AMSD_VGGNet: A Lightweight Multi-Scale Attention Network for High-Accuracy Breast Cancer Histopathological Image Classification

Yulian Li[1*], Ting Chen2[2], Yikang Du[3]

[1] Chengdu College of University of Electronic Science and Technology of China; China; heunglyl@gmail.com

[2] Chengdu College of University of Electronic Science and Technology of China; China; chenting08282022@163.com

[3] Chengdu College of University of Electronic Science and Technology of China; China; 2122414189@qq.com

Abstract. Early diagnosis of breast cancer is crucial for improving patient survival rates, but traditional pathological diagnosis heavily relies on subjective clinical experience, suffering from inefficiency and poor consistency. This paper proposes an improved VGG network model integrating multi-scale features and attention mechanisms for automated classification of breast cancer histopathological images. The model introduces a mixed-domain attention mechanism into the VGG16 backbone, enabling dynamic focus on critical pathological feature regions such as nuclear atypia. Simultaneously, it incorporates a dual-scale dilated convolution module to parallelly extract local details and global contextual information, enhancing multi-scale feature representation. Experimental results demonstrate that AMSD_VGGNet achieves classification accuracies of 99.71% on both BreakHis and ICIAR2018 datasets, with only 12.8% of VGG16's parameter count. Heatmap visualization indicates that its decision logic aligns closely with pathological standards. Furthermore, an interactive system interface developed using PySide6 framework supports high-resolution image loading and real-time classification response, providing an efficient and reliable intelligent auxiliary diagnostic tool for early breast cancer screening.

CCS Concepts: *Computing methodologies → Artificial intelligence*

Keywords: Breast cancer; Hybrid-domain attention mechanism; Multi-scale dilated convolution; Lightweight network

* Corresponding Author: Yulian Li (heunglyl@gmail.com)

1. Introduction

Breast cancer has been identified as a major malignant disease that severely threatens human health, and its prevention and control are now facing new challenges. The latest epidemiological data from the International Agency for Research on Cancer (IARC) shows that the number of confirmed breast cancer cases worldwide in 2024 is expected to exceed 2.46 million, an increase of approximately 14% compared to 2020, maintaining its position as the most common malignancy among women.

Pathological images serve as crucial evidence for medical diagnosis, and the development of computer-aided analysis systems for these images faces dual characteristics: on one hand, there are numerous technical bottlenecks, while on the other, they hold significant medical application prospects. In response to these challenges, the research community has proposed various innovative solutions. Literature [1-3] systematically compares optimization strategies for different algorithms, including improvements in feature engineering and innovations in model architecture. These studies provide important technical references for pathological image analysis. Currently, automated classification and recognition methods for breast cancer medical pathological images can be broadly divided into two categories: traditional machine learning methods and deep learning methods.In the field of breast pathological image analysis, traditional machine learning methods typically adopt a two-stage processing pipeline. First, manually designed feature extraction algorithms are applied to the images, with commonly used methods including texture feature extraction techniques such as Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), and morphological-based regional feature analysis. These methods transform high-dimensional image data into representative feature vectors.Subsequently, various classical machine learning algorithms are employed to construct classification models. Support Vector Machines (SVM) achieve sample differentiation by identifying the optimal classification hyperplane; the Random Forest (RF) algorithm integrates prediction results from multiple decision trees; and the K-Nearest Neighbors (KNN) classifier determines categories based on distance metrics between samples. These algorithms each have unique characteristics and can achieve certain classification effects under specific conditions.Spanhol et al. combined features such as Local Binary Pattern, Complete Local Binary Pattern (CLBP), Local Phase Quantization (LPQ), Gray-Level Co-occurrence Matrix (GLCM), ORB (Oriented FAST and Rotated BRIEF), and Parameter-Free Threshold Adjacency Statistics (PFTAS) with classifiers (e.g., SVM, RF, Quadratic Discriminant Analysis (QDA), and nearest-neighbor classifiers) for benign and malignant classification of breast pathological images, achieving an accuracy of 80% to 85% on the BreakHis dataset [4]. Gupta

et al. proposed a method that fuses features such as wavelet features, opponent color local binary patterns, color, and texture, creating a heterogeneous ensemble classifier using a voting mechanism for classification [5]. Shukla et al. utilized morphological features for automatic detection and classification of breast pathological images, employing histogram equalization to improve local image contrast and comparing various classifiers (e.g., RF, Rotation Forest, SMO, Naïve Bayes, J-Rip, and PART decision trees) [6]. Kahya et al. proposed an adaptive sparse support vector model based on discriminative features, assigning weights to features using adaptive L1 norm and selecting high-accuracy informative features, achieving a prediction accuracy of 94.97% in the binary classification task on the BreakHis 40× dataset [7]. Bardon et al. used the Bag-of-Words model and Locality-constrained Linear Coding to extract handcrafted features, employing an SVM classifier to complete pathological image classification [8].

This study enhances breast cancer histopathology image classification by optimizing data preprocessing and network architecture. It introduces a multi-scale sliding window strategy with standardized processing to expand patch diversity and reduce staining variability, improving training stability. The model integrates attention mechanisms and dilated convolutions into VGG16 to boost multi-scale feature extraction and discriminative representation of tissue/cellular structures. The framework aims to achieve higher accuracy and F1 scores on public datasets with efficient parameter usage, while leveraging visualization techniques to interpret model focus areas. An interactive analysis system is also developed to support clinical decision-making.

2. Related Work

2.1 Dilated Convolution

Dilated Convolution, proposed by Yu et al. [9], aims to expand the network's receptive field without increasing computational load or altering feature map dimensions. The receptive field refers to the mapped region of the output feature map in the original image.

This is achieved by inserting holes (zeros) into the convolutional kernel and introducing a hyperparameter—dilation rate (r), which defines the spacing between kernel elements. When r=1, dilated convolution reduces to standard convolution. As rincreases, the kernel's receptive field expands, as shown in Figure 1 (illustrating receptive field changes for r=1, 2, 3).

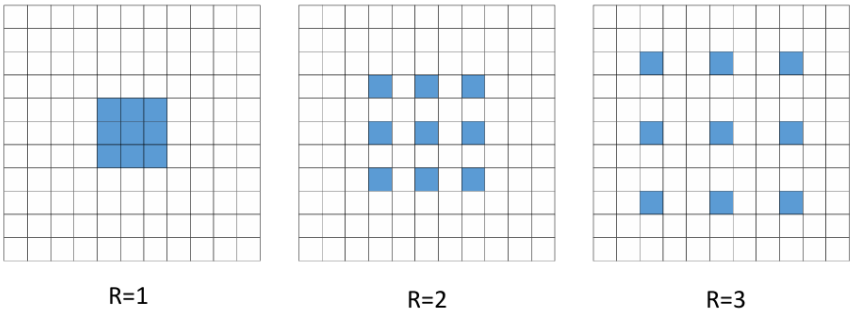


Figure 1 *Receptive Fields Under Different Dilation Rates.*

2.2 CBAM Attention Module

Woo et al. [10] proposed the CBAM attention module, which combines CAM and SAM to adaptively select and weight important information in input feature maps. CBAM is illustrated in Figure 2.

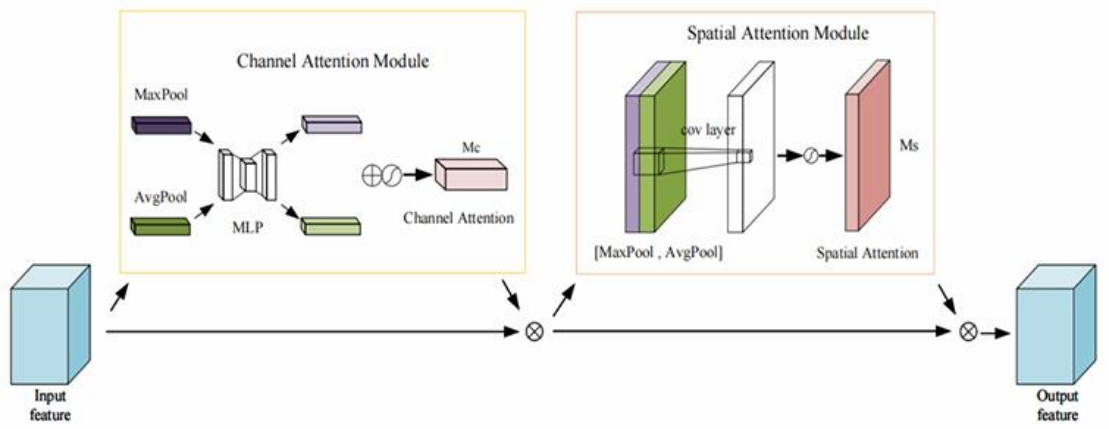


Figure 2 *CBAM Attention Module.*

The channel attention mechanism models global interdependencies across channels, dynamically adjusting their weights to enhance key features. It uses a dual-pooling branch and shared MLP architecture: global average pooling extracts spatial statistics, while max pooling captures salient activations. These features are fed into a shared-weight MLP for nonlinear transformation, generating channel attention weights. The learned weights are multiplied channel-wise with the original features for recalibration. This design preserves global context while highlighting local features, improving discriminative feature extraction in tasks like image classification.

3. Results

3.1. Model Structure Design

The proposed Attention Multi-Scale Dilated VGG Network (AMSD-VGGNet) is an improved network structure based on the classical VGG16 deep convolutional neural network. It integrates the hybrid-domain attention mechanism (CBAM) and a multi-scale dilated convolution module to enhance multi-scale feature extraction and recognition capabilities for medical images. The model retains the first ten convolutional layers and the first four pooling layers of VGG16. CBAM attention modules are introduced before and after these layers to strengthen the focus on critical features. Additionally, a dual-scale dilated convolution module is incorporated at the high-level semantic feature extraction stage to improve the model's ability to recognize lesion regions of varying sizes and morphologies.

The network structure enhances the model's response to effective regions through attention mechanisms, extracts deep features via multi-scale dilated convolutions, and performs feature fusion and fully connected classification at the output. The overall structure balances depth, breadth, and multi-scale expressiveness in feature extraction, meeting the high-precision requirements of medical image analysis. The complete structure is illustrated in Figure 3.

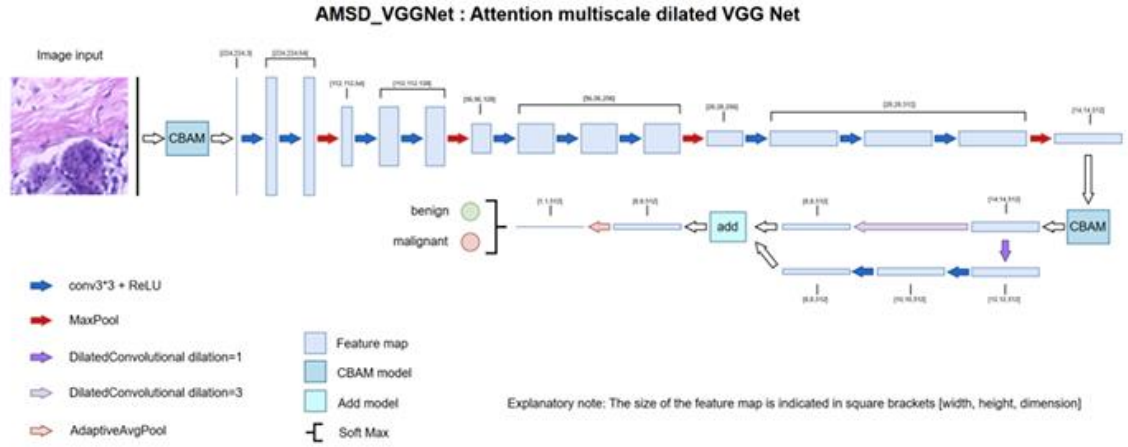


Figure 3 *Improved VGG Network Structure Diagram.*

3.3. Datasets and Data Processing

The BreakHis[10] dataset 1212(2016) contains 7,909 breast cancer histopathology images across four magnifications (40×, 100×, 200×, 400×), classified into benign (4 subtypes) and malignant (4 subtypes) categories. Example images at different magnifications are shown in Figure 4. The BACH Challenge dataset 2828 (ICIAR2018) includes 400 H&E-stained images annotated into four diagnostic classes, with samples illustrated in Figure 5. Both datasets face challenges such as limited data volume and staining/imaging variability.

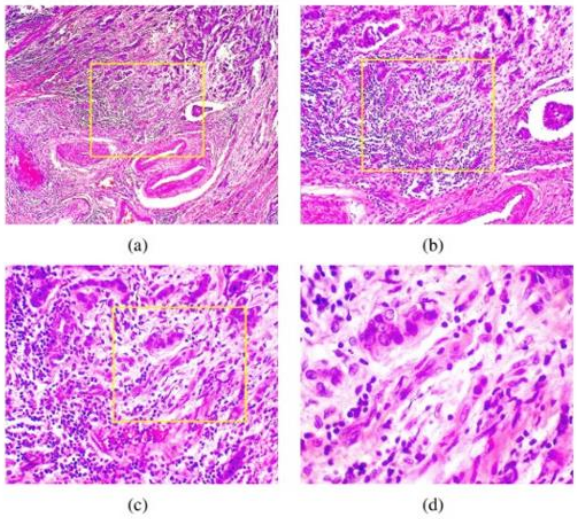


Figure 4 *BreakHis Dataset Samples at Different Magnifications [6]: (a) 40×; (b) 100×; (c) 200×; (d) 400×.*

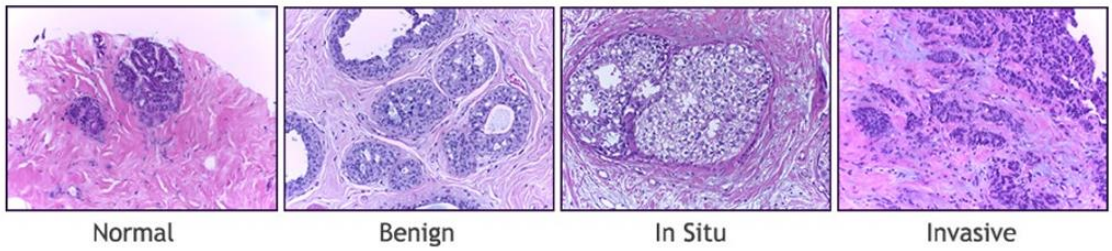
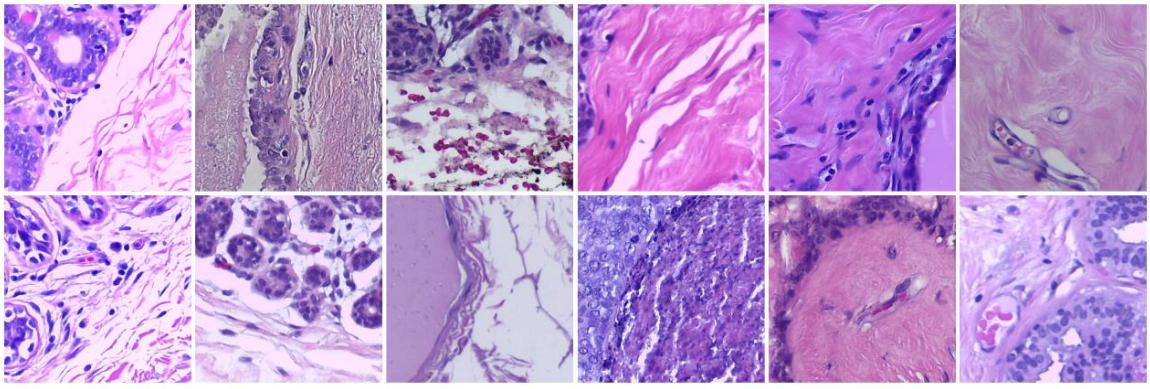


Figure 5 *ICIAR2018 Dataset Samples.*

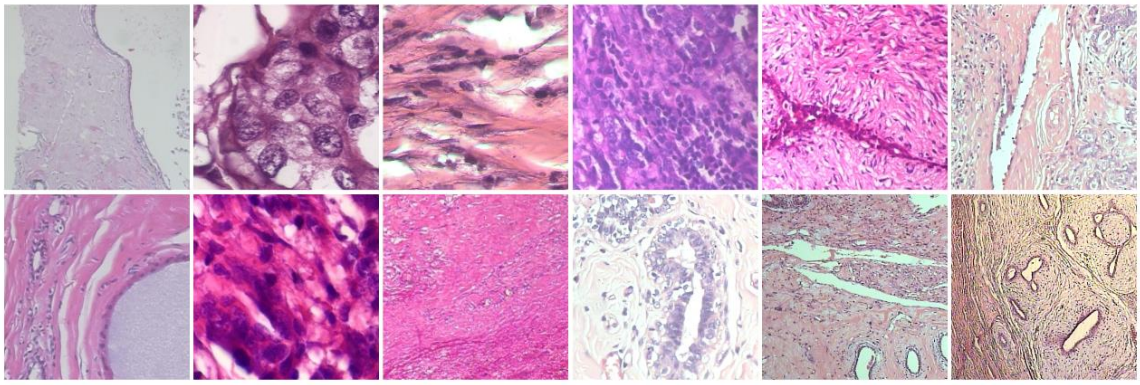
To address these, data augmentation was applied:

- BreakHis images (700×460) were split into two 460×460 patches per image, expanding the dataset to 15,818 patches.
- ICIAR2018[11] images (2048×1536) used multi-scale sliding windows (460×460 to 1024×1024) followed by resizing, generating 24,818 total patches.

Color normalization(demonstrated in Figures 6 and 7) standardized staining variations by aligning pixel-wise mean and standard deviation across datasets. After normalization, images were resized to 224×224 (VGG input standard) and converted to Tensor format for PyTorch compatibility. These steps reduced non-critical interference (e.g., staining inconsistencies), preserved pathological features, and improved model convergence and training stability.



Figures 6 *BreakHis Dataset Differentiated Sampling.*



Figures 7 *ICIAR2018 BACH Dataset Differentiated Sampling.*

3.4 Experimental Environment

The experiments were performed on a high-performance computing system equipped with an Intel Core i7-13620H processor (14 cores, 20 threads), 16GB RAM, and an NVIDIA RTX 4060 Laptop GPU (8GB VRAM with CUDA acceleration) to accelerate deep neural network training. The software stack included PyTorch 2.0 as the core deep learning framework, supported by OpenCV and Pillow for image processing tasks, and torchvision for data augmentation operations, ensuring compatibility and efficiency throughout the training pipeline.

3.5 Model Evaluation Criteria

Accuracy, Precision, Recall, F1 score, ROC curve, PR curve, and confusion matrix are used in this study for comprehensive comparison. Accuracy is a commonly used metric for classification models, indicating the proportion of samples correctly classified by the model on the entire dataset, and its calculation formula is shown in Equation 1:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

In the formula, TP (True Positive) indicates the number of positive class samples correctly identified by the model, and TN (True Negative) indicates the number of negative class samples correctly identified; FP (False Positive) is the number of negative class samples misclassified

as positive, and FN (False Negative) is the number of positive class samples misclassified as negative. The ratio between TP and TN reflect the model's classification accuracy on positive and negative class samples, while FP and FN reflect the model's misclassification.

Precision is used to measure the proportion of samples predicted to be positive classes that actually belong to positive classes by the model, reflecting the accuracy of the model in the prediction of positive classes. Its calculation formula is shown in Equation 2:

$$\text{Precision}=\frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

Recall measures the proportion of all samples that are actually in the positive class that are correctly recognized as positive by the model, reflecting the model's ability to recognize samples in the positive class. The formula for Recall is shown in Equation 3:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

The F1 value is the reconciled average of precision and recall, which integrally reflects the prediction accuracy of the classification model and its ability to recognize actual positive examples. Precision evaluates the proportion of correctness in the prediction results, while recall measures the model's effectiveness in recognizing positive examples. The F1 value is particularly suitable for dealing with classification problems in unbalanced datasets. Its calculation formula is shown in Equation 4:

$$\text{F1} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{4}$$

The vertical coordinate of the ROC curve denotes the True Positive Rate (TPR) and the horizontal coordinate denotes the False Positive Rate (FPR), which is used to evaluate the model's ability to discriminate between positive and negative samples under different decision thresholds. Its calculation formula is shown in 5, 6.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FN}} \tag{6}$$

The vertical coordinate of the PR curve is Precision and the horizontal coordinate is Recall, which is used to measure the model's recognition precision and coverage of positive case samples. Its definition formula is shown in 7, 8.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

3.6 Experimental Design

Five baseline models were selected for comparison: VGG16, ResNet50, DenseNet121, MobileNetV3, and ConvNeXt. The experimental setup ensured uniform data splits (train/validation/test) and preprocessing across all models. Key hyperparameters included a learning rate of 0.0001, batch size of 32, and 10 training epochs. Transfer learning was applied: baseline models utilized pretrained weights, while AMSD-VGGNet initialized with VGG16 weights. Performance was evaluated using accuracy, precision, recall, F1 score, ROC/PR curves, and confusion matrices.

Table 2 *Hyperparameters for Baseline Models*

Parameter	Value
lr	0.0001
batch_size	32
epoch	10
Input size	[224,224]

3.7 Experimental Results

All models trained with identical hyperparameters. After completing the model training, we systematically compared the performance of all models on an independent test set to evaluate their generalization ability in real-world scenarios. The experimental metrics included accuracy, F1 score, precision, recall, and model parameter count. The evaluation results are presented in Table 3.

Table 3 *Comparative experimental results*

Model	accuracy	F1	precision	recall	Params
AMSD_VGGNet	0.9971	0.9971	0.9972	0.9971	17.15M
VGGNet	0.9971	0.9971	0.9972	0.9971	134.27M
MobileNetV3	0.9886	0.9886	0.9886	0.9886	4.20M
DenseNet121	0.8606	0.8594	0.8603	0.8606	7.98M
ResNet50	0.8663	0.8668	0.8683	0.8663	23.51M
ConvNext	0.8989	0.8985	0.8986	0.8989	27.81M

The proposed AMSD_VGGNet achieves near-identical performance to VGG16, with metrics (Accuracy, F1, Precision, Recall) ranging from 0.9971 to 0.9972, while reducing parameters to 17.15M (12.8% of VGG16's 134.27M), enhancing deployability on resource-limited devices. MobileNetV3, as a lightweight alternative, attains 0.9886 across metrics with only 4.20M parameters, balancing efficiency and performance. In contrast, DenseNet121, ResNet50, and ConvNeXt underperform (Accuracy: 0.86–0.90) despite varying parameter counts (7.98M–27.81M). The confusion matrix for AMSD_VGGNet (Figure 3-28) demonstrates exceptional robustness, with >99.7% diagonal predictions and minimal misclassifications, confirming its capability in fine-grained feature recognition.

4. Discussion

The proposed AMSD_VGGNet effectively balances lightweight design and high accuracy (0.9971) in breast cancer histopathology classification, reducing parameters to 17.15M (12.8% of VGG16) through multi-scale dilated convolutions (dilation rates: 1 and 3) and CBAM attention mechanisms. These components synergistically capture cellular/tissue-level features while suppressing staining artifacts and noise. Heatmaps confirm alignment with pathological standards, enhancing clinical trust.

Limitations and Future Directions:

• Data Constraints: Limited dataset size and image quality variations may hinder generalization. Future work should expand datasets with diverse subtypes and staining conditions.

• Clinical Validation: Lack of real-world testing necessitates multi-center trials to assess robustness in clinical settings.

• System Integration: Current tools lack deep collaboration with physician expertise. Developing human-AI collaborative frameworks could improve clinical adaptability by merging model outputs with pathologist insights.

5. Conclusion

This study proposes an improved VGG network model (AMSD_VGGNet) for breast cancer histopathology image classification, systematically validating its efficiency and reliability. Key conclusions are as follows:

（1） Lightweight Design and Performance Balance

AMSD_VGGNet reduces parameters by 87.2% (17.15M) compared to VGG16 through global average pooling, while maintaining a classification accuracy of 0.9971. It significantly outperforms lightweight models like MobileNetV3 (0.9886), achieving an optimal balance between precision and efficiency.

（2） Validation of Innovative Structures

The CBAM attention mechanism effectively focuses on critical pathological features such as nuclear atypia and disordered cell arrangements, suppressing non-diagnostic interference.

The dual-scale dilated convolution module enhances joint recognition of cellular-level morphology and tissue-level structures through multi-scale feature fusion.

（3） Experimental and Practical Value

On the BreakHis and ICIAR2018 datasets, AMSD_VGGNet achieves core metrics (accuracy, F1 score, etc.) exceeding 0.9971, surpassing mainstream models like ResNet50 (0.8663) and ConvNeXt (0.8989). An interactive system developed using the PySide6 framework supports second-level inference and visualization of high-resolution images, providing an efficient tool for clinical auxiliary diagnosis. This study offers a high-precision, lightweight, and interpretable solution for medical image diagnosis. However, its clinical utility requires further validation through large-scale real-world data testing and system optimization.

References

[1] KOMURA D, ISHIKAWA S. Machine learning methods for histopathological image analysis [J]. Computational and Structural Biotechnology Journal, 2018, 16: 34-42.

[2] TUFAIL A B, MA Y K, KAABAR M K, et al. Deep learning in cancer diagnosis and prognosis prediction: A minireview on challenges, recent trends, and future directions [J]. Computational and Mathematical Methods in Medicine, 2021, 2021: 1-15.

[3] CAMPANELLA G, HANNA M G, GENESLAW L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images [J]. Nature Medicine, 2019, 25(8): 1301-1309.

[4] SPANHOL F A, OLIVEIRA L S, PETITJEAN C, et al. A dataset for breast cancer histopathological image classification [J]. IEEE Transactions on Biomedical Engineering, 2015, 63(7): 1455-1462.

[5] GUPTA V, BHAVSAR A. Breast cancer histopathological image classification: Is magnification important? [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, HI, USA: IEEE, 2017: 769-776.

[6] SHUKLA K, TIWARI A, SHARMA S. Classification of histopathological images of breast cancerous and non-cancerous cells based on morphological features [J]. Biomedical and Pharmacology Journal, 2017, 10(1): 353-366.

[7] KAHYA M A, AL-HAYANI W, ALGAMAL Z Y. Classification of breast cancer histopathology images based on adaptive sparse support vector machine [J]. Journal of Applied Mathematics and Bioinformatics, 2017, 7(1): 49-64.

[8] BENHAMMOU Y, ACHCHAB B, HERRERA F, et al. BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights [J]. Neurocomputing, 2020, 375: 9-24.

[9] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[10] BENHAMMOU Y, ACHCHAB B, HERRERA F, et al. BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights [J]. Neurocomputing, 2020, 375: 9-24.

[11] ARESTA G, ARAÚJO T, KWOK S, et al. BACH: Grand challenge on breast cancer histology images [J]. Medical Image Analysis, 2019, 56: 122-139.