CIT. Journal of Computing and Information Technology, Vol. 32, No. 2, June 2024, 127–142
doi: 10.20532/cit.2024.1005848

SSResNeXt: A Novel Deep Learning Architecture for Multi-class Breast Cancer Pathological Image Classification

Bo Xu[1], Faming Li[1] and Ying Wu[2]

[1]School of Information Science, Guangdong University of Finance and Economics, China
[2]Department of Ultrasound, The First Affiliated Hospital of Jinan University, Jinan University, China

Multi-class classification of breast cancer pathological images remains challenging due to complex image features and limited datasets. This study proposes SSResNeXt, a novel deep learning architecture incorporating a new Small-SE-ResNeXt Block with asymmetric convolutions and channel attention mechanisms. Evaluated on the BreaKHis dataset, SSResNeXt achieves state-of-the-art performance with accuracies of 95.2%, 94.0%, 92.7%, and 93.5% for 40X, 100X, 200X, and 400X magnification scales, respectively. Comparative experiments demonstrate SSResNeXt's superior performance over existing models, including ResNet and Swin Transformer variants. The proposed architecture offers improved feature extraction capabilities for breast cancer pathological images without significantly increasing model complexity, providing a promising tool for computer-aided diagnosis systems.

ACM CCS (2012) Classification: Computing methodologies → Machine learning → Machine learning approaches → Neural networks

Keywords: deep learning, multi-classification, breast cancer, pathological images

1. Introduction

Breast cancer is one of the leading causes of morbidity and mortality among women worldwide, claiming countless lives each year. Timely detection in its early stages is crucial for effective treatment. Clinical screening methods include X-ray (Lim *et al.*, 2023) [1], CT, MRI, and biopsy, with biopsy being the gold standard for confirming cancer (Horvat *et al.*, 2019) [2]. However, these methods are time-consuming and rely heavily on the expertise of experienced doctors (Proctor *et al.*, 2017) [3], increasing workload and diagnostic challenges.

In recent years, advancements in computer hardware and computer vision algorithms have significantly enhanced computer-aided diagnosis (CAD) technology (Araújo *et al.*, 2017) [4]. This technology has improved diagnostic accuracy, reduced doctors' workload, and provided more precise and objective test results (Waks *et al.*, 2019) [5]. While machine learning and deep learning in computer vision continue to evolve, training highly accurate models with large datasets is feasible for general natural images due to ample training data and computational power. However, in medical imaging, achieving highly accurate models is challenging due to privacy concerns and high costs of data collection and annotation by professionals. Therefore, improving model accuracy with limited data remains a critical issue.

To address this problem, various model architectures have been explored, from initial multi-layer neural networks to advanced deep convolutional models [34, 35]. New learning methods, such as transfer learning, have also been proposed. Transfer learning leverages existing knowledge from the source domain to enhance learning in the target domain [36, 37]. While deep learning methods have achieved

promising results in classifying breast cancer pathological images, they often require extensive pre-training on large non-target domain datasets, massive model structures, and significant computing resources [38]. This reliance on pre-training and complex model architectures can lead to diminishing returns, where additional data and model complexity result in only marginal improvements in performance.

Model improvements have enhanced breast cancer diagnosis by advancing feature extraction and optimizing deep learning architectures. Studies have also applied data processing techniques to boost model generalization. Aldakhil *et al.* [30] propose ECSAnet, which integrates a Convolutional Block Attention Module (CBAM) into EfficientNetV2, improving classification accuracy by focusing on critical features and using stain normalization and data augmentation. ECSAnet significantly outperforms several state-of-the-art models in multi-class breast cancer classification at various magnifications. Joseph *et al.* [31] achieved the best performance by combining handcrafted feature extraction with DNN and utilizing data augmentation, reaching an average accuracy of 97.10%. Chattopadhyay *et al.* [32] used a dual residual block in a multi-scale dual residual recurrent network (MTRRE-Net74), achieving superior accuracy at four magnification levels compared to previous models. ORTEGA-RUÍZ *et al.* [33] propose DRD-UNet, a deep learning architecture that enhances semantic segmentation of breast cancer tissue slices by introducing a DRD processing block with dilated convolutions, residual connections, and dense layers. This architecture performs well among various UNet models, achieving the highest results in the Jaccard similarity index, Dice coefficient, and overall segmentation accuracy.

Our research aims to design a deep learning classification model for multi-class breast cancer pathological images by exploring various convolutional neural network architectures. The model seeks to improve feature extraction from these images without increasing the model size compared to traditional models. By enhancing classification performance while increasing data volume, it aims to provide a valuable tool for computer-aided diagnosis systems.

The research objectives of this article are as follows:

1. Explore the impact of different current model architectures on breast cancer image classification;
2. Based on the residual structure, design a module for breast cancer pathological image classification;
3. Based on the idea of multi-layer neural network, a network structure that does not need to be too deep is designed for multi-classification tasks of breast cancer pathological images.

2. Related Work

2.1. Exploration and Research on Various Model Structures

AlexNet (Krizhevsky *et al.*, 2012) [6] was the first to use deep convolutional neural networks for high-resolution and large-scale image classification. The model innovatively introduced convolutional neural networks, ReLU activation functions, and dropout regularization for feature extraction and classification. Its five-layer structure became the foundation for subsequent deep learning models. LeCun *et al.* (2015) [7] used AlexNet to classify benign and malignant breast cancer pathological images, achieving a classification accuracy that was 6 percentage points higher than traditional machine learning methods.

GoogLeNet (Szegedy *et al.*, 2015) [8] is an image classification model inspired by AlexNet, optimized to improve recognition rates by refining network modules. It won the 2014 ImageNet (Beyer *et al.*, 2020) [9] image classification competition. The model's convolution blocks use 1x1 convolutions for dimensionality reduction, significantly reducing parameters and enhancing nonlinear representation capability. Fujioka *et al.* (2019) [10] collected 480 images of 96 benign tumors and 467 images of 144 malignant tumors as training data, using GoogLeNet to build a deep learning model. The results showed comparable or improved diagnostic performance, with AUCs of 0.913 and a range of 0.728–0.845, respectively.

VGGNet (Simonyan *et al.*, 2014) [11] is a deep convolutional neural network developed by the Oxford University Vision Group for the 2014 ImageNet image classification competition. It represents the depth and performance limits of traditional sequential convolutional neural networks. This model's simple structure excels at extracting deep semantic information from images and has been applied in areas such as transfer learning and interpretability analysis. Jahangeer *et al.* (2021) [12] integrated VGG-16 deep learning technology with cascade networks to segment filtered images, achieving good results on a breast cancer dataset. Agarwal *et al.* [13] analyzed performance on the publicly available BreakHis dataset using deep CNN and four popular CNN architectures (VGG16, VGG19, MobileNet, and ResNet 50) for histopathology image classification. Among these, VGG16 performed best with an accuracy of 94.67%, a precision of 92.60%, an F1 score of 85.21%, and a recall of 80.52%. Clement *et al.* (2022) [14] proposed a multi-scale pooled image feature representation (MPIFR) deep learning architecture combined with one-to-one support vector machines for classifying 8-category breast cancer histopathology images.

The hybrid CNN-LSTM model proposed by Srikantamurthy *et al.* [15] in 2023 was compared with existing CNN architectures for breast histopathology image classification, including VGG-16, ResNet50, and Inception models. All models were built using three optimizers—adaptive moment estimation (Adam), root mean square propagation (RMSProp), and stochastic gradient descent (SGD)—and trained with varying numbers of epochs.

Patel *et al.* (2023) [16] proposed GARL-Net, a graph-based adaptive regularized deep network for more accurate breast cancer classification, using DenseNet121 as the backbone network for transfer learning. They redesigned the loss function with adaptive regularization and complementary cross-entropy loss to address issues like uneven data distribution in pathological image datasets. To tackle these challenges, Agbley *et al.* (2023) [17] introduced a method that combines different magnification factors of histopathology images using residual networks and information fusion in federated learning (FL).

2.2. Residual-based Module Exploration

Gradient explosion, overfitting, and other issues caused by increasing model depth and width highlight that simply adjusting these dimensions may no longer significantly impact model performance. Such changes can introduce new problems. The introduction of ResNet (He *et al.*, 2016) [18] addressed these issues effectively. Its core idea involves residual blocks with identity mapping, which allow certain layers to skip connections and weaken the strong inter-layer connections, thereby mitigating network degradation. To tackle gradient disappearance or explosion, ResNet also utilizes data preprocessing and Batch Normalization (BN) layers.

Al-Haija *et al.* (2020) [19] used ResNet50 to classify benign and malignant samples from the BreakHis dataset (Lowe *et al*. 2014) [20]. They employed ResNet-50 pre-trained on ImageNet and applied transfer learning to classify the dataset. The model achieved an excellent classification accuracy of 99%, outperforming other models on the same dataset.

Xie *et al.* (2017) [21] developed ResNeXt, building on ResNet and incorporating parallel convolutions from Inception (Szegedy *et al.*, 2016). This model, which uses group convolutions within residual blocks, achieved second place in the ILSVRC 2016 classification task (Russakovsky *et al.*, 2016) [22], demonstrating the scalability of the ResNet structure and the impact of different model architectures on image processing tasks. DenseNet (Huang *et al.*, 2017) [23] extends ResNet's residual idea from local to global, where each layer's input includes outputs from all previous layers, improving information and gradient transmission efficiency.

Jiang *et al.* (2019) [24] designed the SE-ResNet module, incorporating attention mechanisms into ResNet, resulting in BHC-Net. This model achieved a classification accuracy of 0.93 on the BreakHis dataset, significantly outperforming models without attention mechanisms. RK *et al.* (2023) [25] proposed a redesigned ResNet-based model to minimize parameters and improve computational efficiency, using enhanced connections instead of traditional identity shortcuts.

Leow *et al.* (2023) [26] utilized multiple deep convolutional networks as feature extractors to extract and classify image features with random forest (RF) and k-nearest neighbor (KNN) algorithms. Similarly, Sahu *et al.* (2023) [27] used a pre-trained ResNet18 model for feature extraction from X-ray images and applied SVM for cancer diagnosis. To enhance performance, image quality was improved through contrast enhancement, followed by tumor segmentation using a histogram-based K-means technique.

Exploring different model structures and optimizing breast cancer pathological image classification remains challenging. It involves more than just increasing model depth; it requires a comprehensive consideration of parameters, structural complexity, and other design aspects. We aim to explore various deep learning architectures to achieve high performance with smaller model sizes and enhance computer-aided diagnosis and treatment systems.

3. SSResNet

3.1. Model Process

Based on the AlexNet framework, we designed a deep network using a Small SE-ResNeXt Block, named SSResNet, as shown in Figure 1. The model's feature extraction component consists of four layers, each utilizing our designed module. When a feature tensor enters this module, it passes through 32 parallel convolution branches, each performing non-linear convolution. After convolution, a channel attention mechanism re-evaluates feature importance. The feature tensors from these branches are then concatenated to complete the feature extraction process within the module. The specific model process is as follows:

1. **Image Preprocessing**: Images are resized to 224×224 after data augmentation, including cropping and inversion, before being input into the model.
2. **Initial Convolution and Pooling**: The image undergoes a standard convolutional pooling layer, increasing the channel count to 64 and reducing the image size to 56×56.
3. **Feature Extraction**: The processed (64, 56) feature tensors are passed through four consecutive Conv_Layers, each containing multiple Small SE-ResNeXt Blocks. The first block in each layer is responsible for downsampling the input feature tensor. To achieve this, we designed a Downsample Block based on the Small SE-ResNeXt structure. Additionally, we incorporated dense connections, inspired by DenseNet, between the first and last blocks of each layer to mitigate the vanishing gradient problem and enhance feature reuse.
4. **Final Convolution and Classification**: After passing through five convolutional layers, the resulting (2048, 7) tensor is reduced to a (2048, 1) tensor via average pooling (Avgpool) and then fed into a fully connected layer to output multi-classification results.

3.2. Samll-SE-ResNeXt Block

We designed Small SE-ResNeXt Block based on ResNeXt(Xie *et al.*, 2017). The detailed process of the module is as follows:

1. Using asymmetric parallel convolution, 1*3, 3*1 convolution units and 3*3 convolution units share the same weight, enhancing the square convolution kernel while reducing the amount of model parameters, superimposing asymmetric convolutions as shown in Figure 2.
2. Add the SE channel attention module after convolution. After the feature tensor is processed by the channel attention module, each feature channel is assigned a weight, and the neural network will focus on certain channels with large weight values. The entire channel attention module does not change the dimension and size of the feature tensor, but only implements 'assigning different weight values to different channels'. The channel attention mechanism is shown in Figure 3.
3. After completing asymmetric convolution and channel attention, concatenate each branch to obtain an output with the same dimension as the input feature tensor. The complete module structure is shown in Figure 4.

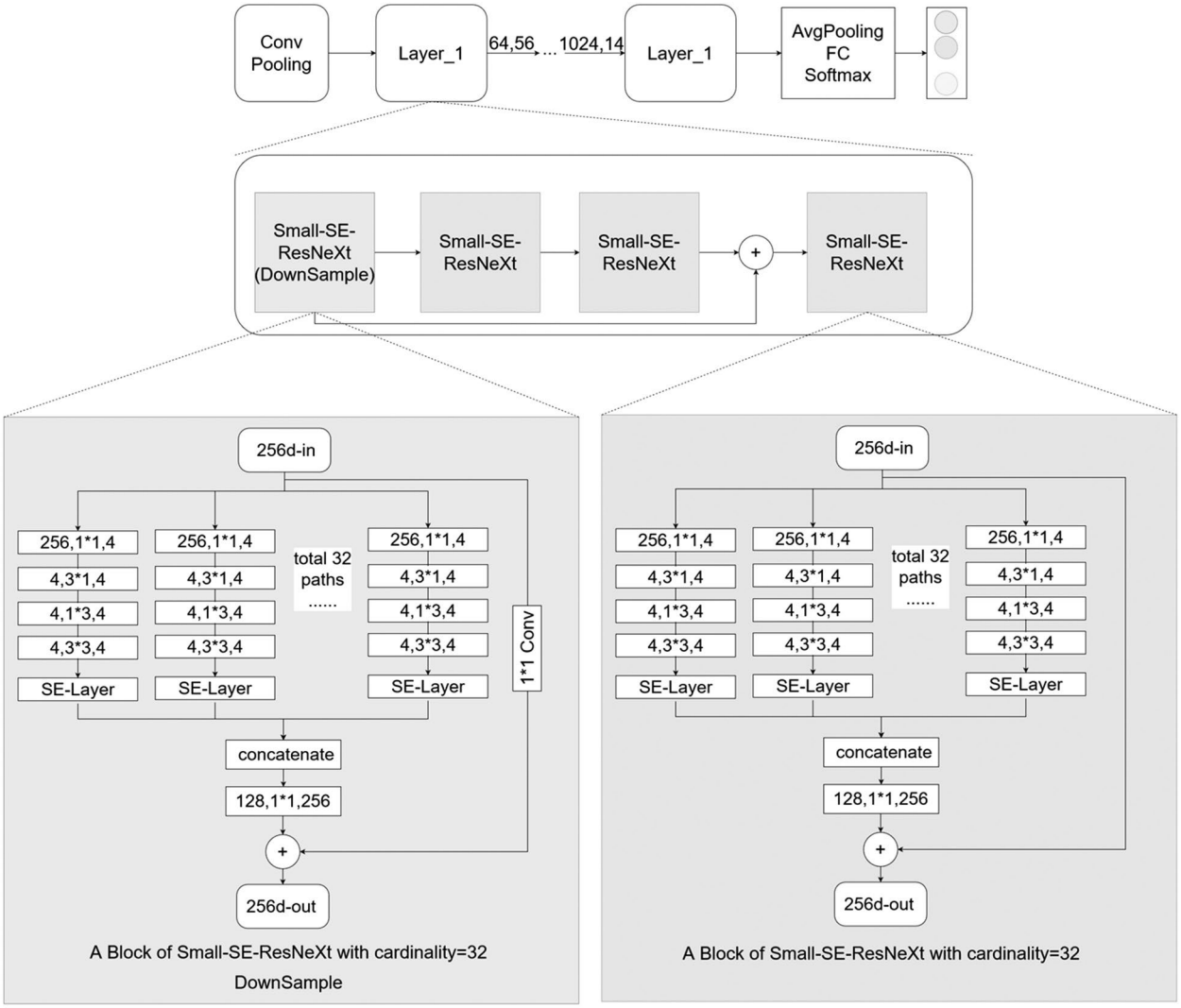

Figure 1. SSResNeXt model structure diagram.

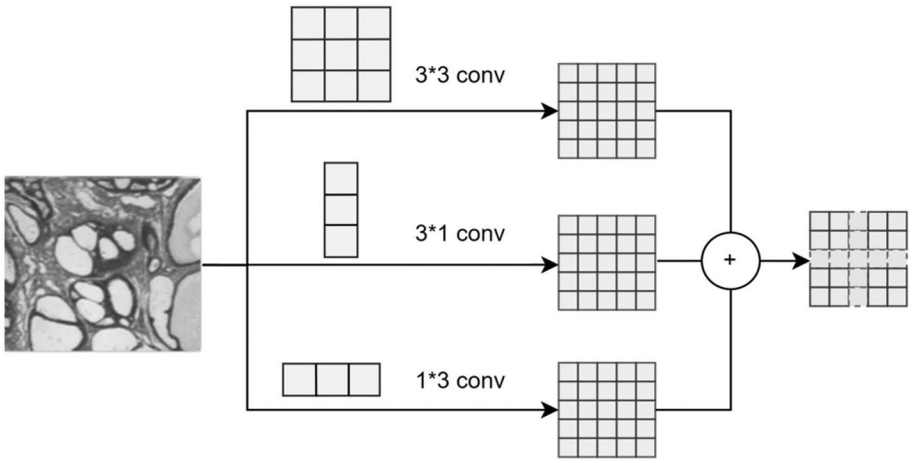

Figure 2. Superposition asymmetric convolution.

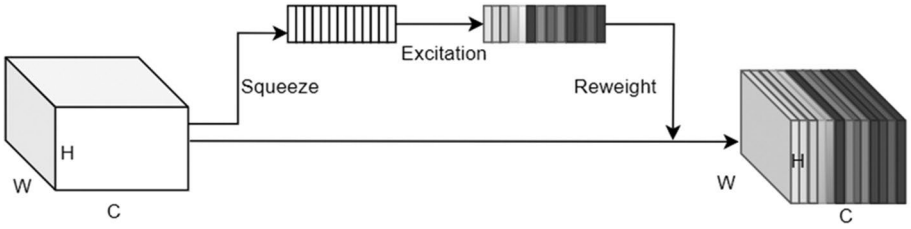

Figure 3. Channel attention mechanism.

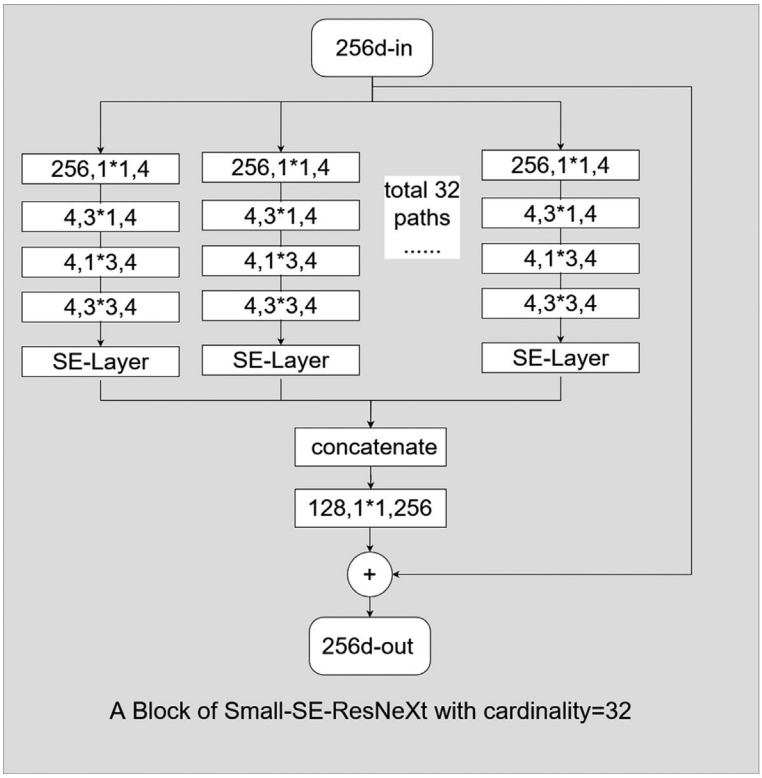

Figure 4. Small SE-ResNeXt Block.

We assume that the input tensor is X, $O_p(p \in [1, ..., 32])$ represents the output of each branch, O represents the output of a complete Small SE-ResNeXt Block.The complete convolution process of the module is expressed as:

$$O = \sum_{p=1}^{32} O_p + X \tag{1}$$

$$O_p = U_p * s = U_p * \delta\left(W_{p1}\, \partial\left(W_{p2} X\right)\right) \tag{2}$$

s represents the one-dimensional weight vector obtained after processing in the channel attention calculation, the sigmoid function is expressed as δ, the relu operation is ∂ The Excitation process is expressed as:

$$s = \delta\left(W_2\, \partial\left(W_1 z\right)\right) \tag{3}$$

W_{1*h} represents the length and width of the asymmetric convolution kernel, X represents the input tensor that has undergone serial asymmetric convolution, U_p represents the tensor obtained by serial convolution. The convolution process of a branch in Block is:

$$U_p = \partial W_{1*1}\left(\partial\left(W_{3*3}\left(W_{1*1} X\right) + \left(W_{1*3} + W_{3*1}\right)\left(W_{1*1} X\right)\right)\right) \tag{4}$$

3.3. Training

To optimize the performance of the pathological image multi-classification model, we employed a combination of loss functions. Specifically, we utilized Cross-Entropy Loss in conjunction with Dice Loss. Cross-Entropy Loss is suitable for multi-class classification tasks as it quantifies the difference between predicted and actual class labels. Dice Loss, typically used in segmentation tasks, is particularly effective for handling imbalanced classes by focusing on small sample categories.

The loss function is defined as follows:

$$Loss = \alpha \times Cross - Entropy\ Loss + \beta \times Dice\ Loss \tag{5}$$

where α and β are the weighting coefficients. Based on experimental results, both α and β were set to 0.5 to balance the impact of each loss function on the overall optimization.

The adaptive momentum estimation (ADAM) optimization method was used in the experiment, in which the initial value of the learning rate is 0.01 and is attenuated to 90% of the original value every 5 epochs. The dynamic learning rate attenuation can make the training process converge as quickly as possible. The Batch_size of the model is set to 8.

4. Experimental results and analysis

4.1. Dataset

To assess the effectiveness of our model in breast cancer pathology image classification, we utilized the BreaKHis dataset, published by the A Lowe team. The dataset comprises 7,909 pathological images, categorized by magnification into four levels: 40X, 100X, 200X, and 400X. Each image is stored with a resolution of 700×460 pixels. The dataset includes eight categories, as illustrated in Figure 5, covering both benign and malignant tumors. The benign tumors consist of four subtypes: adenosis (A), fibroadenoma (FA), tubular adenoma (TA), and phyllodes tumor (PT). The malignant tumors are divided into four subtypes: ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC).

This dataset has an imbalanced distribution, with malignant tumor images being overrepresented. This is a common issue in designing computer-aided diagnostic tools for medical imaging. Our experiment involves multi-class classification based on four different magnification factors. To prevent overfitting, we enhanced the dataset using techniques such as multiple-angle flipping, random cropping, and horizontal flipping within the PyTorch framework, aiming to improve the model's robustness in recognizing pathological images.

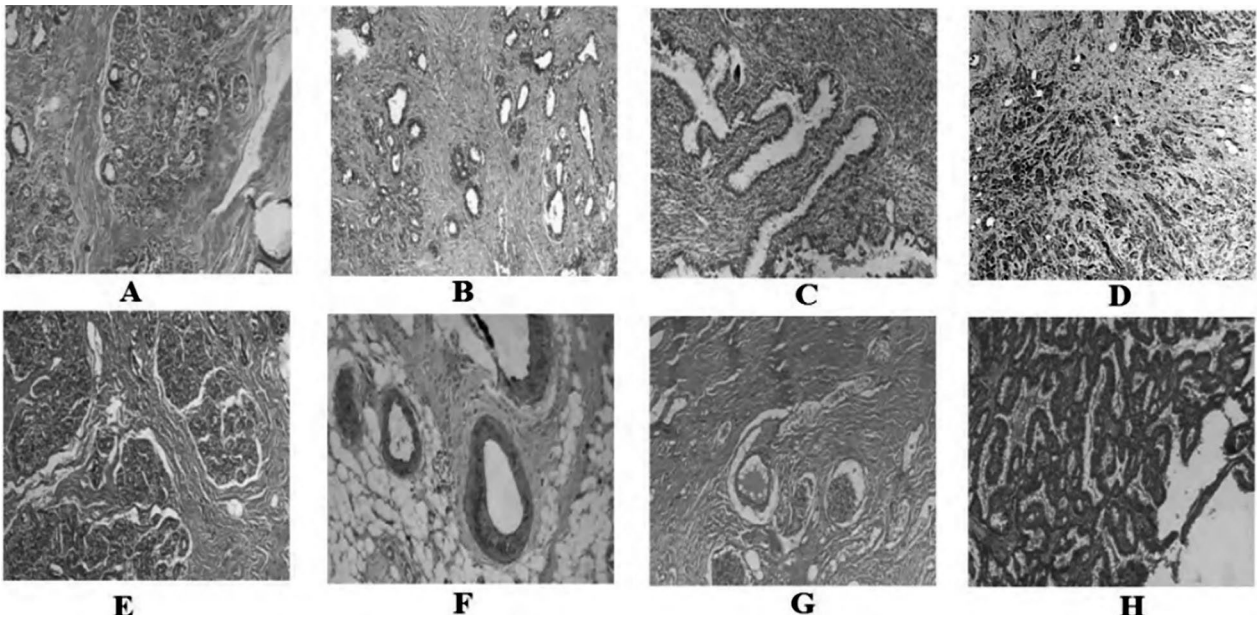

Figure 5. Data set example.

4.2. Experimental Result Evaluation Index

We adopt accuracy, precision, recall, and F1 score as evaluation metrics to evaluate the effectiveness of various classification networks. These indicators are calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

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

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

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{9}$$

These four values correspond to different categories in the binary confusion matrix. The confusion matrix is shown in Table 1.

Table 1. Confusion matrix.

	Predicted Negative	**Predicted Positive**
Actual Negative	TN	FP
Actual Positive	FN	TP

4.3. Ablation Experiment

4.3.1. Parametric Design

Obtaining good model parameters requires not only effective models and data but also proper adjustment of training parameters. These parameters impact convergence speed and model performance, and their effects can vary across different hardware environments. To evaluate the optimization of our model for pathological image classification, we conducted an ablation study. This study compared four models: the original ResNet50, ResNeXt50 with group convolution, SE-ResNeXt50 with an attention mechanism, and our designed SSResNeXt with asymmetric convolution. Each model underwent multi-class classification on data at four different scales.

The computer operating system used is Ubuntu16.04, the PyTorch version is 1.11, the CUDA version is 11.3, the python version is 3.9, the CPU is Intel(R) Core(TM) i9-12900k @3.50 GHz, the memory is 32 GB, and the GPU is NVIDIA GeForce RTX3090 (24 GB).

4.3.2. Analysis of Results

We incorporate modules based on ResNet50, and each model variation is evaluated on four different scale datasets, as shown in Table 2. The classification performance varies with scale, generally decreasing as the scale increases. This decline occurs because increasing the scale leads to a loss of global information, which is critical for pathological images lacking clear contours where texture details are important. At the same scale, adding more modules does not necessarily improve classification ability. In fact, stacking SE channel attention blocks without considering image characteristics can degrade model performance.

Our SSResNeXt model first uses asymmetric convolution to enhance texture details in different areas of the feature map, followed by applying attention weights. This approach maintains the model's computational efficiency while enhancing its performance.

In order to more conveniently compare the performance differences of various models in ablation experiments, we performed visual data analysis on f1-score and drew Figure 6.

As shown in the figure, the SSResNeXt model performs optimally across different scales, with F1-scores of 0.941, 0.926, 0.922, and 0.927 for magnification scales of 40X, 100X, 200X, and 400X, respectively. The limitations of the original ResNet model for multi-class classification of pathological images highlight that no single model fits all scenarios. Appropriate network structures should be chosen based on the specific characteristics of the data and scenario results.

Table 2. Comparison of F1 of each model in the ablation experiment.

	Groups	**SE_Layer**	**Asymmetry**		**Precision**	**Recall**
ResNet50				40X	0.855	0.871
ResNeXt50	✓				0.941	0.915
SE-ResNeXt50	✓	✓			0.9	0.886
SSResNeXt (our proposed)	✓	✓	✓		0.952	0.931
ResNet50				100X	0.916	0.903
ResNeXt50	✓				0.932	0.914
SE-ResNeXt50	✓	✓			0.926	0.911
SSResNeXt (our proposed)	✓	✓	✓		0.94	0.915
ResNet50				200X	0.893	0.893
ResNeXt50	✓				0.909	0.901
SE-ResNeXt50	✓	✓			0.897	0.9
SSResNeXt (our proposed)	✓	✓	✓		0.927	0.919
ResNet50				400X	0.891	0.881
ResNeXt50	✓				0.914	0.892
SE-ResNeXt50	✓	✓			0.927	0.912

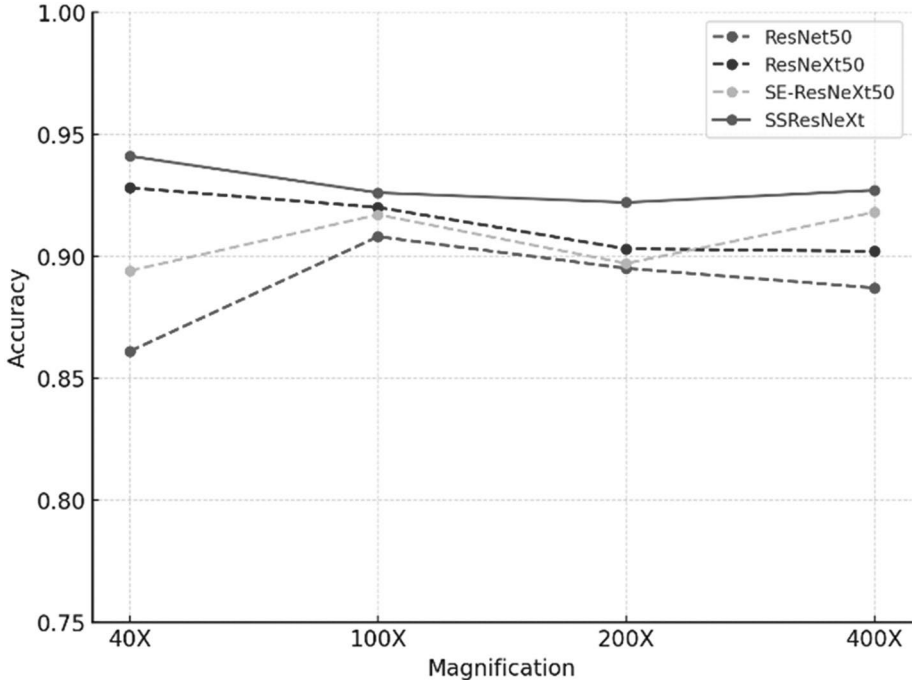

Figure 6. Comparison of F1 of each model in the ablation experiment.

In order to prove the validity of our ablation experimental results, we repeated the experiment for each module at different scales ten times and calculated the mean and standard deviation. At the same time, we provide statistical significance tests for performance comparisons as Figure 7. Let us take the 40X scale experimental statistical analysis in the figure below as an example: the box plot represents the distribution of results from ten repeated experiments, and '*' represents the significance of the differences between different models. The SSResNeXt we designed is significantly better than the other three models, and colleagues' repeated experiments also ensured the stability of the model. It can be found that the SSResNeXt we designed performs best at 40X scale. The performance of the same model at the 40X scale is also better than the other three scales. This is because a smaller magnification can better retain the global information of the image, which is especially important for unnatural images.

4.4. Comparative Experiment

Since this dataset is public and widely used for training various models, differences between models in binary classification tasks are minimal. In the field of medical diagnosis, improving tumor classification can enhance diagnostic support for doctors. Therefore, we compared our model with several recently prominent models.

Recently, vision transformers (Dosovitskiy *et al.*, 2020) [28] have gained attention in medical imaging due to their success in various computer vision tasks. The Swin Transformer, a variant of the Vision Transformer, uses non-overlapping windows and self-attention within local windows to effectively model image data. Tummala *et al.* (2022) [29] evaluated the SwinT ensemble for binary classification of benign and malignant tumors using the BreaKHis dataset. Jiang *et al.* (2022) [24] proposed HBCNet, a multi-classification network for medical images based on SE-ResNet, combining the residual structure of ResNet with the SE module to enhance feature extraction.We compare these similar models with the model we designed. The experimental results are as Table 3.

We will compare the experimental results of each model to assess the F1-scores for pathological images at different scales. As shown in Figure 8, our SSResNeXt model performs best in multi-class classification of pathological images. Compared to the unoptimized ResNet and the Swin Transformer, our model demonstrates superior performance.

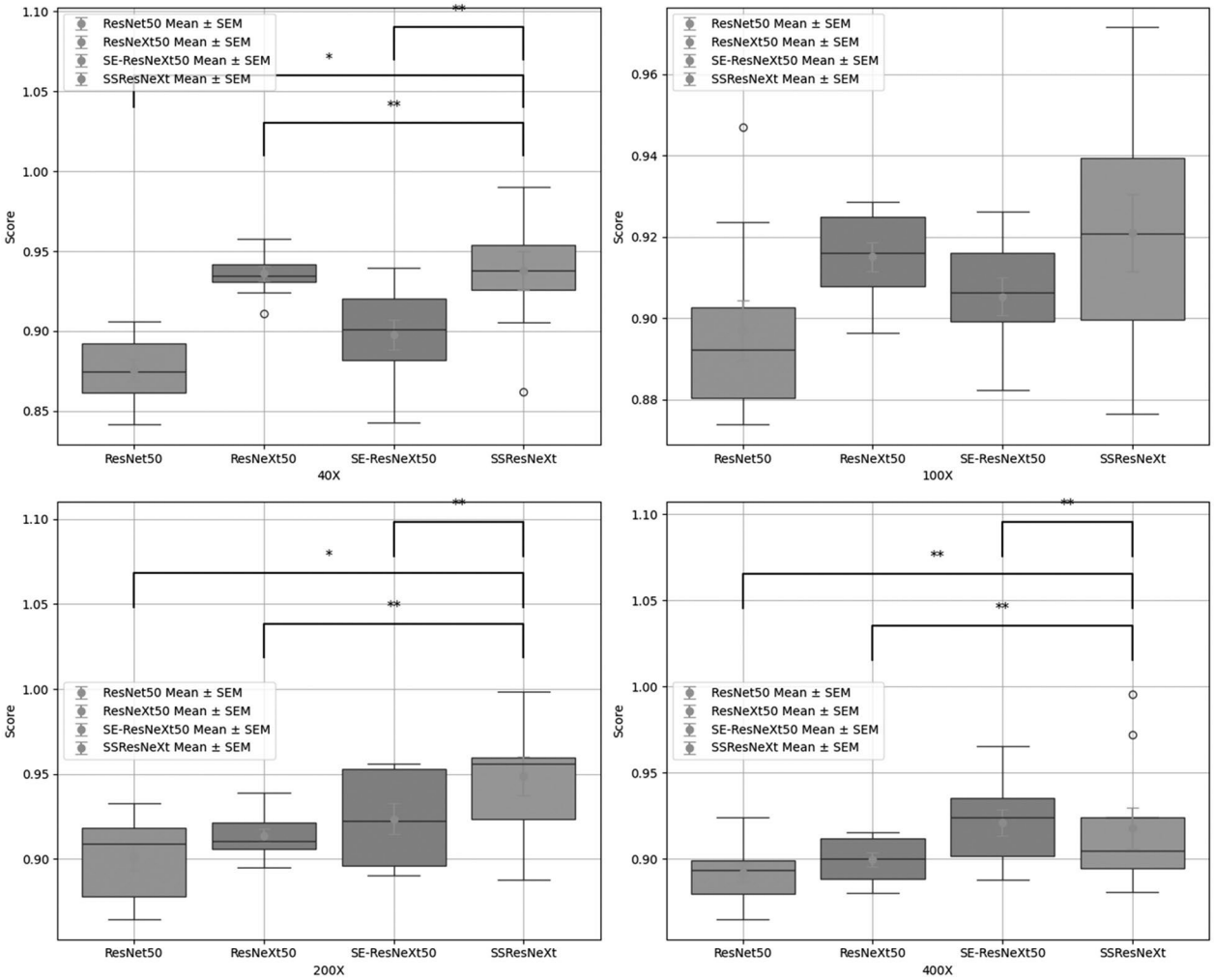

Figure 7. Statistical Significance Tests & Error Bars of F1-Score Across Different Scales.

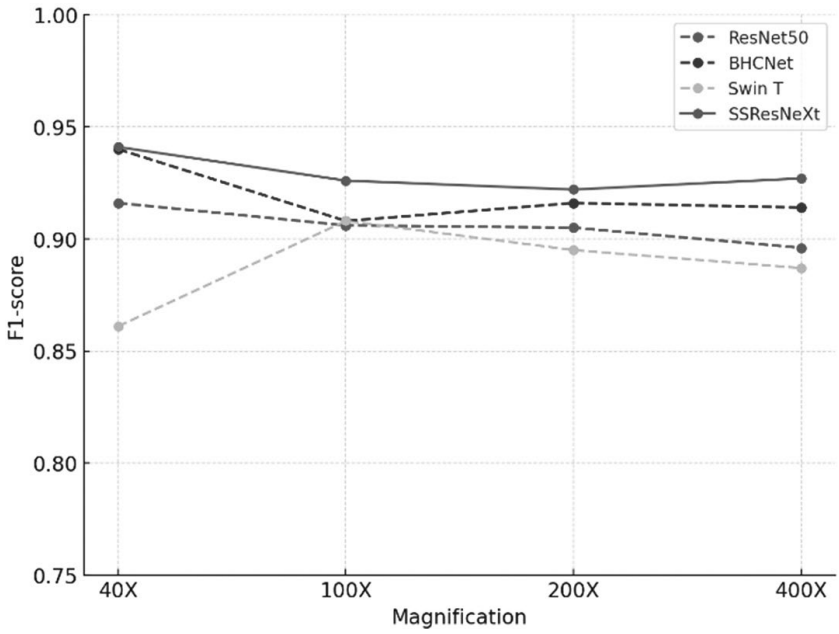

Figure 8. Comparison of F1 of each model in the comparative experiment.

Table 3. Comparison of F1 of each model in the ablation experiment.

	scale	**Precision**	**Recall**	**F1-score**
BHCNet	40X	0.936	0.901	0.916
Swin T		0.942	/	0.94
ResNet		0.855	0.871	0.861
SSResNeXt (our proposed)		0.952	0.931	0.941
BHCNet	100X	0.918	0.896	0.906
Swin T		0.909	/	0.908
ResNet		0.916	0.903	0.908
SSResNeXt (our proposed)		0.94	0.915	0.926
BHCNet	200X	0.921	0.893	0.905
Swin T		0.916	/	0.916
ResNet		0.893	0.893	0.895
SSResNeXt (our proposed)		0.927	0.919	0.922
BHCNet	400X	0.913	0.879	0.896
Swin T		0.915	/	0.914
ResNet		0.891	0.881	0.887
SSResNeXt (our proposed)		0.935	0.919	0.927

4.5. Model Complexity Comparison

Our proposed SSResNeXt model achieves excellent performance in multi-class classification of pathological images without significantly increasing model parameters or computational cost. Model complexity is shown in Table 4. Our model maintains a parameter count within the same order of magnitude as the base model, with potential for further optimization while improving the model structure. Params represents the number of model parameters, and FLOPs indicates the number of floating-point operations required for inference or training.

4.6. Discussion

We used the proposed Small-SE-ResNeXt Block to design a deep network model and achieved better F1-scores on the BreakHiS dataset, showing significant performance improvements compared to the base model. Our approach to module design for pathological images merits consideration in future experiments. In this work, we focused on designing a module specifically for pathological image feature extraction. The SSResNeXt built with this module performs exceptionally well, especially across different image scales, demonstrating its potential to enhance computer-aided diagnosis systems in clinical environments.

However, due to the limited availability of datasets, we have only conducted experiments with a single dataset. In the future, we plan to validate the model with different types of datasets. Additionally, the designed feature extraction module could be applied not only to classification tasks but also to segmentation tasks. Many excellent classification and segmentation models exist in medical image processing, most of which use convolutional approaches to extract deep features layer by layer, followed by pixel-level classification.

Table 4. Complexity comparison.

stage	**output**	**resnet**	**resnext**	**Our proposed**
conv1	112*112	7*7, 64, stride 2		
conv2	56*56	3*3, max pool, stride 2	3*3, max pool, stride 2	3*3, max pool, stride 2
		residual block * 3	resnext block * 3	Small-SE-ResNeXt Block * 3
conv2	28*28	residual block * 4	resnext block * 4	Small-SE-ResNeXt Block * 4
conv4	14*14	residual block * 6	resnext block * 6	Small-SE-ResNeXt Block * 6
conv5	7*7	residual block * 3	resnext block * 3	Small-SE-ResNeXt Block * 3
	1*1	globa average pool 1000-d fc, softmax	globa average pool 1000-d fc, softmax	globa average pool 8-d fc, softmax
params.		2.56E+07	2.50E+07	2.51E+07
FLOPs		4.13E+09	4.29E+09	4.21E+09

5. Conclusion

This study introduces SSResNeXt, a novel deep learning architecture for multi-classification of breast cancer pathological images. By incorporating asymmetric convolutions and channel attention mechanisms in the Small-SE-ResNeXt Block, our model achieves state-of-the-art performance on the BreaKHis dataset across multiple magnification scales. SSResNeXt's superior performance, particularly in handling different image scales, demonstrates its potential for improving computer-aided diagnosis systems in clinical settings. The success of our approach highlights the importance of tailoring deep learning architectures to the specific challenges of medical image analysis. Future work should focus on validating SSResNeXt on larger, more diverse datasets and exploring its applicability to other types of pathological images. Additionally, investigating the interpretability of the model's decisions could further enhance its utility in clinical practice. As deep learning continues to advance, architectures like SSResNeXt pave the way for more accurate and efficient diagnostic tools in breast cancer detection and classification.

Acknowledgement

This study was funded by Guangdong Provincial Medical Research Fund Project (NO. B2023321), Science and Technology Projects in Guangzhou (NO.2024A03J1043).

References

[1] C. N. Lim *et al.*, "Recent Advances in Breast Cancer Diagnosis Entering an Era of Precision Medicine", *Borneo Journal of Medical Sciences (BJMS)*, vol. 13, no. 1, p. 3, 2019. https://doi.org/10.51200/bjms.v13i1.1178

[2] J. V. Horvat *et al.*, "Calcifications at Digital Breast Tomosynthesis: Imaging Features and Biopsy Techniques", *Radiographics*, vol. 39, no. 2, pp. 307–318, 2019. https://doi.org/10.1148/rg.2019180124

[3] M. Proctor and C. Farquhar, "Diagnosis and Management of Dysmenorrhoea", *Bmj*, vol. 332, no. 7550, pp. 1134–1138, 2006. https://doi.org/10.1136/bmj.332.7550.1134

[4] T. Araújo *et al.*, "Classification of Breast Cancer Histology Images Using Convolutional Neural Networks", *PloS One*, vol. 12, no. 6, 2017. https://doi.org/10.1371/journal.pone.0177544

[5] A. G. Waks and E. P. Winer, "Breast Cancer Treatment: A Review", *Jama*, vol. 321, no. 3, pp. 288–300, 2019. http://dx.doi.org/10.1001/jama.2018.19323

[6] A. Krizhevsky *et al.*, "Imagenet Classification with Deep Convolutional Neural Networks", *Advances in Neural Information Processing Systems*, vol. 60, no. 6, pp. 84–90, 2012. https://doi.org/10.1145/3065386

[7] Y. LeCun *et al.*, "Deep Learning", *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] C. Szegedy *et al.*, "Rethinking the Inception Architecture for Computer Vision", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.

[9] L. Beyer *et al.*, "Are We Done with Imagenet?", arXiv preprint arXiv:2006.07159, 2020. https://doi.org/10.48550/arXiv.2006.07159

[10] T. Fujioka *et al.*, "Distinction Between Benign and Malignant Breast Masses at Breast Ultrasound Using Deep Learning Method with Convolutional Neural Network", *Japanese Journal of Radiology*, vol. 37, pp. 466–472, 2019. https://doi.org/10.1007/s11604-019-00831-5

[11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-scale Image Recognition", arXiv preprint arXiv:1409.1556, 2014. https://doi.org/10.48550/arXiv.1409.1556

[12] G. S. B. Jahangeer and T. D. Rajkumar, "Early Detection of Breast Cancer Using Hybrid of Series Network and VGG-16", *Multimedia Tools and Applications*, vol. 80, pp. 7853–7886, 2021. https://doi.org/10.1007/s11042-020-09914-2

[13] P. Agarwal *et al.*, "Breast Cancer Prediction on Breakhis Dataset Using Deep CNN and Transfer Learning Model", Data Engineering for Smart Systems: Proceedings of SSIC 2021, Springer, 2022, pp. 77–88. https://doi.org/10.1007/978-981-16-2641-8_8

[14] D. Clement *et al.*, "Multi-class Breast Cancer Histopathological Image Classification Using Multi-Scale Pooled Image Feature Representation (MPIFR) and One-versus-one Support Vector Machines", *Applied Sciences*, vol. 13, no. 1, p. 156, 2023. https://doi.org/10.3390/app13010156

[15] M. M. Srikantamurthy *et al.*, "Classification of Benign and Malignant Subtypes of Breast Cancer Histopathology Imaging Using Hybrid CNN-LSTM Based Transfer Learning", *BMC Medical Imaging*, vol. 23, no. 1, p. 19, 2023. https://doi.org/10.1186/s12880-023-00964-0

[16] V. Patel *et al.*, "GARL-Net: Graph Based Adaptive Regularized Learning Deep Network for Breast Cancer Classification", *IEEE Access*, vol. 11, pp. 9095–9112, 2023. http://dx.doi.org/10.1109/ACCESS.2023.3239671

[17] B. L. Y. Agbley *et al.*, "Federated Fusion of Magnified Histopathological Images for Breast Tumor Classification in the Internet of Medical Things", *IEEE Journal of Biomedical and Health Informatics*, 2023. http://dx.doi.org/10.1109/JBHI.2023.3256974

[18] K. He *et al.*, "Deep Residual Learning for Image Recognition", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[19] Q. A. Al-Haija and A. Adebanjo, "Breast Cancer Diagnosis in Histopathological Images Using ResNet-50 Convolutional Neural Network", in *Proceedings of the 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, IEEE, pp. 1–7, 2020.

[20] A. Lowe *et al.*, "Mitos Atypia Grand Challenge 2014", 2014, [online] Available: https://mitos-_atypia-_14.grand-_challenge.org/Dataset/

[21] S. Xie *et al.*, "Aggregated Residual Transformations for Deep Neural Networks", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1492–1500.

[22] O. Russakovsky *et al.*, "Imagenet Large Scale Visual Recognition Challenge", *International Journal of Computer Vision*, vol. 115, pp. 211–252, 2015. https://doi.org/10.1007/s11263-015-0816-y

[23] G. Huang *et al.*, "Densely Connected Convolutional Networks", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.

[24] Y. Jiang *et al.*, "Breast Cancer Histopathological Image Classification Using Convolutional Neural Networks with Small SE-ResNet Module", *PloS One*, vol. 14, no. 3, 2019. https://doi.org/10.1371/journal.pone.0214587

[25] R. K. Chandana Mani and J. Kamalakannan, "ERNet: Enhanced ResNet for Classification of Breast Histopathological Images", *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, vol. 22, no. 2, pp. 53–68, 2023. https://doi.org/10.5565/rev/elcvia.1614

[26] J. R. Leow *et al.*, "Breast Cancer Classification with Histopathological Image Based on Machine Learning", *International Journal of Electrical & Computer Engineering*, vol. 13, no. 5, pp. 2088–8708, 2023. http://dx.doi.org/10.11591/ijece.v13i5.pp5885-5897

[27] Y. Sahu *et al.*, "A CNN-SVM Based Computer Aided Diagnosis of Breast Cancer Using Histogram K-means Segmentation Technique", *Multimedia Tools and Applications*, vol. 82, no. 9, p. 14. https://doi.org/10.1007/s11042-022-13807-x

[28] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", arXiv preprint arXiv:2010.11929, 2020. https://doi.org/10.1007/s11042-022-13807-x

[29] S. Tummala *et al.*, "BreaST-Net: Multi-class Classification of Breast Cancer from Histopathological Images Using Ensemble of Swin Transformers", *Mathematics*, vol. 10, no. 21, p. 4109. https://doi.org/10.3390/math10214109

[30] L. A. Aldakhil *et al.*, "Attention-Based Deep Learning Approach for Breast Cancer Histopathological Image Multi-Classification", *Diagnostics*, vol. 14, no. 13, p. 1402, 2024. https://doi.org/10.3390/diagnostics14131402

[31] N. Praveen *et al.*, "Breast Cancer Histopathological Images Multi-classification using Deep Learning[J]", 2024.

[32] A. Gautam and S. K. Singh, "Analysis of Deep Learning Models to Detect Breast Cancer from Histopathology Images", in *Proceedings of the TENCON 2023-2023 IEEE Region 10 Conference (TENCON). IEEE*, 2023, pp. 438–443. http://dx.doi.org/10.1109/TENCON58879.2023.10322406

[33] M. A. Ortega-Ruíz *et al.*, "DRD-UNet, a UNet-like Architecture for Multi-class Breast Cancer Semantic Segmentation", *IEEE Access*, 2024. http://dx.doi.org/10.1109/ACCESS.2024.3377428

[34] M. M. Rahman *et al.*, "Efficient Multi-scale Convolutional Attention Decoding for Medical Image Segmentation", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 11769–11779. https://doi.org/10.48550/arXiv.2405.06880

[35] T. Lin *et al.*, "Interventional Bag Multi-instance-Learning on Whole-slide Pathological Images", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 19830–19839. https://doi.org/10.48550/arXiv.2303.06873

[36] J. Maurício *et al.*, "Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review", *Applied Sciences*, vol. 13, no. 9, p. 5521, 2023. https://doi.org/10.3390/app13095521

[37] B. Zhang *et al.*, "MATNet: A Combining Multi-attention and Transformer Network for Hyperspectral Image Classification", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023. http://dx.doi.org/10.1109/TGRS.2023.3254523

[38] Y. H. He, "Deep-learning the Landscape", Machine Learning: In Pure Mathematics and Theoretical Physics, 2023, pp. 183–221.

Received: July 2024
Revised: August 2024
Accepted: August 2024

Contact addresses:
Bo Xu
School of Information Science
Guangdong University of Finance and Economics
China
e-mail: xubo807127940@163.com

Faming Li
School of Information Science
Guangdong University of Finance and Economics
China
e-mail: 524241284@qq.com

Ying Wu*
Department of Ultrasound
The First Affiliated Hospital of Jinan University
Jinan University
China
e-mail: wuying1983@jnu.edu.cn
*Corresponding author

Bo Xu received the B.Eng. degree from Hengyang Normal University, China, in 2005, MSc degree in computer science and technology from the Hunan University, China, in 2009, and PhD degree in software engineering from the South China University of Technology, China, in 2017. He is currently an associate professor in the School of Information Science, Guangdong University of Finance and Economics, China. His research interests include machine learning and Medical Image Process.

Faming Li is a master student in the School of Information Science, Guangdong University of Finance and Economics. He received his bachelor's degree from Hunan Institute of Technology in 2022. His current research interests are in artificial intelligence methods for Medical Image Process.

Ying Wu received M.M. degree in Imaging and Nuclear Medicine from Jinan University, China, 2010 and M.D. degree in Imaging and Nuclear Medicine from Jinan University, China, 2019. She is currently an Attending Physician in the First Affiliated Hospital of Jinan University. Her current research interests are in artificial intelligence methods for Medical Image Process.