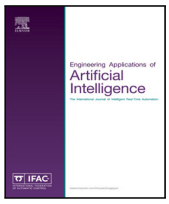
Engineering Applications of Artificial Intelligence 142 (2025) 109908

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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



Research paper

Pseudo-label attention-based multiple instance learning for whole slide image classification

Jing He [a], Ping Wang [b], Jingwen Cai [b], Dan Tang [b], Shaowen Yao [a], Renyang Liu [c,*]

[a] Engineering Research Center of Cyberspace, Yunnan University, Kunming, 650500, Yunnan, China
[b] The National Pilot School of Software, Yunnan University, 650500, Kunming, Yunnan, China
[c] Institute of Data Science, National University of Singapore, Singapore, 117602, Singapore

ARTICLE INFO

Keywords:
Computational pathology
Deep learning
Whole slide image analysis
Weakly supervised classification
Multiple instance learning

ABSTRACT

Automating disease classification in whole slide images (WSIs) is crucial for improving clinical diagnostic efficiency. However, existing multiple instance learning (MIL) approaches for this task often struggle with challenges such as insufficient focus on positive regions and data imbalance between positive and negative regions. These issues can lead to suboptimal performance in practical applications. To address these problems, in this paper, we propose a novel embedding-based MIL technique called pseudo-label attention-based multiple instance learning (PAMIL). PAMIL aggregates each instance's features regarding their contributions to improving downstream classification performance. The key insight of PAMIL involves training the model in a supervised manner by introducing pseudo-labels to emphasize positive regions. Additionally, we propose a fine-tuning strategy to effectively refine the dataset, eliminating the interference of false-positive data and alleviating data imbalance. The effectiveness of PAMIL was demonstrated through comparisons with six state-of-the-art MIL techniques across two large-scale, real-world datasets. Empirical results show that the proposed method outperforms other methods, achieving up to a 2.15% improvement in accuracy and a 1.61% increase in area under the curve (AUC) on the Cancer Genome Atlas Non-Small Cell Lung Cancer (TCGA-NSCLC) dataset, highlighting the superiority of our method in practical applications, such as helping clinicians diagnose quickly.

1. Introduction

Histopathology image analysis plays an essential role in cancer detection, diagnosis, prognosis, and treatment response prediction in patients (Lu et al., 2021; Myronenko et al., 2021; Li et al., 2022). Whole slide image (WSI), as a frequently used form of data in histopathology image analysis, contains a wealth of information about the morphological and functional characteristics of biological systems. They can be used to monitor the underlying mechanisms contributing to disease progression and patient survival outcomes. The widespread use of WSI images underscores the need for automated histopathological image diagnosis (Liu et al., 2017; Li and Ping, 2018; Madabhushi and Lee, 2016; Sirinukunwattana et al., 2017; Chen et al., 2019; Li et al., 2021a).

In many cases, WSIs boast incredibly high resolutions, often as large as 150,000*150,000 pixels. This sheer size renders them nearly impossible for deep learning models to process effectively. Furthermore, the training of supervised deep learning models needs vast datasets with meticulously crafted annotations of high-quality. However, clinical datasets often lack pixel-level annotations for WSIs. Therefore, weakly supervised learning (WSL) has been widely used in the field of histopathology as it can use coarse-grained (image-level) annotations to automatically infer fine-grained (pixel-level or patch-level) information.

Multi-instance learning (MIL), a specific form of WSL, stands as a prominent deep learning technology within digital pathology. MIL-based methods have effectively solved the problem posed by large WSI images by dividing them into numerous patches. Meanwhile, these methods address the limitations associated with pixel-level annotations for tissue phenotyping by using WSI labels or patient-level labels provided through weak supervision. Existing MIL-based methods generally consist of two stages (Dietterich et al., 1997; Maron and Lozano-Pérez, 1997): the first stage involves extracting instance-level feature representations from randomly sampled image patches within a WSI bag (where all patches extracted from a WSI are considered a bag), and the second stage employs an aggregation algorithm to process the bag of instances, yielding a WSI slide-level feature representation for downstream classification tasks.

* Corresponding author.
E-mail addresses: hejing@ynu.edu.cn (J. He), wangping@mail.ynu.edu.cn (P. Wang), tangdan@mail.ynu.edu.cn (D. Tang), yaosw@ynu.edu.cn (S. Yao), ryliu@nus.edu.sg (R. Liu).

https://doi.org/10.1016/j.engappai.2024.109908
Received 22 April 2024; Received in revised form 11 December 2024; Accepted 17 December 2024

Unfortunately, MIL-based WSI analysis approaches still grapple with two main challenges. First, the portion of the region of interest is relatively small when compared to the whole image, resulting in the model exhibiting a bias toward learning negative tissue features (Campanella et al., 2019). Thus, acquiring accurate tissue phenotyping, which encompasses both feature extraction and feature aggregation in the context of MIL, remains a pivotal concern in computational pathology. Furthermore, slide-based classification metrics may yield erroneous assessments if instance-level performance falters, particularly when only slide-level labels are accessible.

To enhance feature extraction performance, recent studies (Li et al., 2021b; Chen et al., 2022) have empowered extractors with self-supervised capabilities at the patch level instead of the traditional pre-training of feature extractors on the ImageNet dataset. This approach has substantially improved downstream task effectiveness. Regarding feature aggregation, existing methods have focused on aggregating instances with consideration of their mutual relationships (Chikontwe et al., 2020; Li et al., 2021b; Shao et al., 2021). However, most of these attention-based MIL methods predominantly emphasize high-scoring instances or prioritize the maximum and minimum instances, often neglecting adequate attention to the positive region (where positive instances reside within the bag). Furthermore, existing MIL methods (Lu et al., 2021; Campanella et al., 2019; Ilse et al., 2018) rely on image-level labels for WSL but lack the ability to discern the actual labels of instances, which could be utilized for supervised learning.

In response to these challenges, this paper proposes a novel embedding-based MIL approach, termed pseudo-label attention-based multiple instance learning (PAMIL). PAMIL explicitly models the contributions of each instance during the feature aggregation process. Unlike existing MIL methods, our focus is not on designing complex feature extractors or classifiers. Instead, we enhance the weights of positive instances to aggregate more accurate bag features. Additionally, we introduce a pseudo-label-based attention strategy to direct the model's focus toward positive regions. Specifically, we predict instance probabilities using an instance probability predictor, assign pseudo-labels to the highest and lowest instances, and subsequently train the instances in a supervised manner to dynamically emphasize positive regions. Furthermore, we devise a fine-tuning strategy to refine the model using features extracted from the enhanced dataset, reducing interference from uncertain instances. Extensive empirical results underscore our method's superiority, showcasing higher area under the curve (AUC) scores and improved classification accuracy. The primary contributions of our proposed method are as follows:

- We propose a novel Multi-instance Learning (MIL) framework, called Pseudo-label Attention-based MIL (PAMIL), which integrates attention mechanisms and a fine-tuning strategy to ensure the model's focus on positive regions while accurately identifying key instances for more effective feature aggregation.
- We introduce an innovative attention module driven by pseudo-labels, guiding the model to emphasize positive regions while retaining complementary information, thereby aggregating richer and more specific features for robust decision-making.
- We establish a dataset fine-tuning strategy based on pseudo-labeled instance probabilities. This strategy prioritizes high and low-scoring instances, eliminating unnecessary data and refining internal attention, leading to a more focused and effective model.
- We validate the effectiveness of the proposed framework using the Cancer Metastases in Lymph Nodes Challenge 2016 (CAMELYON16) and The Cancer Genome Atlas Non-Small Cell Lung Cancer (TCGA-NSCLC) datasets. Our empirical results demonstrate the superior classification performance of PAMIL for Whole Slide Image (WSI) classification tasks.

The remainder of this paper is structured as follows: In Section 2, we provide a brief review of related works in WSL. Section 3 outlines the entire pipeline of our proposed method. Extensive experiments demonstrating the superiority of PAMIL are presented in Section 4. Finally, Section 5 summarizes our work.

2. Related work

2.1. Multiple instance learning

Instance-based methods These approaches focus on training an instance-level classifier to classify individual instances and subsequently aggregate the instance labels to make predictions for the corresponding bag. Typically, instance-based methods employ either average pooling or maximum pooling (Feng and Zhou, 2017; Pinheiro and Collobert, 2015; Zhu et al., 2017), both of which are untrained operations that may limit their applicability. In contrast, instance-based approaches are more inclined to detect maximal instances. However, Liu et al. (2012) have shown that models successfully identifying key instances are more likely to achieve improved bag label predictions.

Embedding-based methods Embedding-based methods focus on training an instance-level feature extractor to generate embedding representations of individual instances. These embeddings are then aggregated to form bag embeddings, which are subsequently used to predict bag labels. This approach helps mitigate the potential issue of undertraining instance-level classifiers, which can occur in instance-based approaches. Wang et al. (2018) have demonstrated that embedding-based methods outperform instance-based ones in accurately predicting bag labels. Ilse et al. (2018) introduced an attention mechanism to aggregate instance features, with the aggregation operation being parameterized by a neural network. Later, Hashimoto et al. (2020) used the attention mechanism to aggregate instance features at different resolutions on multiple scales. Yao et al. (2020) proposed a clustering method followed by the aggregation of instance features from different clusters using the attention mechanism. Additionally, aggregation can also be achieved through the consideration of mutual instance relationships, which can be learned independently by various neural modules such as Recurrent Neural Networks (RNN) (Campanella et al., 2019), Graph Convolutional Networks (GCN) (Tu et al., 2019; Zhao et al., 2020), and Transformers (Li et al., 2019; Shao et al., 2021). Li et al. (2021b) introduced new MIL aggregators that model relationships between instances in a two-fluid structure using trainable distance metrics, demonstrating strong performance in this context.

2.2. Attention mechanism in MIL

The attention mechanism module assesses the significance of various features within an image. In the realm of deep learning, incorporating the attention mechanism module allocates higher weights to essential features, dampening the influence of irrelevant information and thereby enhancing the model's information processing efficiency (Wu et al., 2018a,b). In the context of the MIL problem, the attention weights assigned to each instance in the bag signify the extent of their contribution. Nikolaos Pappas and Andrei Popescu-Beliss (Pappas and Popescu-Belis, 2014) introduced an attention-based MIL approach where instance attention weights are learned using a linear regression model. Qi et al. (2017) employed an attention-based MIL operator, although attention was computed via a dot product, resulting in inferior performance compared to the "max" operator. Ilse et al. (2018) leveraged the attention mechanism to aggregate instances, assessing their importance and assigning distinct weights accordingly, pioneering integration of MIL with the attention mechanism module. Graph-Transformer for Whole Slide Image Classification (GTP) by Zheng et al. (2022) introduces a graph-based transformer model

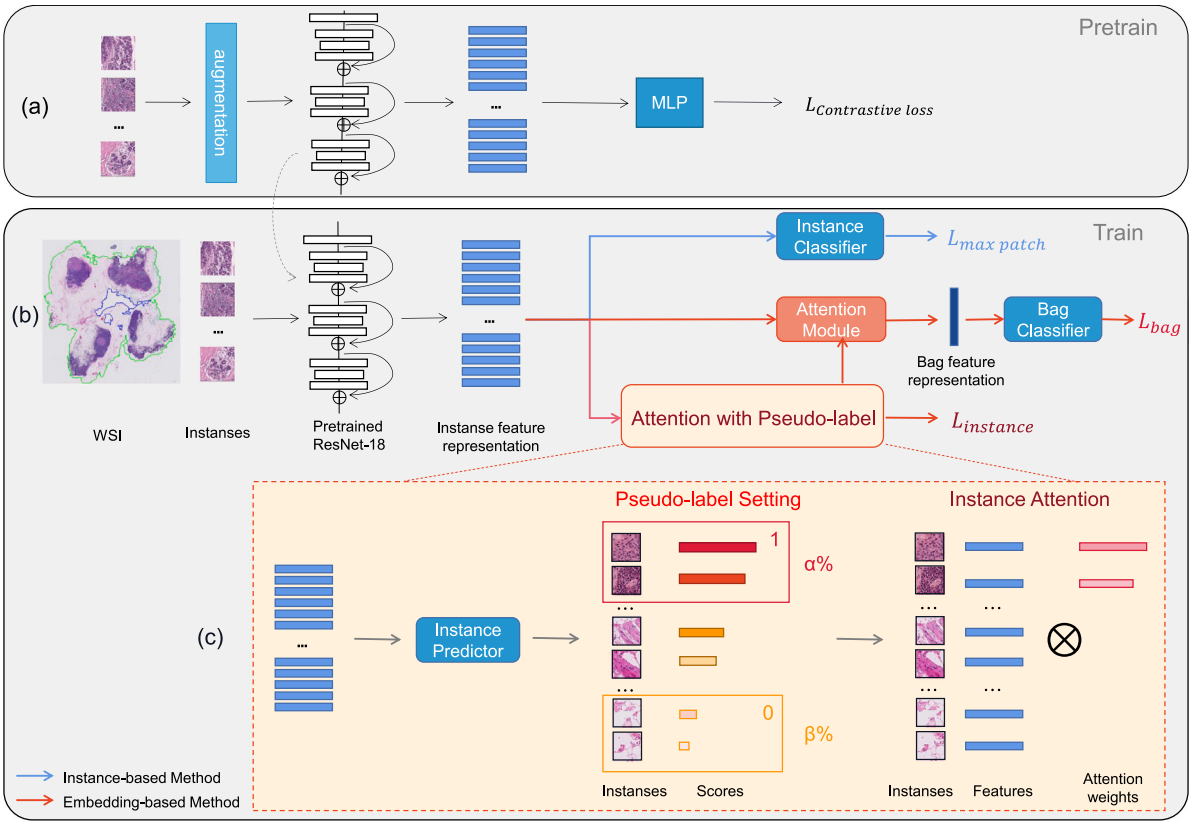


Fig. 1. Overview of PAMIL. (a) Each WSI is cropped into patches for feature extractor training by self-supervised contrastive learning. (b) Trained extractors are used to compute the embedding of each patch. Combination of instance-based and embedding-based methods are used together to obtain bag classification results, where key instance scores and bag "embeddings" aggregation are obtained based on attention with pseudo-label. (c) Attention with pseudo-label of PAMIL. The pseudo-label setting module scores instances by instance predictor and rearranges instances according to prediction scores. Selected top α% instances and bottom β% instances are pseudo-labeled as 1 and 0, respectively. The instance attention module aggregates instance probabilities into an attention matrix to weight top α% instances.

tailored for WSIs, which utilizes the graph structure of WSI patches to enhance the model's interpretability and accuracy. However, these previous efforts did not adequately emphasize the positive regions nor assign high attention weights to key instances. In this paper, we present a method that employs pseudo-labeling to train the model to discern positive regions, thus enhancing classification performance.

2.3. Pseudo-label in MIL

Pseudo-labeling involves creating target labels for unlabeled data to augment and essentially fully annotate a dataset. In the realm of deep learning, incorporating a pseudo-labeling module enables supervised learning with extensive amounts of unlabeled data, thereby enhancing model performance. In histopathology WSI, only slide-level labels are available, and patch-level labels are lacking. The central challenge in applying the pseudo-labeling method within the MIL algorithm is the generation of these pseudo-labels. Campanella et al. (2019) select key instances based on the predicted probabilities from the instance classifier and assign corresponding bag labels to these critical instances. Lu et al. (2021) generate pseudo-labels for clustered instances with high and weak attention, using instance-level feature supervision as signals in the feature space to train the instance-level classifier. Lerousseau et al. (2020) take this a step further by leveraging parameters in the realm of WSI image segmentation to identify instances with high and weak focus, guided by instance prediction probabilities. They then assign pseudo-labels to train instance feature extractors, ultimately achieving advanced segmentation results.

3. Methods

3.1. Problem formulation

In the context of MIL, each slide denoted as W_j from the dataset $W = \{W_1, W_2, \ldots, W_N\}$, comprising N WSIs, is partitioned into smaller patches x_i, where $i = \{1, 2, \ldots, n\}$, and n represents the number of patches extracted from W_j. All the patches x_i originating from a slide W_j collectively form a bag $B = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, where $y_i \in \{0, 1\}$, for $i = \{1, 2, \ldots, n\}$ represents the label of each patch, signifying an instance. To be specific, the training samples in MIL consist of N bags, each labeled with $Y_j \in \{0, 1\}, j = \{1, 2, \ldots, N\}$. This implies that bags containing multiple instances are considered as a set of training samples, with each bag possessing a bag-level ground-truth label denoted as Y_j. Notably, only slide-level labels (i.e., pixel-level labels of WSIs) are accessible, while instance-level labels are absent. In accordance with the standard multiple instance (SMI) assumption (Amores, 2013), if a bag contains at least one positive instance (i.e., one or more instances belong to some target positive class), the bag's label is designated as positive; otherwise, it is considered negative. Consequently, the prediction of the bag label, denoted as $c(B)$, is formulated as follows:

$$c(B) = \begin{cases} 0, & if \sum y_i = 0 \\ 1, & otherwise \end{cases}. \tag{1}$$

Depending on the specific transformations applied to the instances, the bag labels predicted by MIL can be further expressed as:

$$c(B) = g(f(x_0), \ldots, f(x_n)), \tag{2}$$

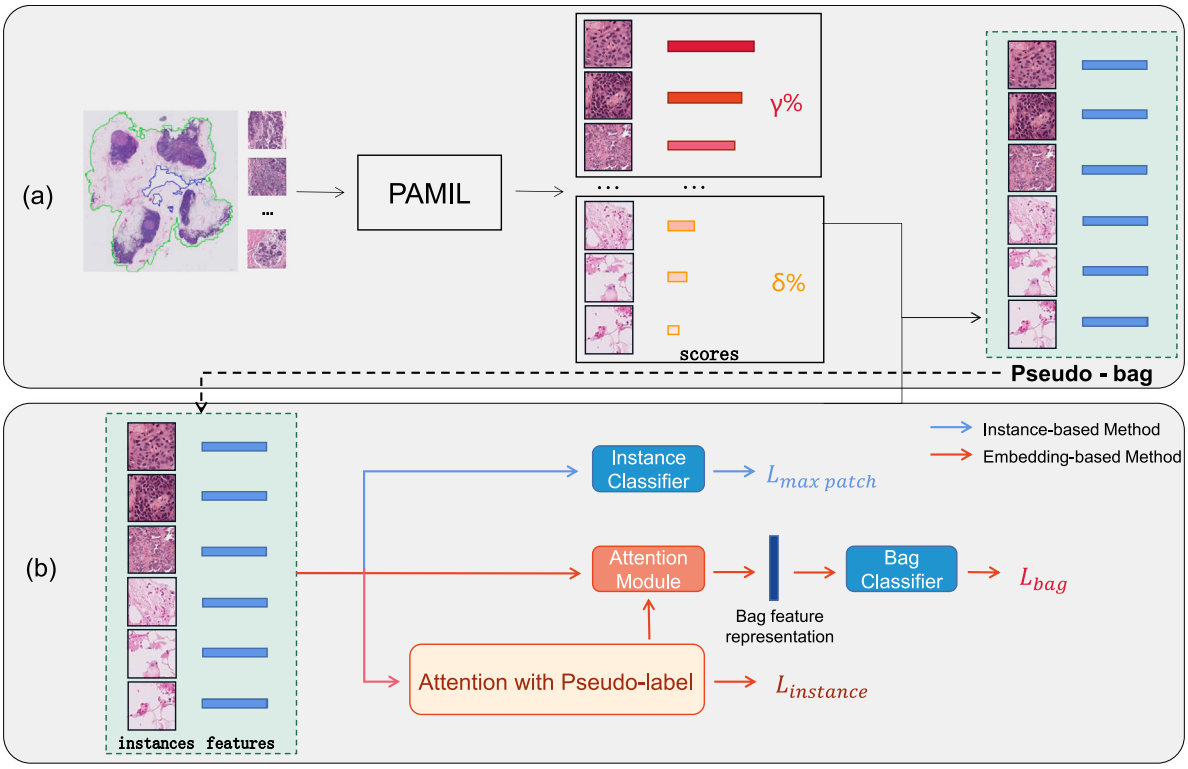


Fig. 2. Workflow of fine-tuning strategy. (a) Training images are first fed into PAMIL to select instances to constitute pseudo-bags. We learn attention weights for instances, which can be used to select top a% (high attentions) and bottom b% (low attentions) instances in each bag. (b) The model is trained again with pseudo-bags to fine-tune internal attention and further improve classification performance.

where, in the context of embedding-based MIL methods, the function f serves as an instance-level feature extractor, responsible for obtaining the embedding representation of each instance. The function g operates as an aggregation operator, tasked with aggregating instance embeddings into a bag-level embedding. MIL initially maps instances to low-dimensional embeddings and aggregates them to generate bag embeddings. These bag embeddings are subsequently processed by a slide-level classifier, yielding a bag representation that is independent of the instance counts.

3.2. Framework overview

The overall framework of the proposed PAMIL is illustrated in Fig. 1, which can be divided into two main phases: pre-training and training. In the pre-training phase, we employ the SOTA self-supervised contrastive learning method, SimCLR (Chen et al., 2020), to pre-train the patch feature extractor using histopathological image data. In the training phase, we utilize the well-trained patch feature extractor to extract instance-level features. The final classification results are generated through two classification branches, which include instance- and bag-level embeddings. In addition, to ensure that the model focuses extensively on positive region features and aggregates more precise bag embeddings for accurate bag classification, we constructed a novel attention module based on pseudo-labels within the bag embedding classification branch.

3.3. PAMIL

3.3.1. Pseudo-label setting

Compared to weakly supervised learning-based or unsupervised learning-based approaches, supervised learning-based methods demonstrate superiority in computational pathology (CPATH) image classification tasks. However, training a model in a fully supervised manner necessitates accessible patch labels. Unfortunately, acquiring such annotations for the numerous patches extracted from WSI which can number in the thousands, is impractical yet crucial for achieving good generalization. PAMIL introduces a novel approach to address the challenge of missing patch-level ground-truth labels. Instead, it constructs a set of pseudo-labels, which are generated by exploiting the properties of the available slide-level labels denoted as Y_j. These pseudo-labels serve as substitutes for patch-level ground-truth labels, enabling the supervised training of the instance predictor g_p. The instances are rearranged based on their prediction scores in descending order as part of this process.

In this study, different pseudo-labels were assigned to negative and positive WSIs: (1) Patches extracted from negative WSIs are all labeled as negative, setting their pseudo-labels to 0. (2) Patches obtained from positive WSIs may contain both positive and negative instances, making it challenging to accurately determine their pseudo-labels. To mitigate the influence of false-positive samples, we consider only the first α percent instances and the last β percent instances for setting pseudo-labels (α% represents the assumed minimum relative area of tumor tissue in the WSI, and β% represents the range for normal tissue). Pseudo-labels for the first α% (high-scoring instances) are assigned as 1, while pseudo-labels for the last β% (low-scoring instances) are assigned as 0, with the constraint that $\alpha + \beta \leq 1$. The instance loss function $L_{instance}$ for setting pseudo-labels can be written as follows:

$$L_{instance} = c_0 \times \sum_{\substack{W_j \in W; \\ Y_j = 0}} \Big[\sum_{x_i \in X} L(g_p(x_i), 0) \Big] + c_1 \times \sum_{\substack{W_j \in W; \\ Y_j = 1}} \Big[\sum_{\substack{x_i \in X; \\ x_i \in P(g_p(x_i);\alpha,100)}} L(g_p(x_i), 1) + \sum_{\substack{x_i \in X; \\ x_i \in P(g_p(x_i);0,\beta)}} L(g_p(x_i), 0) \Big], \tag{3}$$

where c_0 and c_1 represent the predicted probabilities of a negative or positive image W_j from the dataset W, as defined in Eq. (1). L denotes

the cross-entropy loss function. Y_j is the slide-level ground-truth label for W_j, where $Y_j = 0$ represents a normal image, and $Y_j = 1$ represents a positive image. x_i represents an instance from the bag X, cropped from W_j. $g_p(x_i)$ is the predicted probability of the instance x_i generated by the instance predictor g_p. $x_i \in P(g_p(x_i); \alpha, 100)$ signifies instances with instance probabilities falling in the range $(\alpha, 100)$. Similarly, $x_i \in P(g_p(x_i); 0, \beta)$ represents instances with instance probabilities within the range $(0, \beta)$.

3.3.2. Instance attention

The objective of the instance attention mechanism is to enhance the model's focus on positive regions while preserving information from other regions. This is achieved by dynamically adjusting attention weights based on the obtained instance probabilities.

Specifically, the instance probabilities obtained from the supervised instance predictor g_p are aggregated into an attention matrix, which is used to assign higher weights to the top α% instances. Concurrently, other regional features are retained through the introduction of a residual module. Consequently, the attention to positive regions is enhanced, while other information is preserved. This process can be formulated as:

$$score = predictor(\ x_i\ ), \tag{4}$$

$$AttnScore = Softmax(\ top_k\ (\ score\ )\ ), \tag{5}$$

$$x_i{}' = x_i + x_i\ *\ AttnScore, \tag{6}$$

where the $score$ denotes the prediction score of instance x_i, top_k denotes the operation of selecting the first k instances, $*$ denotes matrix multiplication, and $x_i{}'$ is the final output.

Attention Module Subsequently, the effect of the largest instance on the aggregation of bag features is considered. The largest instance plays a significant role in determining whether a bag is positive or negative among all instances. Following the approach in Li et al. (2021b), we utilize the distance between instances and the largest instance as the attention weight for each instance:

$$a_i = D_{(h_i,h_m)} = \frac{exp(\langle h_i, h_m\rangle)}{\sum_{k=0}^{N-1} exp(\langle h_i, h_m\rangle)}, \tag{7}$$

where h_i is the feature of instance x_i and h_m is the feature of the largest instance x_m.

3.4. Fine-tuning strategy

To achieve more accurate aggregation of bag features, we introduce a fine-tuning strategy named PAMIL-Fine Tuning (FT) in this paper, which is illustrated in Fig. 2, aimed at refining the dataset for model training. PAMIL-FT reprocesses the dataset to fine-tune internal attention. Specifically, it leverages the instance probabilities provided by the probability predictor g_p to generate a pseudo-bag, comprising the top γ% (high attention) and the bottom δ% (low attention) instances, for fine-tuning the model. This refined dataset enables the re-ranking of selected positive and negative instances within each bag, with key instances that are more representative of positive tissue receiving higher rankings and greater attention weights. As a result, more accurate aggregated bag features are obtained. Ultimately, the model not only distinguishes positive tissues but also identifies key instances within them, enhancing classification performance.

PAMIL-FT offers the following advantages: (1) Elimination of confounding data (false-positive instances within bags) to improve the model's ability to distinguish positive tissues. (2) Solving the tissue imbalance by adjusting the values of γ and δ to balance the ratio of positive and negative instances within the refined dataset.

4. Experiments

In this section, we present the performance of the proposed method on two challenging datasets: CAMELYON16 and TCGA lung cancers. We compare it with recent MIL-based methods for the histopathological WSI classification task. Additionally, we perform ablation experiments to investigate the proposed methods under different settings.

Implementation. We use ResNet-18 (He et al., 2016) as the backbone network, pre-trained in a self-supervised comparative learning (SimCLR) manner, to extract instance features in the MIL framework. We use the Adam optimizer (Kingma and Ba, 2014) with a constant learning rate of 0.0001 to update the model weights during training. All experiments were conducted on a GPU server equipped with an NVIDIA A100-PCIE-40 GB, utilizing Python 3.8 and PyTorch v1.13.1 as the software environment.

Baselines. To demonstrate the effectiveness of PAMIL, we include the following baselines, encompassing both traditional instance-based deep MIL methods and SOTA deep MIL methods: (1) Conventional instance-based MIL methods, including mean-pooling and max-pooling. (2) The classic AB-MIL (Ilse et al., 2018). (3) RNN-based RNN-MIL (Campanella et al., 2019). (4) Three variants of AB-MIL: non-local attention pooling DSMIL (Li et al., 2021b), single-attention-branch CLAM-SB (Lu et al., 2021), and multi-attention-branch CLAM-MB (Lu et al., 2021). (5) Transformer-based MIL, Trans-MIL (Shao et al., 2021). (6) Double-tier MIL framework, DTFT (Zhang et al., 2022). (7) Graph-Transformer-based, GTP (Zheng et al., 2022). We reproduce the empirical results from their officially released code using the same settings.

Metrics. In our experiments, we use classification accuracy and the area under the receiver operating characteristic curve (AUC) as the primary evaluation metrics. Classification accuracy measures the overall performance of the model, and is calculated as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{8}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

AUC measures the model's ability to distinguish between positive and negative samples at various thresholds. A higher AUC indicates better performance. It is calculated by plotting the ROC curve and computing the area under it:

$$\text{AUC} = \int_{\text{FPR}=0}^{1} \text{TPR}(\text{FPR})\, d\text{FPR}. \tag{9}$$

Together, these metrics provide a comprehensive evaluation of PAMIL's effectiveness in WSI classification.

4.1. Breast cancer classification result

We first introduce the CAMELYON16 dataset and present the experimental results of the proposed method and baselines.

Dataset. The CAMELYON16 breast cancer lymphocyte dataset (Litjens et al., 2018) is a publicly available dataset for detecting breast cancer metastases. It comprises 270 training images and 129 test images with gigapixel resolution. The WSIs have multiple resolutions, which can be divided into approximately 3.2 million patches at 20× magnification and 0.65 million patches at 10× magnification. The task is formulated as a weakly supervised binary classification problem, using only slide-level labels to determine whether a specific bag is positive or negative. If a WSI contains all negative tissue (i.e., all instances in the bag are negative), the bag is labeled as negative. Conversely, if the bag contains positive tissue (i.e., some instances in the bag are positive), it is labeled as positive.

Unfortunately, there exists a significant data imbalance in the WSIs of the CAMELYON16 dataset, with negative tissue being much more abundant than positive tissue. This imbalance can potentially bias the

Table 1
Performance comparison of PAMIL and state-of-the-art methods at 20× and 10× magnification on the CAMELYON16 dataset with Accuracy and AUC, highlighting the advantages of the proposed PAMIL and PAMIL-FT techniques in comparison to other conventional approaches. The results emphasize the robustness and improved classification performance of the PAMIL-based methods in cancer detection tasks.

Methods	CAMELYON16 20×		CAMELYON16 10×	
	Accuracy	AUC	Accuracy	AUC
Mean Pooling	0.639 ± 0.007	0.465 ± 0.010	0.636 ± 0.011	0.539 ± 0.012
Max Pooling	0.806 ± 0.028	0.833 ± 0.022	0.826 ± 0.029	0.858 ± 0.030
AB-MIL (Ilse et al., 2018)	0.835 ± 0.011	0.855 ± 0.006	0.841 ± 0.012	0.860 ± 0.007
RNN-MIL (Campanella et al., 2019)	0.825 ± 0.026	0.873 ± 0.002	0.844 ± 0.024	0.875 ± 0.002
DSMIL (Li et al., 2021b)	0.863 ± 0.013	0.893 ± 0.011	0.925 ± 0.008	0.951 ± 0.009
CLAM-SB (Lu et al., 2021)	0.846 ± 0.028	0.855 ± 0.015	0.859 ± 0.025	0.866 ± 0.013
CLAM-MB (Lu et al., 2021)	0.813 ± 0.025	0.877 ± 0.017	0.823 ± 0.027	0.878 ± 0.019
Trans-MIL (Shao et al., 2021)	0.849 ± 0.010	0.896 ± 0.032	0.860 ± 0.011	0.893 ± 0.031
DTFD-MIL (Zhang et al., 2022)	0.874 ± 0.016	0.891 ± 0.013	0.867 ± 0.125	0.944 ± 0.004
PAMIL (Ours)	0.876 ± 0.010	0.905 ± 0.004	0.938 ± 0.011	0.965 ± 0.003
PAMIL-FT (Ours)	**0.888 ± 0.011**	**0.916 ± 0.005**	**0.953 ± 0.009**	**0.991 ± 0.001**

Table 2
Performance comparison of PAMIL and state-of-the-art methods at 20× and 10× magnification levels on the TCGA Lung Cancer dataset with Accuracy and AUC metrics for various methods, highlighting the advantages of the proposed PAMIL and PAMIL-FT approaches. Results are shown for both 20× and 10× magnifications, specifically for the TCGA-NSCLC subset, emphasizing the improvements in classification performance.

Methods	TCGA-NSCLC 20×		TCGA-NSCLC 10×	
	Accuracy	AUC	Accuracy	AUC
Mean Pooling	0.833 ± 0.014	0.901 ± 0.015	0.728 ± 0.011	0.840 ± 0.012
Max Pooling	0.859 ± 0.030	0.946 ± 0.032	0.847 ± 0.029	0.901 ± 0.033
AB-MIL (Ilse et al., 2018)	0.869 ± 0.033	0.942 ± 0.028	0.772 ± 0.032	0.866 ± 0.028
RNN-MIL (Campanella et al., 2019)	0.862 ± 0.027	0.911 ± 0.026	0.845 ± 0.024	0.895 ± 0.025
DSMIL (Li et al., 2021b)	0.866 ± 0.016	0.926 ± 0.020	0.893 ± 0.013	0.962 ± 0.019
CLAM-SB (Lu et al., 2021)	0.818 ± 0.042	0.882 ± 0.024	0.800 ± 0.041	0.873 ± 0.023
CLAM-MB (Lu et al., 2021)	0.842 ± 0.044	0.938 ± 0.022	0.840 ± 0.043	0.912 ± 0.019
Trans-MIL (Shao et al., 2021)	0.877 ± 0.025	0.930 ± 0.014	0.867 ± 0.022	0.923 ± 0.013
DTFD-MIL (Zhang et al., 2022)	0.889 ± 0.032	0.938 ± 0.026	0.877 ± 0.029	0.937 ± 0.023
GTP (Zheng et al., 2022)[a]	0.823 ± 0.001	0.929 ± 0.003	–	–
PAMIL (Ours)	0.901 ± 0.013	0.947 ± 0.005	0.918 ± 0.012	0.963 ± 0.004
PAMIL-FT (Ours)	**0.907 ± 0.011**	**0.952 ± 0.004**	**0.932 ± 0.010**	**0.969 ± 0.001**

[a] Data sourced from the original paper (Zheng et al., 2022).

Table 3
Detailed information of the CAMELYON16 dataset (random 8:2 train-test split).

Dataset	Type	Negative	Positive	Total	Data size
CAMELYON16	Training set	159	111	270	700G
	Test set	81	48	129	

model to learn negative features while ignoring positive features. The following experiments show that our proposed method, which enhances the attention to positive region features, overcomes this difficulty and achieves advanced performance. Table 3 provides detailed information about the CAMELYON16 dataset, including the distribution of training and test sets.

Results. The classification results on CAMELYON16 are summarized in Table 1. We conducted two sets of comparison experiments using non-overlapping patches of size 256 × 256 pixels sampled from the tissue regions at 20× and 10× magnification, respectively. Since the positive area in CAMELYON16 accounts for only a small fraction of the total WSI, ensuring that the model sufficiently focuses on the positive region is crucial for correct WSI classification.

Traditional pooling aggregators, such as Mean Pooling and Max Pooling, aggregate instance scores to generate a bag score for classification. As shown in Table 1, Max Pooling outperforms Mean Pooling by better identifying key instances, highlighting the importance of key instance identification for pathology image classification. Improved MIL aggregators, however, consistently outperform traditional ones, as they leverage attention mechanisms to assign different weights to instances, allowing the model to focus on key regions more effectively. This suggests that capturing instance relationships is critical for WSI classification.

In the CAMELYON16 20× experiments, DTFD-MIL constructed a dual-layer MIL framework to reduce the number of instances per bag and aggregated instances based on the ABMIL attention mechanism, achieving promising classification performance. Our proposed method, PAMIL, optimizes the attention mechanism further, leading to the highest scores with an accuracy (ACC) of 87.5% and an area under the curve (AUC) of 90.48%. In the 10× experiments, DSMIL assigned higher weights to key instances, outperforming all other compared methods. Our method, PAMIL, improved upon DSMIL, with ACC and AUC increases of 1.25% and 1.40%, respectively. This superior performance can be attributed to PAMIL's ability to focus better on positive region features and assign higher importance to positive tissue instances.

CAMELYON16 breast cancer classification presents additional challenges due to data imbalance, as the positive regions constitute only a small part of the overall WSI, causing models to learn the dominant features of larger negative areas. The experimental results show that PAMIL-FT outperforms all other methods, improving ACC and AUC by 1.56% and 2.66% at 10× magnification and by 1.25% and 1.08% at 20× magnification, respectively. This demonstrates that PAMIL-FT, by adjusting the data structure, mitigates the issue of data imbalance, allowing the model to aggregate more accurate bag features and further enhance classification accuracy.

4.2. Lung cancer classification result

In this part, we introduce the TCGA lung cancer dataset and show the experimental results of the proposed methods on it.

Dataset. TCGA non-small cell lung cancer (TCGA-NSCLC) dataset comprises a total of 1054 WSIs, encompassing two sub-types of lung cancer: Lung adenocarcinoma (LUAD) (Collisson et al., 2014) and lung

Table 4
Effects of model modules in PAMIL on Accuracy and AUC across CAMELYON16 and TCGA-NSCLC datasets. The table compares the performance of different model configurations.

	Methods	CAMELYON16		TCGA-NSCLC	
		Accuracy	AUC	Accuracy	AUC
(I)	SimCLR+PAMIL+FT(**Ours**)	**0.9531**	**0.9911**	**0.9320**	**0.9697**
(II)	SimCLR+PAMIL(**Ours**)	**0.9375**	**0.9645**	**0.9175**	**0.9627**
(III)	SimCLR+Attention+MIL	0.9250	0.9505	0.8932	0.9620
(IV)	Attention+MIL	0.8682	0.8760	0.7719	0.8656

Table 5
Detailed information of the TCGA lung dataset (random 8:2 train-test split).

Dataset	Type	LUAD	LUSC	Total	Data size
TCGA-NSCLC	Training set	432	410	842	767G
	Test set	109	103	212	

squamous cell carcinoma (LUSC) (Network et al., 2012). We selected 1053 WSI digital slides, consisting of 482 contaminated and 571 uncontaminated slides from TCGA, to construct the TCGA-NSCLC dataset. It includes 541 LUAD slides and 513 LUSC slides. The dataset provides 5.2 million patches at 20× magnification and 1.2 million patches at 10× magnification. The task is framed as a weakly supervised sub-type classification problem, wherein only slide-level labels are used to determine whether a WSI belongs to LUAD or LUSC. The bags contain mixtures of tumor and healthy patches for positive bags, and all healthy patches for negative bags. Positive slides in this dataset contain substantial tumor regions (average total cancer area per slide >80%), leading to a large part of positive patches in positive bags. Consequently, pooling operators can achieve better performance compared to the CAMELYON16 dataset. The following experiments show the substantial improvements in classification results achieved by enhancing the attention to positive region features. Table 5 provides detailed information about the TCGA lung dataset, including the distribution of training and test sets.

Results. The classification results on TCGA-NSCLC are summarized in Table 2. Similarly, we conducted two sets of comparison experiments using non-overlapping patches of size 256 × 256 pixels extracted from tissue regions at 20× and 10× magnification, respectively. Since the tumor region in positive slides is significantly larger, even instance-level approaches perform well on the TCGA lung cancer dataset. Consequently, the key challenge lies in the model's ability to identify key instances that can effectively represent the bag for accurate WSI classification.

Due to the large tumor regions in positive slides of the TCGA lung cancer dataset, even traditional pooling aggregators like Mean Pooling and Max Pooling perform well. Analyzing Table 2, we observe that in the TCGA-NSCLC 20× experiment, our proposed method PAMIL improves ACC and AUC by 1.25% and 0.97%, respectively, compared to DTFD-MIL. Similarly, in the TCGA-NSCLC 10× experiment, PAMIL outperforms DSMIL with ACC and AUC improvements of 2.43% and 0.07%. This demonstrates that PAMIL focuses more on positive regions, aggregating richer and more specific bag features.

Furthermore, in the TCGA lung cancer dataset, experimental results show that PAMIL-FT outperforms all other compared methods. Compared to PAMIL, PAMIL-FT improves ACC and AUC by 1.45% and 0.70% at 10× magnification and by 0.58% and 0.44% at 20× magnification. This indicates that PAMIL-FT further enhances model classification accuracy by alleviating data imbalance and adaptively adjusting instance attention weights.

PAMIL demonstrates superior performance in aggregating positive region features, and PAMIL-FT enhances the model's focus on key instances. The proposed method consistently outperforms other existing MIL methods at both 10× and 20× magnification, achieving remarkable results even with the data at 10× magnification.

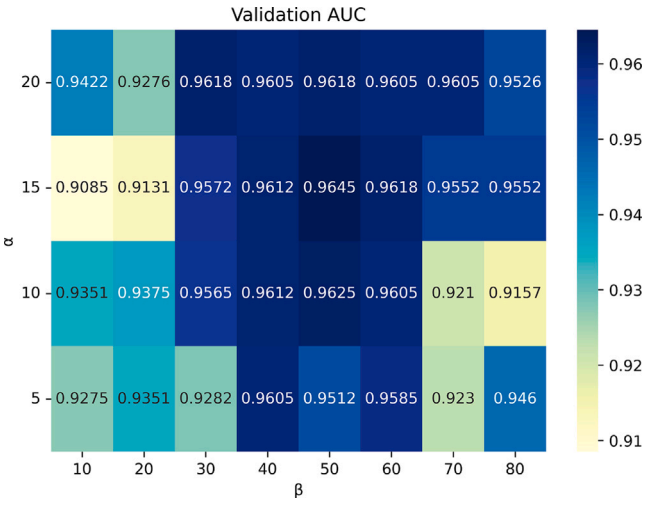


Fig. 3. Effect of α and β on the Validation AUC for PAMIL on the CAMELYON16 dataset. This heatmap shows the AUC values obtained by varying α and β. Darker colors indicate higher AUC values. The figure is used to identify the optimal values of α and β for maximizing the validation AUC on this dataset.

4.3. Ablation study

To further explore the contributions of the PAMIL modules and the fine-tuning strategy in improving performance, we conducted a series of ablation studies. All these experiments were conducted on the CAMELYON16 and TCGA-NSCLC datasets and evaluated using accuracy and AUC.

4.3.1. Effects of pseudo-label-based attention strategy and fine-tuning strategy

Our proposed work compromises two main components: (1) PAMIL and (2) FT strategy (FT). In this context, we compared the effect of the PAMIL module and the fine-tuning strategy with several baselines, all performed on the CAMELYON16 and TCGA-NSCLC datasets at 10× magnification. Experiments (I), (II), and (III) use features learned through self-supervised contrastive learning, while experiment (IV) uses features learned by a feature extractor pre-trained on the ImageNet dataset. Additionally, experiments (III) and (IV) both use attention-based MIL methods.

As shown in Table 4 (I) and (III), both PAMIL and FT contribute to accuracy improvements of 2.81% in CAMELYON16 and 3.88% in TCGA-NSCLC, along with AUCROC improvements of 4.06% in CAMELYON16 and 0.77% in TCGA-NSCLC. Furthermore, the comparison between (I) and (II) highlights the performance boost achieved by the fine-tuning strategy. Additionally, we directly replaced PAMIL with (III), the attention-based MIL method, to confirm whether attention with pseudo-labels can effectively guide the model to focus more on the positive region. Finally, we verified the effectiveness of contrastive learning by comparing (III) and (IV).

4.3.2. Effects of pseudo-labeling method parameters

To further validate the performance of PAMIL on the CAMELYON16 dataset, we varied the values of α and β and measured the AUC. The results under different settings are shown in Fig. 3. Given the data imbalance in CAMELYON16 (positive area ≤20% of the total area), we set the range of α to $0 < \alpha \leq 20$ in the experiments. Since the negative area is ≥80%, the range of β is set to $0 < \beta \leq 80$. As shown in Fig. 3, when $\alpha = 15$ and $\beta = 50$, the proposed method achieves its best performance with an AUC of 0.9645. This indicates that setting pseudo-labels as 1 for the top 15% of instances and as 0 for the bottom 50% during training results in the optimal classification performance for the model.

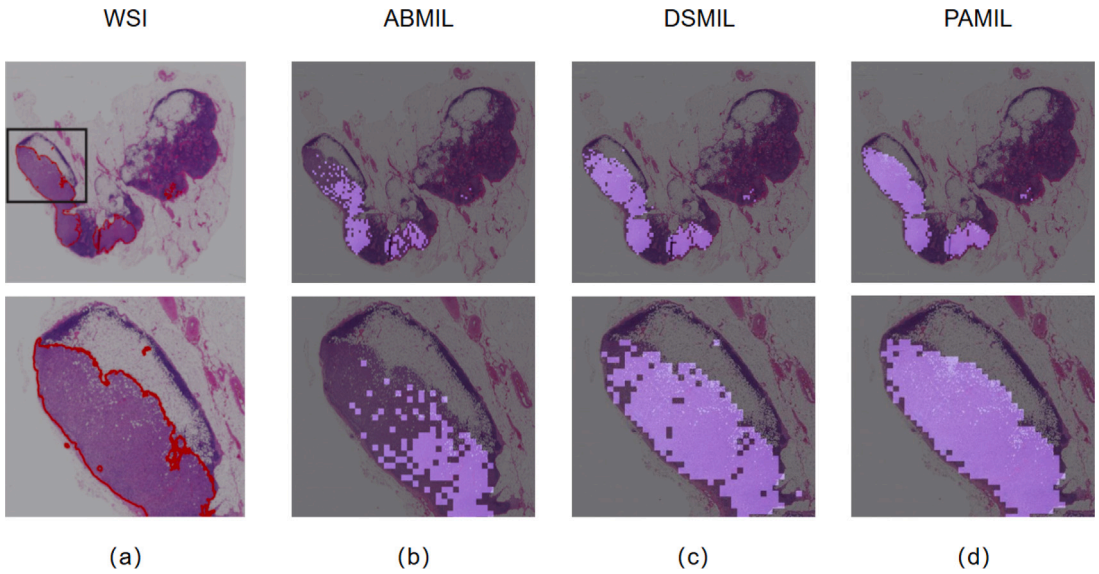


Fig. 4. Interpretability and visualization of attention maps in cancer region on CAMELYON16 cancer dataset. (a) shows ground-truth annotations, with (b–d) showing attention maps of AB-MIL, DSMIL, and PAMIL, respectively.

4.3.3. Interpretability and attention visualization

We will further show the interpretability of PAMIL. As shown in Fig. 4(a), the area within the red curve annotation represents the cancer region, as provided by the annotations in the CAMELYON16 dataset. In Fig. 4(b–d), we visualized the attention scores from AB-MIL, DSMIL, and PAMIL, respectively, as heatmaps to determine the region of interest (ROI) and interpret the important morphology used for diagnosis. Notably, compared to other methods (AB-MIL and DS-MIL), the results from PAMIL exhibit a high level of consistency between the finely annotated area and the heatmap. This observation illustrates that the proposed PAMIL effectively focuses on positive regional tissues, leading to improved classification performance.

4.3.4. Impact of imbalanced data

To investigate whether data imbalance in the CAMELYON16 dataset affects the model's performance, we conducted additional experiments. Specifically, we randomly sampled 200 cases and trained the model with different tumor sample proportions (30%, 40%, 50%, 60%, and 70%), using an 8:2 train-test split. After training, Accuracy, Precision, and F1 scores were computed for each configuration. As illustrated in Fig. 5, the performance differences across these imbalanced settings were minimal, indicating that the proposed method is robust to data imbalance.

4.3.5. Time efficiency analysis

We conducted a comprehensive analysis comparing the efficiency of the proposed model across the preprocessing, training, and prediction stages. In the preprocessing stage, most models employ similar steps, such as pruning, standardization, and feature extraction, resulting in negligible differences in preprocessing time and consistent performance across models. During the training phase, our model exhibited superior efficiency compared to state-of-the-art methods like DSMIL (Li et al., 2021b), primarily due to the use of pre-trained backbone networks as feature extractors, significantly reducing computational overhead. Experimental results revealed that on the CAMELYON16 dataset, training 240 samples required a total of 345.33 s, averaging 1.44 s per sample. Similarly, on the TCGA-NSCLC dataset, training 384 samples took 146.10 s in total, with an average of 0.38 s per sample. In the prediction stage, despite incorporating three predictors, each predictor consists of only two to three lightweight fully connected networks (MLPs), ensuring low computational complexity and maintaining high efficiency. Specifically, for the CAMELYON16 dataset, predicting 60 samples required 84.78 s in total, averaging 1.41 s per sample. On the TCGA-NSCLC dataset, predicting 97 samples took 38.88 s, with an average of 0.40 s per sample, as summarized in Table 6. These findings underscore the competitive time efficiency of our model, demonstrating its potential for practical applications.

Table 6
Training and testing time per sample for TCGA-NSCLC and CAMELYON16 datasets.

Dataset	Train time (s/sample)	Test time (s/sample)
TCGA-NSCLC	1.44	0.40
CAMELYON16	0.38	1.41

5. Conclusion

In this paper, we introduced Pseudo-label Attention-based MIL (PAMIL), a novel embedding-based approach for WSI classification. PAMIL leverages pseudo-labels to highlight positive tissues during the feature aggregation process and use a fine-tuning strategy to mitigate false positives and data imbalance issues. Our experiments demonstrate that PAMIL achieves superior performance even with reduced data requirements by effectively concentrating on relevant regions.

However, this approach has some limitations. PAMIL's reliance on single-modality WSI data overlooks the benefits of multimodal integration, and its performance may degrade in data-scarce or imbalanced situations. Additionally, the current method lacks adaptive segmentation strategies for diverse tissue types and may face efficiency challenges, particularly with large-scale datasets. The reliance on annotated data also poses constraints on scalability.

For future research, several avenues could be explored. First, incorporating multimodal data could enhance the model's robustness and accuracy. Second, investigating adaptive segmentation strategies across tissue types and magnifications within WSIs could improve model adaptability. We also aim to explore real-time processing techniques and optimize performance for larger datasets. Finally, integrating expert pathologist knowledge into the computational process could improve classification accuracy, especially in edge cases. Future studies might also leverage feature pyramids to fully exploit information across multiple layers of WSIs.

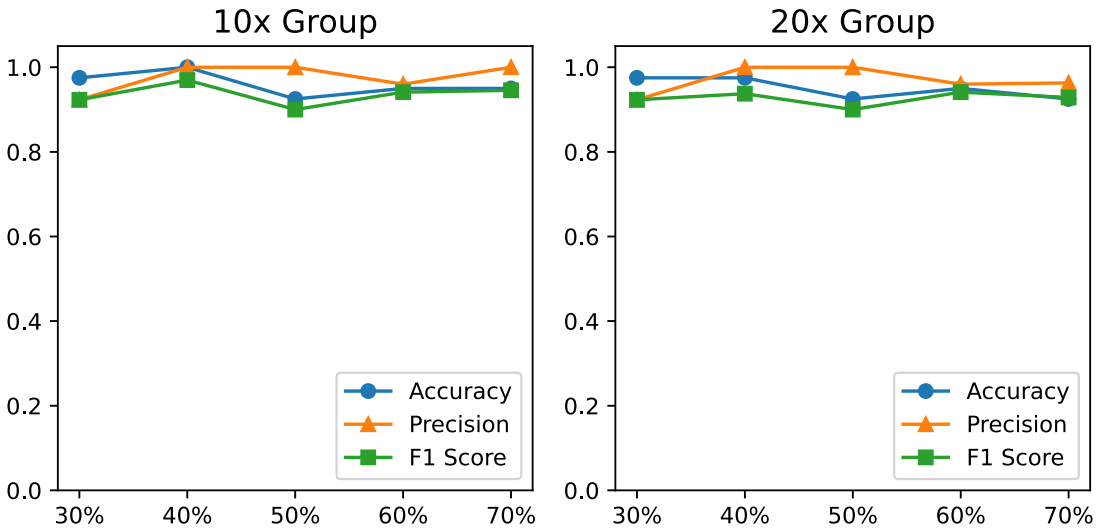


Fig. 5. Accuracy, Precision, and F1 scores under different sample proportions and groups.

CRediT authorship contribution statement

Jing He: Writing – review & editing, Supervision, Funding acquisition. **Ping Wang:** Conceptualization, Writing – original draft. **Jingwen Cai:** Writing – review & editing, Visualization, Validation. **Dan Tang:** Software, Investigation. **Shaowen Yao:** Supervision, Project administration . **Renyang Liu:** Writing – review & editing, Supervision.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is noprofessional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 62162067, 82360280 and in part by the Open Foundation of Key Laboratory in Software Engineering of Yunnan Province under Grant 2020SE310, the Open Foundation of Engineering Research Center of Cyberspace under Grant No. KJAQ202112013, the Science and Technology Plan in Key Fields of Yunnan under Grant 202001BB050076.

Data availability

The data that has been used is confidential.

References

Amores, J., 2013. Multiple instance classification: Review, taxonomy and comparative study. Artificial Intelligence 201, 81–105.

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

Chen, R.J., Chen, C., Li, Y., Chen, T.Y., Trister, A.D., Krishnan, R.G., Mahmood, F., 2022. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In: CVPR. pp. 16144–16155.

Chen, P.-H.C., Gadepalli, K., MacDonald, R., Liu, Y., Kadowaki, S., Nagpal, K., Kohlberger, T., Dean, J., Corrado, G.S., Hipp, J.D., et al., 2019. An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis. Nat. Med. 25, 1453–1457.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations. In: ICML. pp. 1597–1607.

Chikontwe, P., Kim, M., Nam, S.J., Go, H., Park, S.H., 2020. Multiple instance learning with center embeddings for histopathology classification. In: MICCAI. pp. 519–528.

Collisson, E., Campbell, J., Brooks, A., Berger, A., Lee, W., Chmielecki, J., Beer, D., Cope, L., Creighton, C., Danilova, L., et al., 2014. Comprehensive molecular profiling of lung adenocarcinoma: The cancer genome atlas research network. Nature 511, 543–550.

Dietterich, T.G., Lathrop, R.H., Lozano-Pérez, T., 1997. Solving the multiple instance problem with axis-parallel rectangles. Artif. Intell. 89, 31–71.

Feng, J., Zhou, Z.-H., 2017. Deep MIML network. In: AAAI'17.

Hashimoto, N., Fukushima, D., Koga, R., Takagi, Y., Ko, K., Kohno, K., Nakaguro, M., Nakamura, S., Hontani, H., Takeuchi, I., 2020. Multi-scale domain-adversarial multiple-instance CNN for cancer subtype classification with unannotated histopathological images. In: CVPR. pp. 3852–3861.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: CVPR. pp. 770–778.

Ilse, M., Tomczak, J., Welling, M., 2018. Attention-based deep multiple instance learning. In: ICML. pp. 2127–2136.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. CoRR abs/1412.6980.

Lerousseau, M., Vakalopoulou, M., Classe, M., Adam, J., Battistella, E., Carré, A., Estienne, T., Henry, T., Deutsch, E., Paragios, N., 2020. Weakly supervised multiple instance learning histopathological tumor segmentation. In: MICCAI. pp. 470–479.

Li, B., Keikhosravi, A., Loeffler, A.G., Eliceiri, K.W., 2021a. Single image super-resolution for whole slide image using convolutional neural networks and self-supervised color normalization. Med. Image Anal. 68, 101938.

Li, B., Li, Y., Eliceiri, K.W., 2021b. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: CVPR. pp. 14318–14328.

Li, X., Li, C., Rahaman, M.M., Sun, H., Li, X., Wu, J., Yao, Y., Grzegorzek, M., 2022. A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches. Artif. Intell. Rev. 55, 4809–4878.

Li, W., Nguyen, V.-D., Liao, H., Wilder, M., Cheng, K., Luo, J., 2019. Patch transformer for multi-tagging whole slide histopathology images. In: MICCAI. pp. 532–540.

Li, Y., Ping, W., 2018. Cancer metastasis detection with neural conditional random field. CoRR abs/1806.07064.

Litjens, G., Bandi, P., Ehteshami Bejnordi, B., Geessink, O., Balkenhol, M., Bult, P., Halilovic, A., Hermsen, M., van de Loo, R., Vogels, R., et al., 2018. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. GigaScience 7, giy065.

Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G.E., Kohlberger, T., Boyko, A., Venugopalan, S., Timofeev, A., Nelson, P.Q., Corrado, G.S., et al., 2017. Detecting cancer metastases on gigapixel pathology images. CoRR abs/1703.02442.

Liu, G., Wu, J., Zhou, Z.-H., 2012. Key instance detection in multi-instance learning. In: ACML. p. 2012.

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

Madabhushi, A., Lee, G., 2016. Image analysis and machine learning in digital pathology: Challenges and opportunities. Med. Image Anal. 33, 170–175.

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

Myronenko, A., Xu, Z., Yang, D., Roth, H.R., Xu, D., 2021. Accounting for dependencies in deep learning based multiple instance learning for whole slide imaging. In: MICCAI. pp. 329–338.
Network, C.G.A.R., et al., 2012. Comprehensive genomic characterization of squamous cell lung cancers. Nature 489, 519.
Pappas, N., Popescu-Belis, A., 2014. Explaining the stars: Weighted multiple-instance learning for aspect-based sentiment analysis. In: EMNLP. pp. 455–466.
Pinheiro, P.O., Collobert, R., 2015. From image-level to pixel-level labeling with convolutional networks. In: CVPR. pp. 1713–1721.
Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: CVPR. pp. 652–660.
Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al., 2021. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Adv. Neural Inf. Process. Syst. 34, 2136–2147.
Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.-A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., et al., 2017. Gland segmentation in colon histology images: The glas challenge contest. Med. Image Anal. 35, 489–502.
Tu, M., Huang, J., He, X., Zhou, B., 2019. Multiple instance learning with graph neural networks. CoRR abs/1906.04881.
Wang, X., Yan, Y., Tang, P., Bai, X., Liu, W., 2018. Revisiting multiple instance neural networks. Pattern Recognit. 74, 15–24.
Wu, L., Wang, Y., Gao, J., Li, X., 2018a. Where-and-when to look: Deep siamese attention networks for video-based person re-identification. IEEE Trans. Multimed. 21, 1412–1424.
Wu, L., Wang, Y., Li, X., Gao, J., 2018b. Deep attention-based spatially recursive networks for fine-grained visual recognition. IEEE Trans. Cybern. 49, 1791–1802.
Yao, J., Zhu, X., Jonnagaddala, J., Hawkins, N., Huang, J., 2020. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Med. Image Anal. 65, 101789.
Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y., 2022. DTFD-MIL: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: CVPR. pp. 18802–18812.
Zhao, Y., Yang, F., Fang, Y., Liu, H., Zhou, N., Zhang, J., Sun, J., Yang, S., Menze, B., Fan, X., et al., 2020. Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution. In: CVPR. pp. 4837–4846.
Zheng, Y., Gindra, R.H., Green, E.J., Burks, E.J., Betke, M., Beane, J.E., Kolachalama, V.B., 2022. A graph-transformer for whole slide image classification. IEEE Trans. Med. Imaging 41 (11), 3003–3015.
Zhu, W., Lou, Q., Vang, Y.S., Xie, X., 2017. Deep multi-instance networks with sparse label assignment for whole mammogram classification. In: MICCAI. pp. 603–611.

Jing He received the Ph.D. degree from the School of Computer Science and Engineering, University of Electronic Science and Technology, China, 2016. She is currently an associate professor in the School of Software, Yunnan University. Her research interests include medical image processing, multimodal data analysis and recommendation system.

Ping Wang, a Master's graduate from the School of Software at Yunnan University, major in Software Engineering. Her research interests are focused on pathological image analysis.

Jingwen Cai, a Master's student at the School of Software at Yunnan University, majoring in Software Engineering. Her research interests focus on pathological image analysis and multimodal emotion analysis.

Dan Tang, a postgraduate student majoring in Software Engineering in the School of Software at Yunnan University. She focuses on the field of medical image processing and conducts in-depth research on the impact of image staining on image classification.

Shaowen Yao is a Professor at the School of Software, Yunnan University, China. He received his BS and MS degrees in telecommunication engineering from the Yunnan University in 1988 and 1991, respectively, and his PhD degree in computer application technology from University of Electronic Science and Technology of China (UESTC) in 2002. His current research interests include neural network theory and applications, cloud computing and big data computing.

Renyang Liu earned his B.E. in Computer Science from Northwest Normal University in 2017 and his Ph.D. from Yunnan University in 2024. He served as a Joint-training Ph.D. student at the College of Computing and Data Science, Nanyang Technological University, from 2022 to 2023. From 2023 to 2024, he worked as a Research Intern at the School of Cyber Science and Engineering, Sun Yat-sen University. Currently, he is a Research Fellow at the Institute of Data Science, National University of Singapore. His research primarily focuses on security aspects of large language models (LLMs) and large multimodal models (LMMs), with interests spanning AI security, data privacy, and computer vision.