scientific reports

Check for updates

OPEN

Enhanced digital pathology image recognition via multi-attention mechanisms: the MACC-Net approach

Feng Liu[1,2,3], Zheng Wang[1,2,3], Baotian Li[1,2,3], Decai Wang[4]✉, Mingyu Liu[7]✉, Fangfang Gou[5] & Jia Wu[5,6]✉

Digital pathology has revolutionized cancer diagnosis through microscopic analysis, yet manual interpretation remains hindered by inefficiency and subjectivity. Existing deep models for osteosarcoma cell nucleus recognition suffer from the difficulty of capturing hierarchical relationships in single-dimensional attention mechanisms, leading to inaccurate edge recognition. Furthermore, the fixed receptive field of CNNs limits the aggregation of multi-scale information, hindering the differentiation of overlapping cells. This study introduces MACC-Net, a novel multi-attention based method designed to enhance the recognition accuracy of digital pathology images. By integrating channel, spatial, and pixel-level attention mechanisms, MACC-Net overcomes the limitations of traditional single-dimensional attention models, improving feature consistency and receptive field expansion. Experimental results demonstrate a Dice Similarity Coefficient (DSC) of 0.847, highlighting MACC-Net's potential as a reliable auxiliary diagnostic tool for pathologists. Code: https://github.com/GFF1228/MACCNet.

Keywords Multiple attention, Digital pathology images, Multi-scale aggregation, Image segmentation, Auxiliary diagnosis

Pathological diagnosis is the "gold standard" for tumor diagnosis, and its accuracy directly impacts clinical treatment decisions[1]. Pathologists use microscopes to observe cell morphology and distribution, distinguishing between normal cells, inflammatory cells, and cancer cells, and determine the presence, type, grade, and depth of invasion of tumors[2,3]. This is particularly true for the diagnosis of highly malignant bone tumors such as osteosarcoma, where accurate pathological analysis directly influences surgical and chemotherapy planning. Therefore, this study used digital pathology images of osteosarcoma as a test target to validate the model's performance.

Whole-slide imaging (WSI) technology has enabled the digitization of pathology images, but manual interpretation still faces challenges such as low efficiency, high subjectivity, and decreased diagnostic consistency due to fatigue[4–6]. For complex pathology images, pathologists must switch between different magnifications to identify key areas and make judgments, a time-consuming and error-prone process[7,8]. Pathology diagnosis is expensive in terms of equipment, labor, and time, making these challenges even more pronounced in resource-limited settings[9].

Artificial intelligence technology provides a new approach to improve this situation[10–12]. Through image segmentation algorithms, artificial intelligence systems can decompose complex pathological images into different tissue components, such as tumor tissue and normal tissue, thereby improving diagnostic efficiency[13,14]. Among them, convolutional neural networks (CNNs) have achieved remarkable success in tasks such as cell nucleus segmentation, tumor grading, and metastasis detection[15–17]. Attention mechanisms perform well in

[1]School of Information Engineering, Shandong Youth University of Political Science, Jinan, China. [2]New Technology Research and Development Center of Intelligent Information Controlling in Universities of Shandong, Jinan, China. [3]Smart Healthcare Big Data Engineering and Ubiquitous Computing Characteristic Laboratory in Universities of Shandong, Jinan, China. [4]Weifang Engineering Vocational College, Qingzhou, China. [5]State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang 550025, China. [6]Research Center for Artificial Intelligence, Monash University, Clayton, Melbourne, VIC 3800, Australia. [7]Sdu-Anu Joint Science College, Shandong University, Weihai 264209, China. ✉email: 304610599@qq.com; sdumyliu@163.com; jiawu5110@163.com

enhancing the interpretability of models by focusing on diagnostically relevant areas[18–20]. For example, Li Xiaorong et al.[17] proposed a CNN architecture that combines iterative attention feature fusion and residual modules, aiming to accurately segment overlapping and multi-scale cell nuclei.

However, existing deep learning models still have many limitations in the recognition of osteosarcoma pathology images[21]. The single-dimensional attention mechanism cannot fully identify the hierarchical relationship between osteosarcoma cell nuclear features and the global pathological tissue background, resulting in low accuracy in cell nuclear edge recognition[22,23]. Osteosarcoma pathology images contain many overlapping cell regions. The fixed receptive field of CNNs hinders the aggregation of multi-scale background information, making it difficult to distinguish overlapping and aggregated cells and tissues[24].

Based on this, this study is based on the Multiple Attention Guided Assisted Recognition of Digital Pathology Images (MACC-Net), which aims to effectively assist doctors to accurately recognize cancerous tissues in digital pathology images. Its core innovation mainly consists of three modules: hybrid attention feature extraction module (HAFEM), cascade context fusion module (CCIFM), and attention map dynamic balancing module (AMDBM). The HAFEM module maintains nuclear feature consistency through fused channel, spatial, and pixel-point attentions in histopathological imagery. The CCIFM preserves feature uniformity by employing adaptive average pooling and non-local operations to capture global context and expand the model's receptive field for multi-scale integration. Meanwhile, the AMDBM dynamically modulates feature weights via fused foreground, background, and boundary attention maps, prioritizing target-relevant features while suppressing irrelevant information. MACC-Net overcomes the representation bottleneck of single-dimensional attention and expands the model's receptive field, enabling the fusion of multi-scale contextual information and significantly improving the accuracy of cell nucleus edge recognition. It can more accurately identify pathological images of cancerous tissue, providing doctors with a reliable auxiliary diagnostic tool.

The rest of this paper is organized as follows: "Methods" introduces the proposed method in detail, "Results" presents the experimental results, "Discussion" discusses and analyzes the proposed method in relation to current research, and "Conclusion and outlook" summarizes the clinical implications and future directions.

Methods

This study proposes a multi-head attention mechanism-based digital pathology image assisted recognition network (MACC-Net) to achieve accurate segmentation of cell nuclei in pathological sections. The overall network structure is shown in Fig. 1. MACC-Net mainly consists of the following four modules.

(1) The Multi-scale Feature Extraction Module (MFEM) extracts semantic features at different scales using ResNet, capturing tissue representations across various hierarchical levels.
(2) The Hybrid Attention Feature Enhancement Module (HAFEM) simultaneously enhances lesion features using channel and spatial attention mechanisms, suppressing background interference.

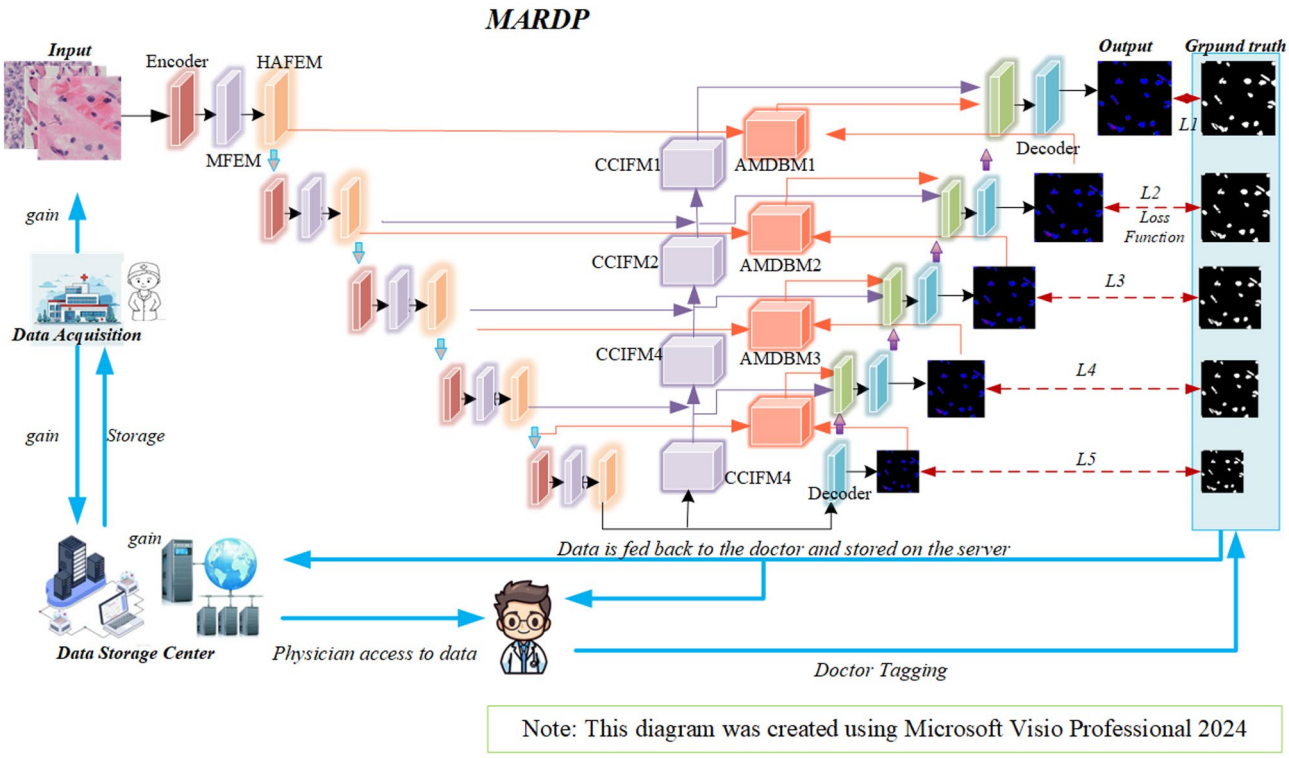


Fig. 1. Overall architecture of MACC-Net.

(3) The Cascaded Context Integration and Fusion Module (CCIFM) performs semantic cascading of features at each scale, expands the receptive field via dilated convolutions, and enables large-scale contextual modeling.
(4) The Attention-Modulated Dynamic Balancing Module (AMDBM) dynamically adjusts feature weights by calculating and fusing foreground, background, and boundary attention maps. This enhances focus on task-relevant features while suppressing irrelevant information.

Multi-scale feature extraction module (MFEM)

In digital pathological images, cells exhibit significant morphological heterogeneity, manifested not only as structural variations across multiple scales but also as prominent characteristics (e.g., blurred nuclear edges and discontinuous textures). These complexities substantially increase the difficulty of segmentation tasks[25–27]. To address this challenge, effectively extracting and fusing multi-scale feature information has become a critical research focus. In this paper, we propose a Multi-scale Feature Extraction Module (MFEM) to enhance the model's ability to perceive tumor regions at diverse scales.

The MFEM module is based on a lightweight backbone network and adopts a hierarchical feature extraction approach to capture multi-scale information from low-level textures to high-level semantics layer by layer. Assuming the input pathological image is $I \in \mathbb{R}^{H\times W\times 3}$, the backbone network is first used to extract four-level semantic feature representations:

$$\{F_1, F_2, F_3, F_4\} = Backbone\,(I) \tag{1}$$

Among them, $F_i \in \mathbb{R}^{H_i\times W_i\times C_i}$ respectively represent the feature maps extracted at different scales. As the network hierarchy deepens, the size of the feature maps gradually decreases while the semantics are increasingly enhanced. Specifically, F_1 captures the edge and texture features in the original image, F_2 and F_3 focus on the local structures and mid-level semantics, and F_4 represents the global high-level semantic abstraction.

The MFEM enhances multi-scale perception by alternating dilated and strided convolutions, balancing receptive field size and computational efficiency. This adaptability is crucial for lesions with varying morphologies and densities. Additionally, shallow features are retained to improve nuclear boundary detection, particularly for small or irregular nuclei.

Additionally, to avoid excessive loss of information flow during the downsampling process, the MFEM module draws inspiration from the U-Net architecture design concept and retains a skip connection mechanism during feature extraction. This enables shallower-layer features to re-participate in the feature fusion process during subsequent fusion stages, effectively preserving edge and texture information of tumor regions.

Batch Normalization and ReLU activation functions are introduced after the extraction of each level of feature maps, ensuring stable gradient propagation and nonlinear enhancement during the training process. The specific feature extraction process for each level can be expressed as:

$$F_i = ReLU\,(BN\,(Conv_{3\times 3}\,(F_{i-1}))) \tag{2}$$

Among them, $Conv_{3\times 3}$ denotes standard convolution operations, BN represents batch normalization operations, and ReLU refers to activation functions.

Multi-attention enhancement module (HAFEM)

Digital pathology images contain a rich background of non-lesional tissues[28–30]. CNN's receptive field and local feature extraction capabilities are insufficient, making it difficult to effectively distinguish tumor areas. To this end, this paper introduces a multi-attention enhancement module (HAFEM) to enhance the network's response to salient cell tissue regions and improve the expression selectivity of feature channels. As shown in Fig. 2, The HAFEM module is primarily composed of three complementary attention mechanisms: Channel Attention focuses on "which feature channels are important," while Spatial Attention emphasizes "which positions in the image are more critical," extracting nuclear boundary features. Pixel-wise Attention captures features of smaller nuclei or inter-nuclear gaps. This module applies weighted calculations of channel and spatial dimensions to each input feature map $F_i \in \mathbb{R}^{H_i\times W_i\times C_i}$, resulting in an enhanced representation F_i'.

$$F_i' = \mathrm{Con}_{1\times 1}\left(F_i^{(s)} \oplus F_i^{(c)} \oplus F_i^{(p)}\right), \mathrm{i} = 1,2,3,4,5 \tag{3}$$

Among them, F_i' denotes the unified attention feature map generated by the multi-attention module, while $\mathrm{Con}_{1\times 1}$ reduces feature channel dimensionality following the activation layer, $F_i^{(s)}$ is the spatial attention feature of the pathological image, $F_i^{(c)}$ is the channel attention feature of the pathological image, and $F_i^{(p)}$ is the pixel-wise attention feature of the pathological image.

The channel attention module leverages global contextual information to guide the importance modeling of feature channels. Specifically, global average pooling and max pooling are first performed on each channel to form two one-dimensional descriptive vectors.

$$f_{avg} = \mathrm{AvgPool}\,(F_i) \tag{4}$$

$$f_{max} = \mathrm{MaxPool}\,(F_i) \tag{5}$$

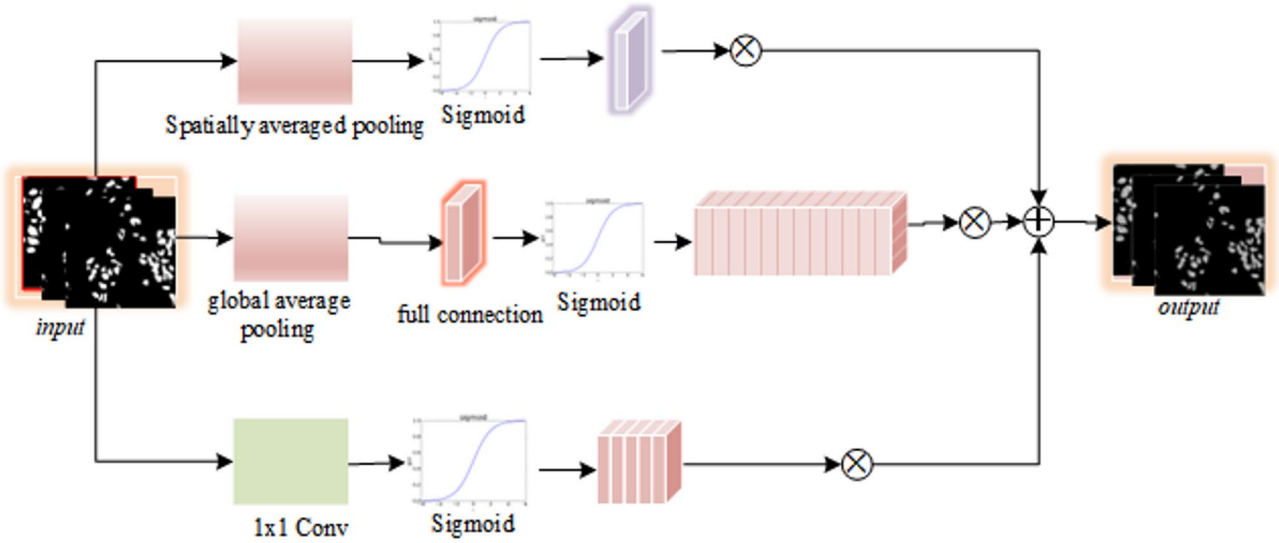


Fig. 2. HAFEM module structure diagram.

Subsequently, a shared multi-layer perceptron (MLP) is used for non-linear mapping, outputting the channel attention weight vector M_c.

$$M_c = \sigma\ (\mathrm{MLP}\ (f_{avg}) + \mathrm{MLP}\ (f_{max})) \tag{6}$$

The final channel attention response is expressed as:

$$F_i^{(c)} = M_c \otimes\ F_i \tag{7}$$

where σ denotes the Sigmoid activation function, and $\otimes$ represents the element-wise multiplication operation between channels.

To further enhance the discriminability of features in spatial positions, the HAFEM module introduces a Spatial Attention mechanism, which aims to detect the most discriminative regional positions in the image. Channel-attentive feature maps undergo max pooling and average pooling operations along the channel axis, producing two distinct spatial context maps.

$$f_{avg}^s = \mathrm{AvgPool}_c\left(F_i^{(c)}\right) \tag{8}$$

$$f_{max}^s = \mathrm{MaxPool}_c\left(F_i^{(c)}\right) \tag{9}$$

Both maps undergo concatenation before being processed by a 7×7 convolutional layer, yielding the spatial attention map M_s.

$$M_s = \sigma\ \left(\mathrm{Conv}_{7\times\ 7}\left(\left[f_{avg}^s; f_{max}^s\right]\right)\right) \tag{10}$$

The final spatially enhanced feature representation $F_i^{(s)}$ is expressed as:

$$F_i^{(s)} = M_s \otimes\ F_i^{(c)} \tag{11}$$

The pixel-wise attention feature $F_i^{(p)}$ is calculated as:

$$F_i^{(p)} = \sigma\ \left(Conv_{1\times\ 1}\left(e_i\right)\right), i = 1,2,3,4,5 \tag{12}$$

The three attention mechanisms described above respectively model the semantic importance of channel dimensions, the saliency of spatial regions, and the significance of small-scale tissue features, thereby collaboratively enhancing the model's focusing ability on cell nucleus regions and suppressing background interference. In terms of structure, the HAFEM module is lightweight with a small number of parameters, making it easy to couple with the backbone network and suitable for promotion in high-resolution pathological image scenarios.

Cascaded context integration and fusion module (CCIFM)

Cell nuclei exhibit diverse morphologies, irregular distribution, blurred boundaries, and clustered or overlapping distributions, making their differentiation heavily dependent on contextual information[31,32]. After multi-scale feature extraction and attention enhancement, it becomes crucial to further integrate information from different levels to build a globally consistent understanding of the lesion. To this end, this paper designs a Cascaded Context Integration and Fusion Module (CCIFM) to effectively fuse shallow and deep semantics, enhance the modeling of contextual dependencies, and thus improve the structural integrity and boundary accuracy of the final segmentation prediction. The detailed structure of the CCIFM module is shown in Fig. 3, which feeds the multi-scale feature maps F_i' from the HAFEM module into four branches for operations to extract global contextual information. The initial three branches first perform adaptive average pooling on feature map F_{i+1}', with pooling kernel dimensions defined as:

$$F_c AdPdo\,(L, j) = (2j - 1)(6 - L) \tag{13}$$

$$L = 1,2,3,4, j = 1,2,3 \tag{14}$$

Here, L indicates the total layers in the CCIFM module, while j corresponds to the index of its first three branches. Each initial branch employs 1×1 convolutions to compress feature channels to 25% of original dimensionality, followed by upsampling to restore spatial resolution. For branch 4, a 1×1 convolution embeds each feature map pixel into vector space, where non-local operations model pairwise pixel relationships to capture long-range dependencies, ultimately deriving deep semantic feature representations.

Finally, assuming the upsampling operation is $\mathcal{U}(\bullet)$, the fused features can be expressed as:

$$F_c = \text{Concat}\left(\mathcal{U}\left(F_1'\right), \mathcal{U}\left(F_2'\right), \mathcal{U}\left(F_3'\right), F_4'\right) \tag{15}$$

The above process aligns and concatenates representations of different levels in the spatial dimension, enabling the model to simultaneously obtain edge details (from shallow layers) and high-level semantic structures (from deep layers) during the fusion stage.

To avoid information conflicts or redundant interference during the fusion process, the concatenated features F_c are passed through a convolutional module for dimension compression and semantic fusion:

$$F_m = \text{Conv}_{3\times 3}\left(F_c\right) \tag{16}$$

This operation not only reduces the channel dimension but also enhances the coupling and integration capability between features, forming a more discriminative fused representation.

Attention map dynamic balance module (AMDBM)

Digital pathological images often contain interference from fragmented cell nuclei, blank backgrounds, and noise such as shadows and lesions[33–36]. The Attention Map Dynamic Balance Module (AMDBM) calculates attention maps for cell nucleus foregrounds, backgrounds, and boundaries respectively, enabling the network to focus more on the edge details of cell nuclei, eliminating interference from background noise during nucleus segmentation, and making lesion boundaries clearer and smoother. Figure 4 illustrates the AMDBM module's architecture. The lesion boundary attention score SC_{bo} generated by the decoder undergoes pixel-wise multiplication with the HAFEM module's output feature map F_c, producing the boundary-enhanced feature map F_{bo}. This operation is formalized as:

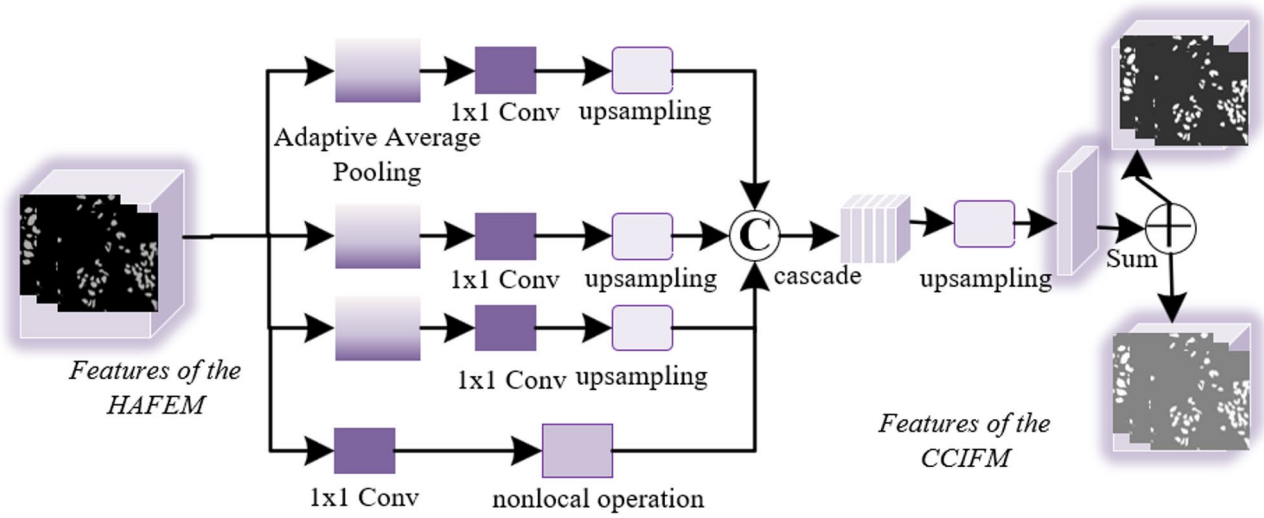


Fig. 3. Structure of the CCIFM module.

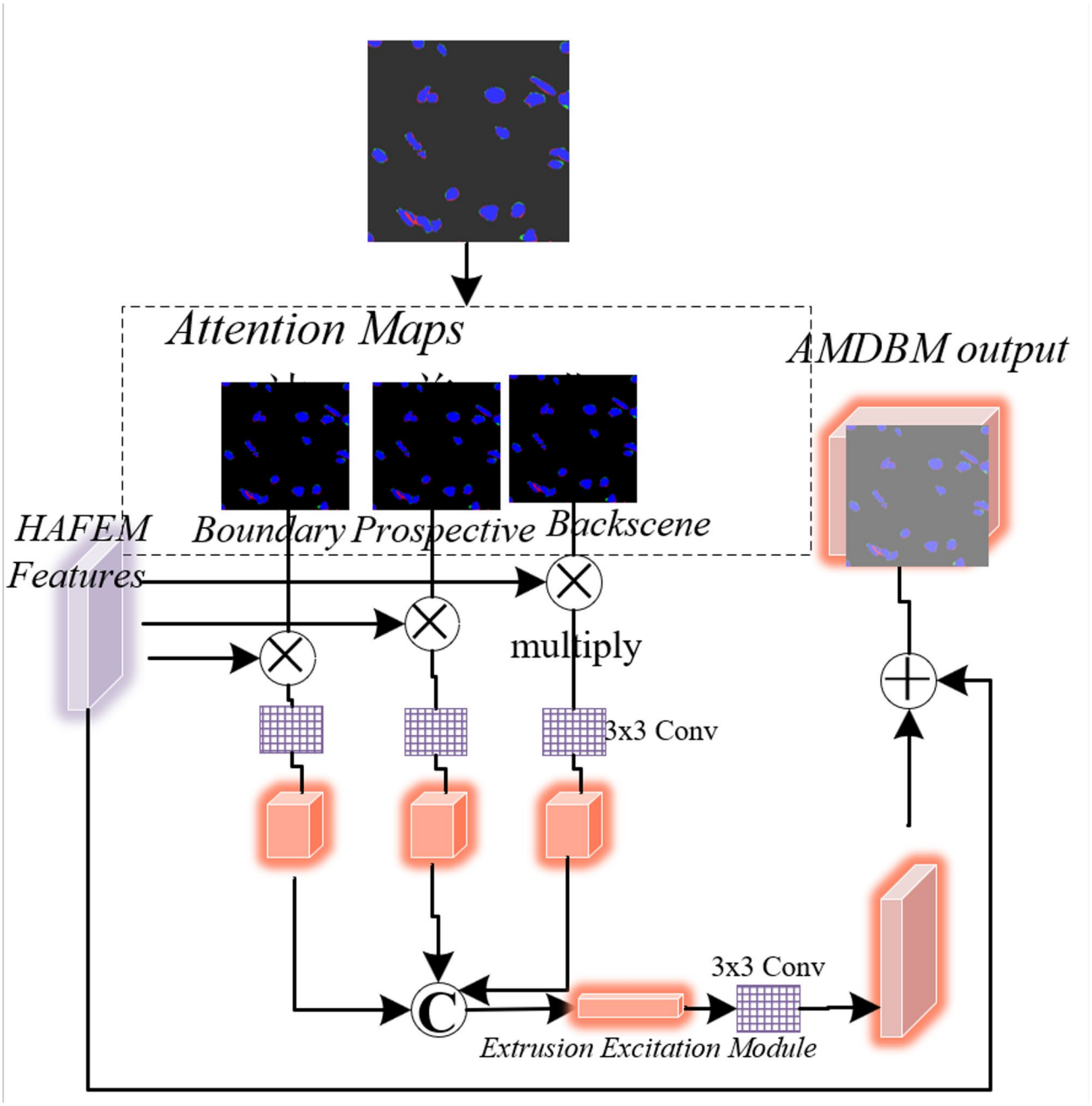


Fig. 4. Structure of the AMDBM module.

$$SC_{bo} = 1 - \frac{|\sigma\ (pred) - 0.5|}{0.5} \tag{17}$$

$$F_{bo} = F_c \otimes\ SC_{bo} \tag{18}$$

where pred represents the nucleus segmentation prediction result output by the decoder.

The decoder-generated foreground attention score SC_{fg} is then multiplied pixel-wise with the HAFEM's output feature map F_c, yielding the foreground-enhanced feature map F_{lg}, formalized as:

$$SC_{fg} = [1 - \sigma\ (pred) - SC_{bo}] \tag{19}$$

$$F_{lg} = F_c \otimes\ SC_{fg} \tag{20}$$

Here, $[]_0^1$ clips attention scores to the [0,1] range. Subsequently, the decoder-derived background attention score SC_{bg} undergoes pixel-wise multiplication with the AMDBM module's output feature map F_c, yielding the background-enhanced feature map F_{bg}, formalized as:

$$SC_{bg} = [1 - \sigma\ (pred) - SC_{bo}] \tag{21}$$

$$F_{bg} = F_c \otimes\ SC_{bg} \tag{22}$$

The three feature maps are concatenated and fused, then processed through a squeeze-and-excitation module coupled with 3×3 convolutions to produce the balanced attention output B_i. The squeeze-and-excitation module reweights feature channels to equilibrate attention across the three regions.

As Fig. 5 depicts, given the significance of local-global context for multi-scale lesions, an Adaptive Feature Selection Module (AFSM) integrates: AMDBM's attention-balanced features B_i, CCIFM's contextual features F_m, and prior decoder outputs $d_i + 1$. And the AFSM's resultant features feed into the decoder to enable adaptive contextual feature selection.

The AFSM module is defined as:

$$F_a = \mathrm{Concat}\left(\mathrm{Conv}_{1\times 1}\left(F_m\right), \mathrm{Conv}_{3\times 3}^{r=6}\left(F_m\right), \mathrm{Conv}_{3\times 3}^{r=12}\left(F_m\right), \mathrm{Conv}_{3\times 3}^{r=18}\left(F_m\right), \mathrm{GlobalAvgPool}\left(F_m\right)\right) \tag{23}$$

where $Conv_{3\times 3}^{r=k}$ represents a dilated convolution with dilation rate *k*. The global average pooling branch is used to introduce image-level context, guiding the network to focus on the overall structure.

Finally, after fusing all scale information, a compression mapping is performed:

$$F_{out} = \mathrm{Conv}_{1\times 1}\left(F_a\right) \tag{24}$$

To obtain the final fused representation F_{out}, which contains rich contextual structure information and multi-scale spatial dependencies, suitable for precise segmentation of complex tumor tissues.

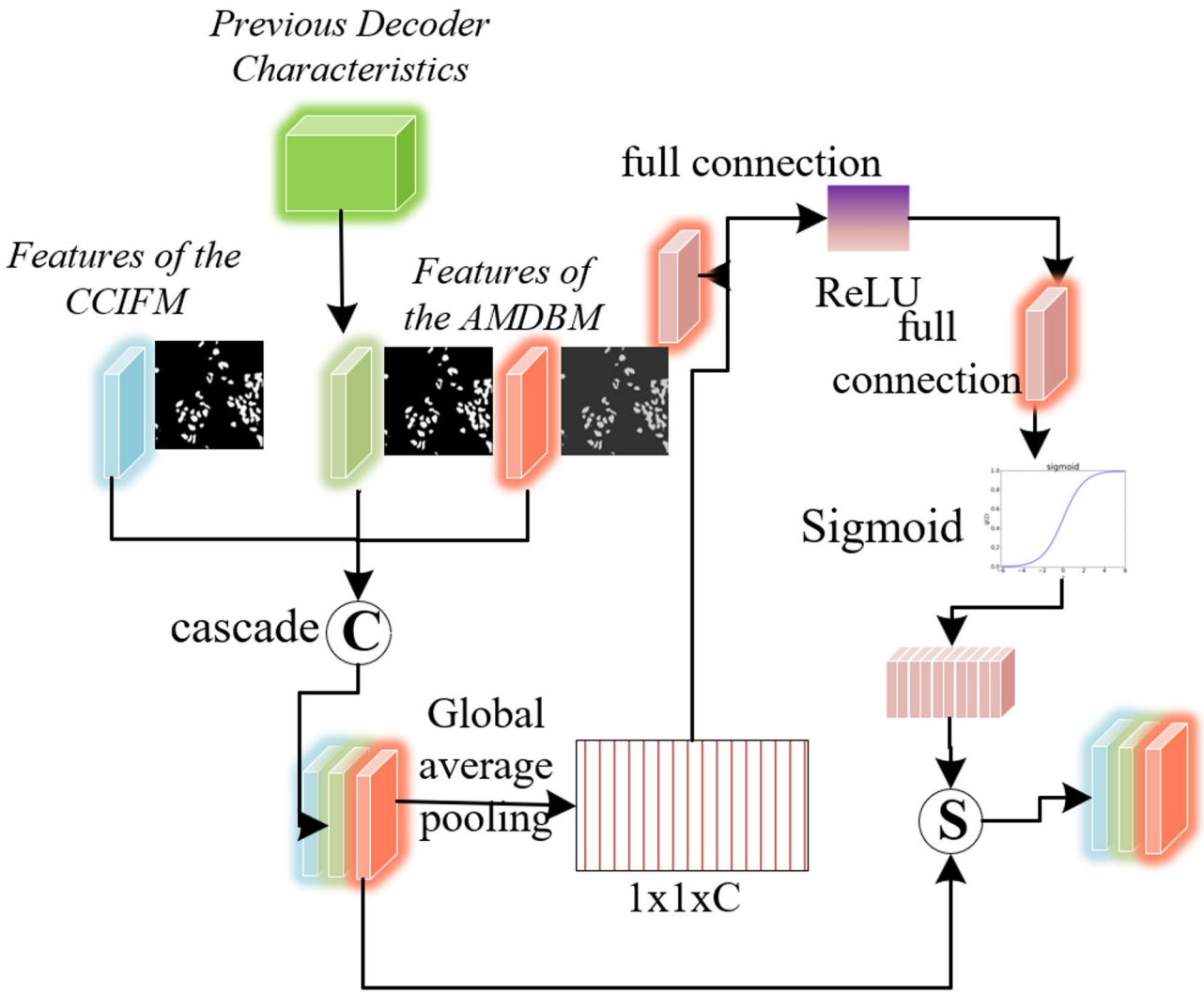


Fig. 5. Structure of the AFSM module.

Loss function

The fused feature F_{out} is mapped to a probability map through the prediction head.

$$P = \sigma\ (\mathrm{Conv}_{1\times 1}(F_{out})) \tag{25}$$

where σ is the Sigmoid activation function, used to output the probability value of each pixel being a lesion. To improve the quality of boundary segmentation, this paper adopts a combined loss function, including Dice loss and BCE (Binary Cross-Entropy) loss.

$$\mathrm{L} = \lambda \cdot L_{\mathrm{Dice}} + (1-\lambda)L_{\mathrm{BCE}} \tag{26}$$

This combined strategy maintains the consistency of the overall region while enhancing the boundary discrimination ability, where $\lambda \in [0,1]$ controls the weight of the two.

Results

Experimental setup

Osteosarcoma digital pathology image dataset: This study used the digital pathology dataset of osteosarcoma from the Institute of Artificial Intelligence at Monash University[37], which contains 1000 whole-slice images. Each slice was magnified 40 times, and 512 × 512 sub-images were obtained by random area sampling. After manually removing contaminated or completely blank images, 2164 images were retained for the experiment. Three pathology students then annotated them to ensure the accuracy of the labels. Finally, this study divided the training set, validation set, and test set into a ratio of 7:1:2.

In addition, in order to verify the performance of the model at lower magnifications, this study magnified the original digital pathology images by 10× and 20× for comparative experiments. Using the same sub-image acquisition method, 1200 10× sub-images and 1200 20× sub-images were obtained respectively. Images of different magnifications are shown in Fig. 6. This study used the same ratio to divide the training set, validation set, and test set.

Public breast cancer dataset TNBC[38]: This data is a collection of digital pathology images selected from triple-negative breast cancer. It contains 33 512 × 512 images.

Experimental Environment: Experiments were conducted on a server with: NVIDIA RTX 3090 GPU, Intel Core i9-12900 K CPU, 64GB RAM, Ubuntu 20.04, and Python 3.8, PyTorch 1.12, CUDA 11.6.

Training process: The Adam optimizer was adopted with an initial learning rate of 1e−4. The training batch size of the model was 32 and a total of 200 training cycles were performed. All input images were normalized before training, while a weighted cross-entropy loss function was chosen to effectively deal with the problem of category imbalance. All experiments were repeated five times and averaged.

Evaluation metrics: In order to provide a more comprehensive and precise quantitative analysis of the segmentation effects of various models, this paper introduces the confusion matrix as an evaluation tool. Based on the above confusion matrix, several key performance indicators can be calculated for evaluating the model performance: accuracy (Acc), precision (Pre), recall (Rec), intersection-union ratio (IoU), Dice Similarity Coefficient (DSC), parameter counts (params), and FLOPs[39].

Analysis of results

Figure 7 visually compares osteosarcoma nuclei segmentation results on the test set. Prediction results from multiple sets of images revealed that models like Swin-Unet and TransUnet suffer from missed detections, especially when cell nuclei are densely distributed. The CE-Net model is also more prone to over-segmentation, misclassifying other tissues as cell nuclei, as shown in the third row of figures. Our proposed model, however, excels in cell nucleus recognition, with predictions highly consistent with ground-truth results. It demonstrates robustness and accuracy when dealing with nuclei with complex morphologies or blurred boundaries.

This study quantitatively analyzes the performance of each model in pathological images. The proposed model demonstrating superior segmentation accuracy and higher prediction-label consistency. As detailed in Table 1 (best-performing values bolded), our approach outperforms all counterparts in overall metrics, notably achieving a DSC of 0.847—markedly exceeding other models. The ChannelNet model also performs well, with a DSC value of 0.841. The CE-Net model has the lowest DSC, at only 0.739.

To verify the impact of magnification on model performance, this study selected some models for comparative analysis. As can be seen from Table 2, when the magnification decreases, the image details are significantly reduced, the cell boundaries and microstructures are blurred, the segmentation difficulty increases, and all indicators will decrease. When the digital pathology image is magnified 10 times, the indicators of each model are poor, among which the DSC values of the U-Net and Att-Unet models are lower than 0.7. When the magnification is 20 times, the performance of the model is relatively improved. The DSC value of the ChannelNet model is only 0.801 at 10×; at 20×, it reaches 0.819; at 40×, it reaches 0.841. The DSC value of the MACC-Net model also increased by 4.3% from 0.803. Overall, when the image is magnified 40 times, the performance of the MACC-Net model is the best, and the resource cost of extracting many high-resolution tiles is reasonable. Therefore, the resource cost of extracting many high-resolution images is reasonable.

Table 3 shows the nucleus segmentation performance of some models in digital pathology images of different tumors. In the TNBC dataset, the performance of each model was relatively low. The CE-Net model achieved a DSC value of only 0.731. In osteosarcoma, the CE-Net model achieved a DSC value of 0.739. The DSC value of the model in this study also dropped to that of the CE-Net model in the TNBC dataset.

To more intuitively and comprehensively demonstrate the differences in segmentation performance among different models, this study used pathology images magnified 40× as a benchmark to analyze the relationship

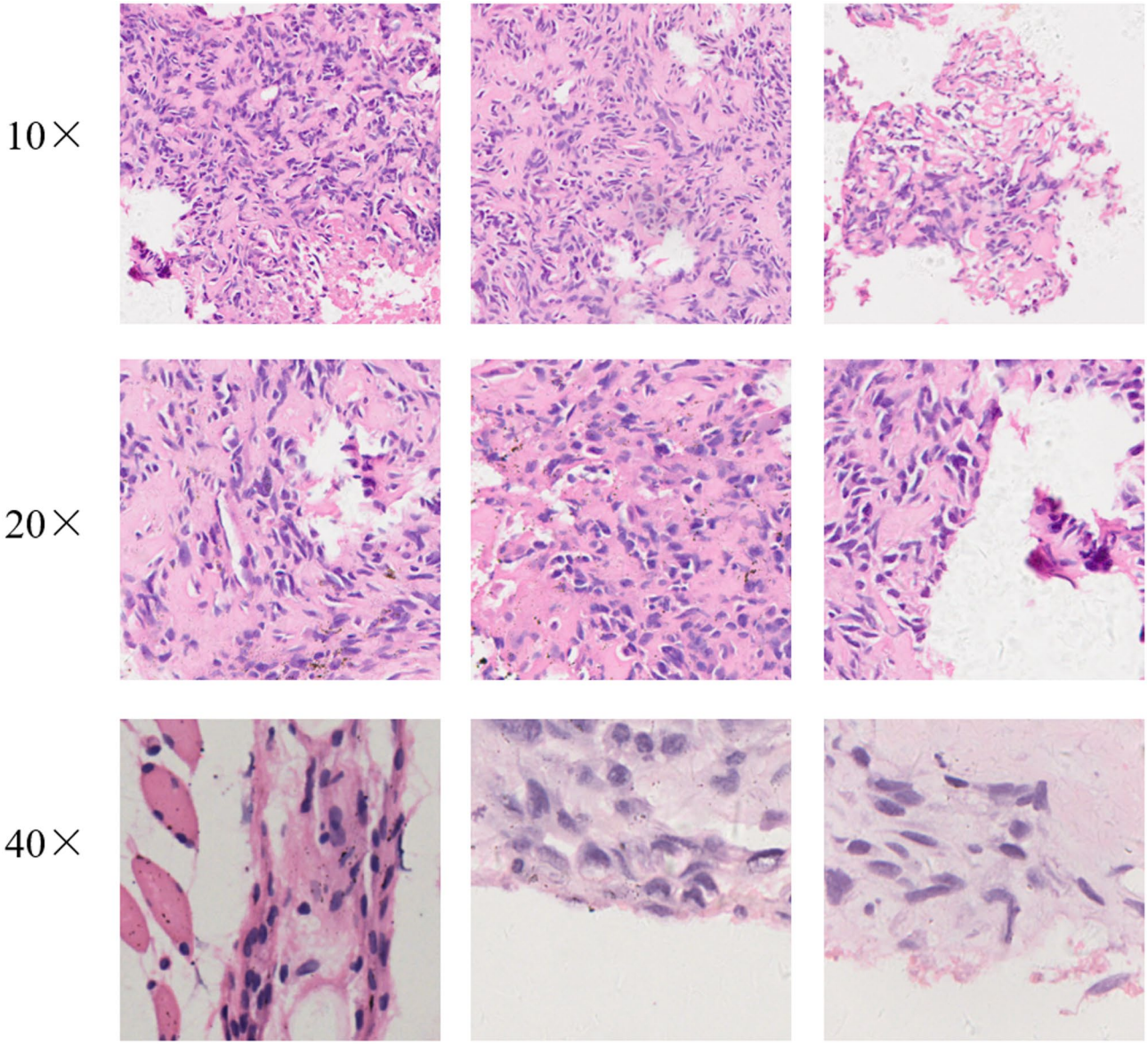


Fig. 6. Digital pathology slides at different magnifications.

between model performance and complexity. Figures 8 and 9 show the relationship between the DSC metric, model parameters, and FLOPs for each segmentation model. As shown in Fig. 8, the MACC-Net model achieves the highest DSC value of all models, reaching 0.847, demonstrating its strong segmentation performance. Furthermore, the MACC-Net model has only 57.65 parameters. Although the UNet and CE-Net models both have well under 10 M parameters, they also perform relatively poorly in terms of segmentation accuracy (DSC and IOU). Furthermore, while the AttUnet model achieves a DSC of 0.825 and an IOU of 0.707, it uses 533 GB of FLOPs, the highest of all models. The SwinUnet model has the lowest FLOPs, at 11.74 GB. The model in this study is only slightly more computationally complex than SwinUnet, but achieves the highest DSC and IOU values. This relatively small increase in computational overhead yields the greatest improvement in accuracy, effectively balancing model accuracy and computational efficiency.

Ablation experiments

To verify the effectiveness of the modules of the proposed model, we set up eight groups of controlled experiments. Each group of experiments progressively removes or replaces a specific module to observe its effect on the overall performance. The specific experimental settings and results are detailed in Table 4 for the Multiscale Feature Extraction Module (MFEM), the Multiple Attention Enhancement Module (HAFEM), the Cascaded Contextual Fusion Module (CCIFM), and the Balanced Attention Module (AMDBM). From the table, it can be seen that the segmentation performance of the model is improved after adding different modules respectively. The performance improvement of the model is most obvious after combining the three modules. Figure 10 shows the comparison of the prediction effect of the model after adding different modules. The Baseline model

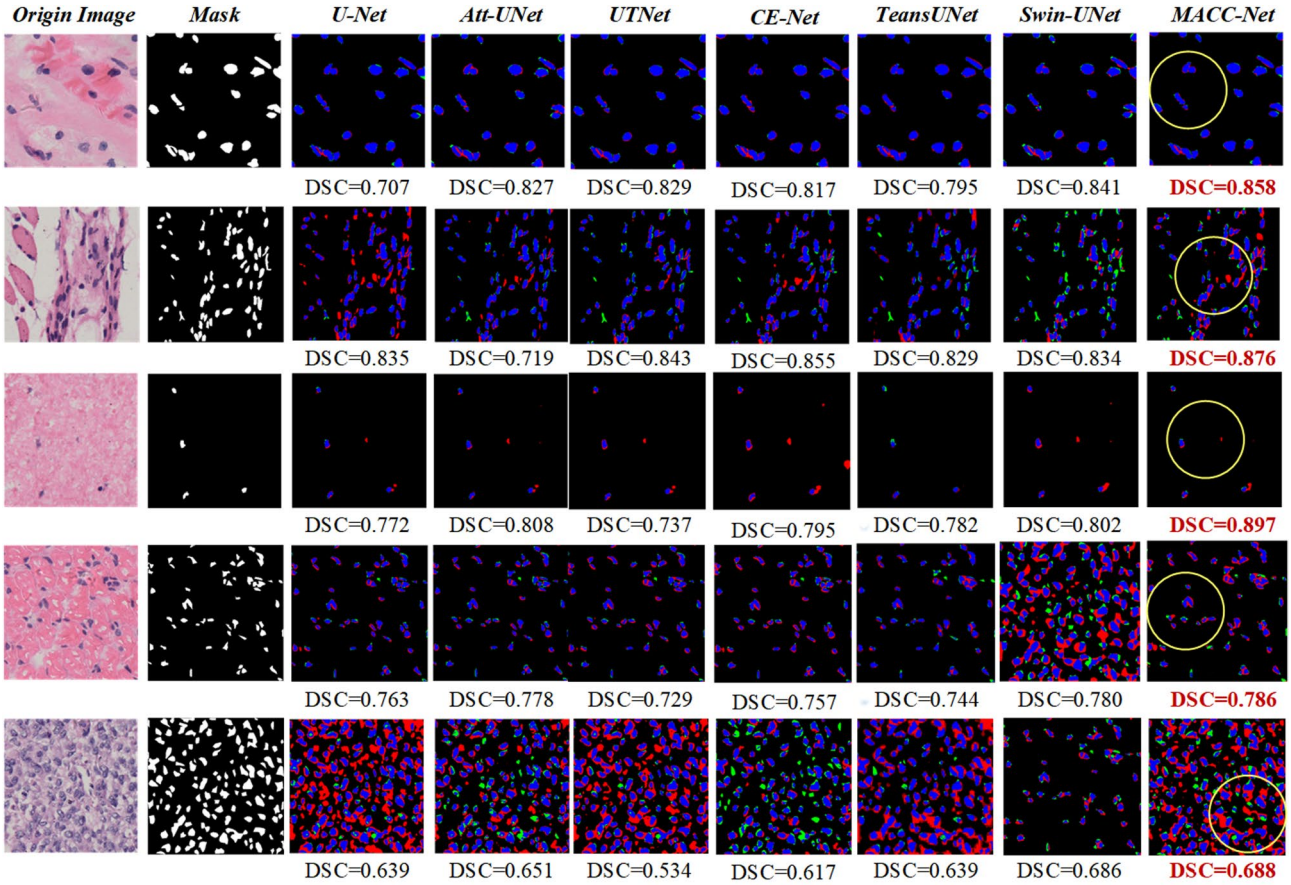


Fig. 7. Comparison of the predictive effectiveness of the models. Column 1 displays original pathology images, Column 2 shows ground-truth annotations, and subsequent columns present segmentation outputs from different models.

Model	Accuracy	Precision	Recall	F1-score	IOU	DSC
U-Net[41]	0.968	0.881	0.889	0.783	0.651	0.783
Att-Unet[42]	0.976	0.841	0.901	0.825	0.707	0.825
SegNet[43]	0.968	0.898	0.887	0.783	0.652	0.783
UTNet[44]	0.971	0.823	0.835	0.781	0.649	0.781
CE-Net[45]	0.945	0.741	0.832	0.739	0.662	0.739
TransUnet[46]	0.973	0.817	**0.929**	0.809	0.684	0.809
Swin-Unet[47]	0.976	0.856	0.923	0.822	0.702	0.822
ResUnet[48]	0.972	0.878	0.916	0.826	0.709	0.826
ChannelNet[49]	0.979	0.901	0.921	0.841	0.718	0.841
MACC-Net (our)	**0.981**	**0.907**	0.931	**0.847**	0.728	**0.847**

Table 1. Comparison of predictive indicators of individual models.

has poorer prediction effect and is more prone to misidentification problem. For the case where the background tissues are similar in color to the nuclei, each model may appear to determine the cell tissues as nuclei, as in the picture in the second row. Although the individual models are subject to errors when segmenting densely distributed cell nuclei, the models are relatively the best predictors with the addition of the HAFEM, CCIFM and AMDBM modules.

The performance effects on the model after the addition of different attention mechanisms are shown in Table 5. On the basis of ResNet 50, after adding spatial attention (SA), the model DSC metric improves by 4.6%; after adding only channel attention (CA), the DSC metric suggests 3.7%; after adding channel attention,

Model	Accuracy			Recall			DSC		
	10×	20×	40×	10×	20×	40×	10×	20×	40×
U-Net	0.932	0.955	0.968	0.791	0.846	0.889	0.698	0.750	0.783
Att-Unet	0.955	0.862	0.976	0.826	0.835	0.901	0.661	0.796	0.825
TransUnet	0.953	0.968	0.973	0.883	0.904	0.929	0.758	0787	0.809
Swin-Unet	0.956	0.961	0.976	0.876	0.900	0.923	0.782	0.807	0.822
ChannelNet	0.951	0.966	0.979	0.878	0.899	0.921	0.801	0.819	0.841
MACC-Net (Our)	0.961	0.975	0.981	0.887	0.908	0.931	0.803	0.826	0.847

Table 2. Model performance on datasets with different magnifications.

Model	Accuracy		Recall		DSC	
	40×	TNBC	40×	TNBC	40×	TNBC
SegNet	0.968	0.957	0.887	0.801	0.783	0.761
Swin-Unet	0.976	0.763	0.923	0.856	0.822	0.735
CE-Net	0.945	0.921	0.832	0.876	0.739	0.731
MACC-Net	0.981	0.972	0.931	0.886	**0.847**	0.823

Table 3. Model performance in different tumor pathology images.

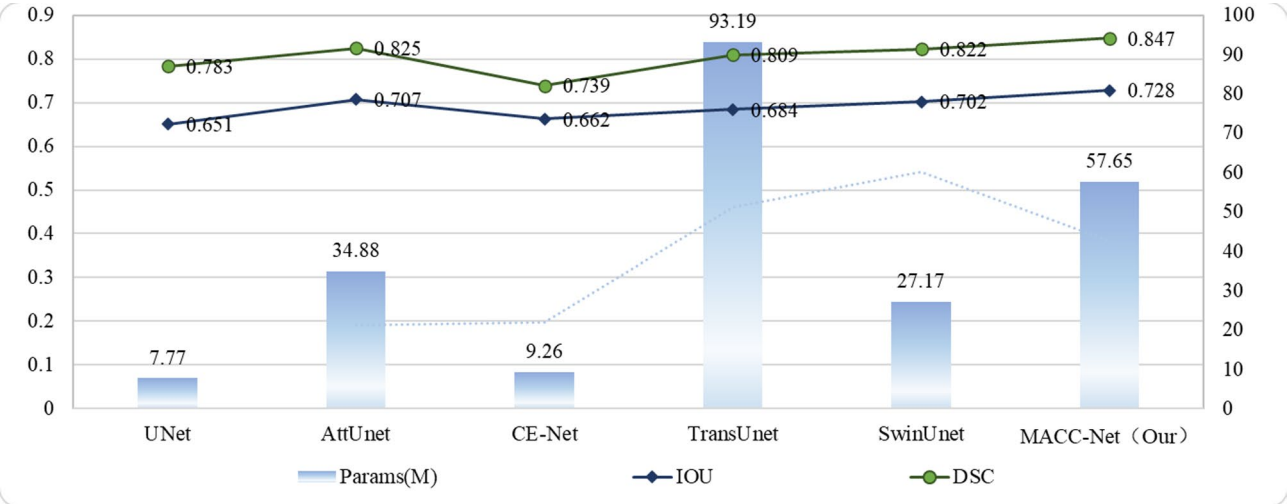


Fig. 8. Evaluation of the accuracy and complexity of each model.

spatial attention, and pixel-point attention mechanism (PA) at the same time, the IOU metric suggests 7.2% and the DSC metric improves by 1.25. From the outside, we show the prediction effect after adding different attention mechanisms to the model in Fig. 11. The ResNet model is more prone to miss segmentation. Overall, the model's prediction performance is better after adding channel attention, spatial attention, and pixel attention mechanism.

Discussion

The experimental results presented in Chap. 3 (graphs and tables) confirm the superior performance of our approach. As visually evidenced in Fig. 7, the proposed MACC-Net model achieves high accuracy in cell nuclei identification, especially for nuclei with varying scales, complex morphologies, or blurred boundaries. Quantitative comparisons in Table 1 further demonstrate that MACC-Net surpasses other models across multiple evaluation metrics. Notably, it achieves the highest recall, indicating its ability to minimize missed detections while maintaining high sensitivity—a critical advantage for segmentation tasks involving overlapping cells and background interference. This capability enhances its suitability for providing reliable clinical decision support.

Table 2 reveals that segmentation performance is significantly influenced by the magnification levels of digital pathology images. The U-Net model, heavily dependent on local details, exhibits pronounced performance degradation at lower magnifications. In contrast, TransUnet and Swin-Unet, benefiting from global

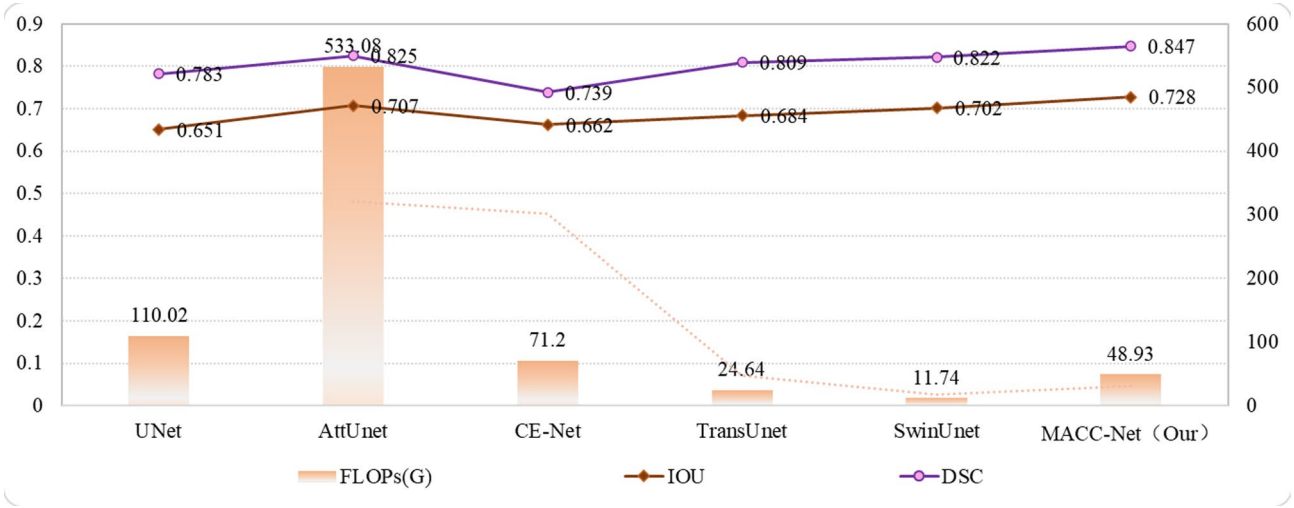


Fig. 9. Evaluation of the accuracy and computational complexity of each model.

Model	Accuracy	IOU	DSC
Baseline	0.942	0.522	0.586
Baseline + HAFEM	0.957	0.633	0.687
Baseline + CCIFM	0.953	0.623	0.678
Baseline + AMDBM	0.955	0.611	0.677
Baseline + HAFEM + CCIFM	0.964	0.688	0.744
Baseline + HAFEM + AMDBM	0.961	0.656	0.722
Baseline + CCIFM + AMDBM	0.966	0.664	0.731
MACC-Net	**0.981**	0.728	**0.847**

Table 4. Comparison of individual metrics in ablation experiments.

modeling, show greater robustness to resolution variations. Our multi-scale approach maintains relatively stable performance across magnifications. While ChannelNet also performs well, its arbitrary input-to-fixed-three-channel design leads to substantial performance drops when processing low-magnification images with blurred structures.

Computational efficiency and model complexity are critical factors in practical model evaluation. As demonstrated in Figs. 8 and 9, the proposed MACC-Net achieves a favorable balance between accuracy and computational cost, delivering significant performance improvements with relatively low hardware requirements. This makes it particularly suitable for resource-constrained environments. In contrast, while the U-Net model has the lowest computational complexity, its inferior segmentation performance (reflected by its low DSC score) limits its clinical applicability. Other models, such as AttUNet, achieve competitive segmentation accuracy but at the expense of high computational complexity, leading to substantial training resource consumption and elevated deployment costs—factors that hinder their practical adoption in real-world clinical scenarios.

Currently, due to the importance, complexity, and large size of pathology images, whole slide imaging (WSI) plays a crucial role in streamlining pathology workflows, ensuring reproducibility, and dissemination[49–52]. Mark D. Zarella et al. [a–54] provide a detailed review of the applications of WSI in pathology practice, medical education, and research, as well as the challenges and prospects of its implementation. Paola Chiara Rizzo et al.[54] summarize the applications of pathology in routine clinical diagnosis, categorizing them as "diagnostic" and "non-diagnostic." Stefano Marletta et al.[55] review and analyze the literature on placental digital pathology. The widespread application of AI in WSI has significantly promoted reproducibility[56–58]. For example, in the analysis of biological specimen data, AI can integrate diverse data, such as digital images and patient health records, to enhance the comprehensiveness of research or diagnosis[59]. At the same time, Oliwia Koteluk et al.[60] also raise the potential threat of AI medical tools to medicine. Therefore, AI tools should enhance rather than replace doctors.

Conclusion and outlook

This study proposes a digital pathology image assisted recognition method (MACC-Net) based on a multi-attention mechanism. This method improves the recognition accuracy of cell nucleus features through multi-dimensional feature extraction and fusion. The results show that this method has better performance in pathology image segmentation and can provide efficient and reliable assistance for clinical pathology diagnosis.

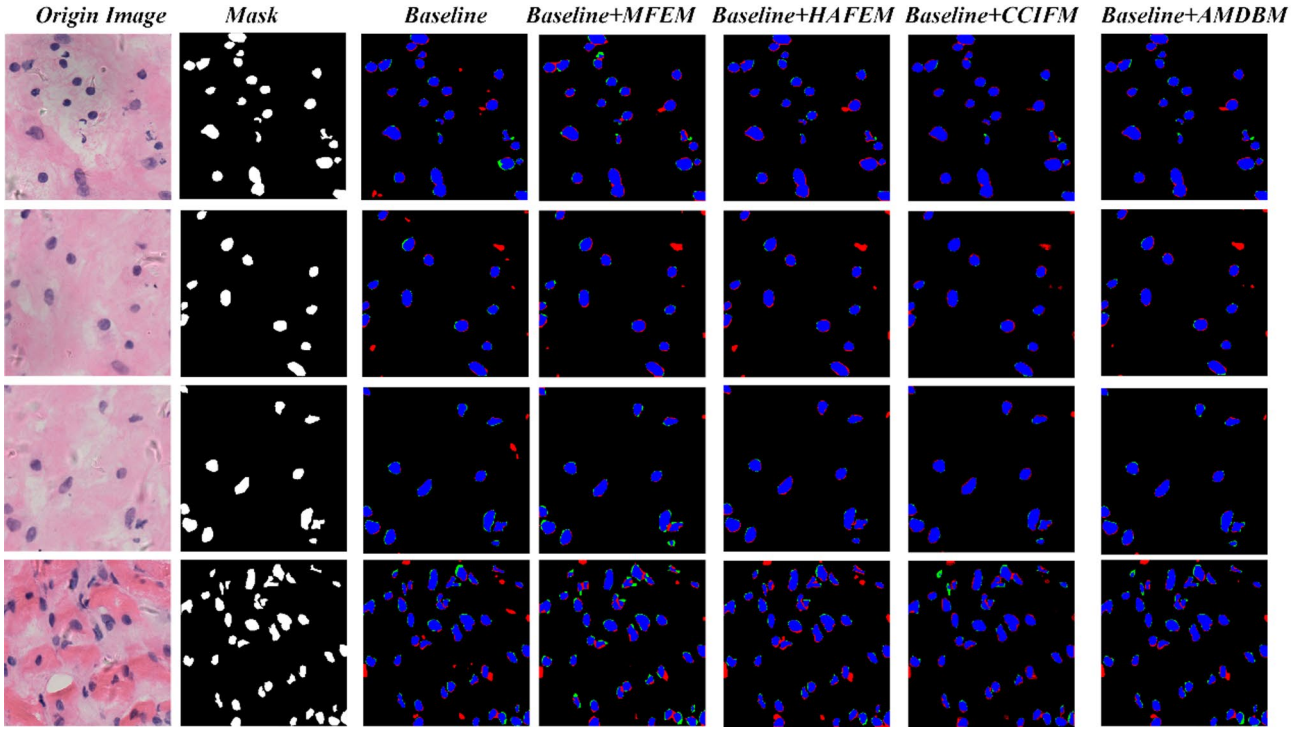


Fig. 10. Predictive effect of the model in ablation experiments.

CA	SA	PA	Accuracy	IOU	DSC
			0.961	0.656	0.722
√			0.965	0.667	0.759
	√		0.966	0.677	0.768
		√	0.957	0.657	0.755
√	√		0.975	0.688	0.812
√		√	0.977	0.694	0.811
	√	√	0.977	0.691	0.807
√	√	√	**0.981**	0.728	**0.847**

Table 5. Impact of different attention mechanisms on the model.

Although this study has a high accuracy rate for cropped standard pathology slices, it is usually necessary to observe complete pathology images in actual clinical environments. This method has problems with splicing artifacts and global information loss. With the improvement of algorithm capabilities and memory, studying full slice processing solutions will be our focus.

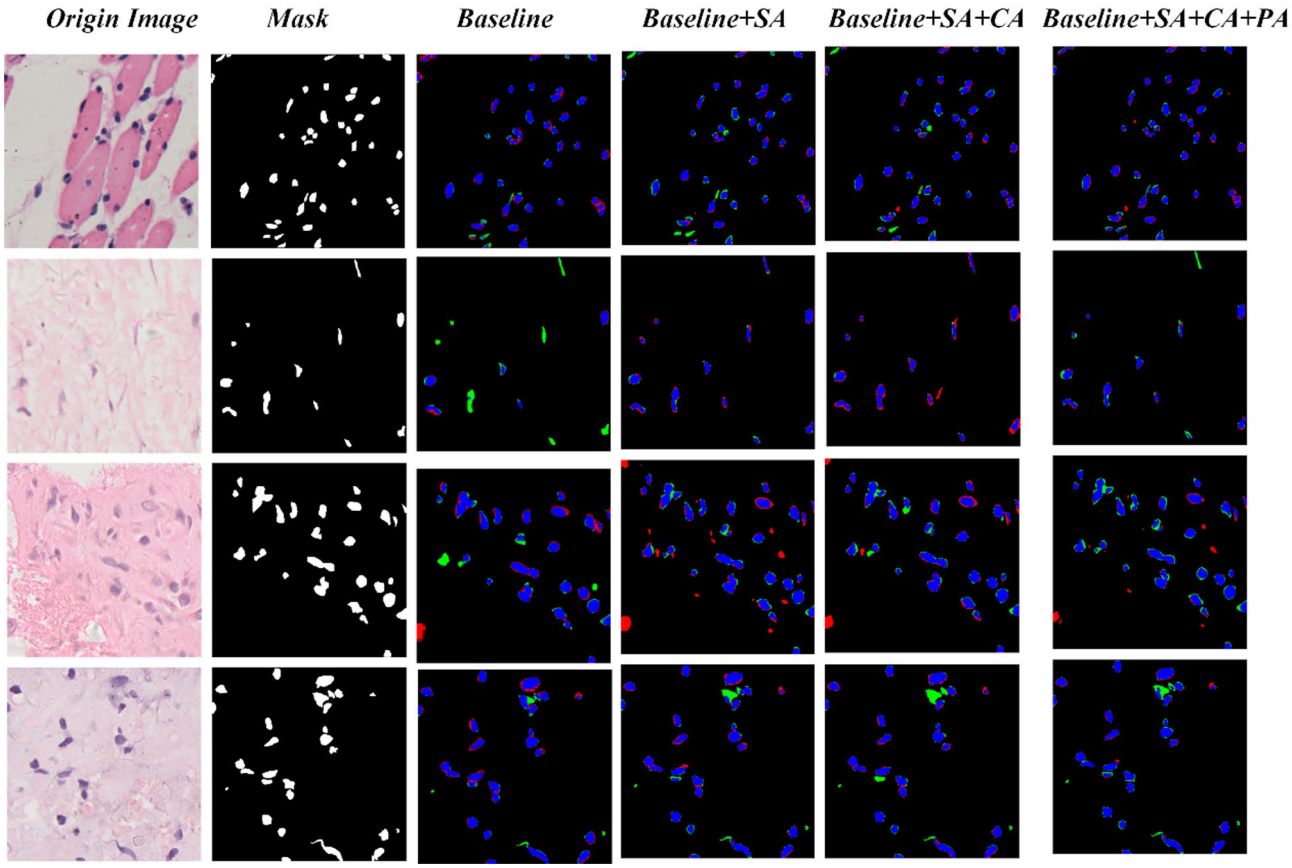


Fig. 11. Prediction effect after adding different attention mechanisms.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to the confidentiality of the data supporting the findings of this study, the results are currently being commercialized, but are available from the corresponding author on reasonable request. The corresponding author will consider data requests 12 months after publication of this article.

Received: 14 July 2025; Accepted: 22 August 2025
Published online: 25 August 2025

References

1. Qing Ye, M., Luo, J., Zhou, C., Cheng, L. & Peng NMD-FusionNet: a multimodal fusion-based medical imaging-assisted diagnostic model for liver cancer. *J. King Saud Univ. Comput. Inf. Sci.* **37**, article number 147. https://doi.org/10.1007/s44443-025-00162-8 (2025).
2. Lu, Y. & Zhou, J. Tumor auxiliary prediction strategy based on multi-scale feature fusion in medical decision-making system. *Adv. Intell. Syst.* **03** https://doi.org/10.1002/aisy.202500455 (2025).
3. He, K., Zhu, J. & Li Two-stage coarse-to-fine method for pathological images in medical decision-making systems. *IET Image Process.* **18**, 175–193. https://doi.org/10.1049/ipr2.12941 (2024).
4. Li, B., Zhou, J. & Gou, F. TransRNetFuse: a highly accurate and precise boundary FCN-transformer feature integration for medical image segmentation. *Complex Intell. Syst.* **11**, 213. https://doi.org/10.1007/s40747-025-01847-3 (2025).
5. Dai, T. & Guan, P. FedAPT: joint adaptive parameter freezing and resource allocation for communication-efficient. *IEEE Internet Things J.* **11** (11), 19520–19536. https://doi.org/10.1109/JIOT.2024.3367946 (2024).
6. El Nahhas, Omar, S. M. et al. From whole-slide image to biomarker prediction: end-to-end weakly supervised deep learning in computational pathology. *Nat. Protoc.* **20** (1), 293–316 (2025).
7. Tang, X. & Liu, J. Artificial intelligence multiprocessing scheme for pathology images based on transformer for nuclei segmentation. *Complex & Intell. Syst.*. **10**, 5831–5849. https://doi.org/10.1007/s40747-024-01471-7 (2024).
8. Lagree, A. et al. A review and comparison of breast tumor cell nuclei segmentation performances using deep convolutional neural networks. *Sci. Rep.* **11** (1), 8025 (2021).
9. Korzynska, A. et al. A review of current systems for annotation of cell and tissue images in digital pathology. *Biocybern. Biomed. Eng.* **41** (4), 1436–1453 (2021).
10. Wu, Y. et al. Predicting pancreatic diseases from fundus images using deep learning. *Vis. Comput.* **41**, 3553–3564. https://doi.org/10.1007/s00371-024-03619-5 (2025).
11. Li, X. & Quan, H. MVPCL: multi-view prototype consistency learning for semi-supervised medical image segmentation. *Vis. Comput.* **41**, 1841–1854. https://doi.org/10.1007/s00371-024-03497-x (2025).

12. Burkle, M. Jr & Khorram-Manesh, A. Artificial intelligence assisted decision-making in current and future complex humanitarian emergencies. *Disaster Med. Pub. Health Prep.* **19**, e64 (2025).
13. Guerroudji, M. et al. A 3D visualization-based augmented reality application for brain tumor segmentation. *Comput. Animat. Virtual Worlds*. **35** (1), e2223 (2024).
14. Li, B., Liu, F. & Lv, B. Cytopathology image analysis method based on high-resolution medical representation learning in medical decision-making system. *Complex. Intell. Syst.* **10**, 4253–4274. https://doi.org/10.1007/s40747-024-01390-7 (2024).
15. John, B. Ethical considerations in AI-assisted decision-making for end-of-life care in healthcare. (2025).
16. Xiao, Z. et al. Human action recognition in immersive virtual reality based on multi-scale spatio-temporal attention network. *Comput. Animat. Virtual Worlds*. **35**, e2293 (2024).
17. Li, X. et al. Multi-level feature fusion network for nuclei segmentation in digital histopathological images. *Visual Comput.* **39** (4), 1307–1322 (2023).
18. Liu, R. et al. Dec., NHBS-Net: A feature fusion attention network for ultrasound neonatal hip bone segmentation. In *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, 3446–3458. https://doi.org/10.1109/TMI.2021.3087857 (2021).
19. Wen, Y. et al. SAT-Net: Structure-aware transformer-based attention fusion network for low-quality retinal fundus images enhancement. In *IEEE Trans. Multimedia*. https://doi.org/10.1109/TMM.2025.3565935
20. Cao, W. et al. Semi-supervised intracranial aneurysm segmentation via reliable weight selection. *Vis. Comput.* **41**, 5421–5433. https://doi.org/10.1007/s00371-024-03730-7 (2025).
21. Li, Z., Lu, Z. & Yin, M. Modeling human trust and reliance in ai-assisted decision making: A markovian approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, No. 5 (2023).
22. Liu, Y. et al. Where to focus: Investigating hierarchical attention relationship for fine-grained visual classification. In *European Conference on Computer Vision* (Springer Nature Switzerland, 2022).
23. Gong, M. et al. GKC-Net: gated KAN with Channel-Position attention mechanism for image deraining. *Pattern Recogn.* 112080 (2025).
24. Jacobsen, J. H. et al. Structured receptive fields in CNNs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016).
25. Zhou, Y. et al. GAMNet: a gated attention mechanism network for grading myopic traction maculopathy in OCT images. *Vis. Comput.* **41**, 1097–1108. https://doi.org/10.1007/s00371-024-03386-3 (2025).
26. Liu, X., Wu, Z. & Wang, X. Validity of non-local mean filter and novel denoising method. *Virtual Real. Intell. Hardw.* **5** (4), 338–350 (2023).
27. Xu, Z. et al. An intelligent MRI assisted diagnosis and treatment system for osteosarcoma based on super-resolution. *Complex Intell. Syst.*. **10**, 6031–6050. https://doi.org/10.1007/s40747-024-01479-z (2024).
28. Liang, M. et al. Multi-scale self-attention generative adversarial network for pathology image restoration. *Visual Comput.* **39** (9), 4305–4321 (2023).
29. He, K., Zhu, J. & Li, L. FASNet: Feature alignment-based method with digital pathology images in assisted diagnosis medical system. *Heliyon* **10**, e40350. https://doi.org/10.1016/j.heliyon.2024.e40350 (2024).
30. Yuxia Niu, Z., Ling, J. & Zhu Pathological image segmentation method based on multiscale and dual attention. *Int. J. Intell. Syst.* https://doi.org/10.1155/int/9987190 (2024)
31. Gou, F. et al. Optimization of edge server group collaboration architecture strategy in IoT smart cities application. *Peer-to-Peer Netw. Appl.* https://doi.org/10.1007/s12083-024-01739-2 (2024).
32. Dai, T., Guan, P., Chen, Z. & Taherkordi, A. Opportunistic routing for mobile edge computing: A community detected and task priority aware approach. *Comput. Netw.*. **258**, 111000. https://doi.org/10.1016/j.comnet.2024.111000 (2025).
33. Li, W. et al. Artificial intelligence auxiliary diagnosis and treatment system for breast cancer in developing countries. *J. X-Ray Sci. Technol.* **32** (2), 395–413. https://doi.org/10.3233/XST-230194 (2024).
34. Guan, P., Yu, K., Wei, W. & Tan, Y. Big data analytics on lung cancer diagnosis framework with deep learning. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **21** (4), 757–768. https://doi.org/10.1109/TCBB.2023.3281638 (2024).
35. Wei, H., Lv, B., Liu, F., & Tang, H. A tumor MRI image segmentation framework based on Class-Correlation pattern aggregation in medical decision-making system. *Mathematics*. **11**, 1187. https://doi.org/10.3390/math11051187 (2023).
36. Zhixun Zhou, P. & Xie, Z. Self-supervised tumor segmentation and prognosis prediction in osteosarcoma using multiparametric MRI and clinical characteristics. *Comput. Methods Progr. Biomed.* **244**, 107974. https://doi.org/10.1016/j.cmpb.2023.107974 (2024).
37. Wu, J., Luo, T. & Zeng, J. Continuous refinement-based digital pathology image assistance scheme in medical decision-making systems. *IEEE J. Biomed. Health Inf.* **28** (4), 2091–2102. https://doi.org/10.1109/JBHI.2024.3351287 (2024).
38. Naylor, P., Laé, M., Reyal, F. & Walter, T. Nuclei segmentation in histopathology images using deep neural networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, Australia*, 933–936. https://doi.org/10.1109/ISBI.2017.7950669 (2017).
39. Li, L., He, K., Zhu, X. & Part, A. A pathology image segmentation framework based on deblurring and region proxy in medical decision-making system. *Biomed. Signal Process. Control.* **95**, 106439. https://doi.org/10.1016/j.bspc.2024.106439 (2024).
40. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, 234–241 (Springer, 2015).
41. Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018).
42. Badrinarayanan, V., Kendall, A. & Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39** (12), 2481–2495 (2017).
43. Gao, Y., Zhou, M. & Metaxas, D. N. UTNet: a hybrid transformer architecture for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 61–71 (Springer, 2021).
44. Gu, Z. et al. Ce-net: context encoder network for 2d medical image segmentation. *IEEE Trans. Med. Imaging.* **38** (10), 2281–2292 (2019).
45. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., … Zhou, Y. (2021). Transunet:Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306.
46. Cao, H. et al. Swin-unet: Unet-like pure transformer for medical image segmentation. *ArXiv Preprint arXiv* 210505537 (2021).
47. Diakogiannis, F. I. et al. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.* **162**, 94–114 (2020).
48. Goldsborough, T. et al. A novel channel invariant architecture for the segmentation of cells and nuclei in multiplexed images using InstanSeg. *bioRxiv* 2024-09 (2024).
49. Liu, J. & Xiao, C. R. On artificial-intelligence-assisted medicine: A survey on medical artificial intelligence. *Diagnostics* **14**, 1472. https://doi.org/10.3390/diagnostics14141472 (2024).
50. He, K., Li, L., Zhou & Part, A. Asymptotic multilayer pooled transformer based strategy for medical assistance in developing countries. *Comput. Electr. Eng.* **119**, 109493. https://doi.org/10.1016/j.compeleceng.2024.109493 (2024).
51. Pan, Y. & Ye, Q. Intelligent cell images segmentation system: based on SDN and moving transformer. *Sci. Rep.* **14**, 24834. https://doi.org/10.1038/s41598-024-76577-6 (2024).
52. Hanna, M. G. et al. Future of artificial intelligence (AI)-machine learning (ML) trends in pathology and medicine. *Modern Pathol.* 100705 (2025).

53. Zarella, M. D. et al. A practical guide to whole slide imaging: a white paper from the digital pathology association. *Arch. Pathol. Lab. Med.* **143**(2), 222–234. https://doi.org/10.5858/arpa.2018-0343-RA (2019).
54. Rizzo, P. C. et al. Digital pathology world tour. *Digit. Health*. **9**, 20552076231194551 (2023).
55. Marletta, S. et al. Application of digital imaging and artificial intelligence to pathology of the placenta. *Pediatr. Dev. Pathol.* **26** (1), 5–12 (2023).
56. Gou, F., Dai, T., Wu, J., A novel feature alignment-based cell nucleus identification scheme for medical decision-making systems. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Lisbon, Portugal*, 2024, 3222–3225. https://doi.org/10.1109/BIBM62325.2024.10822119 (2024).
57. Zhitao et al. Medical Assisted-segmentation system based on global feature and Stepwise feature integration for feature loss problem. *Biomed. Signal Process. Control.* **89**, 105814. https://doi.org/10.1016/j.bspc.2023.105814 (2024).
58. Zhan, X. & Long, H. A semantic fidelity interpretable-assisted decision model for lung nodule classification. *Int. J. Comput. Assist. Radiol. Surg.* **19**, 625–633. https://doi.org/10.1007/s11548-023-03043-5 (2024).
59. Frascarelli, C. et al. Revolutionizing cancer research: the impact of artificial intelligence in digital biobanking. *J. Pers. Med.* **13** (9), 1390. https://doi.org/10.3390/jpm13091390 (2023). PMID: 37763157; PMCID: PMC10532470.
60. Koteluk, O., Wartecki, A., Mazurek, S., Kołodziejczak, I. & Mackiewicz, A. How do machines learn?? Artificial intelligence as a new era in medicine. *J. Pers. Med.* **11** (1), 32. https://doi.org/10.3390/jpm11010032 (2021).

Acknowledgements

All data analyzed during the current study are included in the submission.

Author contributions

For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, F.L., Z.W. and F.G.; writing—original draft preparation, B.L., M.L. and D.W.; writing—review and editing, F.G. and J.W. All authors have read and agreed to the published version of the manuscript."

Funding

This work was supported in the Doctoral Research Initiation Project of Shandong Youth University of Political Science, XXPY21024.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.W., M.L. or J.W.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2025