



The devil is in the details: a small-lesion sensitive weakly supervised learning framework for prostate cancer detection and grading

Zhongyi Yang^{1,2} · Xiyue Wang³ · Jinxi Xiang² · Jun Zhang² · Sen Yang² · Xinran Wang⁴ · Wei Yang² ·
Zhongyu Li¹ · Xiao Han² · Yueping Liu⁴

Received: 27 June 2022 / Revised: 10 January 2023 / Accepted: 26 January 2023 / Published online: 23 February 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Prostate cancer (PCa) is a significant health concern in aging males, and the diagnosis depends primarily on histopathological assessments to determine tumor size and Gleason score. This process is highly time-consuming, subjective, and relies on the extensive experience of the pathologists. Deep learning based artificial intelligence shows an ability to match pathologists on many prostate cancer diagnostic scenarios. However, it is easy to make mistakes on some hard cases with small tumor areas considering the extensively high-resolution of whole slide images (WSIs). The absence of fine-grained and large-scale annotations of such small tumor lesions makes this problem more challenging. Existing methods usually perform uniform cropping of the foreground of WSI and then use convolutional neural networks as the backbone network to predict the classification results. However, cropping can damage the structure of tiny tumors, which affects classification accuracy. To solve this problem, we propose an Intensive-Sampling Multiple Instance Learning Framework (ISMIL), which focuses on tumor regions and improves the recognition of small tumor regions by intensively sampling the crucial regions. Experiments of prostate cancer detection show that our method achieves an area under the receiver operating characteristic curve (AUC) of 0.987 on the PANDA sets, which improves recall by at least 33% with higher specificity over the current primary methods for hard cases. The ISMIL also demonstrates comparable abilities to human experts on the prostate cancer grading task. Moreover, ISMIL has shown good robustness in independent cohorts, which makes it a potential tool to improve the diagnostic efficiency of pathologists.

Keywords Prostate cancer · Computational pathology · Deep learning · Classification

Introduction

Prostate cancer is a severe disease and the second leading cause of cancer death in men [1]. The standard procedure for diagnosing prostate cancer is observing, grading, and

staging biopsy specimens under microscopy [2]. According to the size and stage of the cancerous area, the Gleason score [3] could be evaluated for diagnosis and prognosis. This process is highly time-consuming and relies on the subjective judgment of pathologists, thus leading to inevitable variability and lack of reproducibility.

As digital slide scanners become available, pathological tissue slides can be digitized into high-resolution whole slide images (WSIs). Feature engineering-based computer-aided diagnosis (CAD) systems [4, 5] have also been developed and applied to the diagnostic workflow. In recent years, thanks to computer hardware development, deep learning has been able to rely on very deep networks [6] to extract complex and relevant features. Therefore, deep learning based CAD systems [7–9] have been proposed and applied in detecting and grading of the prostate. Several studies [2, 10, 11] have shown that with the aid of AI, pathologists can make the diagnosis more efficiently and more accurately.

Zhongyi Yang, Xiyue Wang, and Jinxi Xiang contributed equally to this work.

✉ Jun Zhang
xdzhangjun@gmail.com

Zhongyu Li
zhongyuli@xjtu.edu.cn

Yueping Liu
annama@163.com

Extended author information available on the last page of the article.

Although existing methods can achieve high accuracy on most samples, the performance drops significantly when dealing with small tumor lesions which are difficult to distinguish from other samples [12–14]. The reason is that, despite the increasing number of WSIs available, most are annotated with slide-level labels obtained from diagnostic reports (e.g., cancerous versus non-cancerous or Gleason scores) rather than fine-grained patch-level labels. Thus, it is difficult for the model to learn information from weak labels. Besides, the large scale of WSIs makes it challenging to directly input into the convolutional neural networks (CNNs) [6, 15] or vision transformer (ViT) [16, 17]. The current approaches [12, 18–22] cut the WSI uniformly into multiple patches and then feed them sequentially into deep neural networks with weak supervision at the WSI level. These weakly supervised methods are developed based on the multiple instance learning (MIL) paradigm [23].

Traditionally, Feng et al. [24] and Pinheiro et al. [25] used hand-crafted aggregation functions to aggregate the features of all instances (i.e., patches) to represent the whole WSI. Campanella et al. [12] used recurrent neural networks (RNNs) [26] to treat the MIL problem as sequence analysis. However, when the number of instances increases, the RNNs tend to forget the previous information. Using attention mechanisms [27] is the primary and highly interpretable approach. Ilse et al. proposed ABMIL [21], which is the first very effective approach to using the attention mechanism in MIL application. Building on this, Lu et al. [18, 28] separated the feature extraction and

aggregation steps of MIL and distinguished the cancerous tissue from the normal tissue by clustering.

The problem with these approaches is that when a lesioned area in a WSI is cut exactly in half, neither patches contain complete structural information about the diseased tissue (shown as Fig. 1). It leads to the fact that traditional MIL methods are susceptible to poor recall on samples that contain small tumor tissues. An idea that readily comes to mind is that the overlap of adjacent patches would effectively reduce this occurrence, but it would increase the quantity of computation several times. We expect an intensive sampling of only those regions that may contain cancerous tissue to reduce information loss without excessively increasing the quantity of computation.

Motivated by this idea, we developed an Intensive-Sampling MIL framework (ISMIL). ISMIL can adaptively sample WSI to increase the information density in critical regions and reduce the information loss due to cropping. This method first crops the foreground of the WSI uniformly as other Methods. After that, it uses an encoder combining CNNs [6, 15] and Transformer [16, 17, 29], and an attention-based network to find the region of interest and give a rough prediction. Subsequently, this method intensively samples the region of interest and improves the predictions from the previous step. The process is like a real pathologist making a diagnosis.

Our method is primarily used for PCa detection and grading. On the task of prostate cancer detection (i.e., classification of samples as normal or tumor), our model is

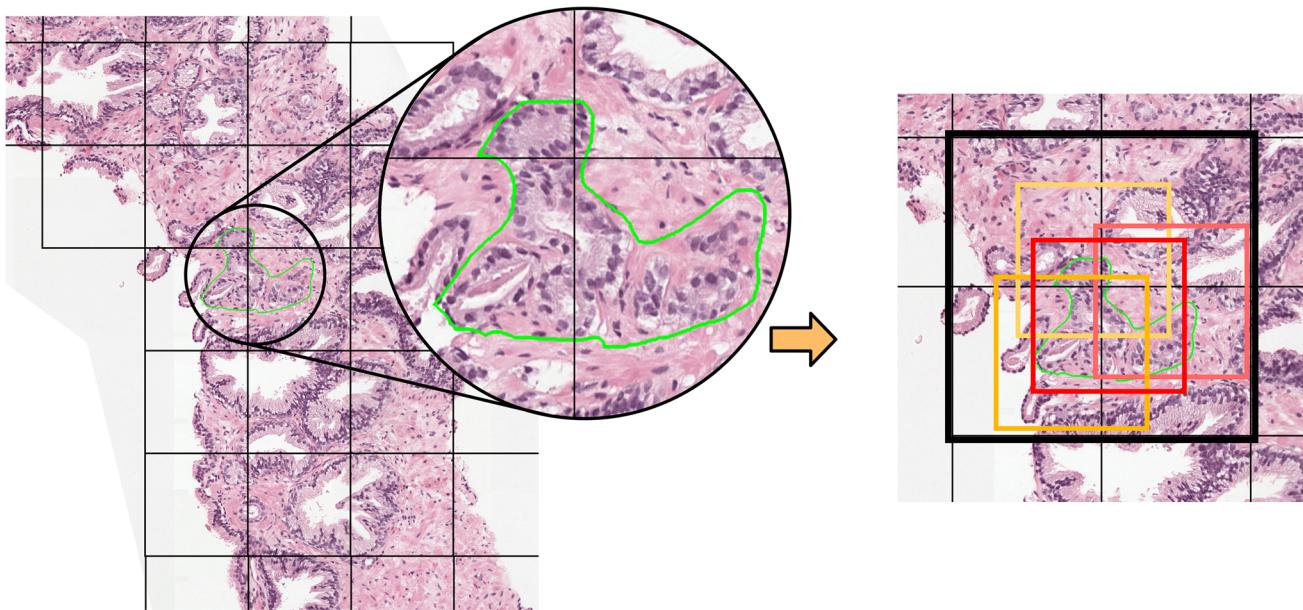


Fig. 1 Schematic diagram of intensively sampling. **Left:** Small cancerous regions are truncated due to uniform cropping of patches. **Right:** Intensively sample the ROI (highlighted in a bold black box) to increase the probability of a complete gland in a single patch. The

thin black box represents the first uniform sampling, and the highlighted color boxes represent four patches obtained from the intensive sampling at the ROI

trained on PANDA [30] dataset (10,613 WSIs of biopsies) with weak labels. It improves the recall rate of small tumor regions by 33% (about 14 points) and the overall AUC in the test set reached 0.987, with a sensitivity and specificity of more than 0.95. Furthermore, on two independent test

cohorts of biopsy specimens (DiagSet-B [31] with 4626 WSIs and HEBEI with 844 WSIs), the models trained on PANDA achieved AUCs of 0.971 and 0.982. Our model also performs well on the PAIP 2021 challenge dataset (50 cancerous WSIs of surgical specimens) and can recall 100%

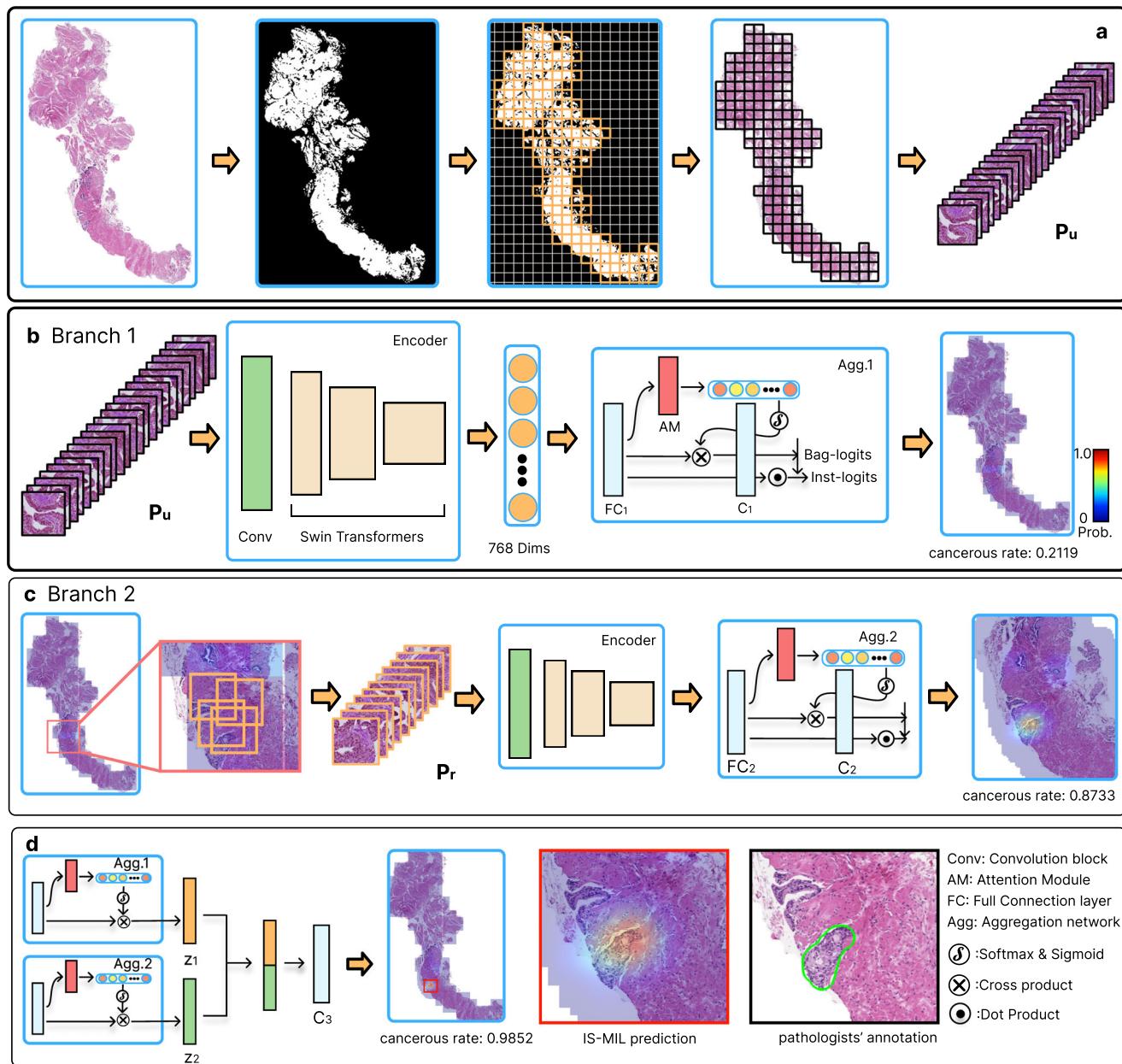


Fig. 2 ISMIL workflow. **(a)**, Separating the foreground of the WSI from the background and then cropping the foreground evenly into patches. **(b)**, The patch bag is fed into a feature extractor consisting of convolutional blocks and Swin Transformer blocks, encoded into 768-dimensional vectors, and the probability that the bag contains cancer and the probability that each patch contains cancer are inferred by an attention-based aggregated classification network. By re-patching the patch back to the WSI, the local probability distribution of containing cancer can be obtained. This process can be regarded as a traditional MIL method, and although it does not directly detect cancer

in hard cases, it provides a reference for intensive sampling. **(c)**, Based on the local probability distribution obtained from the previous step, ISMIL intensively resamples the focal region and infers it to obtain a new cancer-containing probability distribution for the focal region. **(d)**, Finally, the results of the two steps are concatenated and fed into another classifier to obtain the final detection results. The sample in the figure was predicted to be non-cancerous in the first step (traditional attention-based MIL), but tiny tumors were found after further intensive sampling of suspicious areas. Thus, ISMIL has an advantage in dealing with hard cases containing small tumors

of cancer samples. In addition, our method can also be used to predict Gleason scores. With only ISUP labels on the PANDA dataset, ISMIL can achieve $k_{quad} = 0.860$ (i.e., quadratically weighted Cohen's kappa [32]) on the test set, which means that our model reaches a level similar to human experts [33]. Moreover, ISMIL infer faster than CLAM [18] (15.50s vs. 45.34s for generating heatmap of biopsy specimens). With the help of ISMIL, the diagnostic efficiency of pathologists could be potentially improved.

Materials and methods

Overview

In this work, we propose a weakly supervised learning framework called ISMIL to automatically detect tumor tissue in prostate WSI with an attention mechanism and an intensive sampling trick. We used four datasets (PANDA [30], DiagSet-B [31], HEBEI, and the prostate dataset from the PAIP 2021 challenge) to validate the performance of the ISMIL. The overall schematic diagram of the ISMIL is shown in Fig. 2, which is formulated using a two-stage approach. The first stage performs feature extraction, in which the foreground of a WSI is first cropped into several patches, and then these patches are converted into feature vectors by a well-trained feature extractor. The second stage performs feature aggregation based on two attention-based branches (Branch 1 and Branch 2). Branch 1 takes these uniformly cropped patches as input to obtain the preliminary tumor prediction and the distribution of tumor regions. Based on the regions with high attention scores predicted in Branch 1, Branch 2 resamples these regions using our intensive sampling strategy and then predicts a new tumor distribution. Finally, our network combines the outputs of the two branches to give a final prediction. The details of the data, methods, and experiments are described below.

Dataset

The PANDA dataset [30] is used to train the model because of its breadth and diversity. This dataset is one of the largest public datasets of PCa, including 10,613 available WSIs of digitized biopsies stained with hematoxylin and eosin (H&E) originating from two centers. These WSIs are scanned with 3DHISTECH, Hamamatsu Photonics, and Leica Biosystems and have a maximum resolution of 20 \times (0.5 $\mu\text{m}/\text{pixel}$). ISUP grade for each WSI is also provided in the dataset, and its correspondence with the Gleason score can be referred to the WHO standard [34]. Three cohorts from different sources are used for independent validation. DiagSet-B [31] containing 4626 available WSIs scanned using Hamamatsu C12000-22 digital slide scanners, which

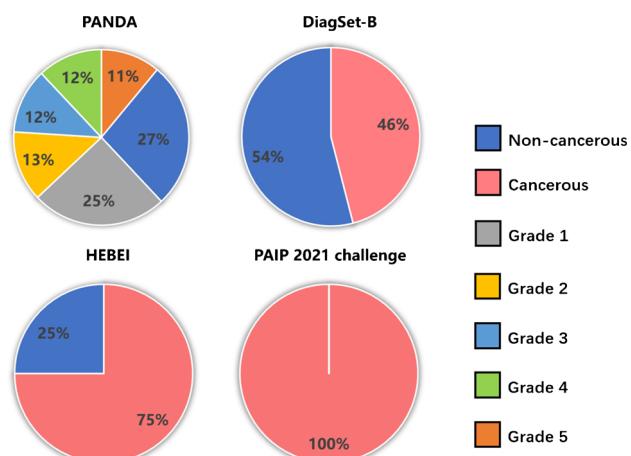


Fig. 3 Label distribution of datasets: The PANDA dataset has credible Gleason grading labels, which are represented by ISUP grading; the Diagset-B and HEBEI datasets have only binary diagnoses, i.e., cancer or non-cancer; all WSIs in the PAIP 2021 challenge dataset are cancerous. The total number of WSIs contained in the four datasets are 10,613, 4626, 844, and 50

are assigned binary labels. HEBEI dataset was provided by The Fourth Hospital of Hebei Medical University, with 844 WSIs scanned with the Leica Aperio system. These WSIs also were given the binary labels extracted from the diagnostic report. The highest magnification level of the WSIs in DiagSet-B and HEBEI dataset is 40 \times (0.25 $\mu\text{m}/\text{pixel}$). All WSIs stained with H&E. Besides, the PAIP challenge 2021 dataset¹ containing 50 prostate surgical specimens with cancer was used to evaluate the performance of the model trained on the biopsy dataset on the surgical specimens. The label distribution of these datasets is shown in Fig. 3.

WSI preprocess

The pathological image of a prostate biopsy specimen always consists of tissue against a background. The pathologist determines whether the gland is cancerous and the grade by observing the color and structure of the gland. It means that the background does not contain any meaningful information, which is harmful, so the first step in preprocessing is removing the background. WSI backgrounds always have high RGB values and brightness, while tissue slides contain more high-frequency signals and show a predictable color distribution. Besides, pathologists always mark the glass slide with a black marker, so areas of low brightness need to be excluded. We designed a dedicated, efficient, and highly accurate segmentation algorithm based on the above characteristics to extract tissue foregrounds. We use the sliding window to crop down

¹<https://paip2021.grand-challenge.org/>

the fixed-size area with a foreground share of more than 15%. The window size was set to $224\text{px} \times 224\text{px}$ at $10\times$ resolution (i.e., $224\mu\text{m} \times 224\mu\text{m}$) for the minor maximum magnification in the three datasets is $20\times$, and structural information of the glands can be preserved at this size.

Feature extraction

The MIL methodology for pathological whole slide images is two-stage, first extracting features with a pre-trained backbone and then classifying them by an aggregation network. If the backbone is trained with WSI labels, it will lead to difficult converging or overfitting because the WSI-level labels are too weak. Therefore, it needs to be trained using other labels or a self-supervised approach. Convolutional neural networks (CNNs) [6, 15, 35, 36] are the most common approach in computer vision to reach state-of-the-art on multiple tasks. In recent years, vision transformer (ViT) [16, 17] has become a popular research direction because of its advantages for learning long-range dependencies. In short, CNN is better at learning the representation of local features, while ViT is better at learning the global features of an image. We used CTransPath [37] as the feature extractor for our weakly supervised classification task. CTransPath has been proven to be a state-of-the-art self-supervised feature extraction in the field of H&E stained histopathological images, which is trained based on the contrast learning paradigm and uses a semantic correlation strategy to construct more positive sample pairs. The training data are from two largest publicly available histopathological image datasets (TCGA² and PAIP³). These WSIs are cropped into around 15 million unlabeled patches. CTransPath combines a convolutional neural network and a transformer model to perceive local and global information in images. The fixed input size of the feature extractor is 224×224 , which is why we set the patch size to 224×224 . The feature extractor converts the patch image into instance features, a 768-dimensional vector x . This process is expressed by the equation

$$x = \text{Encoder}(\mathcal{P}) \quad (1)$$

where \mathcal{P} for a single patch of a WSI.

Aggregation methods

The aggregation and classification of instance features can be expressed like

$$\text{Logits}(B) = c\{g(x_0, x_1, \dots, x_i)\} \quad (2)$$

²<https://portal.gdc.cancer.gov/>

³<http://www.wisepaip.org/paip>

where $\text{Logits}(B)$ represents the prediction of the set of instances from the same WSI, c is a classifier, and most importantly, g is an aggregation method for instance features x_i . Here, g is for the attention module, that is,

$$z = \sum_{k=0}^K a_k x_k \quad (3)$$

where z can be defined as bag-features and

$$a_k = \frac{\exp \{W^T \tanh (V x_k^T)\}}{\sum_{j=1}^K \exp \{W^T \tanh (V x_j^T)\}} \quad (4)$$

where $W \in \mathbb{R}^{L \times 1}$, $V \in \mathbb{R}^{L \times M}$. The attention mechanism is widely used for feature aggregation [18, 21] as a good interpretable module of learnable parameters. The module assigns weights to each instance and weighted sums them as a feature vector representing the bag. In the end, the fused features are fed to the classifier to predict the classification labels. The attention module can be viewed as another binary classifier. Instance features are fed directly into the attention module and classification layer, and the predictions are mapped to probability distributions between 0 and 1 using either *sigmoid* or *softmax*. Then, multiplying them, the result $p(x_i)$ can represent the cancer probability of instances with high robustness. The process is expressed as

$$p(x_i) = \text{sigmoid} \left\{ w^T \tanh \left(V x_k^T \right) \right\} \text{softmax} \{c(x_i)\} \quad (5)$$

c_1, c_2, c_3 are the classifiers of Branch 1, Branch 2, and final the classifier, which is expressed as

$$c(x_i)_n = W_n X_i^T, \quad n \in \{1, 2, 3\}, \quad W_n \in \mathbb{R}^{L \times N} \quad (6)$$

where N is a number of categories.

The proposed method, ISMIL, has two functions: (1) predict the probability that the WSI contains cancer; (2) predict the heatmap to locate the tumor if the WSI contains cancer. To make the prediction more accurate, ISMIL uses two branches and divides the inference into two stages.

In the first stage, Branch 1 first scans the WSI uniformly and a preliminary cancer probability distribution. The process is shown in Fig. 2b. In this step, the probability that the uniformly cropped patch \mathcal{P}_u contains a lesion is $p_u = p\{\text{Encoder}(\mathcal{P}_u)\}$ according to Eq. 5.

Then, in the second stage, the k patches with the highest probability of containing tumors are found according to the probability distribution and intensively resampled around them according to the coordinates. The sampling method is to sample a given coordinate around it n times and to ensure that the overlap of the patch obtained from the n samples is more than τ . The resampled patches \mathcal{P}_r are again

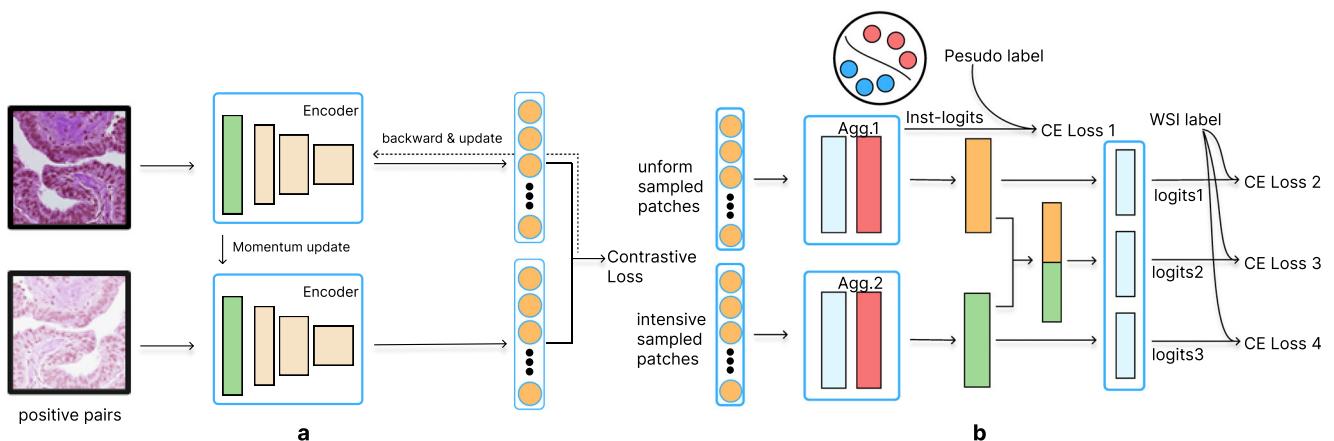


Fig. 4 ISMIL training method. **(a)** We use the CTransPath as the encoder that is trained by a self-supervised contrast learning method. The contrast loss aims to pull the positive pair together and push negative pairs away. The top encoder is updated by gradient descent and the bottom one is updated momentarily based on the parameters of the top one. **(b)** The aggregator of ISMIL consists of two aggregation blocks, the first aggregator uniformly sampled instance features and the second aggregator intensively sampled instance features. Moreover, for

converted into features by the encoder and then fed into the aggregation network Branch 2 as shown in Fig. 2c. That is

$$\mathcal{P}_r = \text{Resample} \{ \text{coords}[\text{argmax}(p_u, k)], n \} \quad (7)$$

Branch 1 and Branch 2 each output aggregated features $z_1 = \sum a_1 x_1$ and $z_2 = \sum a_2 x_2$. Then z_1 and z_2 are concatenated and fed into the final classifier c_3 (Fig. 2d). $\text{Logits}(B)$ is given by:

$$\text{Logits}(B) = c_3\{\text{concat}(z_1, z_2)\} \quad (8)$$

With the two-stage approach described above, ISMIL can correctly predict the probability that a WSI contains a tumor and has a good recall for including microscopic tumors that are difficult to detect.

Training method

Our model mainly comprises one feature extractor (i.e., encoder) and two aggregation branches (Branch 1 and Branch 2), which is trained using gradient descent methods [38]. CTransPath [37] is used as our feature encoder and its code and weights are available at the author's website.⁴ For details of aggregation branches training, Branch 1 is trained using non-overlapped patches uniformly cropped from the WSI foreground, which calculates a cancer probability for each patch. The set of patches with high cancer probability is used as our ROI, and Branch 2 is trained based on those patches resampled with a 50% overlap from the ROI. The training method of ISMIL is shown in Fig. 4.

the first aggregation network, we use clustering to generate instance-level pseudo-labels to train its ability to classify patches (CE loss 1). The CE loss 2 and The CE loss 4 are used as two assisted classification losses for WSI prediction in the first and second aggregation branches, respectively. The CE loss 3 combines the features in both branches to drive the WSI prediction, which is taken for the final prediction in the inference stage

Theoretically, both branches also have enough information to determine whether the WSI is cancerous. Thus, the WSI label can be used both as a label for the final prediction and as an auxiliary signal to supervise the training process of the model. All three classifiers use WSI-level labels with cross-entropy to calculate the loss. Inspired by Lu et al.'s study [18], we trained Branch 1 semi-supervised by generating pseudo-labels. We regard the k instances with the highest probability of positivity in the sample with cancer as positive and the k instances with the lowest probability as negative to generate pseudo-labels and then constrain them with cross-entropy loss. All the losses can be expressed by

$$\mathcal{L}_i = -y_i * \log(p_i) - (1 - y_i) * \log(1 - p_i) \quad (9)$$

where y_2 to y_4 are WSI-level classification labels and y_1 are pseudo-labels. p_i is the probability of the output of the corresponding classifier. \mathcal{L}_1 to \mathcal{L}_4 are the four correspondent loss values and the total loss is

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_2 + \lambda_3 \mathcal{L}_3 + \lambda_4 \mathcal{L}_4 \quad (10)$$

$\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are factors of loss value. Our model is trained using the Adam optimizer [39] with an initial learning rate of 1e-3 and a weight decay of 1e-5. Cosine annealing with a maximum number of 30 iterations is utilized to adjust the learning rate. We choose the best checkpoint based on the loss of the validation set as the final model for testing. A balanced random sampling strategy is used to reduce the impact of imbalanced data distribution.

⁴<https://github.com/Xiyue-Wang/TransPath>

Experimental results

Experimental setups

The proposed method is mainly used for two processes in prostate diagnosis: task 1, tumor detection, and task 2, tumor grading.

Task 1: Tumor detection To test the performance of our model on the tumor detection task, we stratified the PANDA dataset by sampling it according to its labels and sources, then divided it into the training set, validation set, and test set in the proportion of 7:1:2. We treated the tumor detection task as a binary classification task and trained the binary classification model using ISMIL with binary classification labels (with cancer vs. without cancer). The models are trained to converge by observing their performance on the validation set during training, and the model that performs best on the validation set is selected. The models were tested on the PANDA test set, DiagSet-B, HEBEI, and PAIP datasets, and we evaluated their performance using F1-score, AUC, Sensitivity (SEN, i.e., Recall), and Specificity (SPE). We trained several current major models using the same criteria and process and compared their performance. All experiments were repeated five times with different random seeds, and the final performance metrics were taken as their mean values to reduce the random factor.

Hard cases definition Since the difference in model performance is mainly in predicting hard cases, we selected a subset of hard cases from the three test sets for further experiments. According to Beluch et al. [40], ensemble learning can be used to calculate the uncertainty of model predictions and therefore can be used to filter difficult samples. Therefore, we used model screening with secondary confirmation by pathologists to identify hard cases. First, we trained five third-party models (five ABMIL models with different initial parameters trained on randomly sampled subsets of the PANDA training set). We used these trained models to infer the three test sets (PANDA test set, DiagSet-B, and HEBEI). Then, we calculate the variance of the five model predictions for each test sample, which can reflect the degree

of inconsistency among these models. We screened some samples with highly inconsistent predictions ($\text{Std} > 0.05$). Next, we invited three pathologists to relabel these selected samples. Finally, we define the samples for which the pathologists gave inconsistent labels as hard cases, which are 152 hard cases from the PANDA test set (71 with cancers and 82 normal), 131 hard cases from DiagSet-B (83 with cancer and 48 normal), and 47 hard cases from the HEBEI cohort (16 with cancer and 31 normal). Following observation and statistics, we found that these cancerous hard cases had small tumor areas, all less than $700 \mu\text{m}$ in diameter with a median of $138 \mu\text{m}$. Figure 5 shows some hard cases of examples that are characterized by their small size and are easy to ignore. We compare the performance differences of the models on these hard case sets in the same way.

Task 2: Tumor grading Pathologists characterize tumors into different Gleason grades based on the histological structure of the tumor tissue [41]. Biopsy specimens were divided into one of six groups (ISUP) [42–44] based on the distribution of Gleason grades. In this work, tumor grading was performed as a six-classification task (ISUP 0 to 5). We used quadratic weighted kappa (k_{quad}) [32] as an evaluation metric as suggested by [30].

Finally, we tested the time overhead of the above model for inference and generating visual heat maps.

In the experiments, the hyperparameters n is set to 4, k is set to 3, τ is set to 0.5, $\lambda_1, \lambda_2, \lambda_3$ and λ_4 are set to 0.25. The classification threshold for all tasks is set to 0.5. We use 8 Nvidia Tesla V100 GPUs for training and evaluating the models, and test the time consumption for model inference, and generating heat maps on a personal computer (CPU for Intel i7-10700k, single GPU for Nvidia RTX 3070).

Results

Comparison of experimental results in the tumor detection task We compared our method with three main weakly supervised methods based on the PANDA dataset, which are shown in Table 1 and Fig. 6. It is seen that ISMIL leads in all metrics. In particular, for the hard cases subset,

Fig. 5 Examples of hard cases. Two hard cases are shown, where the black boxes indicate the corresponding local areas in the biopsy tissue. The green line is the outline of the cancerous area as outlined by the pathologist. As can be seen, the tumor areas in the hard cases are small and indistinguishable

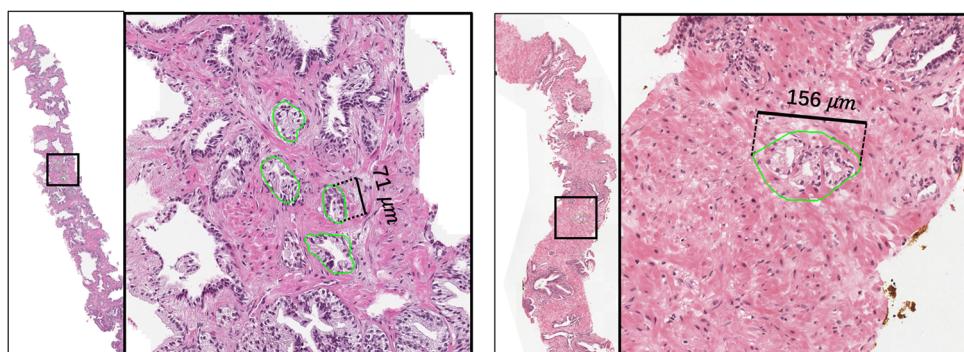


Table 1 The results of comparison experiments on PANDA dataset (a binary classification)

Dataset	Method	F1-score	AUC	SEN	SPE
Full test set	CLAM [18]	0.9205 ± 0.0029	0.9500 ± 0.0014	0.8921 ± 0.0084	0.8789 ± 0.0124
	ABMIL [21]	0.9188 ± 0.0070	0.9535 ± 0.0012	0.8779 ± 0.0180	0.9136 ± 0.0211
	DSMIL [19]	0.9160 ± 0.0032	0.9387 ± 0.0050	0.8998 ± 0.0143	0.8294 ± 0.0484
	ISMIL	0.9654 ± 0.0026	0.9869 ± 0.0008	0.9506 ± 0.0036	0.9512 ± 0.0058
Hard case set	CLAM [18]	0.4383 ± 0.0270	0.5631 ± 0.0089	0.3661 ± 0.0344	0.7358 ± 0.0200
	ABMIL [21]	0.4072 ± 0.0512	0.5605 ± 0.0168	0.3267 ± 0.0566	0.7654 ± 0.0302
	DSMIL [19]	0.4635 ± 0.0418	0.5460 ± 0.0171	0.4253 ± 0.0725	0.6518 ± 0.0778
	ISMIL	0.6339 ± 0.0393	0.7532 ± 0.0314	0.5690 ± 0.0633	0.8049 ± 0.0343

The bolded items represent the highest value in each column

ISMIL leads other methods by more than 33% in sensitivity metrics at higher specificity. The majority of these samples

in the PANDA dataset contain small regions of tumor cells, which suggests that our method has superior advantages in recalling small regions of cancer tissue and thus can improve the overall accuracy.

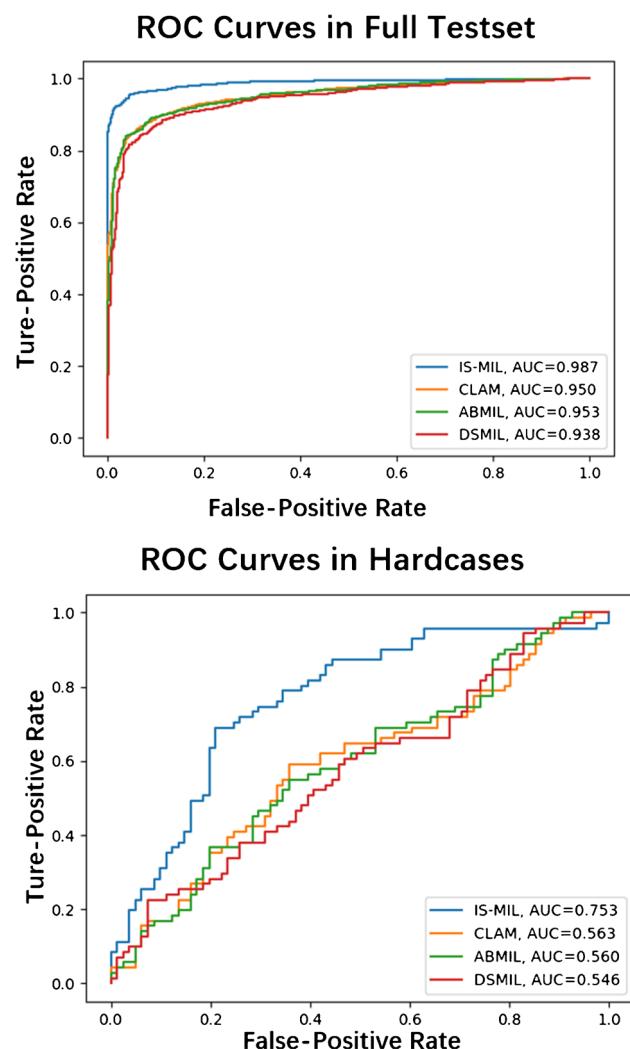


Fig. 6 Comparison of the ROC curves of different methods on the tumor detection task of the PANDA test set. **Top:** comparison on the full test set. **Bottom:** comparison on a subset where most samples are hard cases

Results of ablation experiment in the tumor detection task
We used a B1-only model that only retains Branch 1 (B1) as a control group (i.e., baseline). B2-only is a model using only Branch 2 (B2) predictions, where Branch 1 is only used to output the cancer area probability distribution and Branch 2 uses the additional patch inference obtained from densely sampled ROIs. B1+B2 model used the full ISMIL. The test results of the three models on the PANDA test set and their corresponding hard case subset are shown in Table 2. It is seen that both B1-only and B2-only models achieve inferior performances to the full ISMIL. And the test results on a subset of hard cases show that B2-only has higher performance relative to B1-only for these hard samples.

In addition, we compare our intensive sampling method with the global dense sampling (Global IS) strategy Global IS refers to dense sampling when patches are obtained from WSIs, i.e., there is an overlap between all adjacent patches (50% overlap). It is implemented using Branch 1 (equivalent to CTransPath + CLAM) for training and inference. The results show that the Global IS strategy is less accurate and more time-consuming than ISMIL (refer to Table 4). When tested on the PANDA test dataset, the performance of Global IS was worse than Branch 1. The reason for this is that the Global IS strategy increases the number of patches and therefore increases more normal/noisy patches within each positive WSI, which may dilute the cancer signal and reduce the sensitivity of cancer detection. In contrast, our method benefits from focused sampling, so the input to the network has a higher signal-to-noise ratio.

However, ISMIL did not show a significant improvement over the baseline tested on the entire test set compared to

Table 2 The results of ablation experiments on PANDA dataset (a binary classification task)

Dataset	Method	F1-score	AUC
Full test set	B1 only	0.9575 ± 0.0023	0.9824 ± 0.0007
	B2 only	0.9560 ± 0.0038	0.9832 ± 0.0012
	Global IS	0.9570 ± 0.0021	0.9818 ± 0.0008
	B1 + B2	0.9654 ± 0.0026	0.9869 ± 0.0008
Hard case set	B1 only	0.4750 ± 0.0634	0.6455 ± 0.0430
	B2 only	0.5303 ± 0.0497	0.7313 ± 0.0139
	Global IS	0.6285 ± 0.0180	0.7424 ± 0.0054
	B1 + B2	0.6339 ± 0.0557	0.7532 ± 0.0314

The bolded items represent the highest value in each column

the experimental results tested on the hard cases. In the ablation experiment, the F1-score of B1+B2 improved by 0.8% relative to B1-only. The main reason for this is the very limited number of hard cases in the PANDA test set, accounting for 152 hard samples out of these 2123 test WSIs. On the one hand, Branch 1 (which can be seen as CLAM + CTransPath) can achieve high performance on easy samples, on the other hand, it is difficult to demonstrate the advantages of ISMIL on easy samples. Thus, we additionally supplemented experiments on the hard case set. It can be seen that our ISMIL achieves a much larger performance improvement on the hard case set (around 33% of F1-Score compared with B1-only model).

Results of independent validation in the tumor detection task In specific, our model was trained on the PANDA training set and tested on DiagSet-B and HEBEI for the

prostate cancer detection task. Table 3 shows the results on both the full dataset and the hard case set. Although the metrics are slightly lower compared to the results of the intra-domain test, it still shows better robustness than other methods. These results obtained on the hard cases are also lower than those obtained on the full test sets. ISMIL still shows a significant advantage compared with state-of-the-art methods.

To verify that our method is also applicable to surgical specimen images, we tested the model trained on the PANDA dataset (all biopsy data) using the PAIP 2021 challenge dataset, which contains 50 surgical specimens with prostate cancer. All surgical specimen data were recalled with the classification threshold set to the default value (0.5). We also show that our model trained with biopsy data can work well on WSIs of independent surgical specimens.

Table 3 The results of multi-center independent validation cohorts (a binary classification task)

Dataset	Method	F1-score	AUC
HEBEI	CLAM [18]	0.8519 ± 0.0011	0.9360 ± 0.0088
	ABMIL [21]	0.8624 ± 0.0097	0.9500 ± 0.0016
	DSMIL [19]	0.8550 ± 0.0053	0.9319 ± 0.0099
	ISMIL	0.9079 ± 0.0096	0.9822 ± 0.0020
HEBEI	CLAM [18]	0.5179 ± 0.0098	0.8814 ± 0.0198
	ABMIL [21]	0.5079 ± 0.0000	0.8568 ± 0.0248
	DSMIL [19]	0.5079 ± 0.0000	0.8895 ± 0.0442
	ISMIL	0.6669 ± 0.0490	0.9707 ± 0.0144
DiagSet-B	CLAM [18]	0.7752 ± 0.0127	0.8716 ± 0.0095
	ABMIL [21]	0.7521 ± 0.0228	0.8927 ± 0.0088
	DSMIL [19]	0.7443 ± 0.0539	0.9027 ± 0.0312
	ISMIL	0.8410 ± 0.0273	0.9711 ± 0.0008
DiagSet-B hard cases	CLAM [18]	0.7401 ± 0.0215	0.6126 ± 0.0119
	ABMIL [21]	0.7653 ± 0.0147	0.6371 ± 0.0056
	DSMIL [19]	0.7652 ± 0.0097	0.6341 ± 0.0184
	ISMIL	0.7957 ± 0.0030	0.7348 ± 0.0077

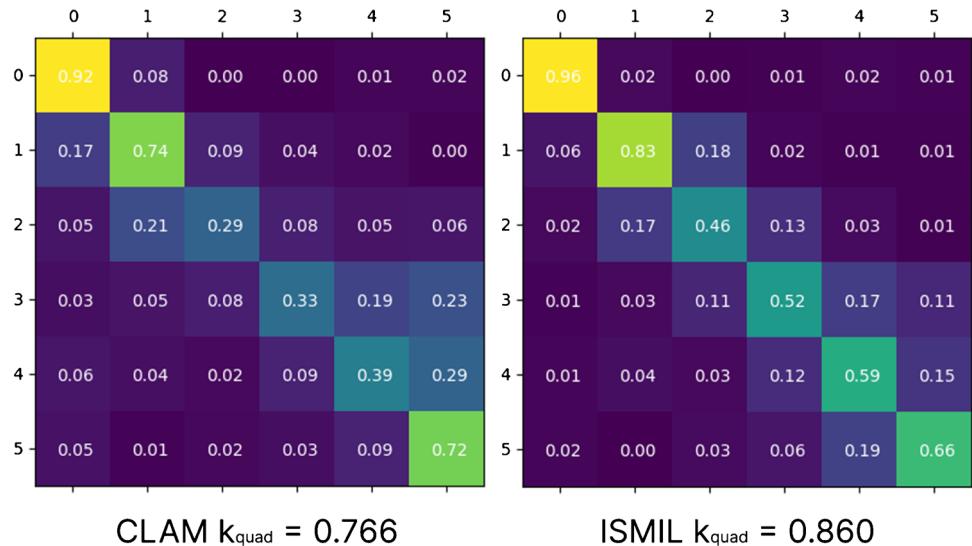
The bolded items represent the highest value in each column

Experimental results in the tumor grading task We used CLAM and ISMIL to train two models and used the quadratically weighted Cohen’s kappa as a metric to evaluate their performance in the ISUP grading task of the PANDA test set. The experimental results are shown in Fig. 7, where CLAM achieves a k_{quad} metric of 0.766 while ISMIL achieves 0.860 on the ISUP grading task. And comparing the confusion matrix shows that our model is better in classifying ISUP 1, ISUP 2, ISUP 3, and ISUP 4. This is because cancer cells of low Gleason-grade are more difficult to distinguish, while our model has better sensitivity.

Tumor visualization ISMIL can draw a heat map of the regional distribution of cancer. Both Branch 1 and Branch 2 can calculate the cancer probability of each patch by Eq. 5, and then the framework reverts the patches to WSI and superimposes the cancer probability to the mean value, and smooth them to obtain the tumor area heat map. As shown in Fig. 8, without relying on local annotation, our method can accurately localize the region of tumor cells. Pathologists can make a quick diagnosis based on this heat map and provide the patients with objective diagnostic evidence.

Time consumption we tested and compared the time consumption of our method with the CLAM method to infer and generate the heat map (overlap of 0.5) on the personal computer, and the experimental results are shown in Table 4. It is seen that our method is faster in inference than the CLAM. This is because, during the visualization phase, CLAM needs to overlap all patches with each WSI, whereas ISMIL only overlaps the focal areas (intensive samples).

Fig. 7 Confusion matrix of model predictions and pathologists’ diagnoses in the ISUP grading task. Brighter colors represent higher consistency. **Left** is the confusion matrix for the prediction of CLAM [18] and the diagnosis of the pathologist, and $k_{quad} = 0.766$. The **right** panel shows the confusion matrix of the predicted results of ISMIL with the pathologist’s diagnosis, and $k_{quad} = 0.860$



Discussion

In this paper, we propose a weakly supervised learning framework called intensive sampling multi-instance learning (ISMIL) to assist pathologists in prostate cancer diagnosis and grading. ISMIL employs a self-supervised pre-trained encoder to extract local features of each WSI and uses an attention-based network to fuse features for the whole WSI representation. ISMIL is divided into two branches. In Branch 1, the foreground of each WSI is uniformly cut into multiple non-overlapped patches, and then the global features and the probability distribution of tumor are outputted. In Branch 2, regions with high probability distribution are further densely sampled (with overlap between patches) to calculate the features of the focal region. Finally, the features of Branch 1 and Branch 2 are concatenated and fed into a classifier to obtain the final prediction results. Meanwhile, a visual heat map can be generated based on the probability distributions of the two branches. In experimental results, we have proved that ISMIL has significant advantages over current methods [18, 19, 21] (especially on hard case sets) and also generates heat maps more efficiently.

Our method uses an attention mechanism to calculate the tumor distribution and then increases the sampling density for regions that may contain tumors. Essentially, our intensive sampling strategy improves the signal-to-noise ratio of the input sample. Recently, some studies have also shown that changing the sampling distribution of images can help improve the network’s performance. Jin et al. [45] proposed a learnable down-sampling module to change the distribution of image samples to improve the segmentation

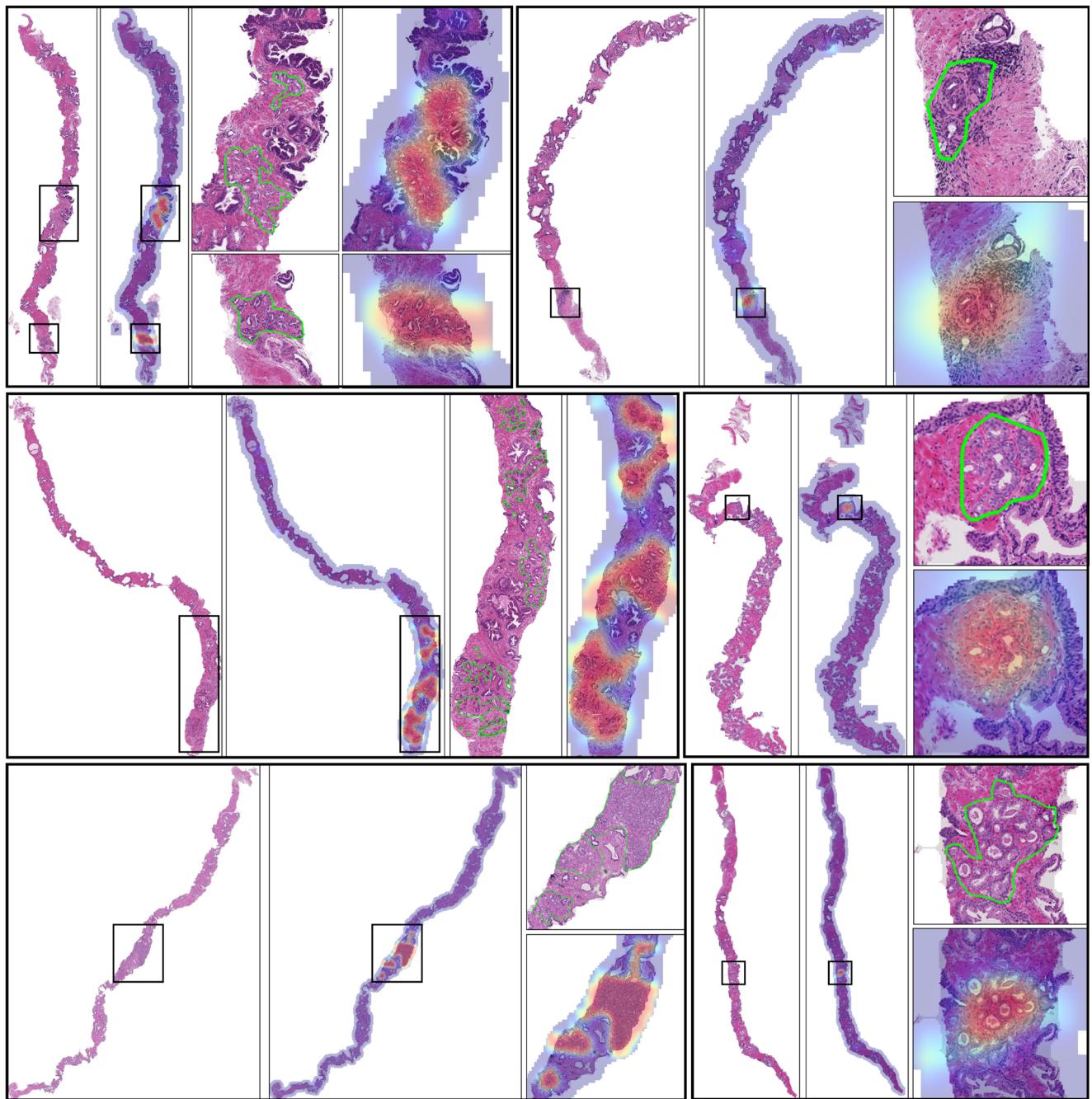


Fig. 8 Heat map of tumor localization. The pathological slides shown are mostly hard samples containing small areas of tumor tissue, which our model was generally able to localize accurately. In each column, the images from left to right represent, in sequence, the original image

of the WSI, a heat map of the entire WSI, and a local image of the tumor region. The green outline is outlined by the pathologist, and the red areas are those predicted by the model to be likely to be cancerous

performance of high-resolution images. Martin et al. [46] used an additional edge detection model to densely sample the edges in the image, improving the balance between accuracy and computational efficiency of the semantic segmentation task. In the field of MIL, Thandiackal et al. proposed ZoomMIL [47], a method that adaptively samples WSI at multiple magnifications. This method achieves high

accuracy with fewer floating-point operations. However, ZoomMIL does not consider the problem of the damage caused by cropping to the image structure. We consider that intensive sampling of key areas can well modify the sampling distribution while avoiding the above problems.

ISMIL has the potential to be applied to practical scenarios of pathology diagnosis [12]. Once the digitalized

Table 4 Time consumption for inference and diagnosis for a single WSI

Stage	Biopsies		Surgical specimens	
	ISMIL	CLAM [18]	ISMIL	CLAM [18]
Preprocessing	1.26s	2.10s	3.48s	3.84s
Inference	11.17s	39.94s	209.04s	493.51s
Heatmap generation and saving	3.06s	3.29s	19.36s	17.60s
Total	15.50s	45.34s	231.88s	514.97s

prostate biopsies and specimens are pre-processed and uploaded to the server of the pathology center, our pathology diagnostic tool can automatically output the probability and type of cancer and generate a tumor distribution using a heat map manner. After that, pathologists could combine AI diagnostic reports with their own pathology knowledge to make more accurate and effective diagnoses. ISMIL also has the potential to be combined with augmented reality microscopy [48–50] to improve the efficiency and accuracy of pathologists' diagnoses while maintaining their diagnostic habits.

This study has several limitations. Our method can generate good tumor localization, but it is difficult to generate fine-grained visual interpretations of cancer subclasses due to weak labeling. In addition, our algorithm does not perform well in Gleason grading. The reason for this may be due to the inherent subjectivity of Gleason grading among pathologists [30, 51]. In the future, our algorithm will be validated in larger and multi-center clinical prostate pathology data, which may help facilitate the establishment of an objective assessment of prostate grading.

Acknowledgements The PAIP dataset used in this research was provided by Seoul National University Hospital by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI18C0316). We also thank Junhan Zhao from Harvard Medical School for helping us complete the comments related to the discussion of clinical applications and polishing our presentation.

Author contribution ZY, XW, JX designed the study. JZ, SY, WY, ZL, XH collected public datasets. XW, YL collected tissue samples and scanned them into WSIs. ZY, XW, JX drafted the manuscript, and all authors approved the final version to be published.

Data availability PANDA, DiagSet-B and PAIP datasets are publicly available (<https://www.kaggle.com/competitions/prostate-cancer-grade-assessment/data>, <https://paperswithcode.com/dataset/diagset>, <http://www.wisepaip.org/paip/>). The other datasets generated and/or analyzed during the current study are available from the corresponding authors on reasonable request.

Code availability Our code will be available online at <https://github.com/YangZyyy/Intensive-sampling-MIL>.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71(3):209–249
- Steiner DF, Nagpal K, Sayres R, Foote DJ, Wedin BD, Pearce A, Cai CJ, Winter SR, Symonds M, Yatziv L et al (2020) Evaluation of the use of combined artificial intelligence and pathologist assessment to review and grade prostate biopsies. JAMA Netw Open 3(11):2023267–2023267
- Delahunt B, Miller RJ, Srigley JR, Evans AJ, Samaratunga H (2012) Gleason grading: past, present and future. Histopathology 60(1):75–86
- Nguyen K, Sabata B, Jain AK (2012) Prostate cancer grading: Gland segmentation and structural features. Pattern Recogn Lett 33(7):951–961
- Doyle S, Feldman MD, Shih N, Tomaszewski J, Madabhushi A (2012) Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer. BMC Bioinforma 13(1):1–15
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
- Arvaniti E, Fricker KS, Moret M, Rupp N, Hermanns T, Fankhauser C, Wey N, Wild PJ, Rueschoff JH, Claassen M (2018) Automated gleason grading of prostate cancer tissue microarrays via deep learning. Sci Rep 8(1):1–11
- Van der Laak J, Litjens G, Ciompi F (2021) Deep learning in histopathology: the path to the clinic. Nat Med 27(5):775–784
- Ayyad SM, Shehata M, Shalaby A, El-Ghar A, Ghazal M, El-Melegy M, Abdel-Hamid NB, Labib LM, Ali HA, El-Baz A et al (2021) Role of AI and histopathological images in detecting prostate cancer: A survey. Sensors 21(8):2586
- Bulten W, Balkenhol M, Belinga J-JA, Brilhante A, Çakir A, Egevad L, Eklund M, Farré X, Geronatsiou K, Molinié V et al (2021) Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists. Modern Pathol 34(3):660–671
- Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammie C, Thng F, Peng L, Stumpe MC (2018) Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. Am J Surg Pathol 42(12):1636
- Campanella G, Hanna MG, Geneslaw L, Mirafiori A, Werneck Krauss Silva V, Busam KJ, Brogi E, Reuter VE, Klimstra DS, Fuchs TJ (2019) Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 25(8):1301–1309
- Wang X, Xiang J, Zhang J, Yang S, Yang Z, Wang M-H, Zhang J, Wei Y, Huang J, Han X (2022) SCL-WC: Cross-slide contrastive learning for weakly-supervised whole-slide image classification. In: Advances in neural information processing systems
- Chen RJ, Lu MY, Weng W-H, Chen TY, Williamson DF, Manz T, Shady M, Mahmood F (2021) Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 4015–4025
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778

16. Kolesnikov A, Dosovitskiy A, Weissenborn D, Heigold G, Uszkoreit J, Beyer L, Minderer M, Dehghani M, Houlsby N, Gelly S et al (2021) An image is worth 16x16 words: Transformers for image recognition at scale
17. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 10012–10022
18. Lu MY, Williamson DF, Chen TY, Chen RJ, Barbieri M, Mahmood F (2021) Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng 5(6):555–570
19. Li B, Li Y, Eliceiri KW (2021) Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 14318–14328
20. 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
21. Ilse M, Tomczak J, Welling M (2018) Attention-based deep multiple instance learning. In: International conference on machine learning, pp 2127–2136. PMLR
22. Sharma Y, Shrivastava A, Ehsan L, Moskaluk CA, Syed S, Brown D (2021) Cluster-to-conquer: A framework for end-to-end multi-instance learning for whole slide image classification. In: Medical imaging with deep learning, pp 682–698. PMLR
23. Carboneau M-A, Cheplygina V, Granger E, Gagnon G (2018) Multiple instance learning: A survey of problem characteristics and applications. Pattern Recogn 77:329–353
24. Feng J, Zhou Z-H (2017) Deep MIML network. In: Proceedings of the AAAI conference on artificial intelligence, vol 31
25. Pinheiro PO, Collobert R (2015) From image-level to pixel-level labeling with convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1713–1721
26. Medsker LR, Jain L (2001) Recurrent neural networks. Des Appl 5:64–67
27. Bahdanau D, Cho KH, Bengio Y (2015) Neural machine translation by jointly learning to align and translate. In: 3rd international conference on learning representations, ICLR 2015
28. Lu MY, Chen TY, Williamson DF, Zhao M, Shady M, Lipkova J, Mahmood F (2021) AI-based pathology predicts origins for cancers of unknown primary. Nature 594(7861):106–110
29. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Adv Neural Inf Process Syst 30
30. Bulten W, Kartasalo K, Chen P-HC, Ström P, Pinckaers H, Nagpal K, Cai Y, Steiner DF, van Boven H, Vink R et al (2022) Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the panda challenge. Nat Med, 1–10
31. Koziarski M, Cyganek B, Olborski B, Antosz Z, Źydak M, Kwolek B, Wasowicz P, Bukala A, Swadźba J, Sitkowski P (2021) DiagSet: a dataset for prostate cancer histopathological image classification. arXiv:2105.04014
32. Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20(1):37–46
33. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, van der Laak J, Hulsbergen-van de Kaa C, Litjens G (2020) Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet Oncol 21(2):233–241
34. Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE (2016) The 2016 WHO classification of tumours of the urinary system and male genital organs—Part B: prostate and bladder tumours. Eur Urol 70(1):106–119
35. Tan M, Le Q (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning, pp 6105–6114. PMLR
36. Wang X, Du Y, Yang S, Zhang J, Wang M, Zhang J, Yang W, Huang J, Han X (2023) RetCCL: clustering-guided contrastive learning for whole-slide image retrieval. Med Image Anal 83:102645
37. Wang X, Yang S, Zhang J, Wang M, Zhang J, Yang W, Huang J, Han X (2022) Transformer-based unsupervised contrastive learning for histopathological image classification. Medical Image Analysis
38. Ruder S (2016) An overview of gradient descent optimization algorithms. arXiv:1609.04747
39. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv:1412.6980
40. Beluch WH, Genewein T, Nürnberger A, Köhler JM (2018) The power of ensembles for active learning in image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 9368–9377
41. Epstein JI (2010) An update of the Gleason grading system. J Urol 183(2):433–440
42. Van Leenders GJ, Van Der Kwast TH, Grignon DJ, Evans AJ, Kristiansen G, Kweldam CF, Litjens G, McKenney JK, Melamed J, Mottet N et al (2020) The 2019 international society of urological pathology (ISUP) consensus conference on grading of prostatic carcinoma. Am J Surg Pathol 44(8):87
43. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA (2016) The 2014 international society of urological pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. Am J Surg Pathol 40(2):244–252
44. Epstein JI, Zelefsky MJ, Sjoberg DD, Nelson JB, Egevad L, Magi-Galluzzi C, Vickers AJ, Parwani AV, Reuter VE, Fine SW et al (2016) A contemporary prostate cancer grading system: a validated alternative to the Gleason score. Eur Urol 69(3):428–435
45. Jin C, Tanno R, Mertzanidou T, Panagiotaki E, Alexander DC (2021) Learning to downsample for segmentation of ultra-high resolution images. In: International conference on learning representations
46. Marin D, He Z, Vajda P, Chatterjee P, Tsai S, Yang F, Boykov Y (2019) Efficient segmentation: Learning downsampling near semantic boundaries. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 2131–2141
47. Thandiackal K, Chen B, Pati P, Jaume G, Williamson DFK, Gabrani M, Goksel O (2022) Differentiable zooming for multiple instance learning on whole-slide images. In: Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T (eds) Computer Vision – ECCV 2022. Springer, pp 699–715
48. Yue M, Zhang J, Wang X, Yan K, Cai L, Tian K, Niu S, Han X, Yu Y, Huang J et al (2021) Can AI-assisted microscope facilitate breast HER2 interpretation? a multi-institutional ring study. Virchows Arch 479(3):443–449
49. Tang H-P, Cai D, Kong Y-Q, Ye H, Ma Z-X, Lv H-S, Tuo L-R, Pan Q-J, Liu Z-H, Han X (2021) Cervical cytology screening facilitated by an artificial intelligence microscope: A preliminary study. Cancer Cytopathol 129(9):693–700
50. Chen P-HC, Gadepalli K, MacDonald R, Liu Y, Kadowaki S, Nagpal K, Kohlberger T, Dean J, Corrado GS, Hipp JD et al (2019) An augmented reality microscope with real-time

- artificial intelligence integration for cancer diagnosis. *Nat Med* 25(9):1453–1457
51. Allsbrook Jr, WC, Mangold KA, Johnson MH, Lane RB, Lane CG, Amin MB, Bostwick DG, Humphrey PA, Jones EC, Reuter VE et al (2001) Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. *Hum Pathol* 32(1):74–80

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Affiliations

Zhongyi Yang^{1,2} · Xiyue Wang³ · Jinxi Xiang² · Jun Zhang²  · Sen Yang² · Xinran Wang⁴ · Wei Yang² · Zhongyu Li¹ · Xiao Han² · Yueping Liu⁴

¹ School of Software Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, China

² Tencent AI Lab, Shenzhen, Guangdong, China

³ College of Computer Science, Sichuan University, Chengdu, Sichuan, China

⁴ Department of Pathology, The Fourth Hospital of Hebei Medical University, Shijiazhuang, Hebei, China