China Instrument and Control Society
Instrumentation
INSTRUMENTATION, Vol. 11, No. 1, March 2024 https://doi.org/10.15878/j.instr.202400058

Article

A Swin Transformer and Residualnetwork Combined Model for Breast Cancer Disease Multi-Classification Using Histopathological Images

Jianjun Zhuang[1, 2, *], Xiaohui Wu[1], Dongdong Meng[1] and Shenghua Jing[3]

[1] School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, NanJing 210044, China

[2] Institute for AI in Medicine, School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing 210044, China

[3] Department of Radiation Oncology, Jinling Hospital, School of Medicine Nanjing University, Nanjing 210002, China

[*] Corresponding author email: jjzhuang@nuist.edu.cn

Abstract: Breast cancer has become a killer of women's health nowadays. In order to exploit the potential representational capabilities of the models more comprehensively, we propose a multi-model fusion strategy. Specifically, we combine two differently structured deep learning models, ResNet101 and Swin Transformer (SwinT), with the addition of the Convolutional Block Attention Module (CBAM) attention mechanism, which makes full use of SwinT's global context information modeling ability and ResNet101's local feature extraction ability, and additionally the cross entropy loss function is replaced by the focus loss function to solve the problem of unbalanced allocation of breast cancer data sets. The multi-classification recognition accuracies of the proposed fusion model under 40X, 100X, 200X and 400X BreakHis datasets are 97.50%, 96.60%, 96.30 and 96.10%, respectively. Compared with a single SwinT model and ResNet101 model, the fusion model has higher accuracy and better generalization ability, which provides a more effective method for screening, diagnosis and pathological classification of female breast cancer.

Keywords: breast cancer pathological image; swin transformer; ResNet101; focal loss

Citation: Jianjun Zhuang Xiaohui Wu,Dongdong Meng1 and Shenghua Jing. "A Swin Transformer and Residual Network Combined Model for Breast Cancer Disease Multi-Classification Using Histopathological Images." *Instrumentation* 11, no. 1 (2024). https://doi.org/10.15878/j.instr.202400058

0 Introduction

Breast cancer is a major public health problem in women's health and has a significant impact on patient health and survival[1]. By 2030, it is expected to become the most common carcer in the US, accounting for 29% of all types[2]. Therefore, early detection and accurate classification are essential for guiding treatment and improving patient outcomes. Due to the complex anatomical structure of breast tissue and diverse tumor morphology, traditional medical imaging methods have certain limitations in breast cancer diagnosis, especially the low sensitivity to early lesions and the discomfort and risk associated with invasive examinations. These limitations have prompted the search for more accurate, convenient and non-invasive diagnostic methods. With the rapid development of computer vision technologies and deep learning algorithms, breast cancer classification methods based on medical images have been widely studied, as they can automatically extract and accurately classify features in support of the early screening and diagnosis of breast cancers.

Recently, Convolutional Neural Network (CNN)[3] have attracted considerable attention due to their excellent feature extraction and adaptability capability to imagery. Owing to local connections and weight-sharing features,

CNN can effectively capture local correlationsin pathological breast cancer images. Swetha and Vadivu[4] used CNN classification method based on the Gabor transform to detect and segment breast cancer regions. Srikantamurthy et al.[5] developed a hybrid CNN and long short-term memory recurrent neural network to identifyfour benign and four malignant breast cancer types. Rafiq et al.[6] proposed three CNN varieties to classify pathological breast cancer images. Sarker et al.[7] proposed a CNN that classifies breast cancer hematoxylin and eosin whole-slice images. However, it is difficult for a CNN to capture long-range dependenciesfrom input data owing to the model's strict limitations on the size and number of receptive fields, which is not conducive to capturing global dependencies images.

Swin Transformer (SwinT)[8] networks abandon convolutional operations and adopt apure attention mechanism to capture multiple global dependencies from input sequences, thereby achieving excellent performance in medical image classification tasks. Tummala et al.[9] used the publicly available BreakHis dataset to study the effectiveness of SwinTin classifying benign and malignant breast cancers and eight specific subtypes under different magnifications. Sunet al.[10] evaluated the performance of aSwinT for lung cancer classification and segmentation, showing that the pretrained Swin Transformer-based(Swin-B) model achieves a maximum classification accuracy of 82.26%, which is superior to other state-of-the-art methods.Cai et al.[11] proposed a multi-domain integrative SwinTmulti-instance learning network that accurately classifies full-section images of colorectal adenomas using only slide-level labels. Khan et al.[12] developed CervixFormer-an end-to-end, multi-scale Swin-B adversarial ensemble learning framework to assess pre-cancerous and cancer-specific cervical malignant lesions on whole-slide images. Wang et al.[13] developed an auxiliary diagnostic algorithm based on SwinT. However, a single model may not fully mine all data features. Therefore, combining the advantages of multiple models can effectively improve classification performance and generalizability.Ayas[14] proposed a SwinT model for multi-category skin lesion classification using a combination of a transformer and a CNN based on end-to-end mapping and requiring no prior knowledge. Lqbal et al.[15] proposed the BTS-ST network, a new Swin-B approach for breast tumor segmentation and classification, by integrating SwinT into U-Net based on conventional CNNs in order to improve global modeling capabilities.

To further improve the accuracy of breast cancer classification, a method combining a Swin transformer and a CNN Residual Network (ResNet) is proposed in this study. Specifically, the ResNet101 is selected, and weighted feature integrations from the two models are carried out to fully exploit the advantages of each model. The fused network can handle global context information viaSwinTand feature extraction capability via the ResNet101, which better captures important breast cancer pathological imageinformations. Simultaneously, to overcome the problem of unbalanced breast cancer dataset distributions, a focus loss function with good processing ability is selected to replace the cross-entropy loss function. The method proposed in this study can effectively improve the accuracy of breast cancer classification, and is of great significance to early diagnosis and treatment.

1 Materials and Methods

1.1 Methods

1.1.1 Fused SwinT and ResNet101 Model

Image processing and analysis is one of the key tasks in today's computer vision field, and different models often have their own strengths and limitations. SwinT and ResNet are two leading image processing models, each of which has demonstrated outstanding performance on specific tasks. SwinT, with its innovative attention mechanism and layering strategy, can efficiently process large-size images and capture rich semantic information, thus achieving impressive results in tasks such as image classification and segmentation. ResNet, as a classic deep convolutional neural network, is known for its powerful feature extraction capability and easy training, and is widely used in image recognition and target detection. However, it is often difficult for a single model to cover the optimal solution in all cases, so in this paper, fusion of different models is proposed with a view to combining their advantages and thus improving the overall performance. A multi-model fusion architecture integrates the outputs of two models, SwinT and ResNet, and by combining their respective feature representations, it is possible to compensate for each other's shortcomings and improve the overall performance. We designed a model that combines ResNet101 and Swin Transformer, and the network structure of this model is shown in Fig.1.

First the image is fed into two branches, ResNet101 and SwinT, respectively, and next, it goes through the feature extraction phase of the ResNet101 and SwinT models. For ResNet101, the image is first passed through a series of convolutional and pooling layers to gradually reduce the size and extract different levels of feature representations.ResNet101 is known for its deep network structure and residual connectivity, which allows it to efficiently capture a wide range of features in an image, from low-level to high-level. For the SwinT model, on the other hand, the image is embedded in chunks and a series of Transformer encoders, which gradually capture global information and multi-level features through a local window attention mechanism and a hierarchical processing strategy. After feature extraction is completed, the feature representations from ResNet101 and SwinT are fed into the feature fusion module. The feature fusion module linearly combines the feature vectors extracted from the two networks through a weighted average fusion

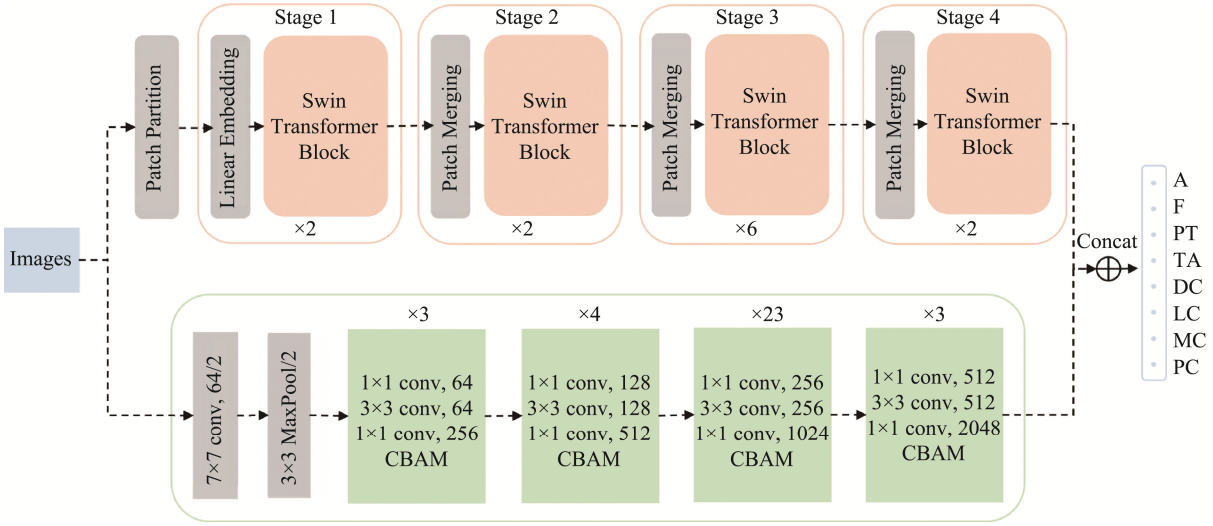

Fig.1 Fusion network model.

method to obtain more comprehensive and rich information. By fusing the ResNet101 and SwinT models, we are able to take full advantage of their respective strengths to improve the overall performance and robustness.The ResNet101 model is used to extract important local features, while Convolutional Block Attention Module (CBAM) is used to capture spatial and channel image feature dependencies while enhancing local information extraction.The SwinT extracts long-range semantic information from breast cancer images, and the feature fusion extracted by the two-branch network effectively overcomes the problem of weak long-range semantic information extracted by the CNN convolutional operation and insufficient local features captured by the SwinT.

1.1.2 Swin Transformer Block

The Swin transformer block is the basic component of the SwinT network, as shown in Fig.2, and it is used for feature extraction and interactions. Each Swin transformer block consists of a multilayer perceptron (MLP)[16] and a window MSA (W-MSA) and a shift window MSA (SW-MSA).The Swin transformer block achieves hierarchical feature interactions and transformations by stacking multiple transformer window attentions and MLPs. The window attention inside each Swin transformer block is used to capture local feature dependencies, whereas the MLP transforms the features in channel dimensions, allowing the network to better capture semantic information. By stacking Swin transformer blocks multiple times, the network effectively extracts and integrates multiscale image features and achieves powerful visual representations. The continuous Swin transformer block is calculated using Eqs. (1)–(4)[8].

$$\hat{z}^{l}=\mathrm{WMSA}(\mathrm{LN}(z^{l-1}))+z^{l-1} \quad (1)$$

$$z^{l}=\mathrm{MLP}(\mathrm{LN}(\hat{z}^{l}))+\hat{z}^{l} \quad (2)$$

$$\hat{z}^{l+1}=\mathrm{SWMSA}(\mathrm{LN}(z^{l}))+z^{l} \quad (3)$$

$$z^{l+1}=\mathrm{MLP}(\mathrm{LN}(\hat{z}^{l+1}))+\hat{z}^{l+1} \quad (4)$$

where $\hat{z}^{l}$ and z^{l} represent the output features of the W-MSA and MLP modules, respectively,for Block l, and $\hat{z}^{l+1}$ and z^{l+1} represent the outputs of the SW-MSA and MLP modules for Blocks l+1, respectively.

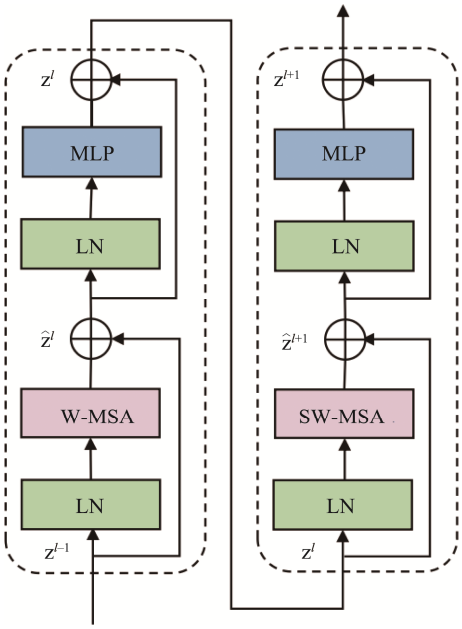

Fig.2 Network structure of Swin transformer module.

1.1.3 Convolutional block attention mechanism (CBAM)

To further capture the spatial and channel dependencies of image features and strengthen local information extraction, a CBAM[17] is added to the ResNet101 network to improve network attention to different image regions by adaptively adjusting each spatial position's importance in the feature map. CBAMs help CNNs better capture important features in images, thereby improving classification, detection, and segmentation performance. As shown in Fig.3, the CBAM module consists of a channel attention module (CAM) and a spatial attention module (SAM), which adaptively filter

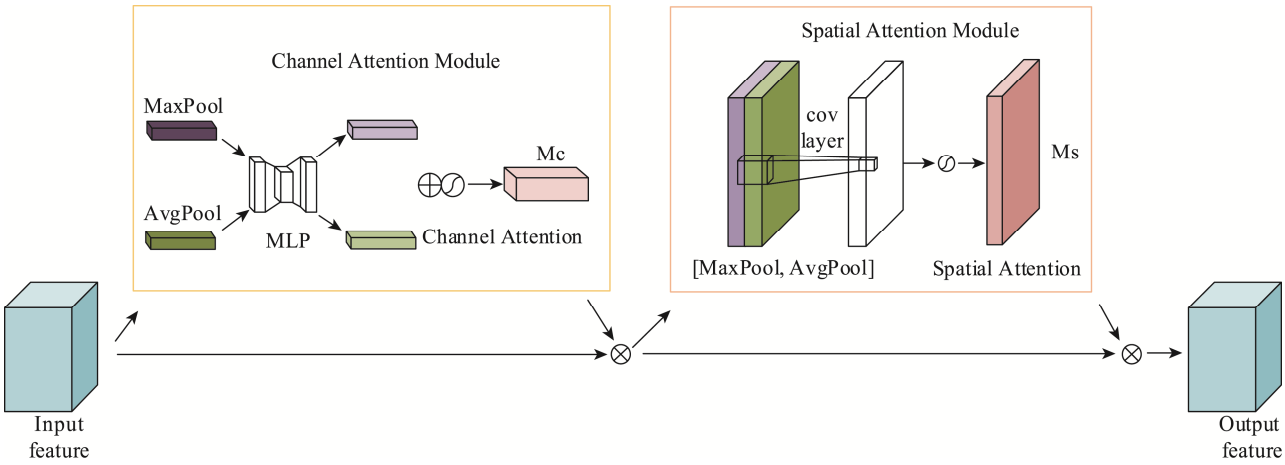

Fig.3 CBAM module.

the input features in the channel and spatial dimensions, respectively. The CAM performs global maximum pooling and global average pooling to obtain the global maximum feature and average feature for each channel. To optimize weights, two weight matrices are used for the same MLP.These output components are combined into a channel attention weighting module. This is shown in Eqs (5) and (6):

$$M_C(F) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))) \quad (5)$$

$$F' = M_C(F) \times F \quad (6)$$

where M_C represents the channel attention feature map, σ isthe Sigmoid activation function, MLP represents two fully connected MLP, F represents the input feature, and AvgPool and MaxPool indicate global average pooling and global maximum pooling, respectively, and F' represents the output features obtained by CAM.

The feature maps obtained after CAM are entered into SAM.The global maximum pooling and average pooling operations are first performed in SAM to obtain the maximum and average values of each pixel point, which are subjected to feature mapping and splicing, and finally the weights of each pixel point are obtained from the convolutional layer, which is subjected to an element-by-element multiplication operation with the original feature map to obtain the augmented feature representation.This is shown in Eqs (7) and (8):

$$M_S(F') = \sigma(f([\mathrm{AvgPool}(F'); \mathrm{MaxPool}(F')])) \quad (7)$$

$$F'' = M_S(F') \times F' \quad (8)$$

where M_S represents the spatial attention feature map, f is a convolutional layer operation and F'' denotes the output features obtained after SAM.

1.1.4 Focal Loss (FL) Function

Focal Loss(FL)[18] is a type of loss function that solves class-imbalance problems and is widely used in target detection and image classification tasks. With traditional cross-entropy loss functions, the loss of each sample is considered equally important. However, when faced with imbalance classes, samples of a few classes often occupy a small part of the entire sample, making it difficult for the model to learn their effective representations. This results in a model that is more biased towards a larger number of categories and has poor predictive performance for a smaller number of categories. The core concept of FL is to reduce the weight of easily classified samples, focus more on difficult samples, and solve the class-imbalance problem by introducing a balance factor and adjustment parameter. The balance factor can be adjusted according to the class weight of the sample, such that the sample of a few classes has a higher weight. The adjustment parameter is used to adjust the difficulty of samples that are easily misclassified. Specifically, the FL formula is shown in Eq (9):

$$\mathrm{FL}(p) = -\alpha(1-p)^{\gamma}\log p \quad (9)$$

where p represents the prediction probability, and $\alpha$is a balance adjustment factor basedon the class weight of the sample. γ is an adjustment parameter used to control the difficulty of the sample.By introducing balance factors and adjusting the parameters, FL effectively solves class-imbalance problems, allowing the model to focus more on difficult samples while improving classification accuracy for a small number of classes. Compared with the traditional cross-entropy loss function, FL significantly improves model in case of class imbalances.

1.2 Datasets and Experimental Programs

1.2.1 BreakHis Dataset

We used the BreakHis public dataset[19] to evaluate the proposed model's performance. The dataset, published in 2016 by Spanholet al., contains 7,909 breast histopathological images from 82 patients based on four magnification factors (i.e., 40X, 100X, 200X, and 400X), including 2,480 benign pathology images and 5,429 malignant pathology images. This dataset provides fine-grained clinical classification information for breast lesions, including adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA)in benign lesions, ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC) in malignant lesions. Hence, the BreakHis dataset trains both benign and malignant clinical significance models. Histopathological images of different breast cancer subtypes are shown in Fig.4, and those under different magnifications are listed in Table 1.

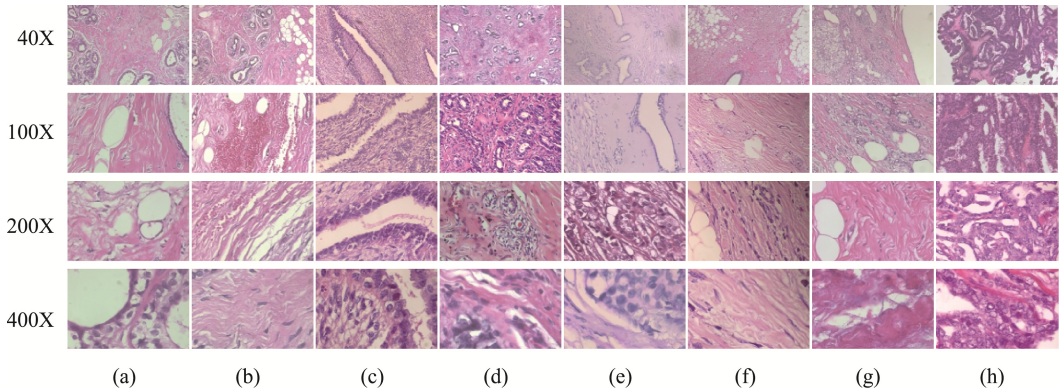

Fig.4 Pathological images of various breast tissue tumor types at different magnifications

Table 1 Number of pathological images of various breast tissue types at different magnifications.

Category	Tumor Type	Magnification Factors				Sum	Total
		40X	100X	200X	400X		
Benign	A	114	113	111	106	444	2,480
	F	253	260	264	237	1014	
	PT	149	150	140	130	569	
	TA	109	121	108	115	453	
Malignant	DC	864	903	896	788	3451	5,429
	LC	156	170	163	137	626	
	MC	205	222	196	169	792	
	PC	145	142	135	138	560	

1.2.2 Parameter Setting and Evaluation Metrics

The processor used in this experiment was an AMD Ryzen 9 7950X 16-Core processor. The Graphics card was an NVIDIA GeForce RTX 4090, andthe operating system was 64-bit Windows. Training was performed in PyCharm, PyCharm version 2022.2.2. The running environment was Python3.8, andthe deep learning framework was PyTorch 11.8. The dataset was randomly divided into training and verification sets ata ratio of 7:3. The Adam optimizer was used to train the model for 100 iterations.The initial learning rate was 0.0001, the batch size was 64,and the hyperparameters of α and γ were 0.25 and 2, respectively.In addition, owing to the small number of images in the dataset, data enhancement techniques were used to expand the training samples using methods of rotation, cropping, and other transformations.

For comparison with existing research results, this study uses Accuracy, Precision, Recall, F1, and confusion matrix logic as evaluation criteria. Accuracy, a classical metric, reflects the total observations based on the number of images correctly determined by the classification model, and the F1 is a measure of the test accuracy combined with recall rate. The role of the confusion matrix to evaluate and analyze model prediction performance in different categories. The specific distribution of each model category was observed by visualizing the confusion matrix whose diagonal represents correctly classified samples, and each column represents the probability that the predicted category is in this class.The sum of each column is one, andeach row represents the category to which the data belongs.The calculation formulas for Accuracy, Precision, Recall, and F1 are shown in Eqs (10)–(13):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

$$\text{F1} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{13}$$

where the true positive (TP) case reflects accurately predicted true cases and the false negative (FN) case reflects mistakenly predicted negative cases.The false positive (FP) case reflects mistakenly predicted positive cases and the true negative (TN) case reflects correctly predicted negative cases.

2 Results

2.1 Model Weight Allocation and Attention Selection

2.1.1 Model Weight Allocation

In multi-model fusion modeling, we target two models, ResNet and SwinT, and feature fusion and adjusting the weights is an effective strategy to improve the model performance. In the experiments, by adjusting the weights of the features extracted from different models, we can explore the impact of weight assignment on the performance of the final model, and by comparing the performance of the model on the test set under different weight assignments, we can find the optimal combination of weights, so as to improve the performance and generalization ability of the overall model. For

thispurpose we conducted experiments on the following weight assignments on the breast cancer dataset at 400X magnification. The experimental results are shown in Table 2.

Table 2 Accuracy of two models with different weights.

SwinT	ResNet101	Accuracy
0.4	0.6	95.5%
0.5	0.5	95.6%
0.6	0.4	96.1%
0.7	0.3	95.7%
0.8	0.2	95.6%

According to Table 2, we can see that the proposed fusion model performs best under the 400X magnification BreakHis dataset when the weight assigned to SwinT is 0.6 and the weight of ResNet101 is 0.4, so given the weight configuration of 0.6 for SwinT and 0.4 for ResNet101 is given a weight configuration of 0.4.

2.1.2 Attention Selection

Attention mechanisms are widely used in deep learning to improve model performance, of which SE (Squeeze-and-Excitation) and CBAM are two common attention mechanisms. Here we compare different attentional mechanisms including Baseline group, SE group and CBAM group to evaluate the performance of these attentional mechanisms.Baseline group is the benchmark model without applying any attentional mechanism, which is used as the basis for comparison.SE group is the experimental group applying the SE module, which is capable of adaptively adjusting the channel feature response to enhance the model's characterization ability. The CBAM group, on the other hand, is the experimental group that has applied the CBAM module, which combines the channel attention and spatial attention mechanisms to capture the correlation between features more comprehensively. By comparing the experimental results of these three combinations, the impact of different attention mechanisms on model performance can be better assessed to provide guidance for model design and optimization.

According to Table 3, It can be seen that Baseline achieved 97.2% accuracy as the baseline model, SE slightly improved to 97.3% with the introduction of channel attention, while CBAM achieved 97.5% accuracy based on the combination of channel and spatial attention. This suggests that CBAM has an advantage on this dataset in that it is able to capture the correlation between features more comprehensively, which helps to improve the performance of the image classification task.

Table 3 Accuracy under different attentional mechanismsin 40X dataset.

Description	Accuracy
Baseline	97.2%
SE	97.3%
CBAM	97.5%

2.2 Experimental Comparison

To verify the effectiveness of the improved model, we compare our results with recent state-of-the-art counterparts on BreakHis dataset[20-24].The comparison results are presented in Table 4. In this study, a ResNet101 fusion model with a SwinTand a CBAM attention mechanism was adopted. During training, the loss function was converted to afocal loss with better processing ability for unbalanced datasets. Ultimately, the accuracy of the eight classification methods at four different magnifications reached 96.10%–97.50%, which is better than the existing classification methods for breast cancers.

To observe the effectiveness of the proposed method more directly, the confusion matrix was analyzed further to understand the incorrect model predictions in certain categories, identify potential problems and patterns, and take corresponding performance improvement actions. Fig.5 shows the confusion matrix of the fusion model for the 40X magnification dataset, with true labels on the vertical axis and predicted labels on the horizontal axis.

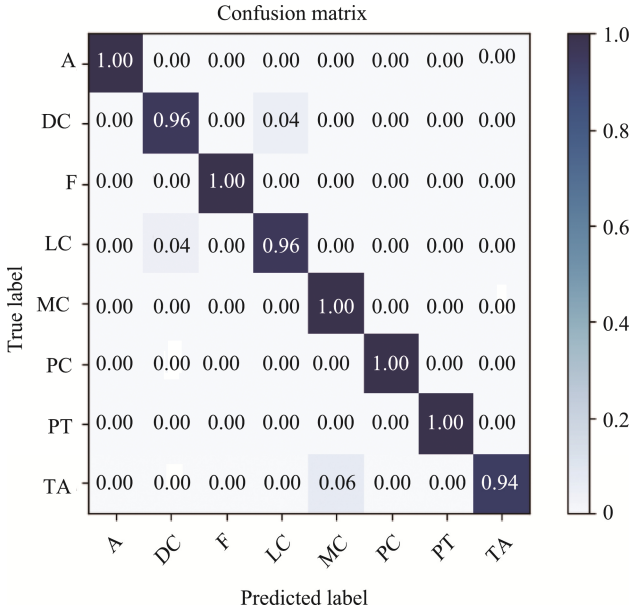

Fig.5 Confusion matrix of the fusion model at 40X magnificationdataset.

2.3 Ablation Experiment

To verify the effectiveness of the proposed fusion model, an ablation experiment was conducted on the BreakHis dataset by controlling for each model component.The Accuracy, Precision, Recall and F1 of a total of six models are listed in Table 5. From the experimental results, we can see that our proposed model shows good results in Accuracy, Precision, Recall, and F1, which indicates that the model achieves good results in several evaluation metrics, and all the three proposed improvement measures can effectively enhance the model accuracy.

Table 4 Comparison with existing breast cancer classification studies with differentmagnificationdataset.

dataset	document	Accuracy	Precision	Recall	F1
40X	Boumarafet al. [20]	94.49%	93.81%	94.78%	94.15%
	Zaalouket al. [21]	97.01%	96.85%	96.17%	96.47%
	Sarkeret al. [22]	95.16%	96.16%	95.82%	95.99%
	Pandeyet al. [23]	96.85%	97.67%	96.91%	97.28%
	Hu et al. [24]	89.72%	90.00%	90.00%	90.00%
	Our	97.50%	96.89%	98.38%	97.63%
100X	Boumarafet al. [20]	93.27%	92.94%	91.59%	92.23%
	Zaalouket al. [21]	95.17%	95.08%	94.02%	94.37%
	Sarkeret al. [22]	94.34%	94.40%	94.87%	94.59%
	Pandeyet al. [23]	96.59%	97.79%	96.36%	96.85%
	Hu et al. [24]	90.84%	91.00%	91.00%	91.00%
	Our	96.60%	96.64%	96.55%	96.59%
200X	Boumarafet al. [20]	91.29%	91.18%	88.28%	89.47%
	Zaalouket al. [21]	91.54%	90.08%	90.16%	89.91%
	Sarkeret al. [22]	86.83%	89.65%	81.33%	84.75%
	Pandeyet al. [23]	95.36%	96.95%	95.42%	96.17%
	Hu et al. [24]	92.04%	93.00%	92.00%	92.00%
	Our	96.30%	96.08%	96.50%	96.29%
400X	Boumarafet al. [20]	89.56%	87.97%	87.97%	87.77%
	Zaalouket al. [21]	90.22%	90.99%	89.87%	89.97%
	Sarkeret al. [22]	93.48%	92.75%	92.02%	91.90%
	Pandeyet al. [23]	94.58%	95.87%	94.68%	95.26%
	Hu et al. [24]	94.21%	94.00%	94.00%	94.00%
	Our	96.10%	96.39%	95.70%	96.04%

Table 5 Accuracy, Precision, Recall, and F1 of ablation experiments with different magnification dataset.

dataset	Model	Accuracy	Precision	Recall	F1
40X	Swin transformer	96.7%	96.91%	97.05%	96.98%
	Swin transformer+FL	97.0%	96.81%	98.30%	97.55%
	ResNet101	96.5%	96.43%	96.80%	96.61%
	ResNet101+FL	96.7%	96.13%	97.99%	97.05%
	Fusion Model	97.2%	97.13%	97.88%	97.50%
	Fusion Model +FL	97.5%	96.89%	98.38%	97.63%
100X	Swin transformer	95.3%	95.29%	94.64%	94.96%
	Swin transformer+FL	96.1%	96.51%	95.69%	96.10%
	ResNet101	94.2%	94.90%	93.48%	94.18%
	ResNet101+FL	94.7%	93.58%	95.38%	94.47%
	Fusion Model	95.8%	95.20%	95.78%	95.49%
	Fusion Model +FL	96.6%	96.64%	96.55%	96.59%
200X	Swin transformer	95.5%	95.25%	96.33%	95.79%
	Swin transformer+FL	96.0%	95.84%	96.29%	96.06%
	ResNet101	94.2%	94.90%	93.48%	94.18%
	ResNet101+FL	94.4%	94.36%	94.43%	94.39%
	Fusion Model	95.8%	95.84%	95.44%	95.64%
	Fusion Model +FL	96.3%	96.08%	96.50%	96.29%
400X	Swin transformer	95.2%	94.48%	94.09%	94.28%
	Swin transformer+FL	95.7%	94.83%	96.61%	95.71%
	ResNet101	93.5%	93.29%	91.41%	92.34%
	ResNet101+FL	93.7%	93.91%	94.16%	94.03%
	Fusion Model	95.9%	95.30%	95.41%	95.35%
	Fusion Model +FL	96.1%	96.39%	95.70%	96.04%

3 Discussion

Analyzing the results of comparative experiments, the best model classification achieved a detection Accuracy of 97.5%, Precision of 96.89%, Recall of 98.38%, and an F1 of 97.63% under the 40X magnification condition. This is because the 40X magnification retains a greater amount of information from the original image compared to higher magnifications, thereby preserving richer details. In breastcancer images, lesions often exhibit small-scale characteristics. Lower magnification aids in preserving these crucial fine structures, facilitating the model's capture of these significant lesion features. Furthermore, lower magnification reduces image noise and eliminates redundant information, allowing the model to concentrate on genuine features and reducing sensitivity to interference. This ultimately enhances the model's classification performance.

As can be seen from Table 5, the proposed method improves model accuracy, as verified under four different magnification datasets: 40X, 100X, 200X and 400X. The fusion model achieved significant improvements in breast cancer classification compared with the single model, showing higher accuracy and better generalizability than ResNet101 or SwinTalone. The local feature extraction capability of ResNet101 and the long-distance semantic information capability of Swin transformer can be effectively utilized by fusing the models so that the focus loss function can help the model focus on the difficult parts of classification while improving the accuracy of breast cancer classification.

As can be seen from the confusion matrix in Fig.5, The fusion model has good classification performance for each category classification under 40X magnification dataset, further reflecting the strong performance of the model in multi-category classification task with high differentiation ability and accuracy for each category.

4 Conclusion

Classifying pathological breast cancer images is a crucial medical task supporting tumor identification and classification. To improve classification performance, we proposed a method that combines ResNet101 and the SwinT. By combining the feature extraction results from the ResNet101 and SwinT, we took full advantage of both models. The features obtained after fusion better captured important information from pathological breast cancer images, thus improving the classification performance and generalizability of the method. Simultaneously, the cross-entropy loss function was replaced by a focus loss function that pays more attention to difficult samples to solve problems of unbalanced breast cancer dataset distributions. Our trained fusion model was evaluated on the test set, and performance indicators, such as Accuracy, Precision, Recall, and F1, were calculated. By analyzing the confusion matrices, we can further understand how the model performs in different categories and discover potential room for improvement.

In summary, our fused ResNet101 and SwinT breast cancer classification model strikes a good balance between image feature extraction and attention mechanisms, thereby improving overall classification performance. This fusion method has broad application prospects in medical imaging diagnoses. However, additional research and experimental validation are needed to determine the best fusion strategy while driving the continued development and innovation of breast cancer classification techniques.

Author Contributions:

Jianjun Zhuang: Writing-review & editing; Funding acquisition. WU Xiaohui: Experimental design; Writing-original draft; Formal analysis; Validation. MENG Dongdong: Data curation; Data preprocessing. JING Shenghua: Supervision.

Funding Information:

By the National Natural Science Foundation of China (NSFC) (No. 61772358), the National Key R&D Program Funded Project (No. 2021YFE0105500), and the Jiangsu University ‘Blue Project’.

Data Availability:

The authors declare that the main data supporting the findings of this study are available within the paper and its Supplementary Information files.

Conflict of Interest:

The authors declare no competing interests.

Dates:

Received 11 January 2024; Accepted 24 March 2024; Published online 31 March 2024

References

[1] Azamjah, N., Soltan-Zadeh, Y., Zayeri, F., Global trend of breast cancer mortality rate: A 25-year study. Asian Pacific Journal of Cancer Prevention: APJCP, 2019, 20 (7): 2015.

[2] Siegel, R.L., Miller, K.D., Wagle, N.S., et al., Cancer statistics, 2023.CA: A Cancer Journal for Clinicians, 2023, 73 (1): 17-48.

[3] Gu, J., Wang, Z., Kuen, J., et al., Recent advances in convolutional neural networks. Pattern recognition, 2018, 77: 354-377.

[4] Swetha, V., Vadivu G., Classifications of benign and malignant mammogram images using Gabor-modified CNN architecture. International Journal of Imaging Systems and Technology, 2023, 33 (5): 1682-1695.

[5] Srikantamurthy, M.M., Rallabandi, V.P., Dudekula, D.B., et al., Classification of benign and malignant subtypes of

breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. BMC Medical Imaging, 2023, 23 (1).

[6] Rafiq, A., Chursin, A., AwadAlrefaei. W., Detection and Classification of Histopathological Breast Images Using a Fusion of CNN Frameworks. Diagnostics, 2023, 13 (10): 1700.

[7] Sarker, M.M.K., Akram, F., Alsharid, M., Efficient breast cancer classification network with dual squeeze and excitation in histopathological images. Diagnostics, 2022, 13 (1): 103.

[8] Liu, Z., Lin, Y.T., Cao, Y., et al., Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF international conference on computer vision, 2021, 2021: 10012-10022.

[9] S. Tummala, J. Kim, S. Kadry, BreaST-Net: Multi-class classification of breast cancer from histopathological images using ensemble of swin transformers. Mathematics, 2022, 10 (21): 4109.

[10] Sun, R., Pang, Y., Li, W., Efficient Lung Cancer Image Classification and Segmentation Algorithm Based on an Improved Swin Transformer. Electronics, 2023, 12 (4): 1024.

[11] Cai, H., Feng, X., Yin, R., et al., MIST: multiple instance learning network based on Swin Transformer for whole slide image classification of colorectal adenomas, The Journal of Pathology, 2023, 259 (2): 125-135.

[12] Khan, A., Han, S., Ilyas, N., et al., CervixFormer: A Multi-scale swin transformer-Based cervical pap-Smear WSI classification framework. Computer Methods and Programs in Biomedicine, 2023, 240: 107718.

[13] Wang, Y., Luo, F., Yang, X., et al., The Swin-Transformer network based on focal loss is used to identify images of pathological subtypes of lung adenocarcinoma with high similarity and class imbalance. Journal of Cancer Research and Clinical Oncology, 2023: 1-12.

[14] Ayas S. Multiclass skin lesion classification in dermoscopic images using swin transformer model[J]. Neural Computing and Applications, 2023, 35 (9): 6713-6722.

[15] Lqbal A, Sharif M. BTS-ST: Swin transformer network for segmentation and classification of multimodality breast cancer images[J]. Knowledge-Based Systems, 2023, 267: 110393.

[16] Zheng, H., Wang, G., Li, X., Swin-MLP: A strawberry appearance quality identification method by Swin Transformer and multi-layer perceptron.Journal of Food Measurement and Characterization, 2022, 16 (4): 2789-2800.

[17] Lin, T.Y., Goyal, P., Girshick, R., et al., Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 99: 2999-3007.

[18] Woo, S., Park, J., Lee, J.Y., et al., CBAM: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV), 2018, 11211: 3-19.

[19] FSpanhol, F.A., Oliveira, L.S., Petitjean, C., et al., A dataset for Breast Cancer Histopathological Image Classification. IEEE Transactions on Biomedical Engineering, 2018, 63 (7): 1455-1462.

[20] Boumaraf, S., Liu, X., Zheng, Z., et al., A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images. Biomedical Signal Processing and Control, 2021, 63: 102192.

[21] Zaalouk, A.M., Ebrahim, G.A., Mohamed, H.K., et al., A deep learning computer-aided diagnosis approach for breast cancer. Bioengineering, 2022, 9 (8): 391.

[22] Sarker, M.M.K., Akram, F., Alsharid, M., et al., Efficient Breast Cancer Classification Network with Dual Squeeze and Excitation in Histopathological Images. Diagnostics, 2023, 13 (1): 103.

[23] Pandey, A., Kumar, A., An integrated approach for breast cancer classification. Multimedia Tools and Applications, 2023, 1-21.

[24] Hu, T., Wu, M., Liu, Y., et al., Classification of breast cancer histopathological images based on SE-DenseNet. Journal of Shaoguan University, 2023, 44 (03): 20-27.