not been included.
5. Conclusion

In this review, we provide a systematic summary of recent studies on weakly supervised learning, semi-supervised learning, and self-supervised learning in the field of computational pathology from the theoretical and methodological perspectives. On this basis, we also present targeted solutions to some current difficulties and shortcomings in this field, and illustrate its future trends. Through a survey of over 130 papers, we find that the field of computational pathology is marching at high speed into a new era, which is automatic diagnosis and analysis with fewer annotation needs, wider application scope, and higher prediction accuracy.

Acknowledgments

This work was supported by National Natural Science Foundation of China under Grant 82072021.

ORCID iDs

Linzhao Qu https://orcid.org/0000-0001-8815-7050

References

Abbet C, Zlobovic I, Bozorgtabar B and Thiran J-P 2020 Divide-and-rule: self-supervised learning for survival analysis in colorectal cancer International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 480–9
Anand D, Yashashi K, Kumar N, Kane S, Gann P H and Sethi A 2021 Weakly supervised learning on unannotated h&E-stained slides predicts braf mutation in thyroid cancer with high accuracy J. Pathol. 255 232–42
Araújo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, Polónia A and Campilho A 2017 Classification of breast cancer histology images using convolutional neural networks PloS One 12 e0177544
Aresta G et al 2019 Bach: grand challenge on breast cancer histology images Med. Image Anal. 56 122–39
Bao H, Dong L and Wei F 2022 Beit: Bert pre-training of image transformers (https://openreview.net/forum?id=p-BhZSe59o4)
Basavanthally A, Ganesan S, Feldman M, Shih N, Mies C, Tomaszewski J and Madabhushani A 2013 Multi-field-of-view framework for distinguishing tumor grade in er− breast cancer from entire histopathology slides IEEE Trans. Biomed. Eng. 60 2089–99
Beck A H, Sangoi A R, Leung S, Marinelli R J, Nielsen T O, Van De Vijver M J, West R B, Van De Rijn M and Koller D 2011 Systematic analysis of breast cancer morphology uncovers stromal features associated with survival Sci. Transl. Med. 3 108ra113–108ra113
Belharbi S, Rony J, Dolz J, Ayed I B, McCaffrey L and Granger E 2022 Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty IEEE Trans. Med. Imaging 41 702–14
Belkin M, Niyogi P and Sindhwani V 2005 On manifold regularizationInt. Workshop Artificial Intell. Stat. 17–24
Belkin M, Niyogi P and Sindhwani V 2006 Manifold regularization: a geometric framework for learning from labeled and unlabeled examples J. Mach. Learn. Res. 7 11
Blum A and Mitchell T 1998 Combining labeled and unlabeled data with co-training Proc. of the Eleventh Annual Conf. on Computational Learning Theory pp 92–100
Boyd J, Liashuhu M, Deutsch E, Paragios N, Christodoulidis S and Vakalopoulou M 2021 Self-supervised representation learning using visual field expansion on digital pathology Proc. of the IEEE/CVF Int. Conf. on Computer Vision pp 639–47
Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, van der Laak J, Hulsbergen-van de Kaa C and Litjens G 2020 Field-of-view framework for automated diagnosis and weakly-supervised segmentation of breast cancer histology images Med. Image Anal. 56 122–39
Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, van der Laak J, Hulsbergen-van de Kaa C and Litjens G 2020 Automated deep-learning system for gleason grading of prostate cancer using biopsies: a diagnostic study Lancet Oncol. 21 233–41
Campanella G, Hanna M G, Geneslaw L, Mira A, Werneck Krauss Silva V, Busam K J, Brogi E, Reuter V E, Klimstra D S and Fuchs T J 2019 Clinical-grade computational pathology using weakly supervised deep learning on whole slide images Nat. Med. 25 1301–9
Caron M, Misra I, Mairal J, Goyal P, Bojanowski P and Joulin A 2020 Unsupervised learning of visual features by contrasting cluster assignments Adv. Neural Inf. Proces. Syst. 33 9912–24
Caron M, Touvron H, Misra I, Jegou H, Mairal J, Bojanowski P and Joulin A 2021 Emerging properties in self-supervised vision transformers Proc. of the IEEE/CVF Int. Conf. on Computer Vision pp 9650–60
Chang C-C and Lin C-J 2011 Libsvm: a library for support vector machines ACM Trans. Intell. Syst. Technol. (TIST) 2 1–27
Chaudhary K, Poirion O B, Lu L and Garnimre I X 2018 Deep learning-based multi-omics integration robustly predicts survival in liver cancer Clin. Cancer Res. 24 1248–59
Chen H, Qi X, Yu L, Dou Q, Qin J and Heng P A 2017 Dcam: deep contour-aware networks for object instance segmentation from histology images Med. Image Anal. 36 135–46
Chen L, Bentley P, Mori K, Misawa K, Fujiwara M and Rueckert D 2019 Self-supervised learning for medical image analysis using image context restoration Med. Image Anal. 58 101539
Chen R and Krishnan R G 2022 Self-supervised vision transformers learn visual concepts in histopathology arXiv:2203.00585
Chen R J, Lu N I and Mahmood F 2020 Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis IEEE Trans. Med. Imaging 41 757–770
Chen T, Kornblith S, Norouzi M and Hinton G 2020 A simple framework for contrastive learning of visual representations Int. Conf. on Machine Learning, PMLR pp 1597–607
Chen X, Fan H, Girshick R and He K 2020 Improved baselines with momentum contrastive learning arXiv:2003.04297
Chen X and He K 2021 Exploring simple siamese representation learning Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition pp 15750–8
Chen Z, Wang T, Wu X, Hua X-S, Zhang H and Sun Q 2022 Class re-activation maps for weakly-supervised semantic segmentation arXiv:2203.00962
Chen Z, Zhang J, Che S, Huang J, Han X and Yuan Y 2021 Diagnostic like a pathologist: weakly-supervised pathologist-tree network for slide-level immunohistochemical scoring XXXV AAAI Conf. on Artificial Intelligence (AAAI-21) (AAAI Press) pp 47–54
Cheng H-T, Yeh C-F, Kuo P-C, Wei A, Liu K-C, Ko M-C, Chao K-H, Peng Y-C and Liu T-L 2020 Self-similarity student for partial label histopathology image segmentation European Conference on Computer Vision (Berlin: Springer) pp 117–32
Chhipa P C, Upadhyay R, Pihlgren G G, Saini R, Uchida S and Liwicki M 2022 Magnification prior: a self-supervised method for learning representations on breast cancer histopathological images at arXiv:2203.07707
Chikontwe P, Kim M, Nam S J, Go H and Park S H 2020 Multiple instance learning with center embeddings for histopathology classification International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 519–28
Chong Y, Ding Y, Yan Q and Pan S 2020 Graph-based semi-supervised learning: a review Neurocomputing 408 216–38
Cifci D, Foersch S and Kather J N 2022 Artificial intelligence to identify genetic alterations in conventional histopathology J. Pathol. 257 430–44
Ciga O, Xu T and Martel A L 2022 Self-supervised contrastive learning for digital histopathology Mach. Learn. Appl. 7 100198
Clark K et al 2013 The cancer imaging archive (tcia): maintaining and operating a public information repository J. Digit. Imaging 26 1045–57
Cong C, Liu S, Ieva A D, Pagnucco M, Berkovsky S and Song Y 2021 Semi-supervised adversarial learning for stain normalisation in histopathology images International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 581–91
Coudray N, Ocampo P S, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, Moreira A L, Razavian N and Tsirigos A 2018 Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning Nat. Med. 24 1559–67
Cruz-Roa A, Basavanahally A, González F, Gilmore H, Feldman M, Ganesan S, Shih N, Tomaszewski J and Madabhushi A 2014 Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks Medical Imaging 2014: Digital Pathology vol 9041 904103 SPIE904103
Cruz-Roa A, Gilmore H, Basavanahally A, Feldman M, Ganesan S, Shih N N, Tomaszewski J, González F A and Madabhushi A 2017 Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent Sci. Rep. 7 1–14
Dai Z, Yang Z, Yang F, Cohen W W and Salakhutdinov R R 2017 Good semi-supervised learning that requires a bad gan

Dong X, Bao J, Zhang T, Chen D, Zhang W, Yuan L, Chen D, Wen F and Yu N 2021 Peco: perceptual codebook for bert pre-training of vision

Doersch C, Gupta A and Efros A A 2015 Unsupervised visual representation learning by context prediction

Doyle S, Rodriguez C, Madabhushi A, Tomaszeweski J and Feldman M 2006 Detecting prostatic adenocarcinoma from digitized histology Proc. of the IEEE Int. Conf. on Medical Imaging pp 1422–30
Donahue J, Krähenbühl P and Darrell T 2016 Adversarial feature learning arXiv:1605.09782
Ding K, Liu Q, Lee E, Zhou M, Lu A and Zhang S 2020 Feature-enhanced graph networks for genetic mutational prediction using histopathological images in colon cancer International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 294–304
Doernburg D, Debeir O and Decaestecker C 2019 Snow: semi-supervised, noisy and one-class unsupervised learning in histology arXiv:1912.03583
Ding K, Liu Q, Lee E, Zhou M, Lu A and Zhang S 2020 Feature-enhanced graph networks for genetic mutational prediction using histopathological images in colon cancer International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 294–304

Ding K, Liu Q, Lee E, Zhou M, Lu A and Zhang S 2020 Feature-enhanced graph networks for genetic mutational prediction using histopathological images in colon cancer International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 294–304

Donahue J, Krähenbühl P and Darrell T 2016 Adversarial feature learning arXiv:1605.09782
Dong X, Bao J, Zhang T, Chen D, Zhang W, Yuan L, Chen D, Wen F and Yu N 2021 Peco: perceptual codebook for bert pre-training of vision transformers arXiv:2111.12710

Doyle S, Hwang M, Shah K, Madabhushi A, Feldman M and Tomaszewski J 2007 Automated grading of prostate cancer using architectural and textural image features 2007 IV IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro, IEEE pp 1284–7

Doyle S, Rodriguez C, Madabhushi A, Tomaszewski J and Feldman M 2006 Detecting prostatic adenocarcinoma from digitized histology using a multi-scale hierarchical classification approach 2006 Int. Conf. of the IEEE Engineering in Medicine and Biology Society, IEEE pp 4759–62

Ehteshami Bejnordi B et al 2018 Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies Modern Pathol. 31 1502–12

Erhan D, Courville A, Bengio Y and Vincent P 2010 Why does unsupervised pre-training help deep learning? Proc. of the Thirteenth Int. Conf. on Artificial Intelligence and Statistics, JMLR Workshop and Conf. Proc. pp 201–8

Feng J and Zhou Z-H 2017 Deep mml network Proc. of the AAAI Conf. on Artificial Intelligence, vol 31

Foucart A, Debeir O and Decaestecker C 2019 Snow: semi-supervised, noisy and one-class unsupervised learning in histology arXiv:1912.03583

Fu Y, Jung A W, Torne R V, Gonzalez S, Vöhringer H, Shmatko A, Yates L R, Jimenez-Linan M, Moore L and Gerstung M 2020 Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis Nat. Cancer 1 800–10

Gelasca E D, Byun J, Obara B and Manjunath B 2008 Evaluation and benchmark for biological image segmentation 2008 XV IEEE Int. Conf. on Image Processing, IEEE pp 1816–9

Gerttych A, Ing N, Ma Z, Fuchs T J, Salman S, Mohanty S, Bhole S, Velásquez-Vacca A, Amin M B and Knudsen B S 2015 Machine learning approaches to analyze histological images of tissues from radical prostatectomies Comput. Med. Imaging Graph. 46 197–208

Gidaris S, Singh P and Komodakis N 2018 Unsupervised representation learning by predicting image rotations arXiv:1803.07728

Goldberg A, Zhu X, Singh A, Xu Z and Nowak R 2020 Multi-manifold semi-supervised learning Arxiv. Intell. Stat., PMLR 169–76

Goodfellow I 2016 Nips 2016 tutorial: generative adversarial networks arXiv:1701.00160

Goodfellow I, Bengio Y, Courville A and Bengio Y 2016 Deep Learn. vol 1

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial nets Adv. Neural Inf. Process. Syst. 27

Grill J-B et al 2020 Bootstrap your own latent-a new approach to self-supervised learning Adv. Neural Inf. Process. Syst. 33 21271–84

Gu F, Burlutskiy N, Andersson M and Wilén L K 2018 Multi-resolution networks for semantic segmentation in whole slide images, computational pathology and ophthalmic Medical Image Analysis (Berlin: Springer) pp 11–8

Guinney J et al 2015 The consensus molecular subtypes of colorectal cancer Nat. Med. 21 1350–6
Mahapatra D, Bozortgarb T, Thiran J-P and Shao L 2020 Structure preserving stain normalization of histopathology images using self-supervised semantic guidance International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 309–19.

Marin N, Otalora S, Müller H and Atzori M 2021 Semi-supervised training of deep convolutional neural networks with heterogeneous data and few local annotations: an experiment on prostate histopathology image classification Med. Image Anal. 73 102165.

Maron O and Lozano-Pérez T 1997 A framework for multiple-instance learning Adv. Neural Inf. Process. Syst. 10.

Marti A, Nolteh-Mozes S, Salama S, Akrab S and Peikari M 2019 Assessment of residual breast cancer cellularity after neoadjuvant chemotherapy using digital pathology Cancer Imaging Archive.

Meric C, Aksoy S, Mercan E, Shapiro L G, Weaver D L and Elmore J G 2017 Multi-instance multi-label learning for multi-class classification of whole slide breast histopathology images IEEE Trans. Med. Imaging 37 316–25.

Muhammad H et al 2019 Unsupervised subtyping of cholangiocarcinoma using a deep clustering convolutional autoencoder International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 604–12.

Murthy V, Hou L, Samaras D, Kurc T M and Saltz J H 2017 Center-focusing multi-task cnn with injected features for classification of glioma nuclear images 2017 IEEE Winter Conf. on Applications of Computer Vision (WACV), IEEE pp 834–41.

Myronenko A, Xu Z, Yang D, Roth R H and Xu D 2021 Accounting for dependencies in deep learning based multiple instance learning for whole slide imaging International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 329–38.

Nagral K et al 2019 Development and validation of a deep learning algorithm for improving gleson scoring of prostate cancer NPI Digit. Med. 2 1–10.

Naik N, Madani A, Esteva A, Keskar N S, Press M F, Ruderman D, Agus D B and Socher R 2020 Deep learning-enabled breast cancer hormonal receptor status determination from base-level h&E stains Nat. Commun. 11 1–8.

Naylor P, Léa M, Reyal F and Walter T 2018 Segmentation of nuclei in histopathology images by deep regression of the distance map Int. J. Comput. Assist. Radiol. Surg. 13 1483–91.

Norozi M and Favaro P 2016 Unsupervised learning of visual representations by solving jigsaw puzzles European Conference on Computer Vision (Berlin: Springer) pp 69–84.

Odena A 2016 Semi-supervised learning with generative adversarial networks arXiv:1606.01583.

Pan J, Bi Q, Yang Y, Zhu P and Bian C 2022 Label-efficient hybrid-supervised learning for medical image segmentation arXiv:2203.03956.

Parag T, Plaza S and Scheffer L 2014 Small sample learning of superpixel classifiers for esm segmentation International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 389–97.

Pathak D, Krahenvukh P, Donahue J, Darrell T and Efros A A 2016 Context encoders: Feature learning by inpainting Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition pp 2536–44.

Peikari M, Salama S, Nolteh-Mozes S and Martel A L 2018 A cluster–then-label semi-supervised learning approach for pathology image classification Sci. Rep. 8 1–13.

Petrick N et al 2021 Spie-aapm-nci breastpathq challenge: an image analysis challenge for quantitative tumor cellularity assessment in breast cancer histology images following neoadjuvant treatment J. Med. Imaging 8 034501.

Quiser T, Sirinukunwattana K, Nakane K, Tsang Y-W, Epstein D and Raipoot N 2016 Persistent homology for fast tumor segmentation in whole slide histology images Proc. Comput. Sci. 90 119–24.

Qu H, Wu P, Huang Q, Yi J, Yan Z, Li K, Riedflinger G M, De S, Zhang S and Metaxas D N 2020 Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images IEEE Trans. Med. Imaging 39 3653–66.

Qu L, Luo X, Liu S, Wang L and Song Z 2022 Dgmil: distribution guided multiple instance learning for whole slide image classification arXiv:2206.08961.

Quiros A C, Coudray N, Yeaton A, Sunhem W, Murray-Smith R, Tsirigos A and Yuan K 2021 Adversarial learning of cancer tissue representations arXiv:2108.02223.

Quiros A C, Murray-Smith R and Yuan K 2019 Pathogeny: learning deep representations of tissue cancer type arXiv:1907.02644.

Qureshi H, Sertel O, Rajpoot N, Wilson R and Gurcan M 2008 Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 196–204.

Radford A, Metz L and Chintala S 2015 Unsupervised representation learning with deep convolutional generative adversarial networks arXiv:1511.06434.

Raipoot K and Raipoot N 2004 Svm optimization for hyperspectral colon tissue cell classification International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 829–37.

Ramamon J and De Raedt L 2000 Multi instance neural networks Proc. of the ICML-2000 Workshop on Attribute-value and Relational Learning pp 53–60.

Rifai S, Dauphin Y N, Vincent P, Bengio Y and Muller X 2011 The manifold tangent classifier Adv. Neural Inf. Process. Syst. 24.

Rifai S, Vincent P, Muller X, Glorot X and Bengio Y 2011 Contractive auto-encoders: explicit invariance during feature extraction Int. Conf. on Machine Learning.

Rony J, Belharbi S, Dolz J, Ayed I B, McCaffrey L and Granger E 2019 Deep weakly-supervised learning methods for classification and localization in histology images: a survey arXiv:1909.03354.

Ru L, Zhan Y, Yu B and Du B 2022 Learning affinity from attention: end-to-end weakly-supervised semantic segmentation with transformers arXiv:2203.02664.

Sahasrabudhe M, Christoudoulidis S, Salgado R, Michie S, Loi S, André F, Paragios N and Vakalopoulou M 2020 Self-supervised nuclei segmentation in histopathological images using attention International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 393–402.

Saillard C, Dehaene O, Marchand T, Moindrot O, Kamoun A, Schmauch B and Jegou S 2021 Self supervised learning improves dmmr / msi detection from histology slides across multiple cancers arXiv:2109.05389.

Saillard C et al 2020 Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides Hepatology 2020–13.

Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A and Chen X 2016 Improved techniques for training gans Adv. Neural Inf. Process. Syst. 29.

Shaban M, Khurram S A, Fraz M M, Alsabuia N, Masood I, Mushitaq S, Hassan M, Loya A and Raipoot N M 2019 A novel digital score for abundance of tumour infiltrating lymphocytes predicts disease free survival in oral squamous cell carcinoma Sci. Rep. 9 1–13.

Shao Z et al 2021 Transmil: transformer based correlated multiple instance learning for whole slide image classification Adv. Neural Inf. Process. Syst. 34.
Sharma Y, Shrivastava A, Ehsan I, Moskaluk C A, Syed S and Brown D 2021 Cluster-to-conquer: a framework for end-to-end multi-instance learning for whole slide image classification Med. Imaging Deep Learn., PMLR 682–98
Shaw S, Pajak M, Lisowska A, Tsafuris S A and O’Neill A Q 2020 Teacher-student chain for efficient semi-supervised histology image classification arXiv:2003.08797
Shi X, Sapkota M, Xing F, Liu F, Cui I and Yang L 2018 Pairwise based deep ranking hashing for histopathology image classification and retrieval Pattern Recognit. 81 14–22
Shi X, Su H, Xing G and Yang L 2020 Graph temporal ensembling based semi-supervised convolutional neural network with noisy labels for histopathology image analysis Med. Image Anal. 60 101624
Shi X, Xing F, Xie Y, Zhang Z, Cui I and Yang L 2020 Loss-based attention for deep multiple instance learning Proc. of the AAAI Conf. on Artificial Intelligence vol 34, pp 5742–9
Shurbs S and Duwairi R 2021 Semi-supervised learning methods and applications in medical imaging analysis: a survey arXiv:2109.08685
Singh S, Janoos F, Pécot T, Caserta E, Leone G, Rittschier J and Machiraju R 2011 Identifying nuclear phenotypes using semi-supervised metric learning Biennial International Conference on Information Processing in Medical Imaging (Berlin: Springer) pp 398–410
Sirinukunwattana K et al 2017 Gland segmentation in colon histology images: the glas challenge contest Med. Image Anal. 35 489–502
Sirinukunwattana K, Raza S E A, Tsang Y-W, Smead D R, Cree I A and Rajpoot N M 2016 Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images IEEE Trans. Med. Imaging 35 1196–206
Skrede O J et al 2020 Deep learning for prediction of colorectal cancer outcome: a discovery and validation study Lancet 395 550–60
Spanhol F A, Oliveira I S, Petitjean C and Heutte L 2015 A dataset for breast cancer histopathological image classification Computer Imaging
Biomed. Eng. 63 1455–62
Sparks R and Madabhushi A 2016 Out-of-sample extrapolation utilizing semi-supervised manifold learning (ose-sml): content based image retrieval for histopathology images Sci. Rep. 6 1–15
Srinidhi CL, Ciga O and Martel A L 2021 Deep neural network models for computational histopathology: a survey Med. Image Anal. 67 101813
Srinidhi CL, Kim S W, Chen F-D and Martel A L 2022 Self-supervised driven consistency training for annotation efficient histopathology image analysis Med. Image Anal. 75 102256
Stacke K, Unger J, Lundstroem C and Elhettet G 2021 Learning representations with contrastive self-supervised learning for histopathology applications arXiv:2112.05760
Su H, Shi X, Cai J and Yang L 2019 Local and global consistency regularized mean teacher for semi-supervised histopathology image analysis International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 559–67
Su H, Yin Z, Huh S, Kanade T and Zhu J 2015 Interactive cell segmentation based on active and semi-supervised learning IEEE Trans. Med. Imaging 35 762–77
Su L, Liu Y, Wang M and Li A 2021 Semi-hic: a novel semi-supervised deep learning method for histopathological image classification Comp. Biol. Med. 137 104788
Swiderska-Chadaj Z et al 2019 Learning to detect lymphocytes in immunohistochemistry with deep learning Med. Image Anal. 58 101547
Tajbakhsh N, Jeyaseelan L, Li Q, Chiang J N, Wu Z and Ding X 2020 Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation Med. Image Anal. 63 101693
Tarvainen A and Valpola H 2017 Mean teacher chain: a better role models: Weight-averaged consistency targets improve semi-supervised deep learning results Adv. Neural Inf. Process. Syst. 30
2019 The Cancer Genome Atlas (TCGA). https://www.cancer.gov/tcga
Team N L S T R et al 2011 The national lung screening trial: overview and study design Radiology 258 243
Tellez D, Litjens G, van der Laak J and Ciompi F 2019 Neural image compression for gigapixel histopathology image analysis Proc. of the XXV Int. Conf. on Machine Learning pp 1096–103
Tolkach Y, Dohngorgon T, Toma M and Kristiansen G 2020 High-accuracy prostate cancer pathology using deep learning Nat. Mach. Intell. 2 411–14
Tomitaka N, Mase T, Sugimoto Y, Matsuda Y, Kato S, Ogawa N and Matsumoto F 2016 Predicting lung cancer risk using deep neural networks based on CT images and smoking history Sci. Rep. 6 1–15
Veta M et al 2020 Predicting breast tumor proliferation from whole-slide images: the tupa16 challenge Med. Image Anal. 43 1–14
Vincent P, Larochelle H, Bengio Y and Manzagol P-A 2008 Extracting and composing robust features with denoising autoencoders Proc. of the XXV Int. Conf. on Machine Learning pp 1006–103
Wang C-W, Lee Y-C, Galista E, Zhou F, Zhu H, Suzuki R, Komura D, Ishikawa S and Cheng S-P 2018 A benchmark for comparing precision medicine methods in thyroid cancer diagnosis using tissue microarrays Bioinformatics 34 1767–73
Wang X, Chen H, Gan C, Lin H, Dou Q, Tsougenis E, Huang Q, Cai M and Heng P-A 2019 Weakly supervised deep learning for whole slide lung cancer image analysis IEEE Trans. Cybernetics 50 3950–62
Wang X, Yan Y, Tang P, Bai X and Liu W 2018 Revisiting multiple instance neural networks Pattern Recognit. 74 15–24
Wang X, Yang S, Zhang J, Wang M, Zhang J, Huang J, Yang W and Han X 2021 Transtaph: transformer-based self-supervised learning for histopathological image classification International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 186–95
Ward R L and Hawkins N J 2015 Molecular and cellular oncology (mco) study tumour morphology UNSW Australia
Wei J, Tafe I, Linnik Y A, Vaikuks I T, Tomita N and Hassanpour S 2019 Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks Sci. Rep. 9 1–8
Wessels F et al 2021 Deep learning approach to predict lymph node metastasis directly from primary tumour histology in prostate cancer BJU Int. 128 352–60
Weston J, Ratle F, Mohabi H and Collobert R 2012 Deep learning via semi-supervised embedding Neural Networks: Tricks of the Trade (Berlin: Springer) 639–55
Wolf G-C et al 2020 Deep learning predicts molecular subtype of muscle-invasive bladder cancer from conventional histopathological slides Eur. Urol. 78 256–64
Wu J, Yu Y, Huang C and Yu K 2015 Deep multiple instance learning for image classification and auto-annotation Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition pp 3460–9
Xie Q, Dai Z, Hovy E, Luong T and Le Q 2020 Unsupervised data augmentation for consistency training Adv. Neural Inf. Process. Syst. 33 6256–68
Xie X, Chen J, Li K and Zheng Y 2020 Instance-aware self-supervised learning for nuclei segmentation International Conference on Medical Image Computing and Computer-assisted Intervention (Berlin: Springer) 341–50
Xu G, Song Z, Sun Z, Ku C, Yang Z, Liu C, Wang S, Ma J and Xu W 2019 Camel: a weakly supervised learning framework for histopathology image segmentation Proc. of the IEEE/CVF Int. Conf. on Computer Vision pp 10682–91
Xu J, Hou J, Zhang Y, Feng R, Ruan C, Zhang T and Fan W 2020 Data-efficient histopathology image analysis with deformation representation learning 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE pp 857–64
Xu K, Su H, Zhu J, Guan J-S and Zhang B 2016 Neuron segmentation based on cnn with semi-supervised regularization Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops pp 20–8
Xu L, Ouyang W, Bennamoun M, Boussaid F and Xu D 2022 Multi-class token transformer for weakly supervised semantic segmentation arXiv:2203.02891
Yalniz I Z, Jegou H, Chen K, Paluri M and Mahajan D 2019 Billion-scale semi-supervised learning for image classification arXiv:1905.00546
Yan Y, Wang X, Guo X, Fang J, Liu W and Huang J 2018 Deep multi-instance learning with dynamic pooling Asian Conf. on Machine Learning, PMLR pp 662–77
Yang J et al 2022 Prediction of her2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning Comput. Struct. Biotechnol. J. 20 33–42
Yang L, Zhang Y, Chen J, Zhang S and Chen D Z 2017 Suggestive annotation: a deep active learning framework for biomedical image segmentation International Conference on Medical Image Computing and Computer-assisted Intervention (Berlin: Springer) 399–407
Yang P, Hong Z, Yin X, Zhu C and Jiang R 2021 Self-supervised visual representation learning for histopathological images Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, Springer pp 47–57
Yang P, Zhai Y, Li L, Li H, Wang J, Zhu C and Jiang R 2020 A deep metric learning approach for histopathological image retrieval Methods 179 14–23
Yao J, Zhu X, Jonnagaddala J, Hawkins N and Huang J 2020 Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks Med. Image Anal. 65 101789
Yu K-H, Zhang C, Berry G J, Altman R B, Ré C, Rubin D L and Snyder M 2016 Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features Nat. Commun. 7 1–10
Zhang R, Isola P and Efros A A 2016 Colorful image colorization European Conference on Computer Vision (Berlin: Springer) pp 649–66
Zhang Y, Yang L, Chen J, Fredericksen M, Hughes D P and Chen D Z 2017 Deep adversarial networks for biomedical image segmentation utilizing unannotated images International Conference on Medical Image Computing and Computer-assisted Intervention (Berlin: Springer) pp 408–16
Zhao Y et al 2020 Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition pp 4837–46
Zheng H, Yang L, Chen J, Han J, Zhang Y, Liang P, Zhao Z, Wang C and Chen D Z 2019 Biomedical image segmentation via representative annotation Proc. of the AAAI Conf. on Artificial Intelligence vol 33, pp 5901–8
Zhou Y, Chen H, Lin H and Heng P-A 2020 Deep semi-supervised knowledge distillation for overlapping cervical cell instance segmentation International Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 521–31
Zhou Z-H and Li M 2005 Tri-training: exploiting unlabeled data using three classifiers IEEE Trans. Knowl. Data Eng. 17 1529–41
Zhou Z, Soufi V, Pang J, Gotway M B and Liang J 2021 Models genesis Med. Image Anal. 67 101840
Zhu X 2005 Semi-supervised learning literature survey CS Tech. Rep. (CS Technical Reports: University of Wisconsin-Madison Department of Computer Sciences)
Zhu X, Yao J, Zhu F and Huang J 2017 Wsisa: making survival prediction from whole slide histopathological images Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition pp 7234–42