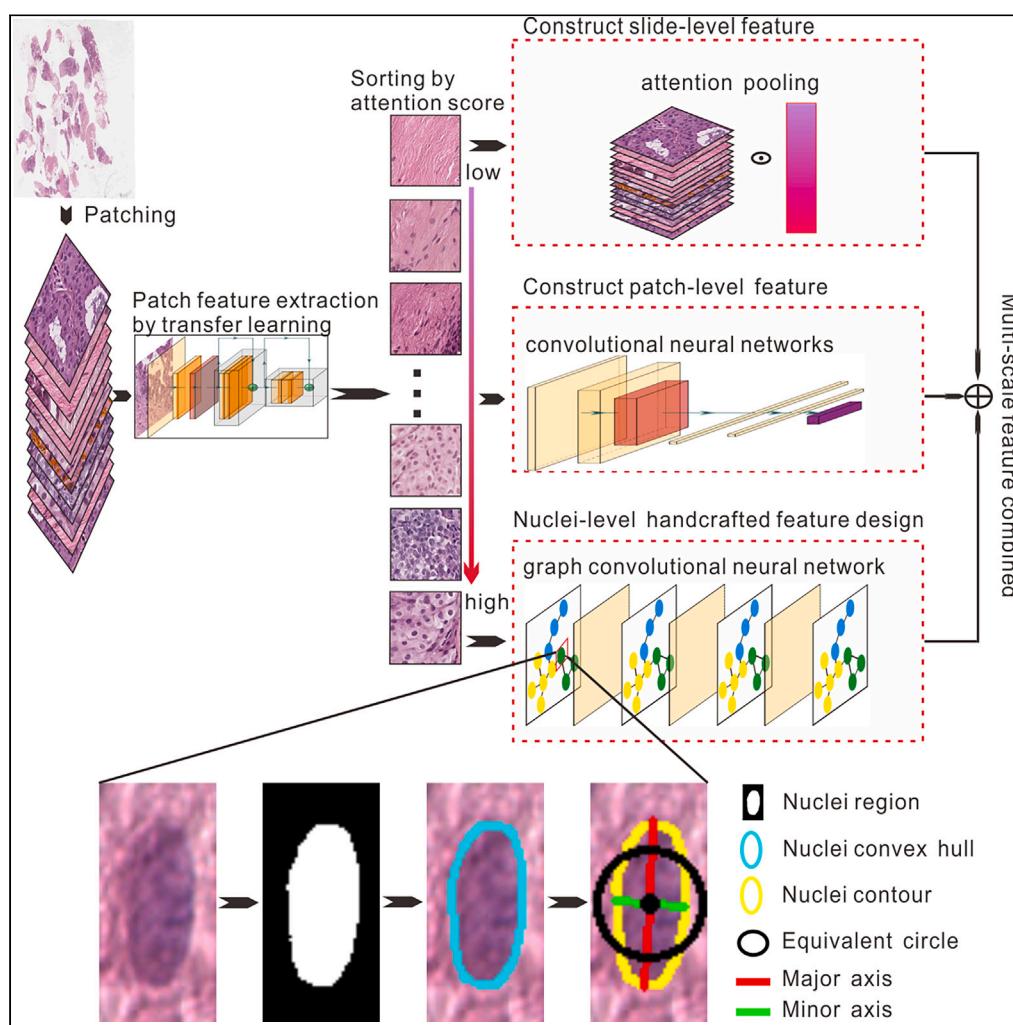


Article

Nuclei-level prior knowledge constrained multiple instance learning for breast histopathology whole slide image classification



Xunping Wang,
Wei Yuan

wei.yuan@tohoku.ac.jp

Highlights
A multiscale multiple instance learning model for whole slide image classification

Integrate prior knowledge into deep learning models to improve interpretability

Expand the analytical dimension of existing whole slide image classification algorithms by introducing nuclei-level features

Wang & Yuan, iScience 27,
109826
June 21, 2024 © 2024 The
Author(s). Published by Elsevier
Inc.
<https://doi.org/10.1016/j.isci.2024.109826>



Article

Nuclei-level prior knowledge constrained multiple instance learning for breast histopathology whole slide image classification

Xunping Wang¹ and Wei Yuan^{2,3,*}

SUMMARY

New breast cancer cases have surpassed lung cancer, becoming the world's most prevalent cancer. Despite advancing medical image analysis, deep learning's lack of interpretability limits its adoption among pathologists. Hence, a nuclei-level prior knowledge constrained multiple instance learning (MIL) (NPKC-MIL) for breast whole slide image (WSI) classification is proposed. NPKC-MIL primarily involves the following steps: Initially, employing the transfer learning to extract patch-level features and aggregate them into slide-level features through attention pooling. Subsequently, abstract the extracted nuclei as nodes, establish nucleus topology using the K-NN (K-Nearest Neighbors, K-NN) algorithm, and create handcrafted features for nodes. Finally, combine patch-level deep learning features with nuclei-level hand-crafted features to fine-tune classification results generated by slide-level deep learning features. The experimental results demonstrate that NPKC-MIL outperforms current comparable deep learning models. NPKC-MIL expands the analytical dimension of WSI classification tasks and integrates prior knowledge into deep learning models to improve interpretability.

INTRODUCTION

In recent years, the incidence of new breast cancer cases has surged rapidly, surpassing lung cancer to become the world's most prevalent cancer type.¹ With the advancement of histological section digitization technology, whole slide image (WSI) offers the advantage of presenting tissue morphology with exceptionally high resolution.² Also, the characteristic of extremely high resolution in WSI makes it possible to classify breast cancer,³ segment lesion areas,⁴ analyze the microenvironment,⁵ and assess prognosis.⁶

The conventional classification process for pathological images adheres to the typical image classification steps: target segmentation, feature extraction, feature selection, and disease diagnosis. These four steps are carried out independently and then integrated for tuning. However, errors are inevitable in each step, and their accumulation can result in unreliable classification outcomes. Additionally, most traditional machine learning algorithms require loading all data into memory at once, which is not feasible for WSIs with gigapixels.⁷

The classification process based on deep learning is end to end.⁸ Thanks to the weight-sharing strategy⁹ and the rapid advancement in computer computing power,¹⁰ deep learning methods have made significant strides in the field of computer vision, propelling advancements in medical image analysis.¹¹

Given the gigantic size of WSI, such as $100,000 \times 100,000$ pixels, the analysis scale of WSI progresses through two stages, from patch level to slide level.

Conventional deep learning methods cannot be directly applied for classifying WSI¹² and must be processed in blocks.^{13–15} Although processing in blocks reduces the algorithm's demands on computer hardware, it disrupts the integrity of the organization and compromises spatial information. The analysis of the entire WSI should fully consider spatial context information rather than being based on arbitrary patches.¹⁶ Therefore, most current studies aggregate patch-level features to the slide-level before making inferences.^{17–19}

At present, the improvement of the deep learning algorithm based on MIL focuses on how to model the spatial relationship of patches to better aggregate patch-level features into slide-level features, while ignoring the influence of nuclear-level features on the final classification results. The distribution and morphology of nuclei are an essential basis for cancer diagnosis.²⁰ Although deep neural networks can theoretically approximate any function model,²¹ it is still not widely recognized by pathologists due to its lack of interpretability.²²

To fully leverage the capacity of deep learning methods in model fitting, while also considering the incorporation of handcrafted features with clear interpretations to enhance the interpretability of the model, this paper intends to investigate a way for classifying WSI constrained by nuclear-level handcrafted features.

¹School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

²Co-Creation Center for Disaster Resilience, International Research Institute of Disaster Science, Tohoku University, Aoba 468-1, Aramaki, Aoba-ku, Sendai 980-8572, Japan

³Lead contact

*Correspondence: wei.yuan@tohoku.ac.jp

<https://doi.org/10.1016/j.isci.2024.109826>



The rest of the paper is arranged as follows: A brief review of some related work is presented below. Detailed ablation experiments and comparison studies are presented in the [results](#) section, followed by the discussion and conclusion part in the [discussion](#) section. Finally, a detailed overview of the methodology is provided in the [STAR methods](#).

The combination of extremely large image sizes and the high expertise threshold in pathology makes it prohibitively expensive to acquire pixel-level WSI labeling data. However, acquiring slide-level labeling is comparatively more accessible. The scenario of 'The class of the entire WSI is known, while the class of all patches that make up the WSI is unknown' aligns precisely with the application of Multiple Instance Learning. In MIL, the training set comprises bags with known class labels, where each bag includes instances with unknown labels. A bag is labeled as positive if it contains at least one positive instance, and if all instances in a bag are negative, the bag is marked as negative.²³

Hence, an increasing number of researchers have approached digital pathological image classification as a weakly supervised classification problem based on the MIL algorithm. Within MIL-based deep learning algorithms, one of the most intuitive ideas is to aggregate patch-level features extracted by Convolutional Neural Network (CNN) into slide-level features through pooling operations.²⁴ Ilse et al.²⁵ take 'the fundamental theorem of symmetric function and the fact that neural network layer has arrangement equivariant'²⁶ as the principle for network design. They propose an attention-based deep MIL algorithm, utilizing the attention mechanism to assign more flexible weights to patches. Li et al.²⁷ innovatively incorporated a non-local attention mechanism to tackle the MIL problem. Through evaluating the similarity between the patch with the highest score and other patches, distinct attention weights were assigned to each patch, thereby optimizing the aggregation of patch-level features into comprehensive slide-level features. This approach aligns with analogous studies reported in existing literature.^{28–30}

Nevertheless, the aforementioned studies are contingent upon the assumption that all instances in each bag are independent and equally distributed. In reality, instances often exhibit correlations,³¹ necessitating the modeling of inter-instance relationships before applying MIL. Shao et al.³² challenge the assumption of independent distribution, proposing a scoring function based on Hausdorff distance to assess correlations between instances. They leverage the Transformer framework³³ to incorporate the spatial topological relationships among similar patches. Philip et al.³⁴ select the c instances with the highest score using the instance classifier. Subsequently, they calibrate the distribution of instances by subtracting the features of these top-scoring c samples from the features of all instances in the bag. However, operating under the assumption that 'cancerous WSI have more information than normal WSI,' their model tends to prioritize features of cancerous WSI, thereby attenuating the recognition of features in normal WSI.

As the bag contains numerous instances (tens of thousands), refining the instances is necessary to reduce the number involved in calculations. Xie et al.³⁵ adopted the feature clustering method to divide the instances in the bag into e categories and then the characteristics of the instance where the centroid of e categories are fused as the characterization of the bag to make the final prediction. Lu et al.³⁶ sorted the instances according to the level of attention scores, and then took the c instances with the highest and lowest attention scores as inputs to train a patch-level classifier to constrain the slide-level diagnosis results, and achieved better results than diagnosis based on slide-level features alone.

In addition, to enrich the diversity of training samples, it is necessary to explore the impact of different data enhancement methods on classification accuracy: Yang et al.³⁷ proposed a "mix-the-bag" data enhancement method, which means adding, replacing, and interpolating among the instances of the bag of the same category to generate new virtual bags. Zhang et al.³⁸ randomly select instances from the parent bag to form sub-bags and assign the label of the parent bag to sub-bags. However, sub-bags derived from positive parent bag may not necessarily contain positive instances, leading to the introduction of sub-bags with incorrect labels. To address this issue, they designed a feature distillation branch to eliminate errors.

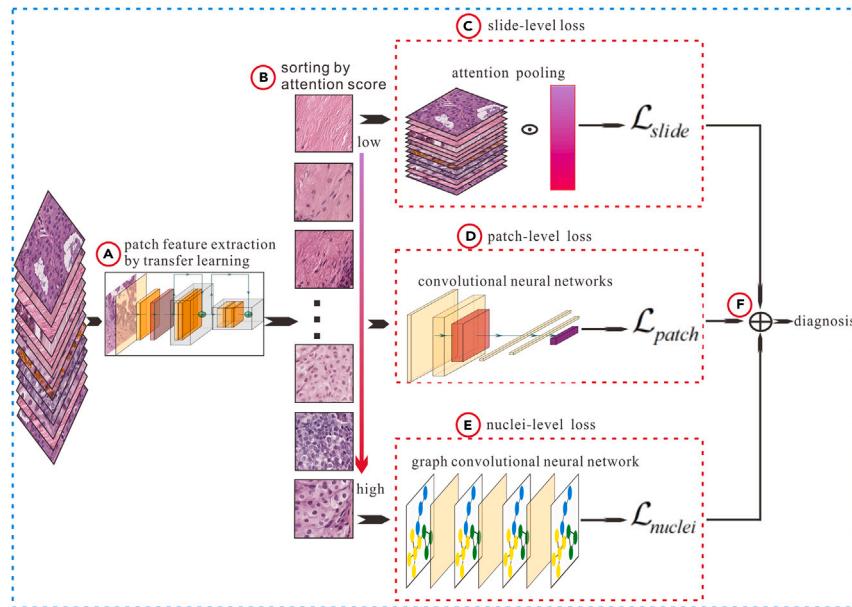
To fully harness the potential of deep learning methods in model fitting and incorporate handcrafted features for building interpretable deep learning models, this paper aims to explore a method for classifying WSI constrained by nuclear-level handcrafted features.

RESULTS

Construction of the NPKC-MIL

The construction process of nuclei-level prior knowledge constrained multiple instance learning (NPKC-MIL) can be summarized as follows: Initially, patch-level features are extracted from WSI and sorted based on the importance of patches. Subsequently, nucleus features (treat extracted nuclei as Graph nodes) are aggregated and updated using the Graph Convolutional Network (GCN). The losses are then calculated at the slide-level, patch-level, and nuclei-level. Finally, patch-level and nuclei-level losses are considered as penalty items and added to the total loss function for model training, as depicted in [Figure 1](#).

[Figure 1](#) outlines the four steps involved in constructing NPKC-MIL: First, utilizing the transfer learning method to extract patch-level features ([Figure 1A](#)) and rank the importance of patches using an attention mechanism ([Figure 1B](#)). Second, aggregating patch-level features into slide-level features through attention pooling and calculating patch-level loss \mathcal{L}_{patch} ([Figure 1C](#)) and slide-level loss \mathcal{L}_{slide} ([Figure 1D](#)) separately. Third, abstracting the extracted nuclei as nodes, constructing the node topology using the K-NN algorithm, extracting nuclei-level handcrafted features, aggregating and updating node information with GCN, and calculating nuclei-level loss \mathcal{L}_{nuclei} ([Figure 1E](#)). Fourth, considering \mathcal{L}_{patch} and \mathcal{L}_{nuclei} as penalty terms, they are added to the total loss function for model training ([Figure 1F](#)).

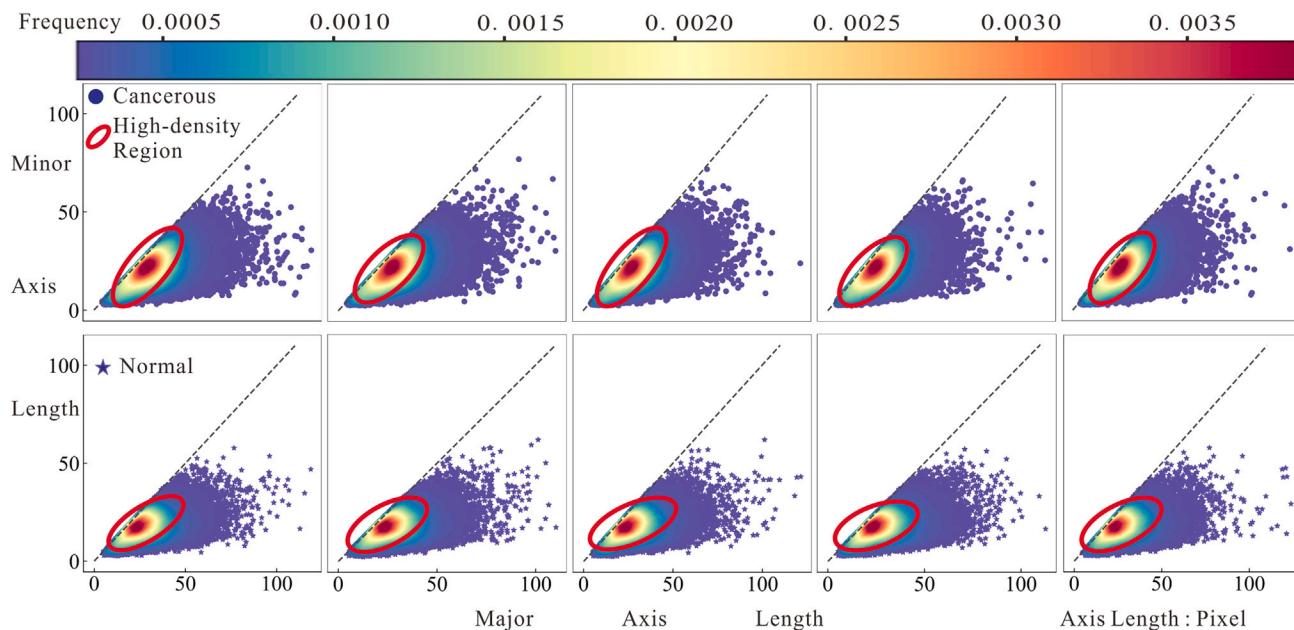
**Figure 1. Flowchart of NPKC-MIL**

- (A) Extracting patch feature by transfer learning.
- (B) Ranking the importance of patches by attention score.
- (C) Calculating slide-level loss.
- (D) Calculating patch-level loss.
- (E) Constructing total loss function.

Nuclei-level feature design

There exists a perceptual understanding that 'cancerous nuclei are larger, rounder, and messier than normal nuclei' in the clinical diagnosis process. To quantitatively articulate these differences and aid in handcrafted feature design, we conducted five randomized trials on selected experimental datasets. Each experiment involved the collection of 100,000 nuclei from both cancerous WSI and normal WSI for statistical analysis.

Figure 2 illustrates the scatter density of normal and cancerous nuclei. A comparison of the high-density region in Figure 2 distinctly reveals that cancerous nuclei exhibit larger areas and a rounder shape.

**Figure 2. Nuclei scatter density map**

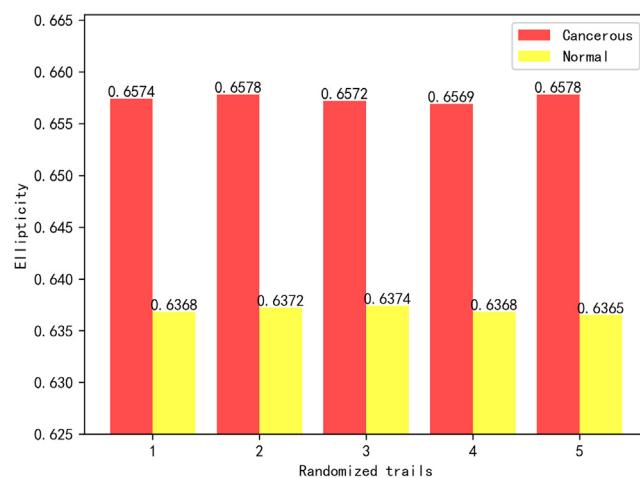


Figure 3. Mean nuclei ellipticity of five randomized trials

The elliptical characteristic in [Figure 3](#) suggests that cancerous nuclei are rounder. Additionally, the quantitative analysis of characteristic parameters such as area and equivalent diameter ellipticity in [Figure 4](#) indicates that cancerous nuclei are larger.

Geometric features

Hence, we extracted a set of 9-dimensional geometric features from the nucleus. These features encompass major axis length (*Major*), minor axis length (*Minor*), area (*Area*), direction θ (the angle between the major axis and horizontal direction), eccentricity (*Ecc*), ellipticity (*Ell*), equivalent diameter (*Dia*, the diameter of a circle with an area equal to the nuclei area), perimeter(*Per*), and convex hull area (*AreaHull*). [Figure 5](#) shows geometric features in part.

Texture features

Based on our understanding, cancerous nuclei exhibit a complex texture, whereas normal nuclei appear smooth. This observation is further supported by [Figure 6](#), which illustrates the complexity of cancerous nuclei texture in contrast to the smooth texture of normal nuclei.

For a quantitative portrayal of the difference, [Figure 7](#) illustrates two features characterizing nucleus texture. Comparative analysis reveals the following insights: 1) Cancerous nuclei exhibit more precise imaging, as indicated by the contrast characteristic reflecting the clarity of nuclei. This aligns with the clinical observation that cancerous nuclei appear hyperchromatic during H&E staining; 2) The texture of cancerous nuclei is more intricate, as evidenced by the entropy characteristic. Entropy serves as a measure of the information content within the image, and a higher value indicates a more complex texture.

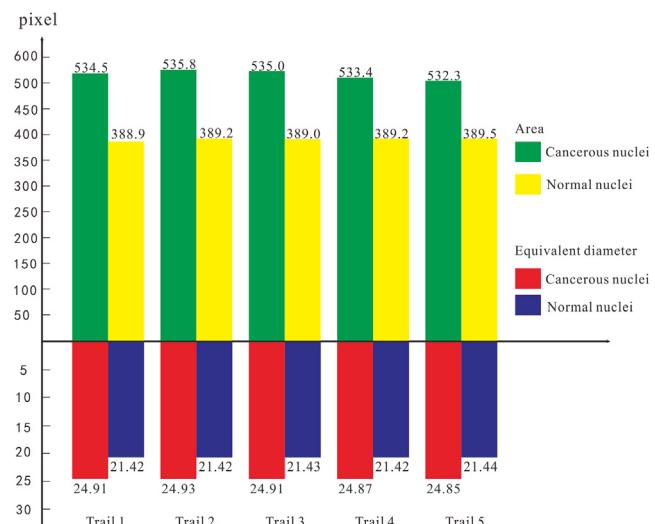


Figure 4. Area and equivalent diameter of normal nuclei and cancerous ones

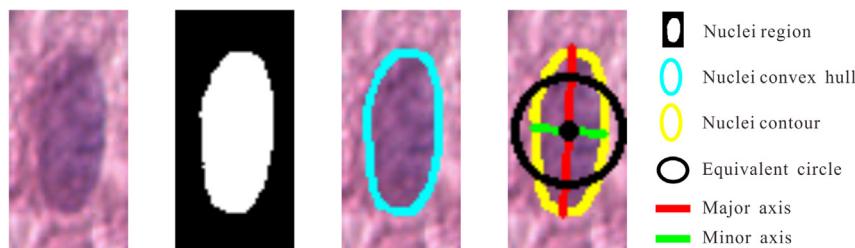


Figure 5. Schematic diagram of the geometric features of the nuclei in part

To this end, we extracted 7-dimensional texture features from the nuclei. These features encompass contrast (Con), dissimilarity (Diss), homogeneity (Hom), entropy (Ent), angular second movement (ASM), roughness (Rou), and dispersion (Disp).

To visually illustrate the distinction between the corresponding characteristics of cancerous nuclei and normal nuclei, Table 1 presents the ratio of 16-dimensional features in 5 randomized trials. The color characteristics are excluded due to the substantial H&E color difference, as evident in Figure 6. A more detailed modeling process is described in the STAR methods section.

Dataset description

Currently, prominent public datasets relevant to auxiliary breast cancer diagnosis include BACH2018,³⁹ BreakHis,⁴⁰ and BRACS.⁴¹ The BACH2018 dataset comprises 10 normal WSIs and 20 cancerous WSIs. BRACS stands out as the most extensive dataset available to the author, with 547 WSIs post-update; however, only 423 WSIs were effectively obtained during this paper's experimentation. Within these, 252 were normal, and 171 were cancerous, resulting in a positive/negative sample ratio of 40.43%:59.57%. BreakHis, offering only patches cropped from WSIs without enabling the construction of slide-level features, is not utilized in this study. The provided data includes labeling information, with '0' denoting normal WSIs and '1' indicating cancerous WSIs.

WSIs exhibit strong heterogeneity and diversity stemming from various sources such as different image acquisition devices, scanning protocols, image gatherers, variations in the H&E staining process before scanning, and inherent differences among patients. These factors contribute to distribution drift in WSI images, and the diverse set of samples significantly influences the model's generalization ability. A model trained on an imbalanced sample set may acquire biased information about sample proportions, leading to predictions that overly focus on certain categories and compromise the model's robustness.⁴²

To enhance sample diversity and maintain a balanced ratio of positive and negative samples, this study incorporates 20 cancerous WSIs from the BACH2018 dataset and 33 self-owned cancerous WSIs in addition to the BRACS dataset. The experimental dataset comprises 476 WSIs with a positive/negative sample ratio of 47.06%:52.94%, as depicted in Table 2.

Implementation details

Typically, human tissue does not cover the entire slide during slide preparation, resulting in WSIs containing significant blank spaces, as depicted in Figure 8. Efficiently extracting the human tissue region is crucial for conserving computing resources and enhancing the algorithm's efficiency.

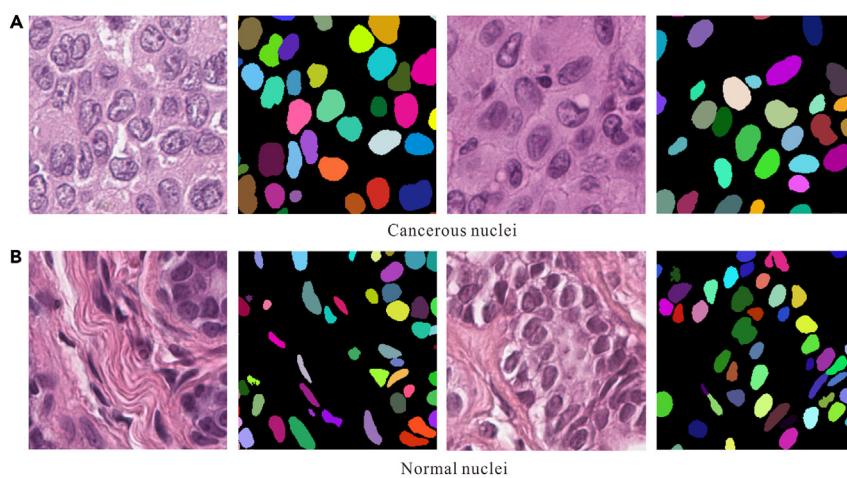


Figure 6. Nuclei texture (the different colors in the black background represent different nuclei)

(A) Cancerous nuclei.
(B) Normal nuclei.

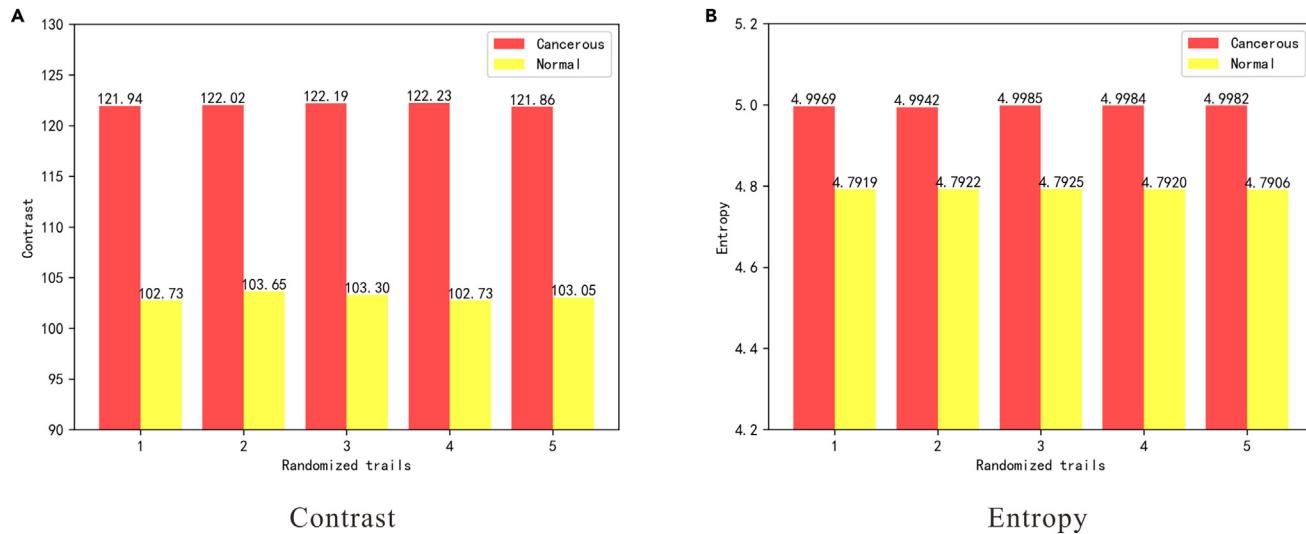


Figure 7. Mean nuclei contrast and entropy of five randomized trials

(A) The contrast difference.
(B) The entropy difference.

The process begins by leveraging the principle that the S-channel of the HSV color model represents a combination of specific spectral colors and white.⁴³ WSIs are then transformed from RGB to HSV in the high pyramid (e.g., layer 4). Subsequently, the human tissue region is extracted using the OSTU method⁴⁴ in the S-channel, and any small holes in the tissue are filled through mathematical morphology closing operations.⁴⁵ The extracted human tissue contour is filtered based on region thresholds, and the silhouettes are restored to the original image layer.

Furthermore, it's impracticable to load an entire WSI into the GPU directly, due to its large size (e.g., 100,000 × 100,000 pixels), so a patch-based processing approach should be employed.⁴⁶ This involves using a sliding window of 512 × 512 pixels on the WSI within the extracted human tissue area, with the top-left coordinates of each sliding window recorded. The final step involves restoring the WSI based on these coordinates. Figure 8 illustrates the step-by-step processing procedure.

According to the percentage of 7:1:2, the experimental images were divided into training datasets, verification datasets, and test datasets. All experiments were performed on a single desktop computer with Window10 operating system, CPU with six cores and six threads, 3.00 GHz i5-8500 (32 GB memory), and the GPU GTX2080 with 8 GB memory.

Table 1. Ratio of handcrafted features between cancerous nuclei and normal ones

Handcrafted features		Mean	STDEV
Geometric Features	Area	1.3726	1.4885
	Convex hull area	1.3664	1.4497
	Eccentricity	0.9751	1.0190
	Equivalent diameter	1.1620	1.2875
	Major axis length	1.1302	1.0945
	Minor axis length	1.1913	1.3453
	Direction	1.4527	0.9916
	Perimeter	1.1607	1.1888
	Ellipticity	1.0321	0.9849
Texture Features	Roughness	0.9962	1.0893
	Contrast	1.1839	0.9701
	Dissimilarity	1.1366	0.9712
	Homogeneity	0.8425	0.8064
	Entropy	1.0429	1.0463
	Angular second moment	0.7735	0.9690
	Dispersion	1.1284	1.0542

Table 2. Construction of experimental datasets

Image source	Collecting Device	Image Format	Pixel Size (μm)	Normal WSIs	Cancerous WSIs
BRACS	Aperio AT2	.svs	0.25×0.25	252	171
BACH2018	Leica SCN400	.svs	0.42×0.42	0	20
Self-owned data	KF-PRO-005-EX	.kfb	0.50×0.50	0	33
Total				252	224

Ablation experiment

To validate the effectiveness of handcrafted feature constraints, three sets of ablation experiments were conducted. The classification result based on slide-level features is denoted as ablation1 (macro feature), patch-level deep learning feature constraints are characterized as ablation2 (meso feature), and nuclei-level handcrafted feature constraints are indicated as ablation3 (micro feature). For simplicity, the model inference probability is denoted as PrePro (Predicted Probability).

As shown in [Table 3](#), for WSIs (WSI-1, WSI-2) characterized by clear imaging and distinct features, both NPKC-MIL and the ablation experiments consistently achieve accurate classification with high probability.

For WSI-3, with a subclass as Flat Epithelial Atypia (FEA), which appears cancerous at the macro view ([Figure 9A](#)), ablation1 misclassified it as cancerous with a very high probability (0.9988). However, ablation2 significantly reduced the misclassification probability to 0.6992 by introducing the 1024-dimensional meso feature constraint. Although ablation3 is constrained by 16-dimensional micro features, its misclassification probability remains high (0.8867). This suggests that micro features play a restrictive role, but their impact is not as pronounced as that of the 1024-dimensional meso features, underscoring that the latter provides more effective constraint information than the former.

For WSI-4, with a subclass as invasive carcinoma (IC), ablation1 misclassified it as normal based on macro features ([Figure 9B](#)). However, upon introducing meso feature constraints (ablation2), the misclassification probability significantly decreased to 0.6487, indicating the effectiveness of the 1024-dimensional meso feature constraints. In contrast, relying solely on the micro 16-dimensional handcrafted feature constraint led to a high probability of misclassification as normal (0.9487). This emphasizes the challenge of exclusively focusing on micro features, encountering the dilemma of ‘seeing the trees but not the forest.’ Therefore, a combination of meso and micro characteristics is essential to constrain macro features correctly. Notably, NPKC-MIL classified it as cancerous with a probability of 0.5220.

The attention score heatmap in [Figure 10](#) reveals that regions with high attention predominantly consist of nuclei-rich areas, while regions with low attention are primarily composed of muscle or fat tissue. Hence, a more comprehensive analysis at the nuclear level is crucial for the areas exhibiting high attention.

Ultimately, as demonstrated in [Table 4](#), the precision and accuracy of ablation2 and ablation3 surpass those of ablation1, underscoring the effectiveness of patch-level or nuclei-level characteristics in constraining slide-level features. However, the sensitivity (SE) of ablation3 is lower than that of ablation1, indicating that the detection of WSI carcinomatosis solely through handcrafted restrictions at the nuclear level is insufficient.

Ablation2 outperforms ablation3 in terms of accuracy (AC), sensitivity (SE), and precision (PC), suggesting that patch-level deep learning features (1024 dimensions) offer more constraint information than nuclei handcrafted features (16 dimensions), albeit with marginal improvement. This analysis also fully illustrates that although deep learning methods can extract features of any dimension theoretically, the dimension of features is not the higher, the better, but rather to extract features with discriminative power. Finally, the amalgamation of nuclei-level and patch-level features enhance the constraint on slide-level features, thereby improving the accuracy of cancerous WSI classification.

Comparative analysis of different methods

The performance of various comparison methods on four test samples is presented in [Table 5](#). For WSIs with clear imaging and distinct features (WSI-1, WSI-2), except for TransMIL’s misclassification of WSI-1 as cancerous, other methods correctly classify WSI-1 with high probability.



Figure 8. Tissue extraction and patching (the green line indicates the tissue contour, and the small rectangular box in step B are sliding window of 512×512 pixels to extract patches for training)

- (A) Extracting tissue.
- (B) Slicing WSI into patches.

Table 3. Performance of different ablation experiments on four test samples

Methods	WSI-1(GT = 0)		WSI-2(GT = 1)		WSI-3(GT = 0)		WSI-4(GT = 1)	
	PrePro	PreCls	PrePro	PreCls	PrePro	PreCls	PrePro	PreCls
NPKC-MIL	0.9249	0	0.7999	1	0.5753	0	0.5220	1
ablation1	0.9996	0	0.9998	1	0.9988	1	0.9267	0
ablation2	0.9997	0	0.9998	1	0.6992	1	0.6487	0
ablation3	0.8872	0	0.6998	1	0.8867	1	0.9487	0

PrePro is an abbreviation for “predict probability”, PreCls is an abbreviation for “predict class”.

Figure 11C reveals that areas with high attention scores exhibit a complex texture comprising red blood cells, muscle cells, lymphocytes, epithelial cells, etc. Aside from DTFD-MIL, other methods assign higher attention to these areas, and the attention heatmap of FRMIL and ReMix is identical, indicating the shared use of the same attention framework.

NPKC-MIL accurately classifies WSI-3 and WSI-4. TransMIL misclassifies WSI-3 as normal but correctly classifies WSI-4. FRMIL incorrectly classifies WSI-3 as cancerous WSI, aligning with its assumption that “cancerous WSIs have more information than normal ones.” CLAM-SB misclassifies WSI-3 and WSI-4, suggesting that for WSIs with complex textures, meso features alone may not sufficiently constrain classification results based on macro features. In comparison to NPKC-MIL, ReMix correctly classifies WSI-3 with a higher probability (0.9466), and DTFD-MIL accurately classifies WSI-4 with a higher probability (0.9022), highlighting the simplicity and effectiveness of data enhancement for small sample data.

Table 6 presents a comparative analysis between NPKC-MIL and other five state-of-the-art methods.^{29,31,33–35} The results highlight that TransMIL, by treating patches as relevant entities and considering their spatial topological information through the Transfer framework, achieves an accuracy of 88.75% and a sensitivity of 90.00%. While FRMIL successfully adjusted the distribution of instances, its underlying assumption that “cancerous WSIs contain more information than normal ones” led the model to prioritize the features of cancerous WSIs, consequently diminishing its ability to recognize features in normal WSIs. This bias resulted in a lower specificity (67.5%) and a higher sensitivity (85%). Within the MIL framework, CLAM-SB employs a clustering concept and adds patch-level constraints with a few representative instances, achieving a sensitivity of 90.00%.

The generalization ability of deep learning models is significantly influenced by the diversity of training samples. ReMix proposed the “mix-the-bag” data enhancement method, which is simple but effective, achieving an accuracy of 91.43%. DTFD-MIL uses pseudo-bags technology for data enhancement, but the child bags derived from the positive parent bags do not always contain positive instances, in which case will introduce a bag with a wrong label, and the feature distillation branch structure used in their study cannot eliminate the artificially introduced errors well, leading to a low SE with 80%.

It is noteworthy that all the aforementioned methods rely on slide-level and patch-level deep learning features from transfer learning models, neglecting micro-level prior knowledge (handcrafted features). In contrast, NPKC-MIL maximizes the potential of deep learning in feature extraction while incorporating meaningful manual features. This approach not only builds interpretable deep learning models but also achieves higher accuracy, specificity, and sensitivity, respectively.

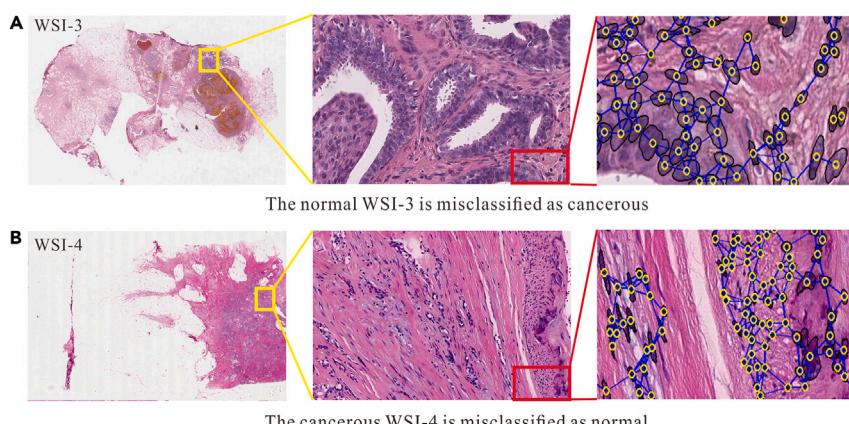


Figure 9. The misclassified test samples and their macro, meso, and micro characteristics diagram (the yellow circle is the nuclei centroid, the black circle is the nuclei contour, and the blue lines represent the adjacency relationship of nuclei)

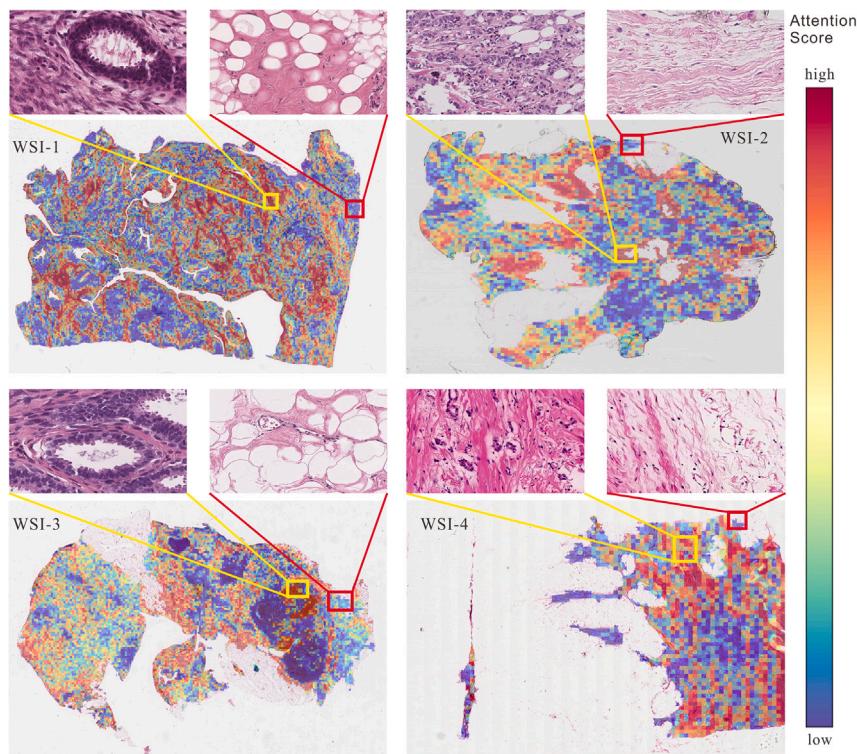


Figure 10. Attention heatmaps of the four test samples (yellow boxes for high attention score areas, red boxes for low attention score areas)

DISCUSSION

How do advancements in AI technologies drive the progress of digital histopathology research?

The image classification process involves a mathematical exploration to identify the mapping relationship between image features and target types. Traditional methods, rooted in feature engineering, rely on effective and comprehensive handcrafted features. In simpler scenarios, the objective function can be determined through sparse regression, and parameters can be solved using least squares.

However, for the challenging task of WSI classification, low-level handcrafted features (including color features, texture features, morphological features, etc.) prove insufficient in capturing the complexity of medical images. As illustrated in Table 4, the impact of 16-dimensional handcrafted feature constraints is inferior to that of deep learning features, underscoring the richer constraint information provided by deep learning features. Additionally, for WSIs exhibiting strong heterogeneity, modeling complex scenes through symbolic mathematical methods poses challenges due to the optimization space's intricate structures and coefficients, making it arduous for optimization algorithms, such as simulated annealing and genetic algorithms, to find solutions. Lastly, many traditional feature-based engineering algorithms necessitate loading all data into memory simultaneously, which is impractical for WSIs with gigapixels.

Why should we pay more attention to prior knowledge at the nucleus level?

Different from parametric models that can be explicitly expressed through mathematical formulas, deep learning operates as a statistical modeling method with unknown structures and coefficients in its equations. It is a data-driven approach optimized by gradient descent algorithms. While the convolution operation can theoretically extract features of any dimension, the lack of interpretability has hindered widespread acceptance among pathologists.

Table 4. Comparison table of ablation experiments

	AC(%)	SP(%)	SE(%)	PC(%)
NPKC-MIL	96.25	97.50	95.00	97.44
ablation1	83.75	85.00	82.50	84.62
ablation2	86.25	90.00	82.50	89.19
ablation3	85.00	90.00	80.00	88.89

Table 5. The performance of different comparison methods on four test samples

Methods	WSI-1(GT = 0)		WSI-2(GT = 1)		WSI-3(GT = 0)		WSI-4(GT = 1)	
	PrePro	PreCls	PrePro	PreCls	PrePro	PreCls	PrePro	PreCls
NPKC-MIL	0.9249	0	0.7999	1	0.5753	0	0.5220	1
TransMIL	0.6714	1	0.9963	1	0.9095	1	0.9662	1
FRMIL	0.5877	0	0.8821	1	0.8752	1	0.6288	0
CLAM-SB	0.9999	0	1.0000	1	0.9916	1	1.0000	0
ReMix	0.9914	0	1.0000	1	0.9466	0	0.9799	0
DTFD-MIL	1.0000	0	1.0000	1	0.9999	1	0.9022	1

PrePro is an abbreviation for "predict probability", PreCls is an abbreviation for "predict class".

As depicted in [Figure 12](#), WSIs exhibit strong heterogeneity and diversity. Various factors, such as different image acquisition devices and variations in the H&E staining process before scanning, can contribute to distribution drift in WSIs. Consequently, extracting nuanced prior knowledge from macro and meso perspectives becomes challenging.

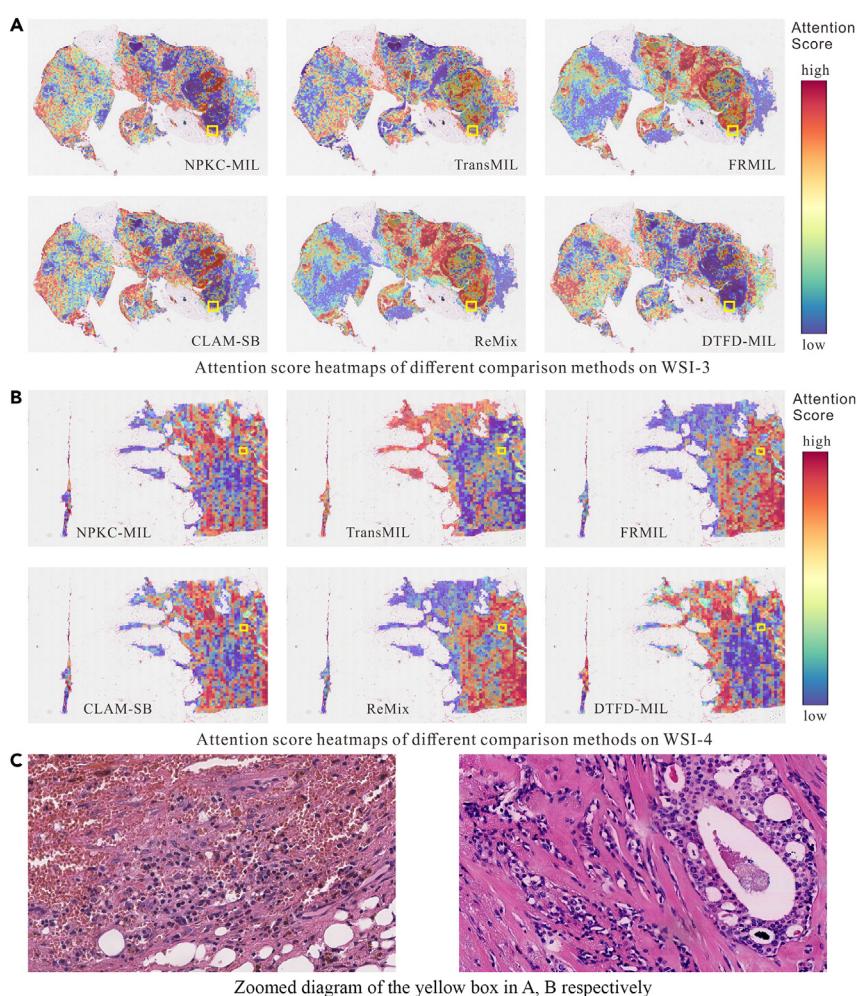


Figure 11. Attention score heatmaps of different comparison methods on WSI-3 and WSI-4 (yellow boxes are areas of high attention score)

(A) Attention score heatmaps of different methods on WSI-3.
 (B) Attention score heatmaps of different methods on WSI-4.
 (C) Zoomed diagram of the yellow box in (A) and (B).

Table 6. Comparison of evaluation criteria of different algorithms

	AC(%)	SP (%)	SE (%)	PC (%)
NPKC-MIL	96.25	97.50	95.00	97.44
TransMIL ²⁹	88.75	87.50	90.00	87.80
FRMIL ³¹	76.25	67.50	85.00	72.34
CLAM-SB ³³	86.25	82.50	90.00	83.72
ReMix ³⁴	86.25	92.50	80.00	91.43
DTFD-MIL ³⁵	83.75	87.50	80.00	86.49

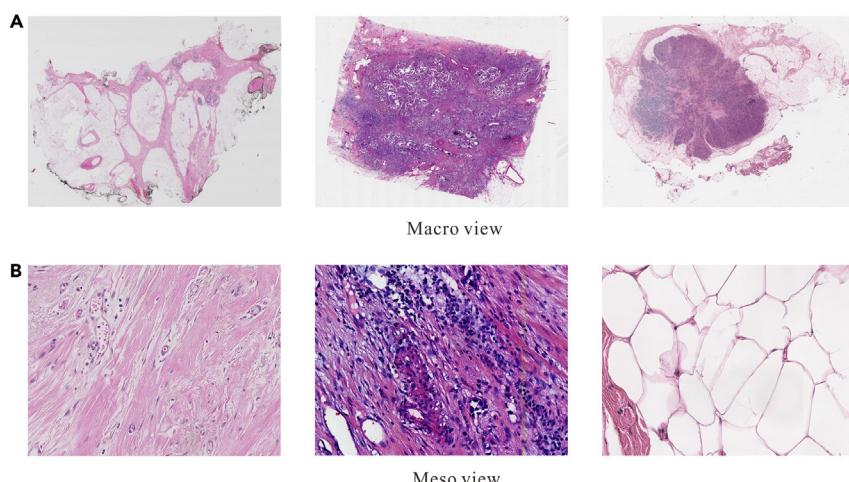
As shown in [Figure 13](#), the nucleus is a concrete object that can be quantitatively described. We can extract various attributes of the nucleus itself and analyze the interactions between nuclei. As indicated in [Figure 10](#), the regions densely populated with nuclei coincide with areas of higher attention scores. Therefore, to enhance the interpretability of the model and broaden the dimension of feature analysis, it is necessary to extract multidimensional prior knowledge of nuclei and conduct quantitative analysis.

How to incorporate prior knowledge into the AI model?

In the construction progress of deep learning modeling, the integration of prior knowledge can occur in three key phases: (1) During the initial design of the model structure, the relationship between differentiation and convolution can be harnessed to shape the convolutional kernel. The network structure can then be tailored using principles from the finite element method. (2) Handcrafted features can be introduced to the deep learning features at the bottom layer to augment feature dimensionality, followed by forward propagation layer by layer. The addition of tens of dimensions of handcrafted features to low-level deep learning features with thousands of dimensions may not necessarily enhance model performance. However, incorporating handcrafted features at the bottom layer becomes viable when the dimensionality of these features is sufficiently high. (3) At the top level of the model, the constraints of handcrafted features (prior knowledge) on the model are realized by adding penalty terms to the loss function. In this paper, we realize the constraints of handcrafted features on the classification results by adding a penalty term $\mathcal{L}_{\text{nuclei}}$ to the model loss function.

Conclusion

Recognizing the limitations of the current breast histopathology WSIs classification algorithm, which solely relies on slide-level and patch-level features while overlooking nuclei-level features, this paper introduces a NPKC-MIL for breast histopathology image classification. NPKC-MIL leverages the advantages of convolutional neural networks in model fitting while also imbuing the model with clear physical significance by the introduction of constraints based on handcrafted features. Experimental results demonstrate that, compared to state-of-the-art methods, the proposed approach exhibits superior specificity, sensitivity, accuracy, and precision in the classification of breast histopathology WSIs. This offers valuable insights for expanding the analytical scope of current WSI classification tasks and incorporating prior knowledge (handcrafted features) into deep learning models.

**Figure 12. Strong heterogeneity of WSI from macro and meso perspectives**

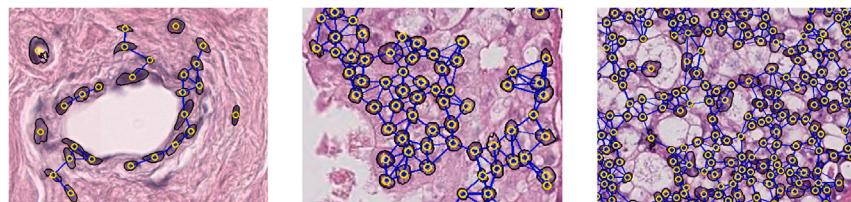


Figure 13. Nuclei in micro perspectives (the yellow circle is the nuclei centroid, the black circle is the nuclei contour, and the blue lines represent the adjacency relationship of nuclei)

Limitations of the study

While we confirmed the effectiveness of integrating prior knowledge into deep learning models, our approach was limited to employing significant 16-dimensional handcrafted features derived from the statistics of 1 million nuclei. Unfortunately, color features and contour features were not considered in NPKC-MIL. Moreover, our experiments were exclusively conducted on breast WSIs. To advance our understanding, future research should focus on exploring additional prior knowledge and conducting tests on diverse cancer types.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- METHOD DETAILS
 - Principle of nuclei-level feature design
 - Nuclei-level topology map construction and node message passing
 - Patch-level feature extraction and slide-level feature construction
 - Classify WSIs based on nuclei-level handcrafted feature constraint
 - Evaluation criteria
- QUANTIFICATION AND STATISTICAL ANALYSIS

ACKNOWLEDGMENTS

This work was supported by the National Plan for the Special Support of Leading Talents under Grant 412500002, and the Japan Society for the Promotion of Science under Grant 23K13419.

AUTHOR CONTRIBUTIONS

All conceptualization, methodology, investigation, manuscript writing, and revisions in the manuscript were completed by X.W. and W.Y.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 22, 2024

Revised: April 17, 2024

Accepted: April 24, 2024

Published: April 26, 2024

REFERENCES

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and 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**, 209–249.
2. Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., and Madabhushi, A. (2016). Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. *IEEE Trans. Med. Imag.* **35**, 119–130.
3. Onishi, Y., Teramoto, A., Tsujimoto, M., Tsukamoto, T., Saito, K., Toyama, H., Imaizumi, K., and Fujita, H. (2020). Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks. *Int. J. Comput. Assist. Radiol. Surg.* **15**, 173–178.
4. Van Rijthoven, M., Balkenhol, M., Siliqa, K., Van Der Laak, J., and Ciompi, F. (2021). HookNet: Multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images. *Med. Image Anal.* **68**, 101890.

5. Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., and Van Valen, D. (2019). Deep learning for cellular image analysis. *Nat. Methods* 16, 1233–1246.
6. Hu, J., Cui, C., Yang, W., Huang, L., Yu, R., Liu, S., and Kong, Y. (2021). Using deep learning to predict anti-PD-1 response in melanoma and lung cancer patients from histopathology images. *Transl. Oncol.* 14, 100921.
7. Komura, D., and Ishikawa, S. (2019). Machine learning approaches for pathologic diagnosis. *Virchows Arch.* 475, 131–138.
8. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444.
9. LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324.
10. Dong, S., Wang, P., and Abbas, K. (2021). A survey on deep learning and its applications. *Comput. Sci. Rev.* 40, 100379.
11. Asgari Taghanaki, S., Abhishek, K., Cohen, J.P., Cohen-Adad, J., and Hamarneh, G. (2021). Deep semantic segmentation of natural and medical images: a review. *Artif. Intell. Rev.* 54, 137–178.
12. Li, X., Li, C., Rahaman, M.M., Sun, H., Li, X., Wu, J., Yao, Y., and 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.
13. Cheng, J., Mo, X., Wang, X., Parwani, A., Feng, Q., and Huang, K. (2018). Identification of topological features in renal tumor microenvironment associated with patient survival. *Bioinformatics* 34, 1024–1030.
14. Cheng, J., Liu, Y., Huang, W., Hong, W., Wang, L., Zhan, X., Han, Z., Ni, D., Huang, K., and Zhang, J. (2021). Computational image analysis identifies histopathological image features associated with somatic mutations and patient survival in gastric adenocarcinoma. *Front. Oncol.* 11, 623382.
15. Zhu, X.L., Yao, J.W., and Huang, J.Z. (2016). Deep convolutional neural network for survival analysis with pathological images. In *Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 544–547.
16. Wang, X., Chen, H., Gan, C., Lin, H., Dou, Q., Tsougenis, E., Huang, Q., Cai, M., and Heng, P.A. (2020). Weakly supervised deep learning for whole slide lung cancer image analysis. *IEEE Trans. Cybern.* 50, 3950–3962.
17. Shao, W., Wang, T., Huang, Z., Han, Z., Zhang, J., and Huang, K. (2021). Weakly supervised deep ordinal cox model for survival prediction from whole-slide pathological images. *IEEE Trans. Med. Imag.* 40, 3739–3747.
18. Iizuka, O., Kanavati, F., Kato, K., Rambeau, M., Arihiro, K., and Tsuneki, M. (2021). Deep learning models for histopathological classification of gastric and colonic epithelial tumors. *Sci. Rep.* 10, 15043.
19. Chikontwe, P., Kim, M., Nam, S.J., Go, H., and Park, S.H. (2020). Multiple instance learning with center embeddings for histopathology classification. In *Proceedings of the Medical Image Computing and Computer Assisted Intervention*, pp. 519–528.
20. Hou, L., Gupta, R., Van Arnam, J.S., Zhang, Y., Sivalenka, K., Samaras, D., Kurc, T.M., and Saltz, J.H. (2020). Dataset of segmented nuclei in hematoxylin and eosin stained histopathology images of ten cancer types. *Sci. Data* 7, 185. <https://doi.org/10.1038/s41597-020-0528-1>.
21. Lu, L., Jin, P., Pang, G., Zhang, Z., and Karniadakis, G.E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.* 3, 218–229.
22. Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., and Zhou, B. (2023). Trustworthy AI: From principles to practices. *ACM Comput. Surv.* 55, 1–46.
23. Quellec, G., Cazuguel, G., Cochener, B., and Lamard, M. (2017). Multiple-instance learning for medical image and video analysis. *IEEE Rev. Biomed. Eng.* 10, 213–234.
24. Kanavati, F., Toyokawa, G., Momosaki, S., Rambeau, M., Kozuma, Y., Shoji, F., Yamazaki, K., Takeo, S., Iizuka, O., and Tsuneki, M. (2020). Weakly-supervised learning for lung carcinoma classification using deep learning. *Sci. Rep.* 10, 9297.
25. Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R., and Smola, A.J. (2017). Deep sets. In *Proceedings of the 31st Annual Conference on Neural Information Processing Systems*, p. 30.
26. Ilse, M., Tomczak, J., and Welling, X. (2018). Attention-based deep multiple instance learning. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 2127–2136.
27. Li, B., Li, Y., and Elciceri, K.W. (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.
28. Hashimoto, N., Fukushima, D., Koga, R., Takagi, Y., Ko, K., Kohno, K., Nakaguro, M., Nakamura, S., Hontani, H., and Takeuchi, I. (2020). Multi-scale domain-adversarial multiple-instance CNN for cancer subtype classification with unannotated histopathological images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (Electr. Network)*, pp. 3852–3861.
29. Tomita, N., Abdollahi, B., Wei, J., Ren, B., Suriawinata, A., and Hassanpour, S. (2019). Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides. *JAMA Netw. Open* 2, e1914645.
30. Sharma, Y., Shrivastava, A., Ehsan, L., Moskaluk, C.A., Syed, S., and Brown, D. (2021). Cluster-to-conquer: A framework for end-to-end multi-instance learning for whole slide image classification. In *Proceedings of the Medical Imaging with Deep Learning*, pp. 682–698.
31. Singer, W., and Gray, C.M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* 18, 555–586.
32. Shao, Z.C., Bian, H., Chen, Y., Wang, Y.F., Zhang, J., and Ji, X.Y. (2021). Transmil: Transformer-based correlated multiple instance learning for whole slide image classification. *Adv. Neural Inf. Process. Syst.* 34, 2136–2147.
33. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st Annual Conference on Neural Information Processing Systems*, p. 30. <https://doi.org/10.48550/arXiv.1706.03762>.
34. Chikontwe, P., Nam, S.J., Go, H., Kim, M., Sung, H.J., and Park, S.H. (2022). Feature Re-calibration Based Multiple Instance Learning for Whole Slide Image Classification. In *Proceedings of the 25th International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 420–430.
35. Xie, C.S., Muhammad, H., Vanderbilt, C.M., Caso, R., Yarlagadda, D.V.K., Campanella, G., and Fuchs, T.J. (2020). Beyond classification: Whole slide tissue histopathology analysis by end-to-end part learning. In *Proceedings of the Medical Imaging with Deep Learning*, pp. 843–856.
36. Lu, M.Y., Williamson, D.F.K., Chen, T.Y., Chen, R.J., Barbieri, M., and Mahmood, F. (2021). Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* 5, 555–570.
37. Yang, J.W., Chen, H.B., Zhao, Y., Yang, F., Zhang, Y., He, L., and Yao, J.H. (2022). Remix: A general and efficient framework for multiple instance learning based whole slide image classification. In *Proceedings of the 25th International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 35–45.
38. Zhang, H.R., Meng, Y.D., Zhao, Y.T., Qiao, Y.H., Yang, X.Y., Coupland, S.E., and Zheng, Y.L. (2022). DTFD-MIL: Double-Tier Feature Distillation Multiple Instance Learning for Histopathology Whole Slide Image Classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18802–18812. <https://doi.org/10.1109/CVPR52688.2022.01824>.
39. Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S.S., Safwan, M., Alex, V., Marami, B., Prastawa, M., Chan, M., Donovan, M., et al. (2019). Bach: Grand challenge on breast cancer histology images. *Med. Image Anal.* 56, 122–139.
40. Spanhol, F.A., Oliveira, L.S., Petitjean, C., and Heutte, L. (2016). A dataset for breast cancer histopathological image classification. *IEEE Trans. Biomed. Eng.* 63, 1455–1462.
41. Brancati, N., Annicello, A.M., Pati, P., Riccio, D., Scognamiglio, G., Jaume, G., De Pietro, G., Di Bonito, M., Foncubierta, A., Botti, G., et al. (2022). Bracs: A dataset for breast carcinoma subtyping in h&e histology images. *Database 2022*, baac093. <https://doi.org/10.1093/database/baac093>.
42. Johnson, J.M., and Khoshgoftaar, T.M. (2019). Survey on deep learning with class imbalance. *J. Big Data* 6, 27.
43. Smith, A.R. (1978). Color gamut transform pairs. *SIGGRAPH Comput. Graph.* 12, 12–19.
44. Jiao, S.H., Li, X.G., and Lu, X. (2006). An improved Ostu method for image segmentation. In *Proceedings of the 8th International Conference on Signal Processing*, Guilin, China, pp. 16–20.

- <https://doi.org/10.1109/ICOSP.2006.345705>.
45. Haralick, R.M., Sternberg, S.R., and Zhuang, X. (1987). Image analysis using mathematical morphology. *IEEE Trans. Pattern Anal. Mach. Intell.* 9, 532–550.
 46. Li, W., Li, J., Polson, J., Wang, Z., Speier, W., and Arnold, C. (2022). High-resolution histopathology image generation and segmentation through adversarial training. *Med. Image Anal.* 75, 102251. <https://doi.org/10.1016/j.media.2021.102251>.
 47. Sklansky, J. (1982). Finding the convex hull of a simple polygon. *Pattern Recogn. Lett.* 1, 79–83.
 48. Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., and Rajpoot, N. (2019). Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image Anal.* 58, 101563.
 49. Cover, T., and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theor.* 13, 21–27.
 50. Kipf, T.N., and Welling, M. (2017). Semi-supervised classification with graph convolutional networks. Preprint at arXiv 1. <https://doi.org/10.48550/arXiv.1609.02907>.
 51. Corso, G., Cavalleri, L., Beaini, D., Liò, P., and Veličković, P. (2020). Principal neighbourhood aggregation for graph nets. *Adv. Neural Inf. Process. Syst.* 33, 13260–13271. <https://doi.org/10.48550/arXiv.2004.05718>.
 52. He, K.M., Zhang, X.Y., Ren, S.Q., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
 53. Ma, X.J., Huang, H.X., Wang, Y.S., Romano, S., Erfani, S., and Bailey, J. (2020). Normalized loss functions for deep learning with noisy labels. In Proceedings of the International Conference on Machine Learning, pp. 6543–6553.
 54. Kingma, D.P., and Ba, J. (2017). Adam: A method for stochastic optimization. Preprint at arXiv 1. <https://doi.org/10.48550/arXiv.1412.6980>.
 55. Francis, K., and Palsson, B.O. (1997). Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. *Proc. Natl. Acad. Sci. USA* 94, 12258–12262.
 56. Hassanin, M., Anwar, S., Radwan, I., Khan, F.S., and Mian, A. (2022). Visual Attention Methods in Deep Learning: An In-Depth Survey. Preprint at arXiv 1. <https://doi.org/10.48550/arXiv.2204.07756>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
BRACS	BRACS	https://www.bracs.icar.cnr.it/
BACH2018	BACH2018	https://iciar2018-challenge.grand-challenge.org/Dataset/
Software and algorithms		
Convex hull	Sklansky,J. ⁴⁷ (1982)	https://doi.org/10.1016/0167-8655(82)90016-2
Hover-net	Graham et al. ⁴⁸ (2019)	https://doi.org/10.1016/j.media.2019.101563
K-NN	Cover et al. ⁴⁹ (1967)	https://doi.org/10.1109/TIT.1967.1053964
Graph convolutional	Kipf et al. ⁵⁰ (2017)	https://doi.org/10.48550/arXiv.1609.02907
Neighborhood aggregation	Corso et al. ⁵¹ (2020)	https://doi.org/10.48550/arXiv.2004.05718
ResNet	He et al. ⁵² (2016)	https://doi.org/10.1109/CVPR.2016.90
APL loss	Ma et al. ⁵³ (2020)	https://doi.org/10.48550/arXiv.2006.13554
Adam	Kingma et al. ⁵⁴ (2017)	https://doi.org/10.48550/arXiv.1412.6980
NPKC-MIL	This study	https://github.com/WxpHB/NPKC-MIL

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Wei Yuan (e-mail: wei.yuan@tohoku.ac.jp).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- BRACS dataset was downloaded <https://www.bracs.icar.cnr.it/>. BACH2018 dataset was downloaded from <https://iciar2018-challenge.grand-challenge.org/Dataset/>.
- All original code has been deposited at Github (<https://github.com/WxpHB/NPKC-MIL>) and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

METHOD DETAILS

Principle of nuclei-level feature design

Geometric features

Mark any binary map of nuclei region in the image as $H_{n \times m}$, its $(p + q)$ -order origin moment $M_{p,q}$ and central moment $\mu_{p,q}$ are:

$$M_{p,q} = \sum_{x=0}^n \sum_{y=0}^m x^p y^q H_{n \times m}(x, y) \quad (\text{Equation 1})$$

$$\mu_{p,q} = \sum_{x=0}^n \sum_{y=0}^m (x - \bar{x})^p (y - \bar{y})^q H_{n \times m}(x, y) \quad (\text{Equation 2})$$

where, $\bar{x} = \frac{M_{1,0}}{M_{0,0}}$, $\bar{y} = \frac{M_{0,1}}{M_{0,0}}$.

Then, the 9-dimensional geometric features can be calculated as follows:

$$\text{Major} = \sqrt{8\left(\bar{\mu}_{2,0} + \bar{\mu}_{0,2} + \sqrt{4\bar{\mu}_{1,1}^2 + (\bar{\mu}_{2,0} - \bar{\mu}_{0,2})^2}\right)} \quad (\text{Equation 3})$$

$$\text{Minor} = \sqrt{8\left(\bar{\mu}_{2,0} + \bar{\mu}_{0,2} - \sqrt{4\bar{\mu}_{1,1}^2 + (\bar{\mu}_{2,0} - \bar{\mu}_{0,2})^2}\right)} \quad (\text{Equation 4})$$

$$\text{Area} = M_{0,0} = \sum_{x=0}^n \sum_{y=0}^m H_{n \times m}(x, y) \quad (\text{Equation 5})$$

$$\theta = \frac{1}{2} \arctan\left(\frac{2\bar{\mu}_{1,1}}{\bar{\mu}_{2,0} - \bar{\mu}_{0,2}}\right) \quad (\text{Equation 6})$$

$$Ecc = \sqrt{1 - \frac{\text{Minor}^2}{\text{Major}^2}} \quad (\text{Equation 7})$$

$$Ell = \frac{\text{Minor}}{\text{Major}} \quad (\text{Equation 8})$$

$$Dia = \sqrt{4 \frac{\text{Area}}{\pi}} \quad (\text{Equation 9})$$

$$Per = N_v + N_h + \sqrt{2}N_{d1} + \sqrt{2}N_{d2} \quad (\text{Equation 10})$$

$$\text{AreaHull} = \sum_x \sum_y T(x, y) \quad (\text{Equation 11})$$

where, $\bar{\mu}_{2,0} = M_{2,0}/M_{0,0} - \bar{x}^2$, $\bar{\mu}_{0,2} = M_{0,2}/M_{0,0} - \bar{y}^2$, $\bar{\mu}_{1,1} = M_{1,1}/M_{0,0} - \bar{x} \cdot \bar{y}$, N_v , N_h , N_{d1} , and N_{d2} represent the number of adjacent pixels of the pixels on the nuclear contour in the directions 0° , 90° , 45° , and 135° , respectively; T is the convex hull⁴⁷ containing the nuclei.

Texture features

Let the GLCM (Gray Level Co-occurrence Matrix, GLCM) of $F_{n \times m}$ as $G_{l \times l}$ and meets:

$$G_{l \times l}(i, j) = \sum_{x=1}^n \sum_{y=1}^m \begin{cases} 1, & \text{if } F_{n \times m}(x, y) = i \text{ and } F_{n \times m}(x + \Delta x, y + \Delta y) = j; \\ 0, & \text{else.} \end{cases} \quad (\text{Equation 12})$$

where, i and j are gray values; $(\Delta x, \Delta y)$ is the offset of the pixel (x, y) . $(\Delta x, \Delta y)$ takes values of $1, 0, 1, 1, 0, 1$, and $-1, 1$, representing the commonly used four directions of 0° , 45° , 90° , and 135° , respectively, $F_{n \times m}$ is the grey-scale map of nuclei region in the image, l is the gray-scale level in $F_{n \times m}$.

Then, 7-dimensional texture features can be outlined as follows:

$$Con = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} (i - j)^2 G(i, j) \quad (\text{Equation 13})$$

$$Diss = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} G(i, j) |i - j| \quad (\text{Equation 14})$$

$$Hom = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} \frac{G(i, j)}{1 + (i - j)^2} \quad (\text{Equation 15})$$

$$Ent = - \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} G(i, j) \log G(i, j) \quad (\text{Equation 16})$$

$$ASM = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} G(i,j)^2 \quad (\text{Equation 17})$$

$$Rou = \frac{P_h}{P_n} \quad (\text{Equation 18})$$

$$Dsip = \sqrt{\sum_{i=0}^{l-1} \sum_{j=0}^{l-1} (G(i,j) - \epsilon)^2} \quad (\text{Equation 19})$$

where, P_h and P_n are the perimeter of the nuclei and its convex hull, respectively, ϵ is the mean of $G_{|X|}$.

Subsequently, the handcrafted features are aggregated and updated following Equation 21. The optimization is performed iteratively using the Adam optimizer to refine the Equation 32.

Nuclei-level topology map construction and node message passing

Based on the fact that "the interaction of nuclei closer in space is stronger, while the interaction of nuclei farther away is weaker",⁵⁵ Firstly, utilize the Hover-Net model⁴⁸ to extract nuclei and abstract them into nodes, denoted as v . Then, employ the K-NN algorithm⁴⁹ to construct an undirected graph of nodes, marked as $G = (V, E)$, where V is a set of nodes and E is a set of edges. The principle of edge construction is as follows: $\forall v_i \in V$ when a node $\forall v_j \in V$ satisfies Equation 20, construct an undirected edge connecting v_i and v_j , denoted as $e_{ij} = (v_i, v_j) \in E$.

$$v_j \in \{w | dist(v_i, w) \leq d_k \wedge dist(v_i, w) < d_{min}, \forall v_i, w \in V\} \quad (\text{Equation 20})$$

where $dist(v, w)$ is the distance function, this paper takes Euclidean distance; d_k is the nearest neighbor of K in $dist(v, w)$; d_{min} is the nearest neighbor threshold. In this experiment, K takes 5 and d_{min} takes 50.

After constructing the node topology graph, extract 16-dimensional handcrafted features such as geometric and texture features of nodes according to the method described in the principle of nuclei-level feature design section. Finally, according to graph convolution theory⁵⁰ and node message passing mechanism,⁵¹ node messages are aggregated and updated according to Equation 21.

$$h^{(t+1)}(v) = U \left(h^{(t)}(v), \bigoplus_{u \in N(v)} M \left(h^{(t)}(v), h^{(t)}(u) \right) \right) \quad (\text{Equation 21})$$

where $h^{(t)}(v)$ and $h^{(t+1)}(v)$ are the feature vectors of the current layer and the next node layer v , respectively. $h^{(t)}(u)$ is the adjacent node feature vectors of the node v ; U , M are multi-layer perceptron; $N(v)$ is the adjacency matrix of the node v ; \bigoplus is a multi-scale scaling and aggregation operator that satisfies:

$$\bigoplus = \begin{bmatrix} I \\ S(D_{r \times r}, \alpha = 1) \\ S(D_{r \times r}, \alpha = -1) \end{bmatrix} \otimes \begin{bmatrix} \mu \\ \sigma \\ max \\ min \end{bmatrix} \quad (\text{Equation 22})$$

where \otimes is the tensor product, I is the identity matrix, $D_{r \times r}$ is the degree matrix of nodes, and r is the number of nodes in a patch; α is the scaling factor, shrinking is negative, magnifying is positive; μ , σ , max , min are the mean value, standard deviation, maximum value and minimum value of node features respectively; S is the degree scaling matrix and meets:

$$S(D_{r \times r}, \alpha) = \left(\frac{\log(D_{r \times r} + I)}{\delta} \right)^\alpha \quad (\text{Equation 23})$$

where δ is the normalized index calculated based on training data, and satisfy:

$$\delta = \frac{1}{T_n} \sum_{i=0}^{r-1} \sum_{j=0}^{r-1} \log(D_{r \times r}(i, j) + 1) \quad (\text{Equation 24})$$

where T_n is the number of nodes for each input of training data.

Patch-level feature extraction and slide-level feature construction

The ResNet 50 network model,⁵² pre-trained on ImageNet, serves as the patch-level feature extractor in our experiment. Following the third residual module of the network, we apply global average pooling to transform the input patch of dimensions 512×512 into a 1×1024 output feature vector, which is then aggregated into slide-level features. Initially, all patches are assigned weights based on the multi-head attention mechanism, as per Equation 25. Subsequently, we select c samples ($c = 8$) with the highest attention values as the training samples for the

nuclei classifier. Finally, the attention pooling method is employed to aggregate patch-level features into slide-level characteristics, as indicated by [Equation 26](#).

$$s_i = \frac{\exp\{\tau(SA_a p_i^\top) \odot \xi(SA_b p_i^\top)\}}{\sum_{i=1}^N \exp\{\tau(SA_a p_i^\top) \odot \xi(SA_b p_i^\top)\}} \quad (\text{Equation 25})$$

$$f_{slide} = \sum_{i=1}^N s_i p_i \quad (\text{Equation 26})$$

where s_i is the weight of the i -th patch ($i = 1, 2, \dots, N$); p_i is features vector of the i -th patch; f_{slide} is the slide-level features vector. SA_a , SA_b indicates the multi-head attention, and meets:

$$SA(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_K}}\right)V \quad (\text{Equation 27})$$

$$\text{softmax}(z) = e^{z_i} / \sum_{i=1}^n e^{z_i} \quad (\text{Equation 28})$$

sofamx is the normalized function. $QK^\top / \sqrt{d_K}$ is the self-attention matrix, performing linear transformations $K = W^K X$, $Q = W^Q X$, $V = W^V X$ on the input image X to obtain Q (Queries), K (Keys), V (Value); W^Q , W^K and W^V are the learnable transformation matrix; $\tau(x)$, $\xi(x)$ are the activation function, and meets:

$$\tau(x) = \frac{1}{1+e^{-x}}; \xi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (\text{Equation 29})$$

Classify WSIs based on nuclei-level handcrafted feature constraint

The attention mechanism serves to identify potential ROI (Regions of Interest, ROIs). However, it's important to note that a region with a high attention score may not necessarily indicate a diseased area, and conversely, a region with a low attention score may not always be normal. This implies that selecting Top-K ROIs based on attention scores introduces noise.⁵⁶ To address this, this paper employs the noise-robust APL loss function⁵³ for both patch-level and nuclei-level, as per [Equation 30](#), and utilizes the cross-entropy error function from [Equation 31](#) for slide-level losses. The slide-level losses are then constrained by incorporating patch-level and nuclei-level losses as penalty terms to formulate the optimization objective function in [Equation 32](#). Finally, a WSI classification model is constructed by approximating the minimum value of [Equation 32](#) using the Adam optimizer.⁵⁴

$$\mathcal{L}_{patch,nuclei} = \log \prod_{k=1}^K p(k|x) - \sum_{k=1}^K p(k|x) \log(q(k|x)) \quad (\text{Equation 30})$$

$$\mathcal{L}_{slide} = - \sum_{i=1}^K q(k|x) \log(p(k|x)) \quad (\text{Equation 31})$$

$$\mathcal{L}_{total} = w_1 \mathcal{L}_{slide} + w_2 \mathcal{L}_{patch} + w_3 \mathcal{L}_{nuclei} \quad (\text{Equation 32})$$

where $q(k|x)$ is the truth distribution; $p(k|x)$ is the predicted value distribution; \mathcal{L}_{slide} , \mathcal{L}_{patch} , \mathcal{L}_{nuclei} , and \mathcal{L}_{total} are slide-level, patch-level, nuclei-level, and total losses, respectively. In this paper, let $w_1 = 0.7$, $w_2 = 0.2$, $w_3 = 0.1$.

Evaluation criteria

The commonly used classification evaluation criteria are accuracy (AC), specificity (SP), which measures an algorithm's ability to recognize normal WSIs, sensitivity (SE), also known as recall rate, which measures the power of the algorithm to identify cancerous WSIs, and precision (PC). Giving GT (Ground Truth) as the truth label, denote normal WSI as GT = 0 and cancerous one as GT = 1. Let PreCls (Predicted Class) be the predicted result of the algorithm, then:

$$AC = \frac{TP+TN}{TP+TN+FP+FN}; SP = \frac{TN}{TN+FP} \quad (\text{Equation 33})$$

$$SE = \frac{TP}{TP+FN}; PC = \frac{TP}{TP+FP} \quad (\text{Equation 34})$$

where TP (True Positive) means GT = 1 and PreCls = 1; TN (True Negative) means GT = 0 and prelude = 0; FN (False Negative) means GT = 1 and PreCls = 0; FP (False Positive) means GT = 0 and PreCls = 1.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical details, such as accuracy and sensitivity, can be located in the results, methods, and/or figure legends.