# Partial-Label Contrastive Representation Learning for Fine-grained Biomarkers Prediction from Histopathology Whole Slide Images

Yushan Zheng, *Member, IEEE*, Kun Wu, Jun Li, Kunming Tang, Jun Shi, *Member, IEEE*, Haibo Wu, Zhiguo Jiang, and Wei Wang

***Abstract*—In the domain of histopathology analysis, existing representation learning methods for biomarkers prediction from whole slide images (WSI) face challenges due to the complexity of tissue subtypes and label noise problems. This paper proposed a novel partial-label contrastive representation learning approach to enhance the discrimination of histopathology image representations for fine-grained biomarkers prediction. We designed a partial-label contrastive clustering (PLCC) module for partial-label disambiguation and a dynamic clustering algorithm to sample the most representative features of each category to the clustering queue during the contrastive learning process. We conducted comprehensive experiments on three gene mutation prediction datasets, including USTC-EGFR, BRCA-HER2, and TCGA-EGFR. The results show that our method outperforms 9 existing methods in terms of Accuracy, AUC, and F1 Score. Specifically, our method achieved an AUC of 0.950 in EGFR mutation subtyping of TCGA-EGFR and an AUC of 0.853 in HER2 0/1+/2+/3+ grading of BRCA-HER2, which demonstrates its superiority in fine-grained biomarkers prediction from histopathology whole slide images. The source code is available at https://github.com/WkEEn/PLCC.**

***Index Terms*—WSI analysis, Gene mutation prediction, Representation learning, Partial-label learning**

## I. Introduction

The incidence of cancer is increasing year by year in various countries [1], [2]. Diagnosis based on histopathological images remains the gold standard for cancer diagnosis [3]. With the development of whole slide imaging and digital pathology, computer-aided cancer diagnosis based on whole slide image (WSI) analysis becomes a viable solution, especially for biomarkers prediction, which is often challenging to detect on the histopathology images [4], [5]. It is promising to improve the diagnostic efficiency of pathologists and reduce misdiagnosis [6]–[8].

The gigapixel characteristic of WSIs determines the current WSI analysis framework highly depends on local feature extraction and representation learning. It has become essential for WSI analysis [9]–[11]. However, learning effective representations for different lesion tissues is a challenging task because the structure and morphology of WSIs are much more complex than natural scene images. Previous WSI analysis frameworks generally utilized the supervised representation learning paradigm, which has been verified effective in various downstream tasks. However, supervised learning highly depends on the quantity and quality of manual annotations while the manual annotation of gigapixel WSIs is labor-intensive and error-prone. The workload of pathological image annotations has become the major bottleneck of WSI analysis development.

To address the shortage of pathological annotations, several studies attempted to reduce the dependence on manual annotation [12], [13]. Typically, self-supervised learning is a bold attempt to completely abandon manual annotation, which uses the image features themselves to guide the network training. Recent successes of contrastive learning in natural images vision tasks show the great priority in the field of representations learning [14], [15], and thereby have been introduced for local representation extraction of WSIs [16], [17]. Correspondingly, weakly supervised learning is widely studied for the WSI level analysis, which neither relies on fine-gained manual annotations by pathologists nor completely discards any annotations. These methods make use of weakly supervised semantic information, such as WSI-level or patient-level labels. Typically, an increasing number of studies formulate the WSI analysis task as a multiple instance learning (MIL)

This work was partly supported by the Beijing Natural Science Foundation (Grant No. 7242270), partly supported by the National Natural Science Foundation of China (Grant No. 62171007, 61901018, and 61906058), partly supported by the Fundamental Research Funds for the Central Universities of China (grant No. YWF-23-Q-1075 and JZ2022HGTB0285), partly supported by Emergency Key Program of Guangzhou Laboratory (Grant No. EKPG21-32), partly supported by Joint Fund for Medical Artificial Intelligence (Grant No. MAI2023C014), partly supported by National Key Research and Development Program of China (Grant No. 2021YFF1201004), and partly supported by Anhui Provincial Health and Medical Research Project (Grant No. AHWJ2023A10143). *(Corresponding authors: Wei Wang and Zhiguo Jiang)*

Yushan Zheng is with the School of Engineering Medicine, Beijing Advanced Innovation Center on Biomedical Engineering, Beihang University, Beijing 100191, China (e-mail: yszheng@buaa.edu.cn).

Kun Wu, Jun Li, Kunming Tang, and Zhiguo Jiang are with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China.

Jun Shi is with the School of Software, Hefei University of Technology, Hefei 230009, China.

Wei Wang and Haibo Wu are with the Department of Pathology, the First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230036, China, and also with the Intelligent Pathology Institute, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230036, China (e-mail: weiwang@hmfl.ac.cn).

problem, which can identify the most meaningful instances (image patches) for the recognition of a WSI.

One of the most popular applications in the field involves the prediction of biomarkers from histopathology WSIs. The status of biomarkers is defined based on the testing results of patients, including immunohistochemistry, fluorescence in situ hybridization, gene sequencing, etc. These states are invisible to the pathologists on hematoxylin-eosin staining images, making it challenging to annotate fine-grained mutated regions. These tasks with only patient-level labels accessible are naturally weakly supervised problems. Notably, Pao et al. [18] developed some deep learning algorithms to assess EGFR status using a real-world advanced lung adenocarcinoma cohort of 2099 patients with histopathology images. Farahmand et al. [19] proposed a CNN-based framework to predict HER2 status and achieved an area under the curve (AUC) of 0.810 on an independent TCGA test set. Schrammen et al. [20] developed a weakly supervised approach for the detection of BRAF mutations using WSIs. The prediction of microsatellite instability (MSI) in colorectal cancer (CRC) stands out as a particularly well-explored area [7], [21]–[23]. Furthermore, Niehues et al. [24] evaluated the efficacy of deep learning-based predictions for MSI, BRAF, KRAS, NRAS, and PIK3CA biomarker statuses from histopathological slides. In addition, Shamai et al. [25] introduced a system that achieved high predictive accuracy for PD-L1 status in breast cancer, with an impressive AUC ranging between 0.910 and 0.930. Other researchers have concentrated on assessing tumor mutational burden (TMB) in lung cancer [26]–[28].

However, it remains a significant problem in fine-grained biomarker prediction using WSI-level labels that current methodologies tend to assume tissue within the WSI can only be classified into two categories, e.g. biomarker-positive or biomarker-negative. The labels to be predicted are often present in a hierarchical structure and exhibit overlapping characteristics. This simplification leads to the generation of conflicting pseudo-labels within the typical binary-categorized, weakly supervised MIL frameworks, introducing noise into the modeling process of the weakly supervised WSI analysis framework. Consequently, most existing MIL methods suffer from this noisy supervision, rendering them less effective for fine-grained WSI classification tasks, particularly in the context of biomarker prediction.

In this paper, we proposed a novel weakly supervised representation learning framework for fine-grained WSI classification. We rethink the local representation learning problem under the weakly supervised paradigm and formulate it as the partial-label learning problem. Different from supervised learning or MIL, the proposed framework assigns a candidate label set for each patch extracted from a WSI based on the coexistence dependencies of the diagnosis prior. During the representation learning, the real label of each patch is identified from the candidate label set using the designed partial-label disambiguation module and then used to improve the ability of the representation to distinguish subtle but significant pathology patterns. The proposed framework was evaluated with WSI classification on 3 large-scale WSI datasets defined by multiple fine-grained histopathology identification tasks. The results have shown the proposed method achieved stably significant improvements than the other representation learning methods with different benchmark WSI classification models.

The contribution of this paper can be summarized as follows:

- We conducted an in-depth study on the label noise problem faced by representation learning in weakly supervised histopathological WSI classification tasks. We proposed a framework based on a partial-label learning paradigm and developed a contrastive representation learning framework with partial-label disambiguation, as shown in Fig. 1. This framework significantly outperforms existing representation learning methods when applied to the fine-grained classification task of biomarker prediction in WSIs. To our knowledge, this is the first research to address the label noise problem introducing partial-label learning in weakly supervised WSI multi-classification tasks at the representation learning level.
- We proposed a partial-label contrastive clustering (PLCC) module that can online achieve partial-label disambiguation during the contrastive learning process. This enables the use of more accurate pseudo-labels for high-quality representation learning. Furthermore, we proposed a dynamic clustering algorithm that continuously samples the most representative features of each category to the clustering queue during the contrastive learning process. This enhances the performance of partial-label disambiguation and finally ensures the model's stable convergence.
- We conducted experiments on three WSI datasets for biomarker predictions and compared our approach with nine methods. Comprehensive experiments have verified the superiority of our method.

## II. Related Works

### A. Contrastive Representation Learning

The framework we proposed is built on contrastive representation learning. Research has consistently shown that representations developed through contrastive learning methods surpass those from supervised learning in effectiveness and robustness for histopathology WSI analysis.

The application of contrastive learning to pathological image analysis is gaining momentum. Huang et al. [29] enhanced survival prediction with representations pre-trained on a pathological dataset via SimCLR, which avoided the use of representations transferred from ImageNet pre-trained models. Wang et al. [30] introduced TransPath, which combines BYOL and Transformer to address the issue of variable input sizes in pathological images, thereby improving classification performance. Ciga et al. [31] found that features pre-trained on a large-scale pathological dataset with SimCLR outperformed those pre-trained on ImageNet in several downstream tasks.

Additionally, some studies have revisited the suitability of contrastive learning methods for pathology image analysis to enhance their adaptability. Liu et al. [32] redefined the contrastive pair sampling strategy by forming positive pairs from adjacent tissue regions, thus capturing local spatial correlations. Yang et al. [33] highlighted the limitations of

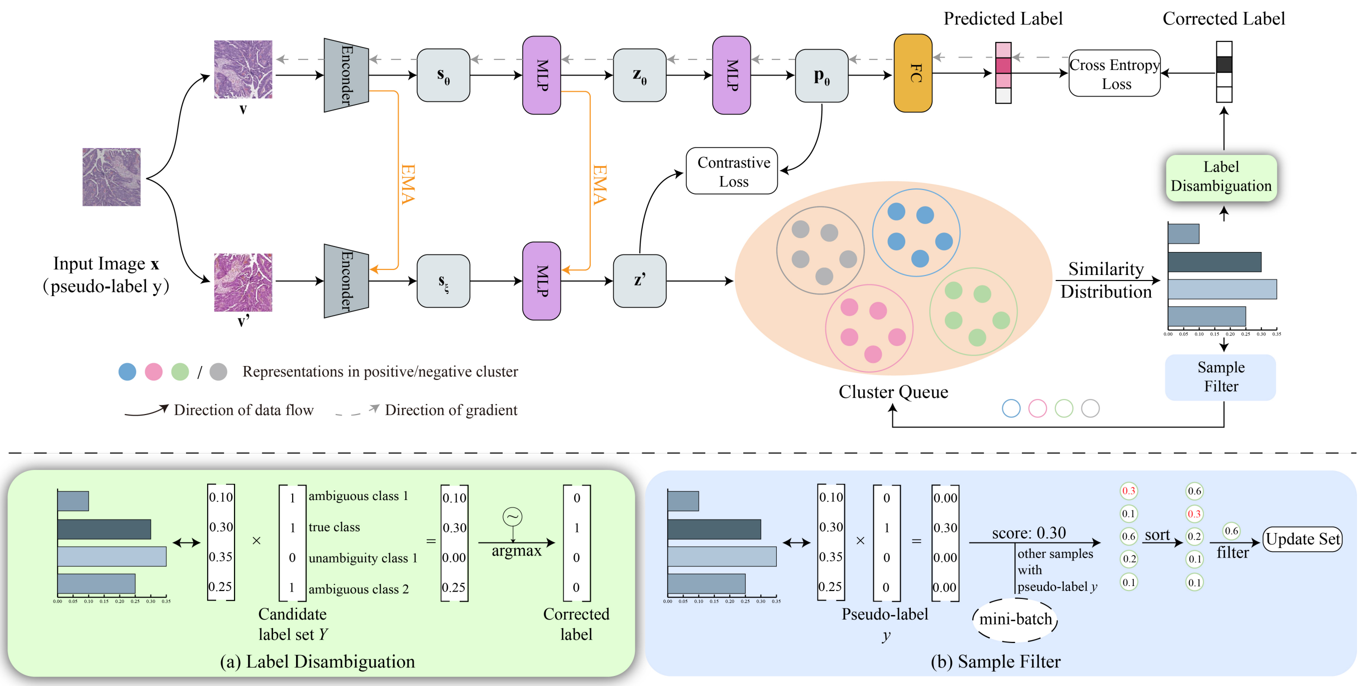


Fig. 1. The overview of our proposed partial-label contrastive learning framework, where (a) is the proposed partial-label disambiguation module, which aims to mine the true category of samples from the candidate labels, and (b) is the sample filter module, which aims to update the cluster queues.

traditional data augmentation methods and proposed a novel augmentation technique, stain vector perturbation, specifically designed for pathological images.

These studies not only affirm the viability of contrastive learning methods in computational pathology but also highlight areas for improvement in the representation learning of pathological images.

### B. MIL-based Weakly Supervised WSI Classification

In weakly supervised learning, the majority of research adopts a multi-instance learning (MIL) framework. This approach regards each WSI as a bag, with sampled patches acting as instances. Some researchers assign the same pseudo-label to each patch as the WSI and train a feature extraction network based on these pseudo-labels. This strategy, however, frequently results in a significant number of false positives.

Campanella et al. [34] introduced a Top-K mechanism to refine training sample selection. This method ranks tiles by their likelihood of being positive and focuses learning on the top-ranked tiles per slide. Following this methodology, Lerousseau et al. [13] suggested that tiles with a low probability of positivity are likely negative, and refined the training dataset during the learning process.

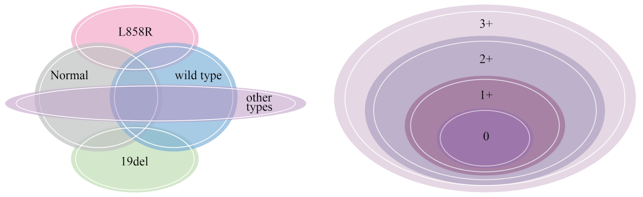
Kalra et al. [35] enhanced tissue representation robustness by using hierarchically arranged target labels for WSIs to fine-tune the feature extractor. Reisenbchler et al. [36] identified the $K$ nearest neighbors for each instance and used these neighbors to compute self-attention scores, which effectively models the patch relationships. Ding et al. [37] improved histopathology image representation learning with MIL by integrating multi-scale information.

The development of Transformer has brought a novel solution and the self-attention mechanism can easily capture the relevant between instances. TransMIL proposed by Shao et al. [38] has verified the effectiveness of feature aggregation by Transformer. Zheng et al. [39] proposed a Transformer framework based on regional anchors to mine the global and local features of slides. Wu et al. [40] introduced self-supervised pre-training into the WSI representation learning. Zhang et al. [41] proposed a double-tier attention-based MIL framework to capture the intrinsic slide features.

### C. Partial-label learning for pathology image analysis

As cancerous tissue differentiates and invades, tissue types in a pathological section are complex and varied. Furthermore, because doctors typically only report the highest level of pathology, the labels we collect are often incomplete. For example, in a tissue section of poorly differentiated adenocarcinoma, there may be areas of inflammation, highly differentiated tissue, and moderately differentiated tissue. Based on this, we can design a candidate label set for different sections, which defines the possible tissue types that may exist in each section. This candidate label set is often consistent among sections of the same category, forming a partial-label learning (PLL) problem.

Label disambiguation is the core of solving the PLL problem. It involves distinguishing the true label from a set of candidates, which becomes a challenging task in the presence of similar pathological patterns. Fan et al. [42] proposed the disambiguation correction network with graph representation to address these challenges by utilizing graph-based representations to enhance the understanding of relationships among candidate labels and refine label prediction accuracy. Recent developments have seen the integration of contrastive learning into the PLL framework to further improve label disambiguation. Contrastive learning can significantly enhance the disambiguation process, which learns to distinguish between similar and dissimilar instances. Wang et al. [43] and Xia et al. [44] have demonstrated how contrastive learning can be employed for effective label disambiguation in PLL and lead to more accurate and robust models. These approaches capitalize on the inherent strengths of contrastive learning to better handle the complexities of partial-label scenarios, particularly in pathology image analysis where tissue heterogeneity and morphological similarities pose substantial challenges. However, we have not identified any research that employs contrastive learning for partial-label disambiguation specifically in the field of pathology image representation learning.



(a) The relationship of EGFR labels (b) The relationship of HER2 labels

Fig. 2. The hierarchical relationships in the labels for biomarker prediction tasks, where (a) illustrates the overlapping relationships among various categories in the EGFR dataset, while (b) displays the progressive relationships between categories in the HER2 dataset.

## III. Methods

### A. Partial-label problem in WSI biomarkers prediction

Biomarkers prediction from WSI inherently exhibits hierarchical relationships among labels. To illustrate, we consider the task of identifying mutations in the epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor-2 (HER2) gene from histopathology WSIs. EGFR is a cell surface receptor activating cell growth and survival [45]. EGFR tyrosine kinase inhibitors (EGFR-TKIs) are the primary tools of targeted therapy for non-small cell lung cancer (NSCLC) of which the efficacy is associated with EGFR mutation status [46]. Such a dataset might include five subtypes: tumor-free, EGFR 19del mutation, EGFR L858R mutation, non-common driver mutations (wild type), and other driver gene mutations (other types). Here, ”wild type” denotes cancer tissues lacking EGFR gene mutations. As depicted in Fig. 2(a), tissues of this category may also appear in WSIs with EGFR-19del mutations. Therefore, a patch from such a WSI could belong to the 19del mutation, wild, and tumor-free categories. We address this as a partial-label issue by creating a label vector to represent the possible tissue types: [tumor-free, wild type, 19del mutation, L858R mutation, other types]. For example, patches from 19del mutation slides are assigned a partial label set $\mathbf{Y} = [1, 1, 1, 0, 0]$, indicating that the slides may also include patches of normal and wild type tissues. Additionally, HER2 is a transmembrane tyrosine kinase receptor with a pathological characteristic of promoting tumor angiogenesis and enhancing tumor invasiveness, including the IHC score of 0 (Normal), 1+, 2+, and 3+. It is a major classifier of molecular subtypes and the therapeutic target of breast cancer with a positive rate of 15%–30% [47]. The grading of HER2 is determined by the protein expression level in the immunohistochemical results. The higher the expression level, the higher the grading of HER2. It is usually compared with a standard cell line expressing a certain level of HER2 receptor, so there is no clear line between adjacent grades, which leads to the possibility that the higher grade class contains the lower grade class, thus forming a hierarchical structure as shown in Fig. 2(b).

The critical challenge lies in determining the true label $\hat{y}$ from the partial label set $\mathbf{Y}$ for each patch and properly utilizing the corrected label to enhance the efficacy of contrastive representation learning.

We built an integrated partial-label-based representation learning framework to tackle these challenges, which is illustrated in Fig. 1. It forms an integrated end-to-end representation learning architecture with the designed module named partial-label contrastive clustering (PLCC). The framework builds on the BYOL network structure. Additionally, a dynamic representation clustering module is placed at the end of the target network branch. These additions enable partial-label correction and weakly supervised representation learning throughout each step of training. We also introduced a category-aware cluster updating module, which ensures continuous updates of clusters during training. The details are presented in this section.

### B. Contrastive representation learning framework

We introduced BYOL as the basic representation learning structure. It’s a Siamese network consisting of two branches, namely an online network and a target network. The online network consists of an encoder $\mathcal{F}_\theta$, a projector $\mathcal{G}_\theta$ and a predictor $\mathcal{Q}_\theta$, which are defined by a set of trainable weights $\theta$. The target branch contains an encoder $\mathcal{F}_\xi$ and a projector $\mathcal{G}_\xi$ that share the same structures of $\mathcal{F}_\theta$ and $\mathcal{G}_\theta$ defined by another set of weights $\xi$. In addition, both projector and predictor are multilayer perceptron (MLP) that are composed of two fully connected layers, a BN layer, and a ReLU layer.

Given a patch $\mathbf{x}$, two different sets of augmentation methods $\mathcal{T}$ and $\mathcal{T}'$ are applied to obtain the augmented views for the patch $\mathbf{x}$, which are $\mathbf{v} = \mathcal{T}(\mathbf{x})$ and $\mathbf{v}' = \mathcal{T}'(\mathbf{x})$. Then, in the online network, the augmented view $\mathbf{v}$ is fed into the network to get representation $\mathbf{s}_\theta = \mathcal{F}_\theta(\mathbf{v})$ and the projection $\mathbf{z}_\theta = \mathcal{G}_\theta(\mathbf{s}_\theta)$. Correspondingly, in the target network, the representation $\mathbf{s}_\xi = \mathcal{F}_\xi(\mathbf{v}')$ and the projection $\mathbf{z}' = \mathcal{G}_\xi(\mathbf{s}_\xi)$ are obtained from $\mathbf{v}'$. Finally, the predictor output $\mathbf{p}_\theta = \mathcal{Q}_\theta(\mathbf{z}_\theta)$ of the online network and the output $\mathbf{z}'$ of the target network are normalized to calculate the loss. The loss function is defined as follows:

$$L_{con} = -\left\langle \mathcal{Q}_\theta(\mathbf{z}_\theta), \mathbf{z}' \right\rangle = -\frac{\mathcal{Q}_\theta^T(\mathbf{z}_\theta) \cdot \mathbf{z}'}{\left\| \mathcal{Q}_\theta(\mathbf{z}_\theta) \right\|_2 \times \left\| \mathbf{z}' \right\|_2}, \quad (1)$$

where $\langle \cdot \rangle$ denotes the cosine similarity measurement function. The above loss function is only used to update the online network weights $\theta$, and the target network branch is updated by the exponential moving average (EMA) mechanism with a hyperparameter $\tau$:

$$\xi = \tau\xi + (1-\tau)\theta, \tau \in [0,1] \tag{2}$$

Additionally, we built a supervised learning path at the end of the online target to utilize the information of partial labels. Specifically, a fully connected layer $\mathcal{FC}_\theta$ is built to predict the real label of the patch, and a cross-entropy loss function is applied to fit this prediction to the clean label determined by the PLCC module, which can be formulated as

$$L_{\text{cls}} = \hat{y}\log(\mathcal{FC}_\theta(\mathbf{p}_\theta)). \tag{3}$$

Finally, the overall optimization objective is

$$L = \lambda L_{\text{con}} + L_{\text{cls}}, \tag{4}$$

where $\lambda$ is a hyper-parameter for weighing the loss functions.

### C. Partial-label disambiguation

Partial-label disambiguation is crucial in the proposed framework, which aims to deduce the true label $\hat{y}$ from the partial label set for a sample $\mathbf{x}$. To achieve this, we established a cluster for each category, denoted as $U_k$, where $k$ ranges from 0, 1, ..., $C-1$, with $C$ representing the number of slide categories in the dataset. These clusters are designed to store the image representations $\mathbf{z}'$ associated with their respective category.

Considering an image $\mathbf{x}$ sampled from a WSI with label $y$, the proposed partial-label disambiguation process based on the category clusters is described as follows.

1) Feed $\mathbf{x}$ into the target branch to obtain the representation $\mathbf{z}'$ through the following modules:

$$\mathbf{z}' = \mathcal{G}_\xi(\mathcal{F}_\xi(\mathbf{v}')), \mathbf{v}' = \mathcal{T}'(\mathbf{x}) \tag{5}$$

2) Calculate the similarity distribution between the representation of $\mathbf{x}$ and the representations in the clusters which can be formulated as

$$d_{ki} = \langle \mathbf{z}', \mathbf{z}'_{ki} \rangle, \mathbf{z}'_{ki} \in U_k \tag{6}$$

where $d_{ki}$ represents the cosine distance between $\mathbf{z}'$ and the representation of the $i$-th sample in the $k$-th category cluster.

3) Highlight the most similar representations in the clusters. This is achieved by softmax operation, which can amplify the differences in the sample representations stored in the clusters. The specific formulas are defined as

$$s_{ki} = \frac{\exp(d_{ki})}{\sum_{m=0}^{C-1}\sum_{i=1}^{N} d_{mi}}, \tag{7}$$

where $N$ denotes the size of each cluster.

4) Measure the overall similarity of the sample to each category by summarizing $s_{ki}$:

$$s_k = \frac{1}{N}\sum_{i=1}^{N} s_{ki}.$$

Then, we represent the similarities between the sample $x$ and the clusters as $\mathbf{S} = [s_0, s_1, \ldots, s_{C-1}]$.

5) Introduce the partial-label a prior. As we have defined the partial label as a vector, we can achieve it by the Hadamard product:

$$\hat{\mathbf{S}} = \mathbf{S} \odot \mathbf{Y}(y), \tag{8}$$

where $\mathbf{Y}(y)$ indicates the partial label for the WSI label $y$. It can be regarded as a mask that indicates which types of tissue are possible in the WSI. Then, the Hadamard product acts as a masking operation to the similarities.

6) Disambiguate to obtain the optimal label, specifically defined as:

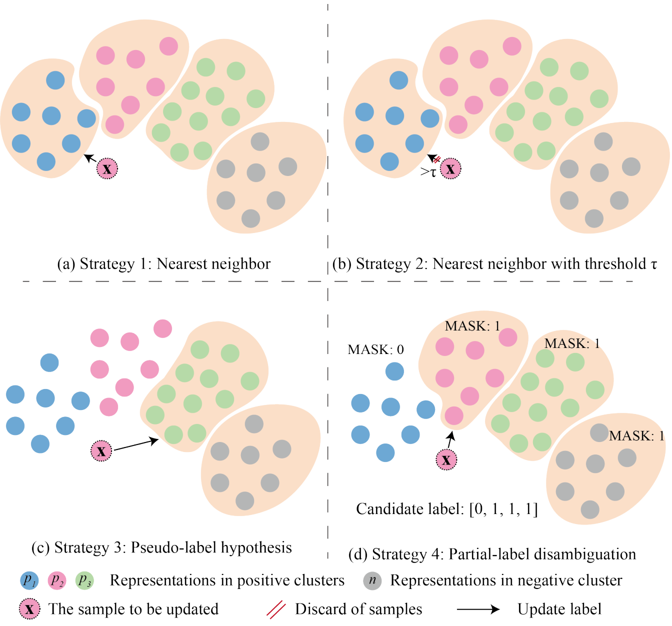
$$\hat{y} = \begin{cases} y & \sum\limits_{k \neq y} \hat{s}_k < \gamma, \\ \arg\max\limits_{k \neq y} \hat{s}_k & otherwise, \end{cases} \tag{9}$$

where $\hat{s}_k$ denotes the $k$-th element of $\hat{\mathbf{S}}$ and $\gamma$ is a threshold for immediately rejecting the patch to be assigned to other classes that are different from the label of its WSI.

Here, we discuss the advantages of the partial label hypothesis to the other non-ideal supervision strategies. Fig. 3 shows the label assignment for a hard positive sample under different strategies, where there are three positive categories $p_1$, $p_2$ and $p_3$ besides the negative category, and $\mathbf{x}$ denotes a patch sampled from a WSI labeled as $p_3$ but its true tissue type belongs to the class $p_2$. Fig. 3(a) illustrates the nearest neighbor strategy. It will assign $\mathbf{x}$ the most similar cluster, which is the most common setting in the semi-supervised learning framework. In this case, $\mathbf{x}$ will be mislabeled as $p_1$. Fig. 3(b) shows the threshold-based strategy where the hard sample will be discarded if the nearest distance is larger than a threshold. Fig. 3(c) is the binary hypothesis, i.e., a patch from a WSI with label $p_3$ is either labeled $p_3$ or negative, which is commonly used in the MIL frameworks. In this strategy, $\mathbf{x}$ will be misclassified as $p_3$. In contrast, the proposed partial-label disambiguation strategy allows the sample to be assigned to one of the pathologically reasonable tissue types, including $p_2$, $p_3$, and negative. As shown in Fig. 3(d), $\mathbf{x}$ will be correctly labeled as $p_2$ based on our disambiguation strategy. Therefore, the proposed method can significantly reduce the label noise in this type of complex weakly supervised learning scenarios and thereby deliver more discriminative representations for the downstream WSI classification tasks.

### D. The category-aware cluster updating

Obviously, the effectiveness of the proposed partial label strategy disambiguation depends on the high quality of the category clusters $\mathbb{U}_k$. It is challenging to build the clusters in the complete absence of patch-level annotations. To ensure $\mathbb{U}_k$ is consistently representative of the corresponding class during the optimization of the network, we propose to update $\mathbb{U}_k$ in each step of training. Specifically, for the $k$-th cluster, a set of samples $\mathbb{U}_k^+$ are recognized from the mini-batch. The representations in $\mathbb{U}_k^+$ are pushed to the cluster queue $\mathbb{U}_k$ and simultaneously pop $|\mathbb{U}_k^+|$ oldest samples from $\mathbb{U}_k$. Let $\mathbb{B}_k = \{\mathbf{x} \mid y = k\}$ be the set of samples in a mini-batch sampled from slides of class $k$. Considering the unbalance of the labels,


x
>τ
x
x
MASK: 0
MASK: 1
MASK: 1
MASK: 1
x
Candidate label: [0, 1, 1, 1]
$p_1$ $p_2$ $p_3$ Representations in positive clusters
$n$ Representations in negative cluster
x The sample to be updated
Discard of samples
Update label

(a) Strategy 1: Nearest neighbor

(b) Strategy 2: Nearest neighbor with threshold τ

(c) Strategy 3: Pseudo-label hypothesis

(d) Strategy 4: Partial-label disambiguation

Fig. 3. The four strategies to update hard positive sample labels, where $\mathbf{x}$ is the sample to be updated which is the patch sampled from the slide of $\boldsymbol{p_3}$ class while belongs to $\boldsymbol{p_2}$ class, (a) signifies updating $\mathbf{x}$ to the most similar cluster, (b) involves introducing a threshold $\boldsymbol{\tau}$ to ensure $\mathbf{x}$ is updated only to the top-k clusters, (c) denotes updating $\mathbf{x}$ to either the sampled WSI category or a negative cluster, and (d) represents the proposed partial-label learning update strategy.

we designed a top $\beta\%$ update strategy for $|\mathbb{U}_k^+|$. The principles are as follows.

- For the cluster corresponding to the negative class, i.e., $k = 0$, $\mathbb{U}_k^+$ samples a portion ($|\mathbb{B}_k| \times \beta\%$) of the sample representations randomly from $\mathbb{B}_k$.
- For the cluster corresponding to the positive class, i.e., $k > 0$, $\mathbb{U}_k^+$ takes the representations that are top $\beta\%$ similar to the cluster from $\mathbb{B}_k$.

The detailed algorithm is summarized in Algorithm 1

By updating in this manner, it is ensured that each cluster is continuously updated during training and maintains the consistency between the representations stored in each category queue and the model's output representations.

### E. Utilizing the trained encoder for WSI analysis tasks

Finally, we followed the MIL paradigm for WSI-level tasks and employed the encoder trained under the PLCC framework as the patch feature extractor. The extracted features are input into WSI classification benchmarks for classification. Our proposed method is a contrastive self-supervised representation learning framework that focuses on the intrinsic features of pathology images. Therefore, it can be applied to various specific downstream tasks and is capable of extracting image representations that are discernible and relevant to specific tasks.

**Algorithm 1:** The pseudo-code for the cluster updating algorithm.

```
Input:
    $\{\mathbb{U}_k\}_{k=0,1,\ldots,C-1}$: The clusters of $C$ categories;
    $\{\mathbb{B}_k\}_{k=0,1,\ldots,C-1}$: The sets of representations in the current
    mini-batch for different categories;
    $\beta\%$: The ratio of random sampling.
Output:
    $\{\hat{\mathbb{U}}_k\}_{k=0,1,\ldots,C-1}$: The updated clusters;
for $k \leftarrow 0$ to $C-1$ do
    if $k = 0$ then
        $\mathbb{U}_k^+ = \text{random}(\mathbb{B}_k, |\mathbb{B}_k| \times \beta\%)$;
        // Randomly select $\beta\%$ of samples from $\mathbb{B}_k$.
        $\mathbb{U}_k.\text{pop}(|\mathbb{U}_k^+|)$;
        // Pop $|\mathbb{U}_k^+|$ oldest samples from $\mathbb{U}_k$.
        $\mathbb{U}_k.\text{push}(\mathbb{U}_k^+)$;
        // Push $\mathbb{U}_k^+$ into $\mathbb{U}_k$ for updating.
    else
        for $\mathbf{z}'_j \in \mathbb{B}_k$ do
            $s_{kj} = \frac{1}{|\mathbb{U}_k|} \sum_{i=1}^{|\mathbb{U}_k|} \frac{\exp(\langle \mathbf{z}'_j, \mathbf{z}'_{ki} \rangle)}{\sum_{k=0}^{C-1} \sum_{i=1}^{|\mathbb{U}_k|} \exp(\langle \mathbf{z}'_j, \mathbf{z}'_{ki} \rangle)}$;
            // Calculate the similarity between the
               sample and clusters.
        end
        $\mathbf{S}_k = \{s_{kj}\}_{j=1,\ldots,|\mathbb{B}_k|}$;
        $index$=sorted(range($|\mathbf{S}_k|$), key=lambda i : $\mathbf{S}_k[i]$, reverse=True);
        // Sort in descending order and return the
           indexes.
        $\mathbb{U}_k^+ = \{\mathbb{B}_k[j] \mid j \in index[: |\mathbb{B}_k| \times \beta\%]\}$;
        // Select top $\beta\%$ similar samples to the
           cluster from $\mathbb{B}_k$.
        $\mathbb{U}_k.\text{pop}(|\mathbb{U}_k^+|)$;
        // Pop $|\mathbb{U}_k^+|$ oldest samples from $\mathbb{U}_k$.
        $\mathbb{U}_k.\text{push}(\mathbb{U}_k^+)$;
        // Push $\mathbb{U}_k^+$ into $\mathbb{U}_k$ for updating.
    end
    $\hat{\mathbb{U}}_k = \mathbb{U}_k$;
end
return $\{\hat{\mathbb{U}}_k\}_{k=0,1,\ldots,C-1}$
```

## IV. Experiments

### A. Experimental Settings

Our experiments were conducted on two in-house WSI datasets and a public dataset collected from the Cancer Genome Atlas (TCGA) program. The details of each dataset are as follows and the data distribution is shown in Table I:

- **USTC-EGFR**[1] contains 754 in-house WSIs from 521 patients of lung adenocarcinoma for epidermal growth factor receptor (EGFR) gene mutation identification, which are categorized into 5 classes including EGFR 19del mutation, EGFR L858R mutation, non-common driver mutations (wild type), other driver gene mutations (other types), and tumor-free tissue (Normal). Notably, we did not deliberately collect normal cases but took negative slides from the paraneoplastic tissue of patients as controls.
- **BRCA-HER2**[2] contains 279 in-house WSIs from 279 patients of human epidermal growth factor receptor-2 (HER2) protein and gene expression in breast cancer patients, which are categorized into 4 subtypes including the IHC score of 1+, the IHC score of 2+, the IHC score of 3+, and the IHC score of 0 (Normal).
- **TCGA-EGFR** is a dataset collected from the TCGA for epidermal growth factor receptor (EGFR) gene mutation

[1] The study was approved by the Medical Research Ethics Committee of the First Affiliated Hospital of the University of Science and Technology of China (Anhui Provincial Hospital) under the protocol No. USTC-2022-RE-454

[2] The study was approved by the Medical Research Ethics Committee of the First Affiliated Hospital of the University of Science and Technology of China (Anhui Provincial Hospital) under the protocol No. 2023KY-378

identification of lung adenocarcinoma contains 785 WSIs from 304 patients, which are categorized into the same 5 subtypes as the USTC-EGFR dataset.

TABLE I
THE DATA DISTRIBUTION OF THE THREE DATASETS.

| **USTC-EGFR** | *Normal* | *19del* | *L858R* | *wild type* | *other types* | Total |
|---|---|---|---|---|---|---|
| Slide | 165 | 118 | 184 | 146 | 141 | 754 |
| Case | – | 108 | 126 | 146 | 141 | 521 |
| **BRCA-HER2** | *HER2 0* | *HER2 1+* | *HER2 2+* | *HER2 3+* | | |
| Slide | 76 | 77 | 95 | 31 | | 279 |
| Case | 76 | 77 | 95 | 31 | | 279 |
| **TCGA-EGFR** | *Normal* | *19del* | *L858R* | *wild type* | *other types* | |
| Slide | 80 | 22 | 20 | 579 | 84 | 785 |
| Case | 74 | 7 | 6 | 190 | 27 | 304 |

TABLE II
CANDIDATE LABEL SETS FOR THE TCGA-EGFR AND USTC-EGFR DATASETS

| candidate label<br>category | Normal | wild type | 19del | L858R | other types |
|---|---|---|---|---|---|
| Normal | 1 | 0 | 0 | 0 | 0 |
| wild type | 1 | 1 | 0 | 0 | 0 |
| 19del | 1 | 1 | 1 | 0 | 0 |
| L858R | 1 | 1 | 0 | 1 | 0 |
| other types | 1 | 1 | 0 | 0 | 1 |

TABLE III
CANDIDATE LABEL SETS FOR THE BRCA-HER2 DATASET

| candidate label<br>category | 0 | 1+ | 2+ | 3+ |
|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 |
| 1+ | 1 | 1 | 0 | 0 |
| 2+ | 1 | 1 | 1 | 0 |
| 3+ | 1 | 1 | 1 | 1 |

The two in-house datasets were captured using a Motic Easyscan Pro 6 under 20X magnification with 0.52 µm/pixel resolution. We used the genetic testing results of the patients as the true label for EGFR mutation status and leveraged the pathologist's diagnosis of the IHC staining results as the HER2 grading truth label. Each of the above three datasets was randomly split into training, validating, and testing sets following the ratio of 6:1:3 at the patient level in our experiments. We cropped these WSIs into non-overlapping image patches in size of $256 \times 256$ pixels as the input of the model. In partial-label learning, each sample requires the allocation of a candidate label set. The candidate label sets for the two datasets are configured as shown in Tables II and III based on prior knowledge provided by pathologists, where each row represents the candidate label set for a specific class of slides. The value 0 indicates that the category is not part of the candidate label set, and the value 1 denotes that the corresponding category is a candidate label. For histopathology slides, the tissue categories present in slides of the same class are fixed. Therefore, samples derived from slides of the same class should share the same candidate label set.

We trained the ResNet50 [48] as the encoder with a batch size of 128 under different methods involved in our experiments. For the self-supervised methods, we train each method by 100 epochs on the training set. For other state-of-the-art weakly-supervised learning methods, we train the model using the early-stop strategy.

All the experiments were implemented in Python with PyTorch and run on a computer with an Intel Xeon Gold 6126 CPU of 2.60GHz and 4 GPUs of Nvidia Geforce 3090.

### *B. Ablation Studies*

We first verified the effectiveness of the various improvements made in PLCC. The ablation studies in this section are evaluated based on the performance of whole slide classification tasks on the UTSC-EGFR dataset. The degraded models in the ablation experiments are detailed as follows:

- **PLCC-w/o-Specificity:** In PLCC, the similarity measurement is achieved by comparing the current sample with every sample in the clusters, and a softmax function is used to amplify the specificity between samples. A prototype measurement method is now adopted as a substitute, involving the calculation of similarity between the current sample and the cluster centers of each category, as described in Equation 10 which calculates the similarity between the $i$-th sample and the $k$-th cluster. Thus, PLCC-w/o-Specificity is a version of PLCC that employs the prototype measurement approach.

$$s_{ik} = \frac{\exp(\langle \mathbf{z}'_i, \bar{\mathbf{z}}'_k \rangle)}{\sum_{j=1}^{C} \exp(\mathbf{z}'_i, \bar{\mathbf{z}}'_j)} \tag{10}$$

- **PLCC-w/o-PLL:** It is a variant of PLCC that uses the binary label disambiguation strategy, where a pseudo-label for one positive tissue cannot be updated to other positive categories but can only be updated to negative labels, i.e., $y = 0$ in our setting, as stated in Equation 11.

$$\hat{y} = \begin{cases} 0 & y = 0 \text{ or } s_0 > T_{\text{neg}} \\ y & otherwise \end{cases}, \tag{11}$$

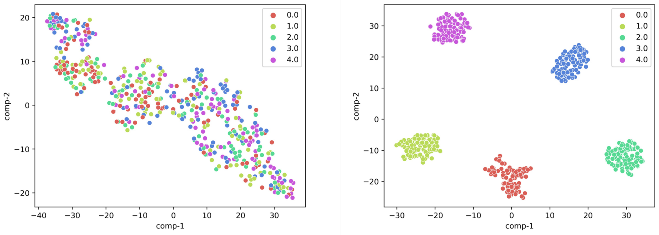
  where the threshold $T_{\text{neg}}$ is established to identify false positive samples.
- **PLCC-w/o-Proportion:** It adopts a different sample updating strategy. Specifically, for the $k$-th cluster, a group of samples $\mathbb{U}_k^+$ is selected based on fixed thresholds from the mini-batch and updated into the cluster $\mathbb{U}_k$, while the earliest $|\mathbb{U}_k^+|$ samples in $\mathbb{U}_k$ are dequeued. In this variance, $\mathbb{U}_k^+$ is selected based on the following strategy:

$$\mathbb{U}_k^+ = \begin{cases} \{\mathbf{z}'_i \mid s_{k0} > T_{\text{neg}}\} & k = 0 \\ \{\mathbf{z}'_i \mid s_{k0} < (1 - T_{\text{pos}})\} & otherwise \end{cases} \tag{12}$$

The ablation experiments results are shown in the Table IV. The results show that the three modifications in PLCC play crucial roles. In PLCC-w/o-Specificity, the loss of perception of differences between samples led to decreases in the AUC metric by 0.019 and 0.013 under the CLAM framework and TransMIL, respectively. The removal of partial-label learning in PLCC-w/o-PLL resulted in a more pronounced decline compared to PLCC-w/o-Specificity, with AUC reductions of 0.034 and 0.015 under CLAM and TransMIL, respectively,

TABLE IV
ABLATION EXPERIMENTS OF THE PLCC METHOD ON THE USTC-EGFR DATASET

| Methods | CLAM | | | TransMIL | | |
|---|---|---|---|---|---|---|
| | ACC | AUC | F1-Score | ACC | AUC | F1-Score |
| PLCC-w/o-Specificity | 0.655 | 0.914 | 0.659 | 0.713 | 0.937 | 0.713 |
| PLCC-w/o-PLL | 0.565 | 0.899 | 0.538 | 0.709 | 0.935 | 0.707 |
| PLCC-w/o-Proportion | 0.336 | 0.664 | 0.236 | 0.251 | 0.631 | 0.130 |
| PLCC | 0.717 | 0.933 | 0.710 | 0.740 | 0.950 | 0.738 |



(a) PLCC-w/o-Proportion (b) PLCC

Fig. 4. T-SNE feature visualization of the five categories from USTC-EGFR dataset, where (a) shows the distribution of feature learned by PLCC-w/o-Proportion and (b) shows the distribution of feature learned by PLCC.

and significant decreases in ACC and F1-Score. This substantiates the effectiveness of partial-label learning in resolving subtype classification tasks. Notably, the reduction in metrics under the CLAM framework was more significant than that under TransMIL, potentially because the CLAM framework focuses more on the representation of local individual instances. Hence, partial-label learning enhances slide classification performance in the CLAM framework, which distinguishes different subtypes of tissues in positive slides. PLCC-w/o-Proportion, which selects update samples through thresholding rather than proportion, was observed to lead to an early influx of noise samples into the cluster queue during training, causing feature collapse and severely impairing whole slide classification performance in downstream tasks.

Furthermore, features of 50 randomly sampled samples from each category's cluster were visualized using T-SNE for both PLCC-w/o-Proportion and PLCC, as shown in Fig. 4. In the visualization of PLCC-w/o-Proportion, features from different clusters were intermingled without distinction, and semantic category information was not correctly embedded into the image features. In contrast, in PLCC's feature visualization, features from each category's storage queue were well clustered into a distinct group. It demonstrates the effectiveness of the proposed method.

### C. Comparison with other representation methods

We compared performance with 9 different representation learning methods on the task of whole slide classification to validate the advanced nature of the proposed method in the field of pathology image feature learning.

Table V displays the results of the whole slide classification under the TransMIL [38] framework. PLCC achieved the optimal performance in ACC, AUC, and F1 score across three gene mutation prediction datasets. For instance, in terms of AUC, PLCC achieved 0.950, 0.853, and 0.919 on USTC-EGFR, BRCA-HER2, and TCGA-EGFR, respectively, which represents increases of 5.3%, 5.8%, and 1.0% over the second-best results. Table VI shows the results of the whole slide classification within the CLAM [12] framework. PLCC also attained the highest metrics across the three tasks. Taking the F1 score as an example, PLCC achieved 0.710, 0.622, and 0.815 on USTC-EGFR, BRCA-HER2, and TCGA-EGFR, respectively, which makes improvements of 9.6%, 3.7%, and 0.4% over the second-best results.

Pseudo-label utilizes pseudo-labels as supervision for training ResNet50 [48], which introduces a significant amount of label noise. The multiple instance learning approach proposed by Lerousseau et al. [13] is notably sensitive to the pre-trained representations and hyperparameter configurations, and it is challenged by the task of subtype classification. MoCo v2 [49], a method dependent on negative sample contrast, is prone to be affected by class imbalance and the limitations of memory bank capacity. PiCO [43] is a partial-label learning method developed for tasks involving the analysis of natural scene images.

BYOL [15] and SimTriplet [32] do not rely on negative sample contrast, but in the context of gene mutation tasks, a single slide can include multiple tissues with various mutations. These tissues often intermingle within the slide and complicate their differentiation. For instance, in the BRCA-HER2 dataset, the 2+ category contains the tissue characteristics of three other categories. SimTriplet [32] can inadvertently introduce false positives with semantic ambiguity, which considered adjacent tissues as positive samples and degraded the representation quality. cTtransPath [52] and RetCCL [51] are contrastive learning pre-trained frameworks specific to histopathology images. cTransPath is a Transformer-based unsupervised representation learning framework pre-trained on the public TCGA and PAIP datasets. RetCCL is a contrastive learning framework with a weighted InfoNCE loss and a group-level InfoNCE loss, which is also pre-trained on TCGA and PAIP datasets. We employed the released pre-trained weights of the two models as the patch feature extractors. Before inference, the two models were exposed to a substantial amount of publicly available data, which proved to be an effective strategy for the TCGA-EGFR dataset. However, it is notable that there are overlaps between the pre-training datasets of RetCCL and cTransPath, which include TCGA Lung data, and the TCGA-EGFR test set used in our experiments. Therefore, the results for RetCCL and cTransPath on the TCGA-EGFR are suffering from data-leakage. Conversely, these approaches did not yield the desired results when applied to the two in-house datasets, where the models failed to identify discriminative representations.

PLIP [50] leveraged text as a form of supervision that is finer-grained and informationally denser, which pre-trained the [53] on a substantial volume of image-text pairs. Consequently, PLIP [50] demonstrates objectively robust performance across the three datasets and surpasses most of the other methods. However, it still exhibits a certain degree of disparity when compared to PLCC.

The proposed PLCC learns the true labels corresponding to different tissues through the use of candidate label sets,

TABLE V

COMPARISON OF METHODS ON TCGA-EGFR, BRCA-HER2, AND USTC-EGFR DATASETS WITH TRANSMIL [38], WHERE THE BEST METRICS ARE DENOTED IN BOLD AND THE SECOND-BEST RESULTS ARE UNDERLINED.

| Methods | USTC-EGFR | | | BRCA-HER2 | | | TCGA-EGFR | | |
|---|---|---|---|---|---|---|---|---|---|
| | **ACC** | **AUC** | **F1 Score** | **ACC** | **AUC** | **F1 Score** | **ACC** | **AUC** | **F1 Score** |
| Pseudo-label [48] | 0.466 | 0.802 | 0.421 | 0.365 | 0.742 | 0.378 | 0.722 | 0.644 | 0.718 |
| Lerousseau et al. [13] | 0.368 | 0.660 | 0.326 | 0.428 | 0.776 | 0.427 | 0.729 | 0.652 | 0.721 |
| MoCo v2 [49] | 0.514 | 0.816 | 0.492 | 0.414 | 0.758 | 0.451 | 0.735 | 0.655 | 0.741 |
| BYOL [15] | 0.534 | 0.826 | 0.517 | 0.451 | 0.764 | 0.402 | 0.751 | 0.903 | 0.745 |
| SimTriplet [32] | 0.529 | 0.841 | 0.510 | 0.439 | 0.780 | 0.463 | 0.753 | 0.896 | 0.753 |
| PiCO [43] | 0.587 | 0.876 | 0.578 | 0.464 | 0.780 | 0.451 | 0.731 | 0.902 | 0.758 |
| PLIP [50] | 0.610 | 0.897 | 0.601 | 0.479 | 0.795 | 0.488 | 0.736 | 0.905 | 0.731 |
| RetCCL [51] | 0.585 | 0.850 | 0.579 | 0.452 | 0.778 | 0.499 | 0.741$^*$ | 0.906$^*$ | 0.744$^*$ |
| cTransPath [52] | 0.556 | 0.836 | 0.550 | 0.512 | 0.782 | 0.505 | 0.757$^*$ | 0.909$^*$ | 0.753$^*$ |
| PLCC | **0.740** | **0.950** | **0.738** | **0.749** | **0.853** | **0.585** | **0.766** | **0.919** | **0.762** |

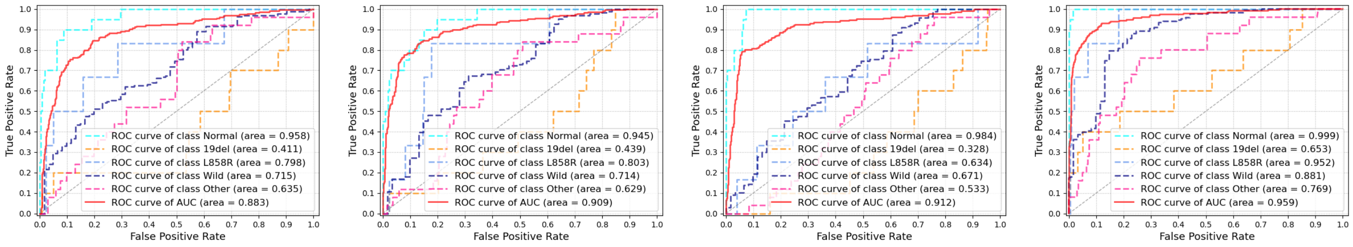
$^*$ The data may appear inflated due to the overlap between the pre-training datasets of RetCCL and cTransPath, which include TCGA Lung data, and the TCGA-EGFR test set used in these experiments.

TABLE VI

COMPARISON OF METHODS ON TCGA-EGFR, BRCA-HER2, AND USTC-EGFR DATASETS WITH CLAM [12], WHERE THE BEST METRICS ARE DENOTED IN BOLD AND THE SECOND-BEST RESULTS ARE UNDERLINED.

| Methods | USTC-EGFR | | | BRCA-HER2 | | | TCGA-EGFR | | |
|---|---|---|---|---|---|---|---|---|---|
| | **ACC** | **AUC** | **F1 Score** | **ACC** | **AUC** | **F1 Score** | **ACC** | **AUC** | **F1 Score** |
| Pseudo-label [48] | 0.426 | 0.776 | 0.403 | 0.414 | 0.753 | 0.415 | 0.726 | 0.696 | 0.727 |
| Lerousseau et al. [13] | 0.390 | 0.710 | 0.339 | 0.439 | 0.759 | 0.428 | 0.732 | 0.701 | 0.731 |
| MoCo v2 [49] | 0.448 | 0.790 | 0.377 | 0.426 | 0.756 | 0.390 | 0.749 | 0.733 | 0.758 |
| BYOL [15] | 0.462 | 0.803 | 0.447 | 0.441 | 0.765 | 0.439 | 0.735 | 0.883 | 0.736 |
| SimTriplet [32] | 0.511 | 0.819 | 0.464 | 0.472 | 0.798 | 0.476 | 0.810 | 0.930 | 0.811 |
| PiCO [43] | 0.529 | 0.866 | 0.493 | 0.465 | 0.784 | 0.463 | 0.765 | 0.912 | 0.762 |
| PLIP [50] | 0.621 | 0.894 | 0.614 | 0.588 | 0.834 | 0.585 | 0.748 | 0.909 | 0.749 |
| RetCCL [51] | 0.551 | 0.828 | 0.552 | 0.536 | 0.804 | 0.533 | 0.788$^*$ | 0.948$^*$ | 0.789$^*$ |
| cTransPath [52] | 0.556 | 0.842 | 0.557 | 0.567 | 0.843 | 0.566 | 0.792$^*$ | 0.950$^*$ | 0.791$^*$ |
| PLCC | **0.717** | **0.933** | **0.710** | **0.634** | **0.882** | **0.622** | **0.819** | **0.959** | **0.815** |

$^*$ The data may appear inflated due to the overlap between the pre-training datasets of RetCCL and cTransPath, which include TCGA Lung data, and the TCGA-EGFR test set used in these experiments.



(a) The result of BYOL (b) The result of PLIP (c) The result of PiCO (d) The result of PLCC

Fig. 5. The ROC curves of classification results on the TCGA-EGFR dataset with CLAM framework.

which can embed accurate semantic information. Coupled with a dynamic queue updating strategy, it further enhances the distinction between representations of different tissue types.

The TCGA-EGFR dataset exhibits a significant data imbalance issue, which comprises 80 WSIs for normal, 22 WSIs for 19del, 20 WSIs for L858R, 579 WSIs for wild type, and 84 WSIs for other types, with the wild type slides constituting over 70% of the data. Fig. 5 presents the ROC curves for various categories in the classification of the TCGA-EGFR dataset within the CLAM [12] framework. BYOL [15] and PLIP [50] exhibit suboptimal performance, primarily due to their diminished ability to discriminate among categories with fewer samples, which results in lower overall metrics as shown in Fig. 5(a) and (b). The improvement in metrics for PiCO [43] is reliant on the identification of a subset of the normal category, yet they still fail to discern the differences in the remaining concentrated subtypes as shown in Fig. 5(c). Fig. 5(d) demonstrates that the proposed PLCC method significantly enhances the ability to differentiate tissue characteristics across categories and yields satisfying classification results even for the sparsely represented categories.

### *D. Visualization*

To further investigate the enhancements brought by the PLCC method, we visualized the top 20% of tissue regions that are of most interest when performing biomarkers prediction tasks under the CLAM framework. The visualization results are shown in Fig.6 and Fig.7.

In the USTC-EGFR and BRCA-HER2 test set, a random slide from each category was selected for visualization. Each

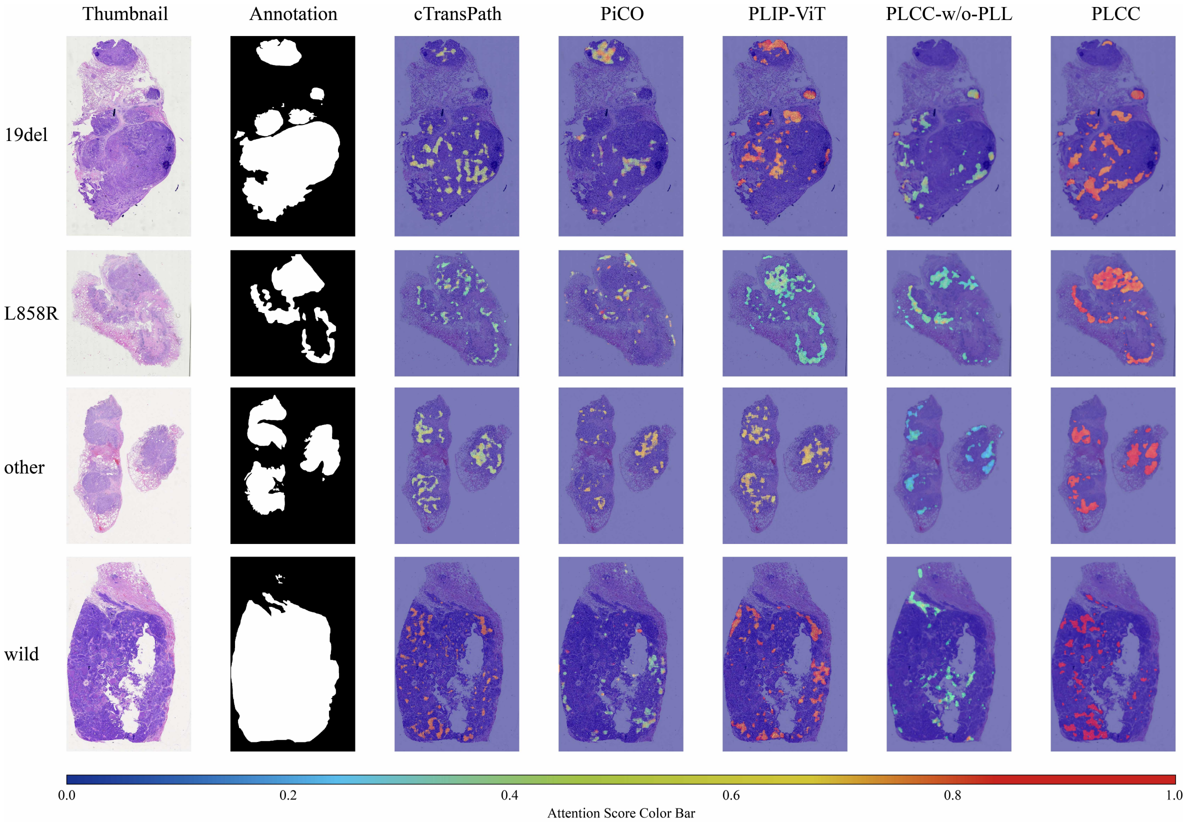


Fig. 6. Heatmap of attention areas and degrees on different category slides in the USTC-EGFR test set.

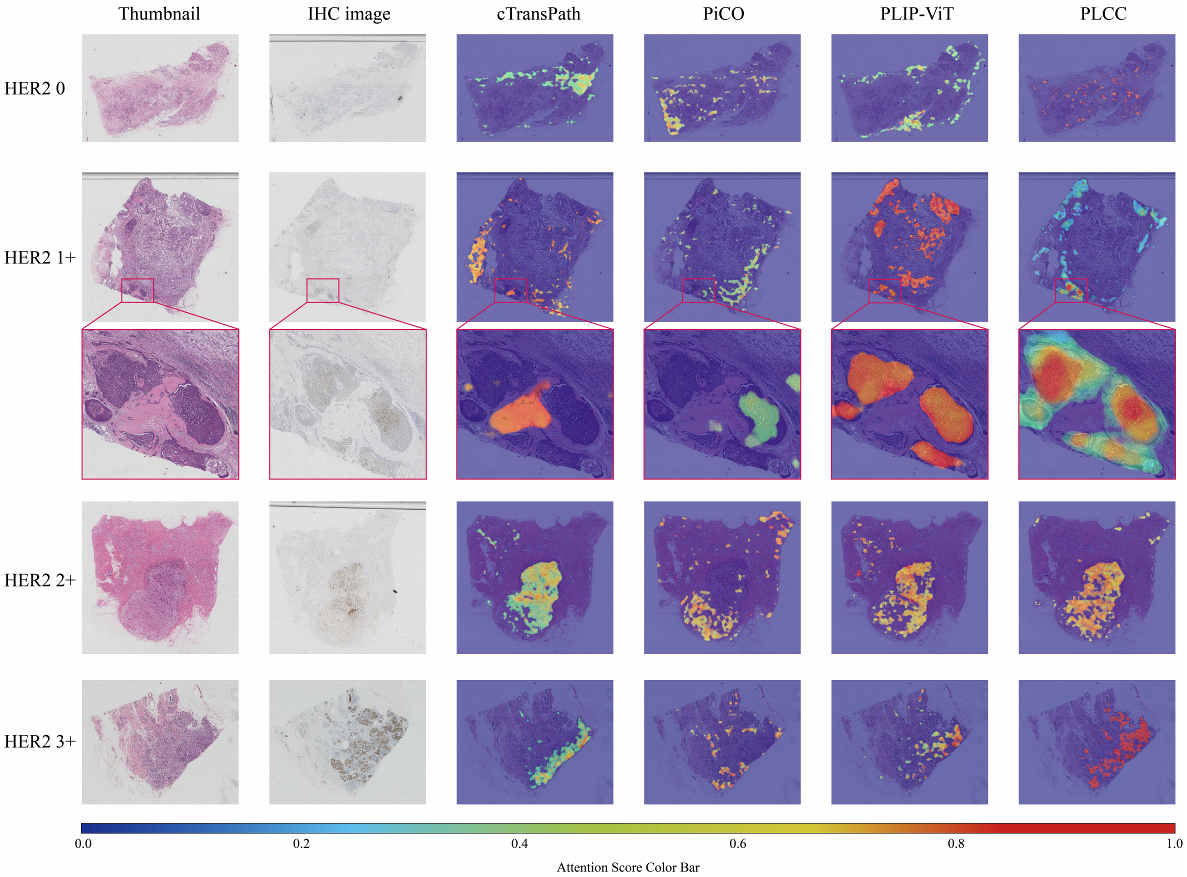


Fig. 7. Heatmap of attention areas and degrees on different category slides in the BRCA-HER2 test set.

column represents different views of the chosen slide, where the first column shows the thumbnails of the slide, and the second column depicts the ground truth of the cancerous region for comparison. The remaining column illustrates the top 20% tissue regions of interest focused by each method.

The attention scores of the regions are represented in varying colors, with blue indicating a low attention score (lower focus by the model) and red indicating a high attention score (higher focus by the model). For a whole slide sample $S = \{\mathbf{x}_i\}_{i=1}^n$ containing $n$ image patches, the attention score $a_i$ for $\mathbf{x}_i$ in a whole slide classification task is determined by the following formulas:

$$\mathbf{A} = \text{Attention}(\{\mathbf{x}_i\}_{i=1}^n) \tag{13}$$

$$a_i = \frac{\exp(a_i)}{\sum_{a \in \mathbf{A}} \exp(a)} \tag{14}$$

where Attention$(\cdot)$ represents the Attention module in the CLAM framework, and $\mathbf{A} = \{a_1, a_2, \ldots, a_n\}$ is the output of the Attention module, with $\mathbf{A} \in \mathbb{R}^n$. Each element in $\mathbf{A}$ represents the attention score of different image patches, indicating the model's focus on different patches.

In Fig. 6, the second column depicts pixel-level tumor region annotations provided by pathologists where white areas represent tumor tissues and the column of PLCC-w/o-PLL illustrates the tissue regions of interest focused by the PLCC-w/o-PLL method using the prototype measurement method. Compared with the PLCC-w/o-PLL method, PLCC shows a higher degree of attention to the top 20% of tissue regions. The column of the PLCC-w/o-PLL method predominantly displays yellow-green blocks, while PLCC shows redder blocks, which indicates a significantly higher focus on the top 20% regions. Additionally, removing PLL results in some normal tissue regions being focused on in the L858R and wild type categories, whereas in the PLCC method, the focused regions are all within the tumor areas annotated by pathologists. PLIP is a pre-trained model based on a substantial quantity of histopathology data. While it is capable of identifying tumor areas, the attention score is typically low and does not highlight crucial regions. The responses of PiCO and cTransPath to the tumor area are found to be less effective.

In Fig. 7, the second column shows the IHC staining images of each HER2 status where the more dark brown areas accumulated represent the higher mutation level. Compared with other methods, PLCC has a lower and balanced response to the negative HER2 0 slide, and a more prominent response and comprehensive coverage of the significant areas of positive slides. Taking HER2 1+ in the second row as an example, the IHC image demonstrates small and concentrated positive regions, and PLCC can accurately capture related regions while reducing attention to other pairs of negative regions. These visualization results further validate the superiority of the proposed PLCC method.

### E. Discussion

It should be noted that the term ”partial-label” is also used in image segmentation. This typically refers to images with annotations that cover only a portion of anatomical structures or image patches, leading to spatially partial or sparse segmentation labels. This concept is generally applied in segmentation tasks involving multi-source datasets [54], [55]. In this paper, ”partial-label” refers to the definition of the category for the same object, focusing on defining the candidate label set and eliminating ambiguous labels [43], [44].

PLIP [50] is a multi-modal model pre-trained on a substantial-volume histopathology dataset containing 208,414 image-text pairs named as OpenPath. PLIP is capable of capturing a more comprehensive and detailed pathological tissue structure distribution, with the guidance of cleaned and refined text information to the image content. However, the OpenPath dataset mainly describes histological features of pathological images and lacks fine-grained biomarker expression information resulting in suboptimal performance of PLIP relative to PLCC.

## V. Conclusion

We proposed the partial-label contrastive learning framework to eliminate label noise in fine-grained histopathology image analysis for biomarker prediction. The partial-label disambiguation with PLCC achieved more discriminative representations from complex tissue subtypes. The proposed dynamic clustering algorithm continuously mines the most representative features of each category to the clustering queue and enhances the performance of partial-label disambiguation. The superiority of this method is validated through extensive experiments that demonstrate its superior performance compared with seven methods on three gene mutation datasets. The effectiveness of PLCC is particularly notable in addressing the challenges of label noise and complex tissue subtypes, which are prevalent in biomarker predictions from pathology images. The visualization results, especially the attention heatmaps, underscore PLCC's precision in identifying relevant tissue regions and reinforce its potential to improve pathology image analysis.

## References

[1] B. S. Chhikara and K. Parang, “Global cancer statistics 2022: the trends projection analysis,” *Chemical Biology Letters*, vol. 10, no. 1, pp. 451–451, 2023, doi:10.3322/caac.21708. [Online]. Available: https://pubs.thesciencein.org/journal/index.php/cbl/article/view/451/293

[2] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, “Cancer statistics, 2022,” *CA: a cancer journal for clinicians*, vol. 72, no. 1, pp. 7–33, 2022, doi:10.3322/caac.21708.

[3] V. Kumar, A. K. Abbas, N. Fausto, and J. C. Aster, *Robbins and Cotran pathologic basis of disease, professional edition e-book*. Elsevier health sciences, 2014.

[4] D. Jin, S. Liang, A. Shmatko, A. Arnold, D. Horst, T. G. Grünewald, M. Gerstung, and X. Bai, “Teacher-student collaborated multiple instance learning for pan-cancer pdl1 expression prediction from histopathology slides,” *Nature Communications*, vol. 15, no. 1, p. 3063, 2024.

[5] S. Volinsky-Fremond, N. Horeweg, S. Andani, J. Barkey Wolf, M. W. Lafarge, C. D. de Kroon, G. Ørtoft, E. Høgdall, J. Dijkstra, J. J. Jobsen *et al.*, “Prediction of recurrence risk in endometrial cancer with multimodal deep learning,” *Nature Medicine*, pp. 1–12, 2024.

[6] B. Acs, M. Rantalainen, and J. Hartman, “Artificial intelligence as the next step towards precision pathology,” *Journal of internal medicine*, vol. 288, no. 1, pp. 62–81, 2020, doi:10.1111/joim.13030.

[7] A. Echle, N. G. Laleh, P. L. Schrammen, N. P. West, C. Trautwein, T. J. Brinker, S. B. Gruber, R. D. Buelow, P. Boor, H. I. Grabsch *et al.*, "Deep learning for the detection of microsatellite instability from histology images in colorectal cancer: a systematic literature review," *ImmunoInformatics*, vol. 3, p. 100008, 2021, doi:10.1016/j.immuno.2021.100008.

[8] M. G. Hanna, O. Ardon, V. E. Reuter, S. J. Sirintrapun, C. England, D. S. Klimstra, and M. R. Hameed, "Integrating digital pathology into clinical practice," *Modern Pathology*, vol. 35, no. 2, pp. 152–164, 2022, doi:10.1038/s41379-021-00929-0.

[9] J. Yang, H. Chen, Y. Zhao, F. Yang, Y. Zhang, L. He, and J. Yao, "Remix: A general and efficient framework for multiple instance learning based whole slide image classification," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II*. Springer, 2022, pp. 35–45, doi:10.1007/978-3-031-16434-7_4.

[10] Y. Zhao, Z. Lin, K. Sun, Y. Zhang, J. Huang, L. Wang, and J. Yao, "Setmil: spatial encoding transformer-based multiple instance learning for pathological image analysis," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II*. Springer, 2022, pp. 66–76, doi:10.1007/978-3-031-16434-7_7.

[11] Y. Zheng, J. Li, J. Shi, F. Xie, and Z. Jiang, "Kernel attention transformer (kat) for histopathology whole slide image classification," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II*. Springer, 2022, pp. 283–292, doi:10.1007/978-3-031-16434-7_28.

[12] M. Y. Lu, D. F. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri, and F. Mahmood, "Data-efficient and weakly supervised computational pathology on whole-slide images," *Nature biomedical engineering*, vol. 5, no. 6, pp. 555–570, 2021, doi:10.1038/s41551-020-00682-w.

[13] M. Lerousseau, M. Vakalopoulou, M. Classe, J. Adam, E. Battistella, A. Carré, T. Estienne, T. Henry, E. Deutsch, and N. Paragios, "Weakly supervised multiple instance learning histopathological tumor segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23*. Springer, 2020, pp. 470–479, doi:10.1007/978-3-030-59722-1_45.

[14] X. Chen and K. He, "Exploring simple siamese representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 15 750–15 758. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2021/html/Chen_Exploring_Simple_Siamese_Representation_Learning_CVPR_2021_paper.html

[15] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar *et al.*, "Bootstrap your own latent-a new approach to self-supervised learning," *Advances in neural information processing systems*, vol. 33, pp. 21 271–21 284, 2020. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/f3ada80d5c4ee70142b17b8192b2958e-Paper.pdf

[16] C. Abbet, I. Zlobec, B. Bozorgtabar, and J.-P. Thiran, "Divide-and-rule: self-supervised learning for survival analysis in colorectal cancer," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23*. Springer, 2020, pp. 480–489, doi:10.1007/978-3-030-59722-1_46.

[17] N. A. Koohbanani, B. Unnikrishnan, S. A. Khurram, P. Krishnaswamy, and N. Rajpoot, "Self-path: Self-supervision for classification of pathology images with limited annotations," *IEEE Transactions on Medical Imaging*, vol. 40, no. 10, pp. 2845–2856, 2021, doi:10.1109/TMI.2021.3056023.

[18] J. J. Pao, M. Biggs, D. Duncan, D. I. Lin, R. Davis, R. S. Huang, D. Ferguson, T. Janovitz, M. C. Hiemenz, N. R. Eddy *et al.*, "Predicting egfr mutational status from pathology images using a real-world dataset," *Scientific reports*, vol. 13, no. 1, p. 4404, 2023.

[19] S. Farahmand, A. I. Fernandez, F. S. Ahmed, D. L. Rimm, J. H. Chuang, E. Reisenbichler, and K. Zarringhalam, "Deep learning trained on hematoxylin and eosin tumor region of interest predicts her2 status and trastuzumab treatment response in her2+ breast cancer," *Modern Pathology*, vol. 35, no. 1, pp. 44–51, 2022.

[20] P. L. Schrammen, N. Ghaffari Laleh, A. Echle, D. Truhn, V. Schulz, T. J. Brinker, H. Brenner, J. Chang-Claude, E. Alwers, A. Brobeil *et al.*, "Weakly supervised annotation-free cancer detection and prediction of genotype in routine histopathology," *The Journal of pathology*, vol. 256, no. 1, pp. 50–60, 2022, doi:10.1002/path.5800.

[21] Q. Wang, J. Xu, A. Wang, Y. Chen, T. Wang, D. Chen, J. Zhang, and T. B. Brismar, "Systematic review of machine learning-based radiomics approach for predicting microsatellite instability status in colorectal cancer," *La radiologia medica*, vol. 128, no. 2, pp. 136–148, 2023, doi:10.1007/s11547-023-01593-x.

[22] A. Echle, H. I. Grabsch, P. Quirke, P. A. van den Brandt, N. P. West, G. G. Hutchins, L. R. Heij, X. Tan, S. D. Richman, J. Krause *et al.*, "Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning," *Gastroenterology*, vol. 159, no. 4, pp. 1406–1416, 2020, doi:10.1053/j.gastro.2020.06.021.

[23] S. H. Lee, I. H. Song, and H.-J. Jang, "Feasibility of deep learning-based fully automated classification of microsatellite instability in tissue slides of colorectal cancer," *International Journal of Cancer*, vol. 149, no. 3, pp. 728–740, 2021, doi:10.1002/ijc.33599.

[24] J. M. Niehues, P. Quirke, N. P. West, H. I. Grabsch, M. van Treeck, Y. Schirris, G. P. Veldhuizen, G. G. Hutchins, S. D. Richman, S. Foersch *et al.*, "Generalizable biomarker prediction from cancer pathology slides with self-supervised deep learning: A retrospective multi-centric study," *Cell Reports Medicine*, vol. 4, no. 4, 2023, doi:10.1016/j.xcrm.2023.100980.

[25] G. Shamai, A. Livne, A. Polónia, E. Sabo, A. Cretu, G. Bar-Sela, and R. Kimmel, "Deep learning-based image analysis predicts pd-l1 status from h&e-stained histopathology images in breast cancer," *Nature Communications*, vol. 13, no. 1, p. 6753, 2022, doi:10.1038/s41467-022-34275-9.

[26] M. S. Jain and T. F. Massoud, "Predicting tumour mutational burden from histopathological images using multiscale deep learning," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 356–362, 2020, doi:10.1038/s42256-020-0190-5.

[27] H. Xu, S. Park, J. R. Clemenceau, N. Radakovich, S. H. Lee, and T. H. Hwang, "Deep transfer learning approach to predict tumor mutation burden (tmb) and delineate spatial heterogeneity of tmb within tumors from whole slide images," *Cold Spring Harbor Lab*, vol. 1, p. 554527, 2020, doi:10.1101/554527.

[28] A. Sadhwani, H.-W. Chang, A. Behrooz, T. Brown, I. Auvigne-Flament, H. Patel, R. Findlater, V. Velez, F. Tan, K. Tekiela *et al.*, "Comparative analysis of machine learning approaches to classify tumor mutation burden in lung adenocarcinoma using histopathology images," *Scientific reports*, vol. 11, no. 1, p. 16605, 2021, doi:10.1038/s41598-021-95747-4.

[29] Z. Huang, H. Chai, R. Wang, H. Wang, Y. Yang, and H. Wu, "Integration of patch features through self-supervised learning and transformer for survival analysis on whole slide images," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*. Springer, 2021, pp. 561–570, doi:10.1007/978-3-030-87237-3_54.

[30] X. Wang, S. Yang, J. Zhang, M. Wang, J. Zhang, J. Huang, W. Yang, and X. Han, "Transpath: Transformer-based self-supervised learning for histopathological image classification," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*. Springer, 2021, pp. 186–195, doi:10.1007/978-3-030-87237-3_18.

[31] O. Ciga, T. Xu, and A. L. Martel, "Self supervised contrastive learning for digital histopathology," *Machine Learning with Applications*, vol. 7, p. 100198, 2022, doi:10.1016/j.mlwa.2021.100198.

[32] Q. Liu, P. C. Louis, Y. Lu, A. Jha, M. Zhao, R. Deng, T. Yao, J. T. Roland, H. Yang, S. Zhao *et al.*, "Simtriplet: Simple triplet representation learning with a single gpu," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*. Springer, 2021, pp. 102–112, doi:10.1007/978-3-030-87196-3_10.

[33] P. Yang, Z. Hong, X. Yin, C. Zhu, and R. Jiang, "Self-supervised visual representation learning for histopathological images," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*. Springer, 2021, pp. 47–57, doi:10.1007/978-3-030-87196-3_5.

[34] G. Campanella, M. G. Hanna, L. Geneslaw, A. Miraflor, V. Werneck Krauss Silva, K. J. Busam, E. Brogi, V. E. Reuter, D. S. Klimstra, and T. J. Fuchs, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature medicine*, vol. 25, no. 8, pp. 1301–1309, 2019, doi:10.1038/s41591-019-0508-1.

[35] S. Kalra, M. Adnan, S. Hemati, T. Dehkharghanian, S. Rahnamayan, and H. R. Tizhoosh, "Pay attention with focus: A novel learning scheme for classification of whole slide images," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October*

*1, 2021, Proceedings, Part VIII 24*. Springer, 2021, pp. 350–359, doi:10.1007/978-3-030-87237-3_34.

[36] D. Reisenbüchler, S. J. Wagner, M. Boxberg, and T. Peng, "Local attention graph-based transformer for multi-target genetic alteration prediction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 377–386, doi:10.1007/978-3-031-16434-7_37.

[37] S. Ding, J. Wang, J. Li, and J. Shi, "Multi-scale prototypical transformer for whole slide image classification," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2023, pp. 602–611, doi:10.1007/978-3-031-43987-2_58.

[38] Z. Shao, H. Bian, Y. Chen, Y. Wang, J. Zhang, X. Ji *et al.*, "Transmil: Transformer based correlated multiple instance learning for whole slide image classification," *Advances in neural information processing systems*, vol. 34, pp. 2136–2147, 2021. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/file/10c272d06794d3e5785d5e7c5356e9ff-Paper.pdf

[39] Y. Zheng, J. Li, J. Shi, F. Xie, J. Huai, M. Cao, and Z. Jiang, "Kernel attention transformer for histopathology whole slide image analysis and assistant cancer diagnosis," *IEEE Transactions on Medical Imaging*, vol. 42, no. 9, pp. 2726–2739, 2023.

[40] K. Wu, Y. Zheng, J. Shi, F. Xie, and Z. Jiang, "Position-aware masked autoencoder for histopathology wsi representation learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 714–724, doi:10.1007/978-3-031-43987-2_69.

[41] H. Zhang, Y. Meng, Y. Zhao, Y. Qiao, X. Yang, S. E. Coupland, and Y. Zheng, "Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18 802–18 812. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2022/html/Zhang_DTFD-MIL_Double-Tier_Feature_Distillation_Multiple_Instance_Learning_for_Histopathology_Whole_CVPR_2022_paper.html

[42] J. Fan, Y. Yu, Z. Wang, and J. Gu, "Partial label learning based on disambiguation correction net with graph representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 8, pp. 4953–4967, 2021, doi:10.1109/TCSVT.2021.3139968.

[43] H. Wang, R. Xiao, Y. Li, L. Feng, G. Niu, G. Chen, and J. Zhao, "Pico: Contrastive label disambiguation for partial label learning," in *International Conference on Learning Representations*, 2022. [Online]. Available: https://openreview.net/forum?id=EhYjZy6e1gJ

[44] S. Xia, J. Lv, N. Xu, G. Niu, and X. Geng, "Towards effective visual representations for partial-label learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 15 589–15 598. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2023/html/Xia_Towards_Effective_Visual_Representations_for_Partial-Label_Learning_CVPR_2023_paper.html

[45] M. D. Siegelin and A. C. Borczuk, "Epidermal growth factor receptor mutations in lung adenocarcinoma," *Laboratory investigation*, vol. 94, no. 2, pp. 129–137, 2014.

[46] C. Arteaga, "Targeting her1/egfr: a molecular approach to cancer therapy," in *Seminars in oncology*, vol. 30, no. 3. Elsevier, 2003, pp. 3–14.

[47] S. Modi, C. Saura, T. Yamashita, Y. H. Park, S.-B. Kim, K. Tamura, F. Andre, H. Iwata, Y. Ito, J. Tsurutani *et al.*, "Trastuzumab deruxtecan in previously treated her2-positive breast cancer," *New England Journal of Medicine*, vol. 382, no. 7, pp. 610–621, 2020.

[48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

[49] X. Chen, H. Fan, R. Girshick, and K. He, "Improved baselines with momentum contrastive learning," *arXiv preprint arXiv:2003.04297*, 2020. [Online]. Available: https://arxiv.org/abs/2003.04297

[50] Z. Huang, F. Bianchi, M. Yuksekgonul, T. Montine, and J. Zou, "Leveraging medical twitter to build a visual–language foundation model for pathology ai," *bioRxiv*, pp. 2023–03, 2023, doi:10.1101/2023.03.29.534834.

[51] X. Wang, Y. Du, S. Yang, J. Zhang, M. Wang, J. Zhang, W. Yang, J. Huang, and X. Han, "Retccl: Clustering-guided contrastive learning for whole-slide image retrieval," *Medical image analysis*, vol. 83, p. 102645, 2023.

[52] X. Wang, S. Yang, J. Zhang, M. Wang, J. Zhang, W. Yang, J. Huang, and X. Han, "Transformer-based unsupervised contrastive learning for histopathological image classification," *Medical image analysis*, vol. 81, p. 102559, 2022.

[53] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PMLR, 2021, pp. 8748–8763. [Online]. Available: https://proceedings.mlr.press/v139/radford21a.html

[54] X. Chen, H. Zheng, Y. Li, Y. Ma, L. Ma, H. Li, and Y. Fan, "Versatile medical image segmentation learned from multi-source datasets via model self-disambiguation," *arXiv preprint arXiv:2311.10696*, 2023.

[55] Z. Wang and C. Yang, "Mixsegnet: Fusing multiple mixed-supervisory signals with multiple views of networks for mixed-supervised medical image segmentation," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108059, 2024.