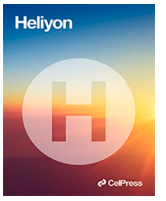
Heliyon 10 (2024) e30889

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Convolutional neural network with parallel convolution scale attention module and ResCBAM for breast histology image classification

Ting Yan [a], Guohui Chen [a], Huimin Zhang [b], Guolan Wang [c], Zhenpeng Yan [a], Ying Li [b], Songrui Xu [a], Qichao Zhou [a], Ruyi Shi [d], Zhi Tian [e,f], Bin Wang [b,*]

[a] *Translational Medicine Research Center, Shanxi Medical University, Taiyuan, China*
[b] *College of Information and Computer, Taiyuan University of Technology, Taiyuan, China*
[c] *Computer Information Engineering Institute, Shanxi Technology and Business College, Taiyuan, China*
[d] *Department of Cell Biology and Genetics, Shanxi Medical University, Taiyuan, Shanxi, 030001, China*
[e] *Second Clinical Medical College, Shanxi Medical University, 382 Wuyi Road, Taiyuan, Shanxi, 030001, China*
[f] *Department of Orthopedics, The Second Hospital of Shanxi Medical University, Shanxi Key Laboratory of Bone and Soft Tissue Injury Repair, 382 Wuyi Road, Taiyuan, Shanxi, 030001, China*

ARTICLE INFO

Keywords:
Breast cancer
Parallel convolution scale attention
Dilation convolution
Feature fusion
ResCBAM

ABSTRACT

Breast cancer is the most common cause of female morbidity and death worldwide. Compared with other cancers, early detection of breast cancer is more helpful to improve the prognosis of patients. In order to achieve early diagnosis and treatment, clinical treatment requires rapid and accurate diagnosis. Therefore, the development of an automatic detection system for breast cancer suitable for patient imaging is of great significance for assisting clinical treatment. Accurate classification of pathological images plays a key role in computer-aided medical diagnosis and prognosis. However, in the automatic recognition and classification methods of breast cancer pathological images, the scale information, the loss of image information caused by insufficient feature fusion, and the enormous structure of the model may lead to inaccurate or inefficient classification. To minimize the impact, we proposed a lightweight PCSAM-ResCBAM model based on two-stage convolutional neural network. The model included a Parallel Convolution Scale Attention Module network (PCSAM-Net) and a Residual Convolutional Block Attention Module network (ResCBAM-Net). The first-level convolutional network was built through a 4-layer PCSAM module to achieve prediction and classification of patches extracted from images. To optimize the network's ability to represent global features of images, we proposed a tiled feature fusion method to fuse patch features from the same image, and proposed a residual convolutional attention module. Based on the above, the second-level convolutional network was constructed to achieve predictive classification of images. We evaluated the performance of our proposed model on the ICIAR2018 dataset and the BreakHis dataset, respectively. Furthermore, through model ablation studies, we found that scale attention and dilated convolution play an important role in improving model performance. Our proposed model outperforms the existing state-of-the-art models on 200 × and 400 × magnification datasets with a maximum accuracy of 98.74 %.

* Corresponding author.
E-mail address: wangbin01@tyut.edu.cn (B. Wang).

https://doi.org/10.1016/j.heliyon.2024.e30889
Received 6 September 2023; Received in revised form 4 May 2024; Accepted 7 May 2024
Available online 8 May 2024

1. Introduction

According to data released in 2021 by the World Health Organization's International Agency for Research on Cancer, breast cancer has become the number one cancer in the world [1]. A study has shown that if breast cancer can be detected early and properly treated, the 5-year survival rate of patients can reach more than 90 % [2]. Therefore, the early diagnosis of breast cancer is particularly important, which can not only significantly improve the survival time and quality of life of patients, but also reduce treatment costs and mortality [3]. In clinical diagnosis, diagnosis based on pathological images is the most reliable method. However, the diagnosis process is time-consuming and labor-intensive. Computer aided detection tools can help pathologists improve the diagnosis speed and reduce the misdiagnosis rate [4–8]. At present, convolutional neural network (CNN) is widely used in medical image researches [8–12]. Previous image analysis methods based on deep learning have shown their potential utility in breast cancer diagnosis [13,14].

CNN-based approaches have been widely used for classification tasks of breast pathological images [15–17]. AlexNet was the first convolutional neural network used to classify breast pathological images on BreakHis dataset, and its recognition accuracy was significantly higher than that of traditional machine learning algorithms [18]. Arau'jo et al. proposed a CNN-based approach and achieved accuracy of 83.3 % for binary classification tasks on the breast cancer Classification Challenge 2015 dataset [19]. Some studies used deeper neural networks on the two-class problem of breast cancer histopathological images, and achieved accuracy of more than 90 % on the BreakHis dataset [13,20,21]. Furthermore, the authors built the model from the perspective of increasing network width and achieved good performance [22,23]. Inception-ResNet considered scaling factors between 0.1 and 0.3 to scale the entire parallel convolution module (Inception), which effectively improved the model performance [24]. Kassami et al. applied different data augmentation methods to optimize the Xception model, and the average classification accuracy in the ICIAR2018 dataset reached 92.50 % [25], while Golatkar et al. fine-tuned the Inception V3 network and trained on the extracted patches. The accuracy of binary classification test reached 93 % [26]. Munien et al. achieved the classification of four subtypes of the ICIAR2018 dataset by fine-tuning EfficientNets. The results showed that the model trained through the transfer learning method had a better recognition ability of images [27].

Some studies have focused on the different capabilities of extracting features with convolution branches, and proposed feature fusion to obtain multi-scale input information and improve the network representation ability. The Inception-series network [24,28] serves as a typical representative of networks designed to capture multi-scale input information through increased network width. It is composed of multiple parallel convolution branch modules stacked.

In previous research, we combined receptive field block (RFB) with a lightweight CNN model to extract deep semantic features from images [29]. Within the receptive field block, we used the multi-branch convolution structure, and assigned convolution kernels of different sizes to correspond with different receptive fields, enabling model to extract the deep and rich details of the image. Compared with the previous related research, the proposed model exhibited higher detection accuracy and lower complexity. The RFB module contributed to the feature extraction of the lightweight CNN model while also minimizing the number of model parameters.

However, the existing models did not consider the different contribution of branches in parallel convolution module. While the information extracted by different branches complements each other, it may also lead to redundancy, adding burden to the network training and reducing the performance of the model. Therefore, while considering adding multi-scale information, it is more important to assign different activation factors, to avoid feature redundancy and improve feature quality.

Several recent studies resized the images to small sizes for training, but this approach inevitably damaged the image quality [22,30,31]. There are also studies cut images into patches and use voting algorithms to calculate the prediction results of the images [20,21,32,33], however, since the cut patches may not contain cancer cells, this can lead to bias in the voting result.

At present, attention mechanisms have been widely used to improve feature extraction in deep learning [34–37]. Moreover, the attention mechanism has two advantages: fewer parameters and stronger interpretability [38,39]. Mesut et al. built a deep convolution neural network BreakNet based on convolutional block attention module (CBAM), convolution blocks, dense blocks and residual blocks, and achieved accuracy of 98.51 % (for 200 × magnification) [30]. The works also showed that CBAM can effectively improve the performance of the model [40–42].

In our study, we proposed the parallel convolutional scale attention module (PCSAM) that incorporated four parallel convolution branches and residual connections to improve the network's ability to capture multi-scale information. Additionally, we introduced the dilated convolution [43] to increase the receptive field. Concurrently, we set a scale attention factor to control the proportion of each branch output in the module, differentiating the impact of different features and optimizing the combination of multi-scale information. By integrating the parallel convolutional scale attention module network and the residual convolutional block attention module network, we proposed the PCSAM-ResCBAM model. We integrated the method of tiled feature fusion in the model to better retain the edge information between different patches. The experimental results demonstrated that our model had better classification performance than other methods, extracted high-quality feature representation and effectively predicted the category of breast cancer pathological images. This contributed to providing supportive suggestions for the diagnosis of pathological images. Our network is a lightweight network with lower parameters and computation, which is optimized for mobile device deployment. The contributions of our research are as follows.

(1) We proposed a lightweight PCSAM-ResCBAM model, including the Parallel Convolution Scale Attention Module network (PCSAM-Net) and Residual Convolutional Block Attention Module network (ResCBAM-Net). The PCSAM-Net was built by our proposed Parallel Convolution Scale Attention Module (PCSAM), which acted as a patch-level network to extract the features of patches extracted from the image and implemented patch-level classification. The ResCBAM-Net built by our proposed Residual

Convolutional Block Attention Module (ResCBAM) was responsible for extracting global features of the image to implement the image-level classification on the basis of fusion of patches.

(2) We proposed a tiling feature fusion method to optimize the fusion of patch-level feature maps.

(3) We evaluated the performance of our model using two public benchmark datasets respectively, and the proposed method outperformed existing state-of-the-art models on high magnification datasets (200 × , 400 ×).

(4) We performed ablation experiments to validate how the PCSAM affected the performance of the classification model.

2. Materials and methods

2.1. Datasets and preprocessing

Fig. 1 shows the overall workflow of the study. We collected histopathological images of breast cancer from the public datasets, the BreakHis dataset [44] and ICIAR2018 dataset [45], for the model training and validation respectively. The BreakHis dataset contains four different magnification factors (40 × , 100 × , 200 × , 400 ×), including 7909 H&E staining images with a resolution of 700 × 460 pixels from 82 breast tumor patients. Due to the problem of image size, we screened 74 images from patient SOB_M_PC_14–12465. These 74 images were actually screened out. When processing images, deep learning models usually need to input images with uniform size and resolution, which can extract features and perform training more effectively. When the input image size is inconsistent, it may increase the difficulty of the model to extract features through convolution operation and reduce the performance of the model. Furthermore, during the training process, the model needs to continuously adjust its internal parameters to adapt to different input sizes, which may increase the difficulty of the model and reduce the training efficiency. The structural details of the BreakHis dataset are shown in Table 1. ICIAR2018 dataset consists of 400 H&E staining images (2048 × 1536 pixels) with 200 × magnification factor, including 200 benign samples and 200 malignant samples. We listed all of the abbreviations in Supplementary Table 1.

For each dataset (BreakHis 40 × , 100 × , 200 × , 400 × , and ICIAR2018 dataset), 80 % of the dataset was randomly selected as the training set (train and validation) and 20 % of the dataset remained for testing. Since images in the BreakHis dataset are original images that have not been normalized, we used the Macenko method [46] to color normalize each image. As shown in Table 2, we found that the number of malignant samples is approximately twice that of benign samples. In order to balance the number of samples, we expanded the number of benign samples by means of data augmentation, such as vertical mirroring and horizontal vertical mirroring. For the ICIAR2018 dataset, we did not pre-process the images but processed the patches obtained from the images. We rotated each patch by 4 multiples of 90°, random color perturbations, with and without mirroring, resulting in 16 valid variations for each patch. Moreover, since the ICIAR2018 dataset images contain many other parts with sparsely located nuclei such as cytoplasm, which

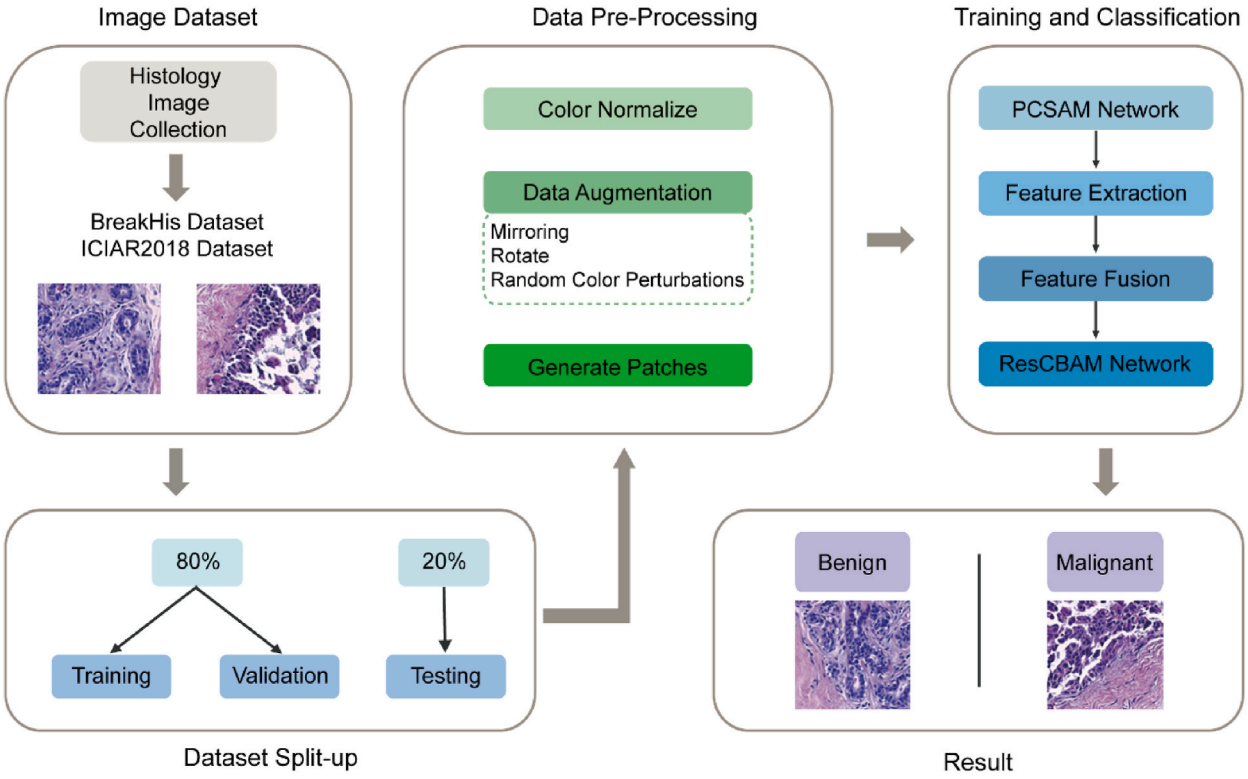


Fig. 1. The overall workflow of the study.

Table 1
Structure of the BreakHis dataset with four magnification factors (40 × , 100 × , 200 × , and 400 ×).

Class	Magnification Factors				Total
	40 ×	100 ×	200 ×	400 ×	
Benign	625	644	623	588	2480
Malignant	1349	1416	1371	1219	5355
Total	1974	2060	1994	1807	7835

Table 2
Classification accuracy of the PCSAM-Net on the BreakHis and ICIAR2018 test sets through three different voting strategies.

Dataset	Maj	Max	Sum	Average
40 ×	92.84	93.61	92.83	93.09
100 ×	95.39	95.87	95.39	95.55
200 ×	95.72	94.21	95.21	95.05
400 ×	96.84	96.26	96.55	96.55
ICIAR2018	94.65	94.25	95.0	94.63

are not relevant for classification, we did not use all the patches extracted from the images when training the PCSAM-Net. Instead, we selected only those patches with high nuclei density using methods proposed in the literature [26].

In addition, we used the sliding window method to extract fixed-size patches during training. According to previous studies [30, 47], in order to make the patch sufficient to cover the relevant tissue structures of different cells, we set the k value (patch size) to 224 for the BreakHis dataset and 512 for the ICIAR2018 dataset. When training the PCSAM-Net, we chose stride (s) of the sliding window to be half the value of k, which caused patches to overlap. Allowing overlap is critical for the patch-wise network to learn features shared between patches. When training ResCBAM-Net, we set s and k to the same value. Equations (1)–(1) is the formula for calculating the number of patches (N=P_{num}) obtained from each image.

$$P_{num} = \lfloor 1 + (I_w - k) \ / \ s \rfloor \times \lfloor 1 + (I_H - k) \ / \ s \rfloor \tag{1-1}$$

Among them, I_H and I_W denote the height and width of the image, respectively.

2.2. Proposed model

The proposed PCSAM-ResCBAM model based on the two-stage convolutional network is shown schematically in Fig. 2. The first-level network PCSAM-Net built by PCSAM is used to extract features from patches, learning not only the overall structure of cells, but also their textures. The second-level network ResCBAM-Net built by ResCBAM extracts the global features based on the feature map after patch feature fusion. Once both networks are trained, we use them jointly to infer image-level class predictions.

2.2.1. PCSAM

The detailed structure of the proposed parallel convolution scale attention module (PCSAM) is shown in Fig. 2(c). In PCSAM, multiple parallel convolution branches are set to increase the width of the network, and different numbers of 3 × 3 convolution layers are superimposed in different branches to realize feature extraction of different dimensions. Dilation convolutions with different dilated rates are set in the module to maximize the effective receptive field of the network and extract different scales image information. The dilated rate in the module is set to 1, 3, or 4 respectively. In addition, the input and output of the convolution block were summed at pixel level through residual connection (short-cut) [48] to preserve low-level image features, and to make the error of the deep model no greater than the error of the shallow model.

In addition, in order to optimize multi-scale information combinations, we proposed scale attention in the module to optimize the weight distribution of the model. Different from the spatial or channel attention mechanism, the scale attention mechanism was implemented by introducing activation scales (AS). This mechanism can not only control the influence degree of each branch on the output feature map to avoid feature redundancy, but also change the influence distribution of pixels in the receptive field region, and improve the feature quality of network extraction. Research has found that the influence distribution in the receptive field is Gaussian and usually decays rapidly from the center [49], which ignores the distribution features of other non-central locations in the receptive field, and is not conducive to improving the feature characterization ability of the network. The proposed scale attention increased the influence of the non-central part of the image on the output. In our setting, the output of each module in the network is no longer the uniform distribution of each convolution branch, but it gathers highly correlated features and weakens irrelevant non-key features. In the training process, we set activation scales as super parameters and selected the optimal values through the grid search algorithm.

2.2.2. First-level network: PCSAM-Net

The detailed structure of the proposed first-level network PCSAM-Net is shown in Fig. 2(b), which is trained based on patches, and extracted patch features to realize the prediction and classification of patches. Once trained, we discard the classifier layer of this

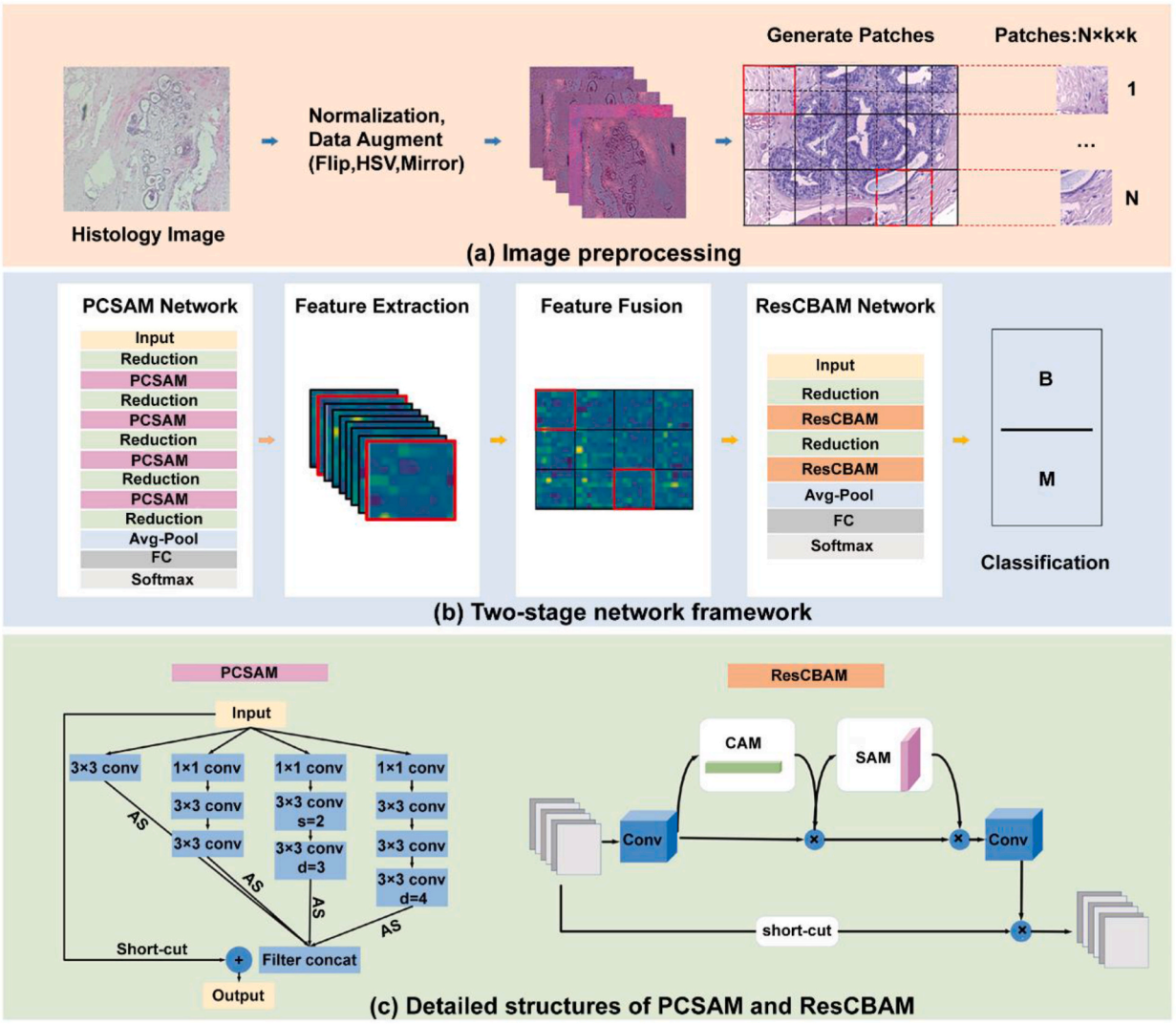


Fig. 2. An overview of the proposed workflow. (a) Image processing. Left panel, the original pathological image of a breast cancer patient. Middle panel, the color normalization and data augmentation of the image. Right panel, the acquisition of the patches. Patches are regularly captured by sliding a window of size $k \times k$ on the image, with a window sliding step of size s, and finally N patches are obtained. (b) Two-stage network framework. The first panel on the left, the framework of the first-level network PCSAM-Net. The second panel, feature extraction. The third panel, the fusion of different patches using tiling fusion method. The fourth panel, the framework of the second level network ResCBAM-Net. The last panel is the result of classification of breast lesions: benign or malignant. (c) Detailed structures of PCSAM and ResCBAM. Left panel, the structure of PCSAM module, in which multiple parallel convolution branches are set, and different numbers of 3×3 convolutional layers and AS values are superimposed to achieve feature extraction of different dimensions. Right panel, the structure of ResCBAM module, which consists of residual connection and CBAM module, explores the importance difference of different features from two dimensions of space and channel.

network and use the last convolutional layer to extract patch feature maps.

Deep networks are necessary for building strong representations, but even very deep networks cause higher computational costs. Given these two factors, we designed and stacked four layers of PCSAM to build PCSAM-Net. With regard to the setting of the convolution layer, we followed the principles that most networks follow: If the size of the output feature map is halved, the number of channels of the output feature map should be doubled to ensure that the amount of information contained in adjacent convolutional layers does not differ too much [50]. The reduction module is composed of a 3×3 convolution layer, with stride and padding set 2. In addition, we set the ReLu layer [51] and batch normalization (BN) layer [52] after each convolutional layer. ReLU is used to complete nonlinear mapping in the network, and batch normalization is used to narrow the distribution gap between each training batch. The classifier layer of the network is composed of global average pooling (GAP), full connection layer (FC), and Softmax function.

2.2.3. Second-level network: ResCBAM-Net

The detailed structure of the proposed second-level network ResCBAM-Net is shown in Fig. 2(b), which is composed of two Residual

Convolutional Block Attention Module (ResCBAM) layer and Reduction layer. Based on the feature map after patches feature fusion, the network extracts image global features and realizes image level classification. The reduction layer and classifier layer structure in this network are consistent with those mentioned above, so we will not repeat them here. The ResCBAM module is described in detail below.

As shown in Fig. 2(c), the proposed ResCBAM contained three sequential sub-modules: channel attention module (CAM), spatial

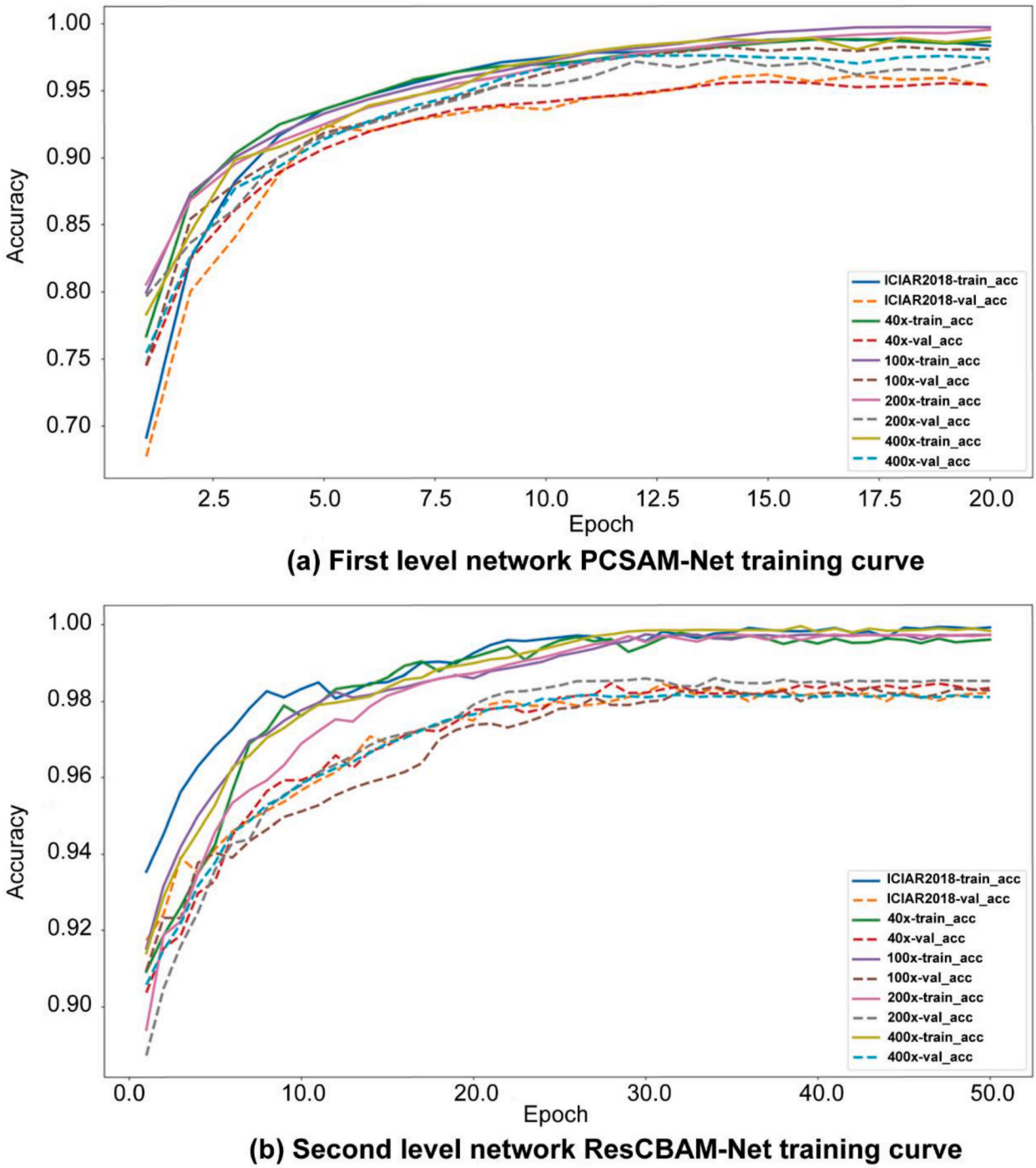


Fig. 3. Accuracy of training sets and verification machines during iterative learning of networks. (a) the accuracy of the training and validation sets of the first level network PCSAM-Net during iterative learning. (b) the accuracy of the training and validation sets of the second level network ResCBAM-Net during iterative learning.

attention module (SAM) and short-cut connection. The channel attention module kept the channel dimension of the input feature map unchanged, compressed the spatial dimension, and focused on meaningful information in the input, while the spatial attention module kept the spatial dimension unchanged, compressed the channel dimension, and focused on location information.

Assuming that the input feature map is: $F \in R^{C\times H\times W}$, and the output feature map is $Out_F \in R^{C\times H\times W}$. ResCBAM is used to derive the one-dimensional channel attention map: $M_C \in R^{C\times 1\times 1}$ and the two-dimensional spatial attention map: $M_S \in R^{1\times H\times W}$ and get original feature map F by short-cut connection, we calculate the output data of the ResCBAM module based on Equations (2)–(1), Equations (2)–(2) and Equations (2)–(3), respectively.

$$F^{'} = M_C(F) \times F \tag{2-1}$$

$$F^{''} = M_S(F^{'}) \times F^{'} \tag{2-2}$$

$$Out_F = F + F^{''} \tag{2–3}$$

2.2.4. Tiling feature fusion

In order to fuse the N patch feature maps from the same image to obtain the global feature map corresponding to the original image, we proposed a tiling feature fusion method. The tiling feature fusion method fused the features of the patches and integrated the edge information between different patches. As shown in Fig. 2(b), the patch feature maps were reassembled according to the order of patch intercepted by the sliding window, and the feature corresponded to the position of the original input image to the maximum extent.

Compared with the method using concatenation, our proposed method was more conducive to the extraction of association information between different patches and integrating the global features of the image to achieve image-level classification. The experimental results in Section 3.3 further demonstrated the superiority of our proposed method.

2.3. Implementation

We used the BreakHis and ICIAR2018 datasets to train and validate the model, with 80 % of each dataset being the training set and validation set, and 20 % being the test set. Accuracy, precision, recall, F1-Score, Matthews correlation coefficient (MCC), and area under the receiver operating characteristic curve (AUC) were used to evaluate the performance of the model on each dataset. Higher values of each index indicated better prediction performance of the model.

All experiments used the initialization parameters: the batch size was 16, the momentum was 0.9, the L2 weight regularization parameter was 4×10^{-4} and the optimizer chose Adam with the parameters β1 and β2 (0.9, 0.999). The initial learning rate and iteration epoch of the two network training sessions were (0.002, 20) and (0.001, 50) respectively. During the training process, we adopted the LambdaLR learning strategy, which made the learning rate iteratively decrease, and finally, the learning rate decreased to 0.1 times of the original. The activation factor value AS in the module was selected and set by the grid search algorithm, and the best model was set as follows: 1.0, 1.0, 0.1, 0.2; 0.7, 0.9, 0.1, 0.9; 0.4, 0.6, 0.4, 0.8; 0.6, 0.6, 0.3, 0.3; these values correspond to the branches of the four modules from top to bottom and left to right. The code was implemented based on PyTorch, and the experiment was conducted on a workstation with NVIDIA 2080-Ti GPU.

In addition, to reduce the impact of test data diversity, we adopt the TTA strategy in the testing phase. Data augmentation was performed on the test set, and this process allowed the model to more fully consider the potential of an image during testing.

3. Results

3.1. Classification results

In order to comprehensively evaluate the model, we used accuracy, precision, recall, F1_score, receiver operating characteristic (ROC) curve, and Matthews Correlation Coefficient (MCC).

In order to evaluate the performance of the first-level network PCSAM-Net, we adopted three different voting strategies to fuse the predicted results of the patch to obtain image-level prediction results, including probability sum (Sum), majority voting (Maj), and maximum probability (Max). The specific introduction of voting strategies is as follows: (1) Sum: Sum the predicted probability values for each category of N patches, and the category corresponding to the maximum value is the predicted category of the image; (2) Maj: Counts the predicted results of N patches, and the category with the highest number of votes is the predicted category of the image; (3)

Table 3
Classification results of PCSAM-ResCBAM model on BreakHis and ICIAR2018 datasets.

dataset	accuracy	precision	recall	F1_score	MCC
40 ×	95.14	96.0	96.0	96.0	89.38
100 ×	97.09	97.0	97.0	97.0	93.43
200 ×	98.74	99.0	99.0	99.0	97.14
400 ×	97.99	97.5	97.5	97.5	95.45
ICIAR2018	97.50	97.50	97.62	97.50	95.18

Max: The category corresponding to the highest predicted probability value among N patches is the predicted category of the image. As shown in Table 2, the average voting classification accuracy of the first-level network PCSAM-Net on the BreakHis and ICIAR2018 test sets was 93.09 %, 95.55 %, 95.05 %, 96.55 %, and 94.63 %, respectively. In addition, Fig. 3 (a) shows the accuracy changes of the training and validation sets during the first-level network training process on different datasets. Through multiple iterations of backpropagation, the accuracy of the training and validation sets continues to increase until the network converges. It is worth noting that the accuracy shown in Fig. 3 (a) is the prediction accuracy of the PCSAM-Net network for patches, rather than the image level accuracy obtained by voting.

Table 3 shows the image-level classification results of the ensemble model PCSAM-ResCBAM on the BreakHis and ICIAR2018 test sets. The PCSAM-ResCBAM model achieved good accuracy and precision, but still achieved high recall and MCC values, indicating that the model can accurately identify sample categories. As shown in Fig. 4, the AUC values obtained by the model tested on the BreakHis 40 × , 100 × , 200 × , 400 × and ICIAR2018 datasets were 0.9690, 0.9725, 0.9963, 0.9783, and 0.9750, respectively. These results show that the proposed ensemble model can effectively excavate the deep features of benign and malignant data in breast cancer histopathologic images, and can predict the types of images more accurately. Fig. 5 shows the confusion matrix on each test set. It can be seen that the classification precision of the model for benign and malignant is similar on the four magnification test sets of BreskHis, both exceeding 94 % [Fig. 5 (a - d)]. On the ICIAR2018 dataset, the ensemble model achieved classification precision of 95 % for benign and malignant types, and was more sensitive to malignant categories [Fig. 5 (e)]. Fig. 3 (b) shows the accuracy changes of the training and validation sets of the second-level network ResCBAM-Net training process. After 35 iterations, the classification accuracy growth on different datasets becomes stable, the classification accuracy of the validation and training sets is close, and the network tends to converge. In addition, the accuracy shown in Fig. 3 (b) is the prediction accuracy of the ensemble model PCSAM-ResCBAM for images, including multiple deformations of the image, rather than the final prediction results. There is a phenomenon of incorrect prediction of the original image but accurate classification of its rotation or flipping deformation, which is consistent with clinical diagnosis experience and requires comprehensive observation and analysis of the image from multiple perspectives for diagnosis.

As shown in Fig. 6, we compared the image level classification accuracy obtained from PCSAM-Net network voting (Sum, Maj, Max) on different datasets with the accuracy predicted by the ensemble model PCSAM-ResCBAM (Ensemble) in the form of a bar chart. Classification accuracy of the ensemble model has been improved compared to the voting results, with an accuracy improvement of about 2 % on the ICIAR2018 [Fig. 6 (e)] and BreakHis 40 × , 100 × , 200 × datasets [Fig. 6 (a - c)]. The improvement is not significant on the BreakHis 400 × dataset [Fig. 6 (d)], but there is still an improvement. These results indicate the necessity and effectiveness of the feature fusion method and the second-level network ResCBAM-Net proposed. The lower improvement effect on the BreakHis 400

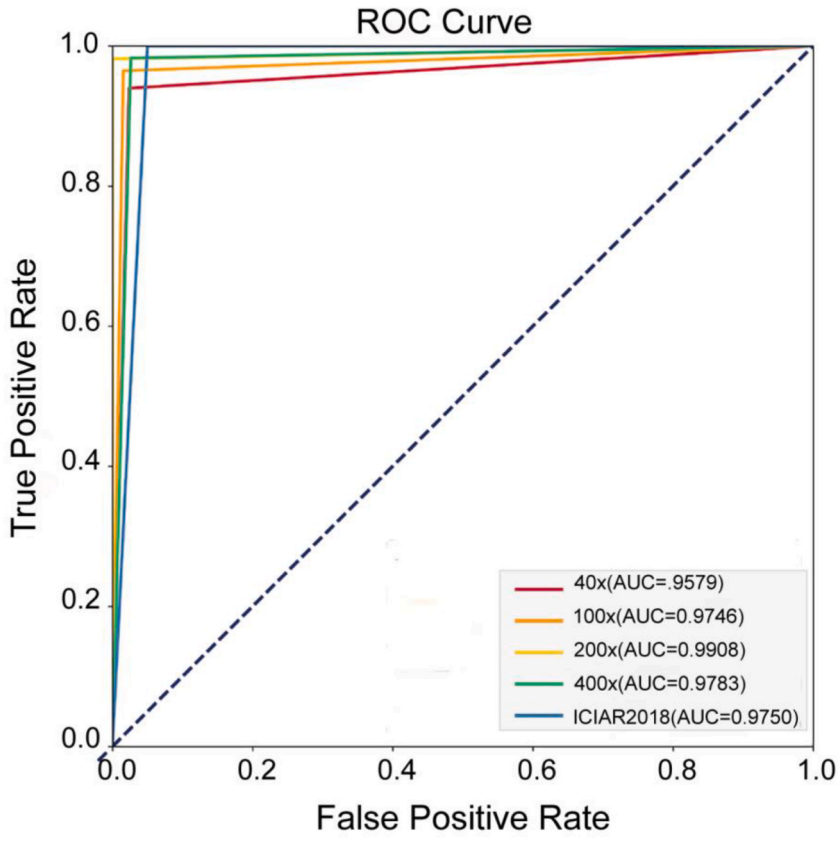


Fig. 4. ROC curves of PCSAM-ResCBAM model testing using BreakHis (40 × , 100 × , 200 × , 400 ×) and ICIAR2018.

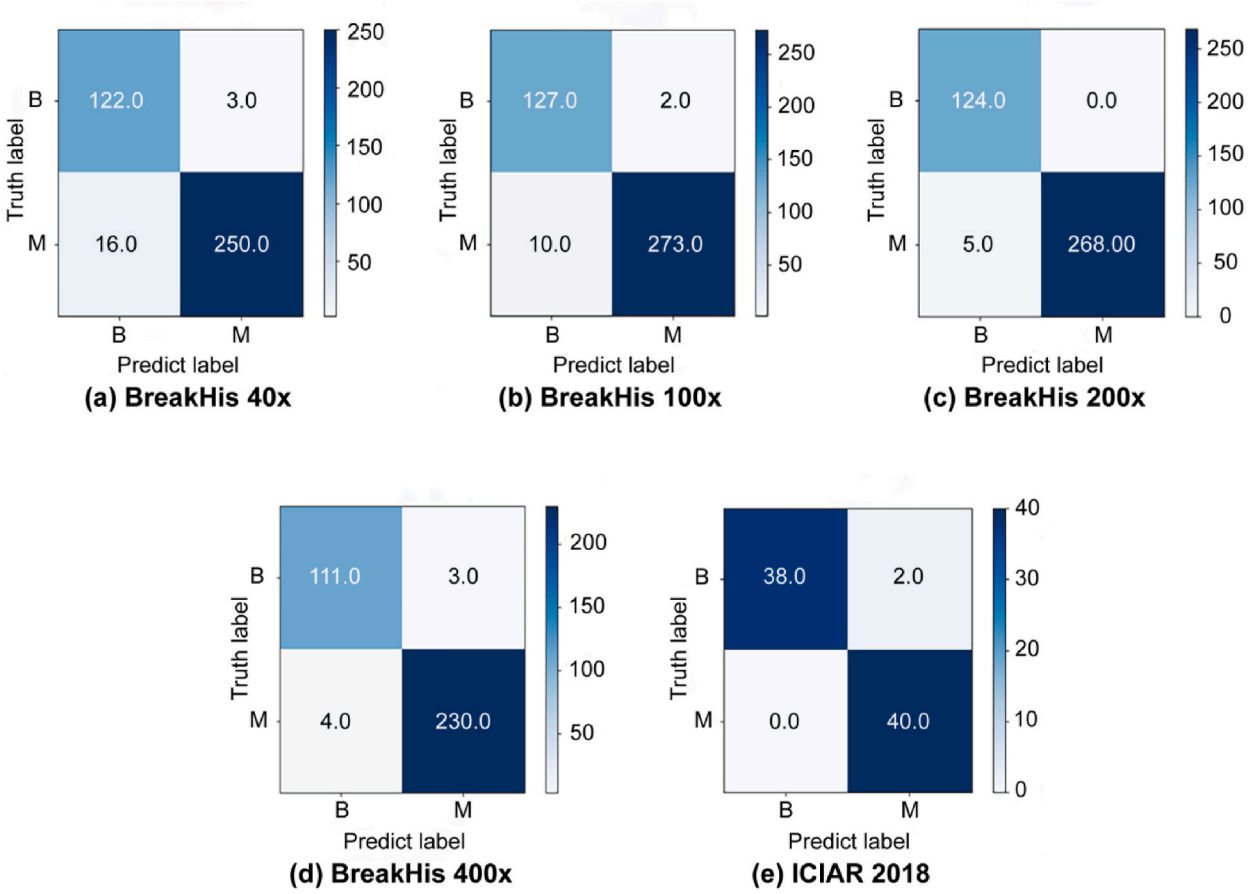


Fig. 5. Classification confusion matrix of PCSAM-ResCBAM model on different datasets. (a) on the BreakHis (40 ×) dataset. (b) on the BreakHis (100 ×) dataset. (c) on the BreakHis (200 ×) dataset. (d) on the BreakHis (400 ×) dataset. (e) on the ICIAR2018 dataset.

× dataset is due to the fact that images with a magnification of 400 × provide higher levels of detail while also containing more complex information and noise that can easily interfere with classification. In addition, clinical diagnosis based on breast pathological images should not only focus on the shape, size, and distribution of the nucleus, but also on the surrounding tissue structure and other information. However, for 400 × magnification images, the information in the image is more local, with less information about surrounding tissue morphology and distribution information, which to some extent limits the performance of the model.

In addition, in order to further understand the feature areas that the model focused on when working, we used the Grad-CAM method [53] to generate class activation maps, including at patch level and image level, on the ICIAR2018 dataset and BreakHis (40 × , 100 × , 200 × , 400 ×) datasets, as shown in Fig. 7 (a - c) and Supplementary Fig. 1 (a - d) in the supplementary materials, respectively. The patch level class activation maps on the ICIAR2018 and BreakHis datasets are composed of 12 and 6 patch class activation maps, respectively. As shown in Fig. 7 and Supplementary Fig. 1, the proposed patch-level network PCSAM-Net focused on the feature of nuclei distribution and shape, and correctly predicted patch class by using these characteristics [Patch-level Visualization, Fig. 7 (b)], while the ensemble model proposed (PCSAM-ResCBAM) focused on a broader region of the image, including information on the nucleus as well as tissue texture features, using which to correctly predict image [Image-level Visualization, Fig. 7 (c)].

3.2. Higher model evaluation metrics compared to state-of-the-art methods

In order to evaluate the performance of the proposed model, we compared the classification accuracy of the PCSAM-ResCBAM model on the BreakHis (40x, 100x, 200x, 400x) dataset and ICIAR2018 dataset with the results of other existing models.

As shown in Table 4, the classification accuracy of the PCSAM-ResCBAM model on the BreakHis 40 × and 100 × datasets is slightly lower than the optimal results of existing models, but it achieved the best classification results on the 200 × and 400 × datasets, with accuracy rates of 98.74 % and 97.99 %, respectively, superior to other existing models. In addition, the classification accuracy of the first level network PCSAM-Net obtained through voting is also better than most of the models in the table, especially the average classification accuracy obtained through voting on the 400 × magnification test set is 96.55 %, which is better than other existing

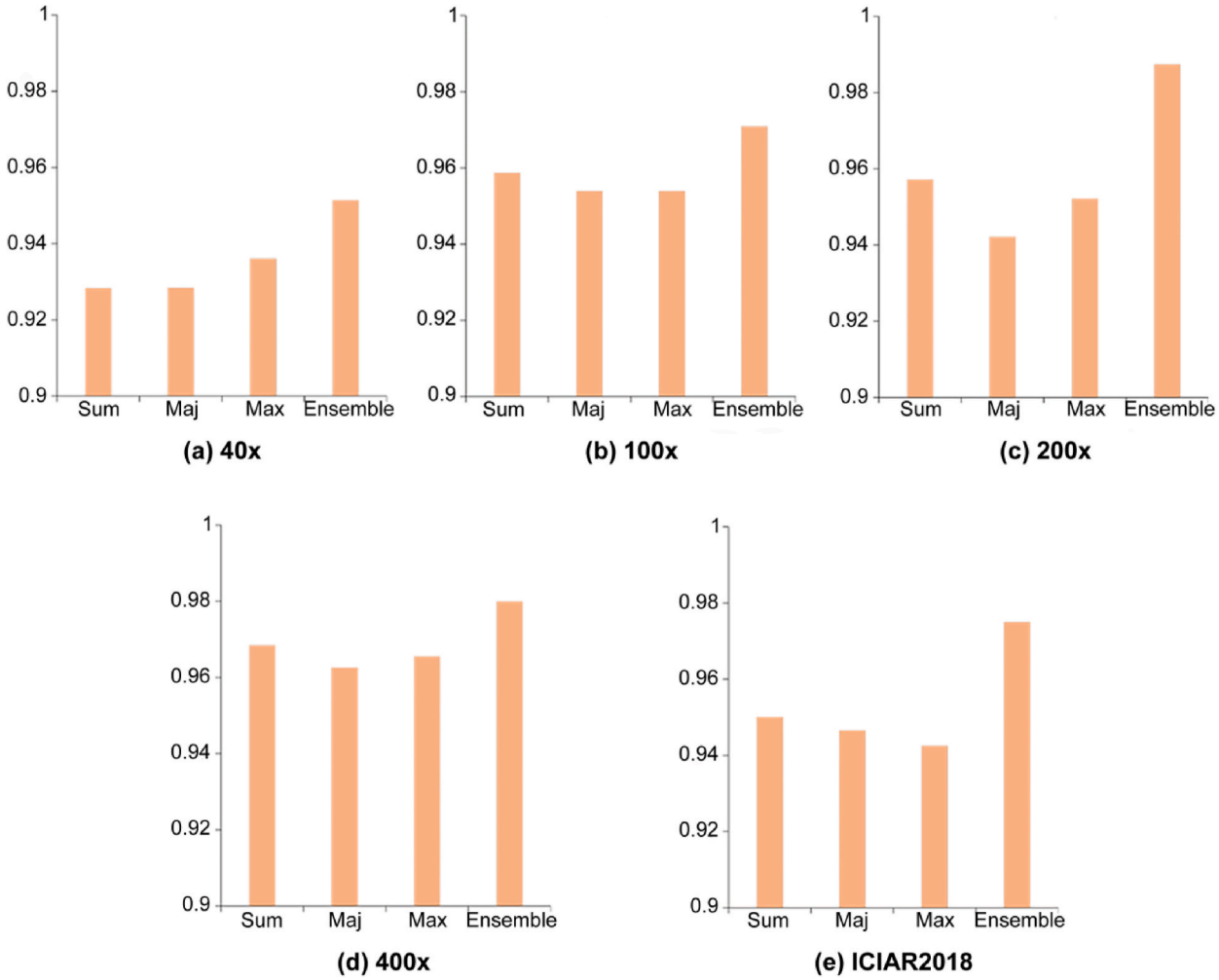


Fig. 6. Comparison of classification accuracy of PCSAM-Net (Sum, Maj, Max) and PCSAM-ResCBAM network (Ensemble) on BreakHis and ICIAR2018 datasets. (a) on the BreakHis (40 ×) dataset. (b) on the BreakHis (100 ×) dataset. (c) on the BreakHis (200 ×) dataset. (d) on the BreakHis (400 ×) dataset. (e) on the ICIAR2018 dataset.

models. The comparison of these results indicates that the proposed model PCSAM-ResCBAM has certain advantages in breast pathology image classification tasks, and is more suitable for predicting and classifying breast pathology images with high magnification (200 × and 400 ×).

Similarly, as shown in Table 5, the classification accuracy of the PCSAM-ResCBAM model was compared with other existing models on the ICIAR2018 (200 ×) dataset. The classification accuracy of the first-level network PCSAM-Net through voting is still better than the results of most models in the table, and the test classification accuracy of the ensemble model PCSAM-ResCBAM reaches 97.5 %, which is better than other existing models. It is worth noting that both the image level classification accuracy obtained by the first level network PCSAM-Net voting and the accuracy obtained from the ensemble model PCSAM-ResCBAM test are better than those obtained from models that also use parallel convolution modules in the network structure (Inception V3, Inception V4, Inception ResNet V2), with a classification accuracy improvement of about 1 %–4 %. These results further demonstrate the effectiveness of the proposed PCSAM module, as well as the necessity and effectiveness of the two-stage network framework. The combination of parallel convolutional branches and scale attention mechanism effectively improves the feature extraction ability of the network. In addition, as shown in Fig. 8, Compared with other existing models, the proposed model has fewer the number of parameters and Flops, and can better adapt to mobile and medical devices, making it easy to deploy and apply.

3.3. Optimal performance of tiling fusion comparing to concatenation methods

To illustrate the advantages of the tiling feature fusion method proposed in this paper, we compared the model classification performance using the tiling fusion method and concatenation method (concatenating multiple feature maps in the channel dimension) on the ICIAR2018 dataset, and compared the impact of different feature map sizes after fusion on network classification accuracy.

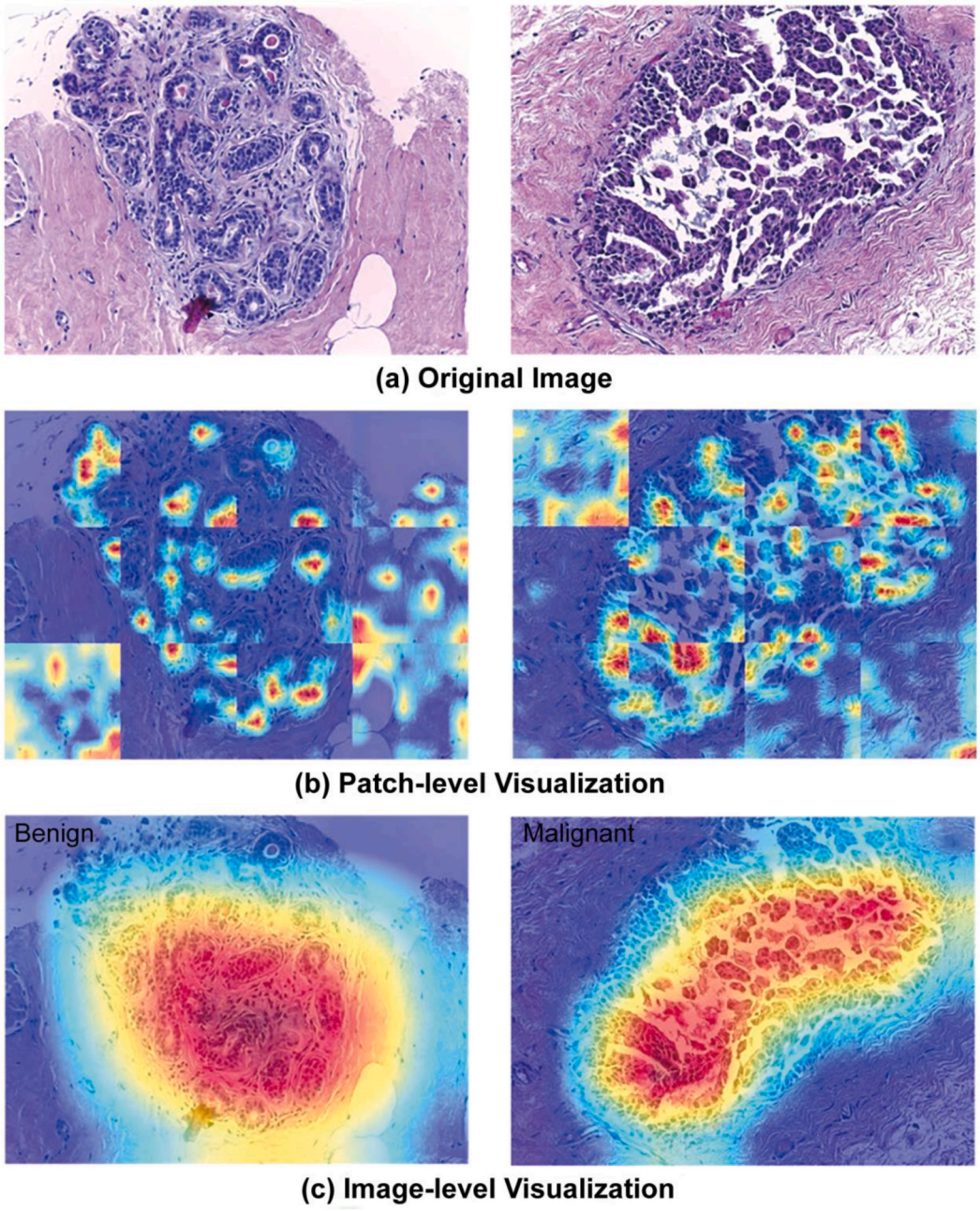


Fig. 7. Representative examples of heat map images generated by using the Grad-CAM method. (a), left figure, an original pathological image of a benign breast lesion. Right figure, an original pathological image of a malignant breast lesion. (b), the left figure, the class activation map of benign breast lesions at the patch level. Right figure, the class activation map of malignant breast lesions at the patch level. (c), the left figure is the class activation map of benign breast lesions at the image level. Right figure, the class activation map of malignant breast lesions at the image level.

As shown in Table 6, the tiling feature fusion method has obvious advantages: When using the tiling feature fusion method, the highest classification accuracy, precision, recall, and F1_score obtained from the ensemble model PCSAM-ResCBAM testing were 97.50 %, 97.50 %, 97.62 %, and 97.50 %, respectively. When using the concatenation method, the highest classification accuracy, precision, recall, and F1_score obtained from the ensemble model PCSAM-ResCBAM testing were 92.50 %, 92.50 %, 92.67 %, and 92.50 %, respectively. In terms of accuracy, our proposed method improved by around 5 %. These results indicate the effectiveness of the proposed tiling feature fusion method, and the edge information fusion of patch feature maps is necessary to effectively improve the classification performance of the model.

Compared with different feature map sizes after tiling fusion, the network performance with $512 \times 24 \times 32$ feature map size is

Table 4
Comparison of classification accuracy between the proposed model and other existing models on BreakHis datasets.

model	Magnification Factors(accuracy(%))			
	40 ×	100 ×	200 ×	400 ×
AlexNet [18]	85.6	83.5	82.7	80.7
Inception&Boosting&Fusion [54]	95.1	96.3	96.9	93.8
Sequential [55]	94.71	95.9	96.76	89.11
CSDCNN [56]	95.8	96.9	96.7	94.9
FE-BkCapsNet [32]	92.71	94.52	94.03	93.54
ResNet50-CBAM [57]	91.2	91.7	92.6	88.9
3PCNNB-Net [22]	92.27	93.07	97.04	92.09
IRRCNN + Aug [33]	97.95	97.57	97.32	97.36
BreastNet [30]	**97.99**	**97.84**	98.51	95.88
DRDA-Net [58]	95.72	94.41	97.43	96.84
PCSAM-Net Voting Average	93.09	95.55	95.05	96.55
PCSAM-ResCBAM	**95.14**	**97.09**	**98.74**	**97.99**

Table 5
Comparison of classification accuracy between the PCSAM-ResCBAM model and other existing models on the ICIAR2018 dataset.

model	ICIAR2018(200 ×)
CNN + SVM [19]	83.30
CNN + APOD [59]	92.50
Based on Inception V3 [60]	93.00
Inception V4 [23]	93.70
VGG16 [61]	93.8
Inception-ResNet V2	93.75
PCSAM-Net Voting Average	94.63
PCSAM-ResCBAM	**97.50**

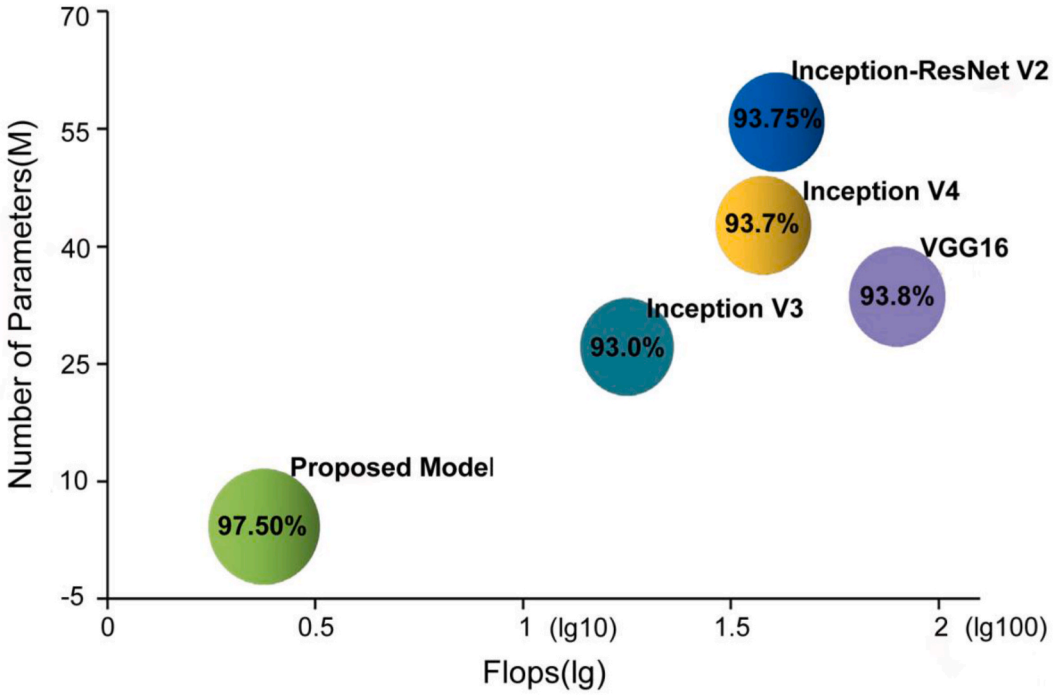


Fig. 8. Comparison of the proposed model with other models in terms of accuracy, number of parameters and Flops.

better than that of network with other sizes, with 97.50 % accuracy, while the accuracy of the feature map size of 1024 × 24 × 32 (96.25 %) is slightly lower than the optimal accuracy. Therefore, without being limited by hardware and ensuring that the convolutional layer of the second-level network can effectively extract features, we believe that the number of feature map channels before

Table 6
Comparison of the impact of different feature fusion methods and different feature map sizes on network performance.

feature fusion method	C × H × W	accuracy	precision	recall	F1-score
Tiling	192 × 64 × 64	91.88	91.88	92.15	91.87
	192 × 32 × 32	90.42	90.42	90.85	90.43
	768 × 64 × 64	82.08	82.08	82.31	82.31
	768 × 32 × 32	92.50	92.50	92.67	92.50
Concatenation	1024 × 24 × 32	96.25	96.25	96.42	96.25
	1024 × 48 × 64	93.75	93.75	93.96	93.74
	512 × 24 × 32	97.5	97.5	97.62	97.5
	512 × 48 × 64	91.87	91.87	92.04	91.87

tiling fusion should be larger, and the height and width should be smaller, so that the feature map information after tiling is stronger and more concentrated, which is conducive to the training and learning of the image-level network.

3.4. Ablation study in scale attention and dilation convolution

Special emphasis is placed on why we use scale attention and dilation convolution in PCSAM, and how their addition improves model performance. Table 7 shows the results of these studies, with or without the use of scale attention and dilation convolution. Experiments were performed on the ICIAR2018 dataset.

Scale Attention (SA): The introduction of the scale attention mechanism changed the influence distribution in the receptive field. In multiple parallel convolution branches, these branches extracted different scales of feature information. Subsequently, the search algorithm selected the optimal branch activation scale. Finally, the feature information was selectively fused based on different activation values. The "Without SA" model corresponded to the activation value of 1, where each convolution branch contributed equally to the output and different scale information was uniformly fused. Under two different settings, the image classification accuracy obtained by voting did not differ significantly. However, the performance evaluation index values of the "ALL" ensemble model were 2 % higher than those of the "without SA" ensemble model: the accuracy, precision, recall, and F1_score of the "ALL" ensemble model were 97.50 %, 97.50 %, 97.62 % and 97.50 %, while those of "without SA" ensemble model were 95.46 %, 95.0 %, 95.1 % and 95.46 % (Table 7). Our proposed method achieved the highest metric results, indicating that scale attention is beneficial for enhancing the network representation ability. The introduction of the scale attention mechanism changed the distribution of influence in the receptive field, and optimized the multi-scale information combination to optimize the model. By selecting the optimal branch activation scales by the search algorithm to apply the scale attention mechanism, the performance of the model can be greatly improved.

Dilation Convolution: The DeepLab series models proposed the dilation convolution. Compared with the traditional convolution operation, the dilation convolution does not require additional parameters, yet it can effectively expand the network receptive field. Furthermore, when multiple dilation convolutions with different inflation rates are used, it becomes possible to capture more context information, thereby improving the understanding ability of the model for the input data. The model that discards the dilation convolution in PCSAM is called the "Non-Dilation" model. As shown in Table 7, the image classification accuracy of the PCSAM-Net with dilation convolution was more than 1 % higher than that of the model discarding dilated convolution. Moreover, the performance of the ensemble model was improved by more than 2 %, and accuracy was 97.50 % and 95% respectively. The comparison of results indicates that setting dilation convolution with different dilated rates in the module can improve the network's representation ability and improve model performance.

In the above ablation experiment comparison, it can be seen that under different settings, the image level classification performance of the PCSAM-Net voting has not changed significantly, and the classification accuracy corresponding to the "ALL" model has a slight advantage. At the ensemble model level, the classification performance of the ensemble model corresponding to "ALL" is significantly better than the other two settings. The comparison of these results indicates that the application of dilation convolution and scale attention can not only improve the classification performance of the first-level network, but also extract and retain more image features to optimize the training and learning of the second-level network to improve the classification performance of the ensemble model.

4. Conclusion

In summary, we first analyzed the limitations of existing breast cancer pathological image classification methods and developed a PCSAM-ResCBAM model based on pathological images to distinguish benign from malignant breast lesions. This model included two levels: PCSAM-Net and ResCBAM-Net. Among them, the PCSAM-Net is based on patches to extract the overall structures and texture features of cells, and ResCBAM-Net extracts the images' global characteristics to implement the image-level classification based on feature fusion. In addition, we proposed a novel tiling feature fusion method. We evaluated model performance on BreakHis and ICIAR2018 datasets respectively, and obtained state-of-the-art results at 200x and 400x magnification. The results showed that our model outperformed other existing models in differentiating benign from malignant breast lesions. Moreover, the proposed model is a lightweight model and is easier to deploy on mobile devices. We also performed an ablation study, which verified the importance of scale attention and dilation convolution in the PCSAM. Our results showed that the use of deep learning methods to classify pathological images could assist clinicians in disease diagnosis and provide better personalized treatment for patients. In future studies, we

Table 7
Comparison of results in the model using or not using dilation convolution or scale attention on the ICIAR2018 dataset.

	PCSAM-Net Voting				PCSAM-ResCBAM			
	Maj	Max	Sum	average	accuracy	precision	recall	F1-Score
ALL	94.65	94.25	95	94.63	97.50	97.50	97.62	97.50
Non-Dilation	93.75	94.25	95	94.3	95	95.0	95.12	95
Without SA	94	94.5	93.75	94.8	95.46	95.0	95.1	95.46

will further improve the model to achieve multiple subtype classification of breast pathological images.

Deep learning methods have been widely applied in the study of breast pathological image classification. However, the existing methods fail to consider the important proportion of multi-scale information. Some studies have attempted to increase the network receptive field and improve the performance of the model by increasing the depth or width of the network [62], which has led to a large number of parameters. However, these studies have failed to consider the unique characteristics of the pathological image, such as the disordered textures of the nucleus and the associated tissues or the presence of numerous cytoplasm and other interference information. Additionally, they have overlooked the correlation among features during feature fusion.

We compared our work with previous studies, which were summarized in Table 4, and the results showed that our proposed model had some advantages in the classification task of breast pathological images. In our study, we proposed a parallel convolutional scale attention module (PCSAM), which fully considered the contribution of multi-scale information extracted by the convolutional module to the output, and set up the receptive fields of the cavernous convolutional network with different dilation rates to expand the receptive fields. Furthermore, to avoid image information loss and improve network performance, we proposed a tiling feature fusion method, which significantly improved the performance of the model compared with the superposition fusion method.

There were two main limitations in this study. First, the currently publicly available breast cancer pathological image datasets are small, and their research development is limited. Second, our model had good performance in realizing the automatic classification of breast cancer pathological images. However, in the clinical diagnosis, after the pathologist determines whether the breast lesion is benign or malignant, it is necessary to further determine the subtype category, and the drug response of different cancer subtypes has a significant impact on the treatment outcome. Due to the subtle differences between the pathological images of breast subtypes, the lightweight model cannot extract more complex features with higher dimensions.

Therefore, our future work will focus on building more advanced algorithms to overcome the impact of image differences in multiple small-scale datasets. These algorithms will aim to capture the spatial structure and context information from images, enabling multi-subtype classification of breast pathological images. Furthermore, the model proposed in our study demonstrates potential for application in real-world scenarios, potentially facilitating the practical utilization of computer-assisted medicine in the diagnosis and treatment of human breast cancer.

Data availability statement

The data that support the findings of this study are openly available at the following URL: https://iciar2018-challenge.grand-challenge.org/Dataset/(ICIAR2018 dataset), https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/(BreakHis dataset).

CRediT authorship contribution statement

Ting Yan: Writing – original draft, Methodology, Investigation. **Guohui Chen:** Formal analysis, Data curation. **Huimin Zhang:** Formal analysis, Data curation. **Guolan Wang:** Formal analysis. **Zhenpeng Yan:** Data curation. **Ying Li:** Formal analysis. **Songrui Xu:** Formal analysis. **Qichao Zhou:** Formal analysis. **Ruyi Shi:** Supervision. **Zhi Tian:** Formal analysis. **Bin Wang:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the Fundamental Research Program of Shanxi province (20210302123292, 20210302123112, 20210302123293), the Central Guidance on Local Science and Technology Development Fund of Shanxi Province (YDZJSX2021A018), the National Natural Science Foundation of China (62176177), and the Open Fund from Key Laboratory of Cellular Physiology (Shanxi Medical University), Ministry of Education, China (No.CPOF202212).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e30889.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: a cancer journal for clinicians 71 (3) (2021) 209–249.
[2] B. Fu, P. Liu, J. Lin, et al., Predicting Invasive disease-free survival for early-stage breast cancer patients using follow-up clinical data, IEEE Trans. Biomed. Eng. (2018).
[3] M. Biswas, V. Kuppili, L. Saba, et al., State-of-the-art review on deep learning in medical imaging, Front Biosci (Landmark Ed) 24 (3) (2019) 392–426.
[4] M. Fraiwan, E. Faouri, On the automatic detection and classification of skin cancer using deep transfer learning, Sensors 22 (13) (2022) 4963.
[5] K. Ma, L. Sun, Y. Wang, J. Wang, Classification of blood cancer images using a convolutional neural networks ensemble, in: Eleventh International Conference on Digital Image Processing (ICDIP 2019), SPIE, 2019.
[6] A. Saber, M. Sakr, O.M. Abo-Seida, et al., A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique, IEEE Access 9 (2021) 71194–71209.
[7] B. Wang, S. Zhang, X. Wu, et al., Multiple survival outcome prediction of glioblastoma patients based on multiparametric MRI, Front. Oncol. (2021) 4870.
[8] T. Yan, L. Liu, Z. Yan, et al., A radiomics nomogram for non-invasive prediction of progression-free survival in esophageal squamous cell carcinoma, Front. Comput. Neurosci. 16 (2022).
[9] L. Zhang, C. Li, D. Peng, et al., Raman spectroscopy and machine learning for the classification of breast cancers, Spectrochim. Acta Mol. Biomol. Spectrosc. 264 (2022) 120300.
[10] D. Liu, X. Sun, A. Liu, et al., Predictive value of a novel Asian lung cancer screening nomogram based on artificial intelligence and epidemiological characteristics, Thoracic Cancer 12 (23) (2021) 3130–3140.
[11] B. NiroomandFam, A. Nikravanshalmani, M. Khalilian, Automatic breast mass detection in mammograms using density of wavelet coefficients and a patch-based CNN, Int. J. Comput. Assist. Radiol. Surg. 16 (10) (2021) 1805–1815.
[12] Y. Wang, X. Ge, H. Ma, et al., Deep learning in medical ultrasound image analysis: a review, IEEE Access 9 (2021) 54310–54324.
[13] Y. Benhammou, B. Achchab, F. Herrera, S. Tabik, BreakHis based breast cancer automatic diagnosis using deep learning: taxonomy, survey and insights, Neurocomputing 375 (2020) 9–24.
[14] G. Murtaza, L. Shuib, A.W. Abdul Wahab, et al., Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges, Artif. Intell. Rev. 53 (3) (2020) 1655–1720.
[15] K. Nazeri, A. Aminpour, M. Ebrahimi, Two-stage convolutional neural network for breast cancer histology image classification, in: Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15, Springer, 2018.
[16] R. Yan, F. Ren, Z. Wang, et al., Breast cancer histopathological image classification using a hybrid deep neural network, Methods 173 (2020) 52–60.
[17] M. Salvi, U.R. Acharya, F. Molinari, K.M. Meiburger, The impact of pre-and post-image processing techniques on deep learning frameworks: a comprehensive review for digital pathology image analysis, Comput. Biol. Med. 128 (2021) 104129.
[18] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, Breast cancer histopathological image classification using convolutional neural networks, in: 2016 International Joint Conference on Neural Networks (IJCNN), IEEE, 2016.
[19] T. Araújo, G. Aresta, E. Castro, et al., Classification of breast cancer histology images using convolutional neural networks, PLoS One 12 (6) (2017) e0177544.
[20] Z. Gandomkar, P.C. Brennan, C. Mello-Thoms, MuDeRN: multi-category classification of breast histopathological image using deep residual networks, Artif. Intell. Med. 88 (2018) 14–24.
[21] S. Kushwaha, M. Adil, M. Abuzar, et al., Deep learning-based model for breast cancer histopathology image classification, in: 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), IEEE, 2021.
[22] A.M. Ibraheem, K.H. Rahouma, H.F. Hamed, 3PCNNB-Net: three parallel CNN branches for breast cancer classification through histopathological images, J. Med. Biol. Eng. 41 (4) (2021) 494–503.
[23] M.I. Sarker, H. Kim, D. Tarasov, D. Akhmetzanov, Inception architecture and residual connections in classification of breast cancer histology images, arXiv preprint arXiv:1912.04619 (2019).
[24] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2017.
[25] S.H. Kassani, P.H. Kassani, M.J. Wesolowski, et al., Breast cancer diagnosis with transfer learning and global pooling, in: 2019 International Conference on Information and Communication Technology Convergence (ICTC), IEEE, 2019.
[26] A. Golatkar, D. Anand, A. Sethi, Classification of breast cancer histology using deep learning, in: Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15, Springer, 2018.
[27] C. Munien, S. Viriri, Classification of hematoxylin and eosin-stained breast cancer histology microscopy images using transfer learning with EfficientNets, Comput. Intell. Neurosci. (2021. 2021) 5580914.
[28] C. Szegedy, W. Liu, Y. Jia, et al., Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[29] L. Zhang, F. Xu, Y. Li, et al., A lightweight convolutional neural network model with receptive field block for C-shaped root canal detection in mandibular second molars, Sci. Rep. 12 (1) (2022) 17373.
[30] M. Toğaçar, K.B. Özkurt, B. Ergen, Z. Cömert, BreastNet: a novel convolutional neural network model through histopathological images for the diagnosis of breast cancer, Phys. Stat. Mech. Appl. 545 (2020) 123592.
[31] X.Y. Liew, N. Hameed, J. Clos, An investigation of XGBoost-based algorithm for breast cancer classification, Machine Learning with Applications 6 (2021) 100154.
[32] P. Wang, J. Wang, Y. Li, et al., Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing, Biomed. Signal Process Control 65 (2021) 102341.
[33] M.Z. Alom, C. Yakopcic, M.S. Nasrin, et al., Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network, J. Digit. Imag. 32 (2019) 605–617.
[34] Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, Neurocomputing 452 (2021) 48–62.
[35] R. Ranjbarzadeh, A Bagherian Kasgari, S Jafarzadeh Ghoushchi, et al., Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images, Sci. Rep. 11 (1) (2021) 1–17.
[36] L. He, J. C-W Chan, Z. Wang, Automatic depression recognition using CNN with attention mechanism from videos, Neurocomputing 422 (2021) 165–175.
[37] H. Aboutalebi, M. Pavlova, H. Gunraj, et al., MEDUSA: multi-scale encoder-decoder self-attention deep neural network architecture for medical image analysis, Front. Med. 8 (2022) 2891.
[38] A. Vaswani, N. Shazeer, N. Parmar, et al., Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[39] V. Mnih, N. Heess, A. Graves, Recurrent models of visual attention, Adv. Neural Inf. Process. Syst. 27 (2014).
[40] R. Li, S. Wang, Z. Wang, L. Zhang, Breast cancer X-ray image staging: based on efficient net with multi-scale fusion and cbam attention, in: Journal of Physics: Conference Series, IOP Publishing, 2021.

[41] J. Wang, Z. Yu, Z. Luan, et al., RDAU-net: based on a residual convolutional neural network with DFP and CBAM for brain tumor segmentation, Front. Oncol. (2022) 210.
[42] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018.
[43] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, arXiv preprint arXiv:1511.07122 (2015).
[44] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, IEEE Trans. Biomed. Eng. 63 (7) (2015) 1455–1462.
[45] G. Aresta, T. Araújo, S. Kwok, et al., Bach: grand challenge on breast cancer histology images, Med. Image Anal. 56 (2019) 122–139.
[46] M. Macenko, M. Niethammer, J.S. Marron, et al., A method for normalizing histology slides for quantitative analysis, in: 2009 IEEE International Symposium on Biomedical Imaging: from Nano to Macro, IEEE, 2009.
[47] K. Nazeri, A. Aminpour, M. Ebrahimi, Two-stage convolutional neural network for breast cancer histology image classification, in: International Conference Image Analysis and Recognition, Springer, 2018.
[48] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[49] W. Luo, Y. Li, R. Urtasun, R. Zemel, Understanding the effective receptive field in deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 29 (2016).
50] X. Cao, A Practical Theory for Designing Very Deep Convolutional Neural Networks, 2015. Unpublished Technical Report.
[51] A.F. Agarap, Deep learning using rectified linear units (relu), arXiv preprint arXiv:1803.08375 (2018).
[52] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, PMLR, 2015.
[53] R.R. Selvaraju, M. Cogswell, A. Das, et al., Grad-cam: visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017.
[54] D.M. Vo, N.-Q. Nguyen, S.-W. Lee, Classification of breast cancer histology images using incremental boosting convolution networks, Inf. Sci. 482 (2019) 123–138.
[55] V. Gupta, A. Bhavsar, Sequential modeling of deep features for breast cancer histopathological image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018.
[56] Z. Han, B. Wei, Y. Zheng, et al., Breast cancer multi-classification from histopathological images with structured deep learning model, Sci. Rep. 7 (1) (2017) 4172.
[57] X. Zhang, Y. Zhang, B. Qian, et al., Classifying breast cancer histopathological images using a robust artificial neural network architecture, in: Bioinformatics and Biomedical Engineering: 7th International Work-Conference, IWBBIO 2019, Granada, Spain, May 8-10, 2019, Proceedings, Part I 7, Springer, 2019.
[58] S. Chattopadhyay, A. Dey, P.K. Singh, R. Sarkar, DRDA-Net: dense residual dual-shuffle attention network for breast cancer classification using histopathological images, Comput. Biol. Med. 145 (2022) 105437.
[59] K. Roy, D. Banik, D. Bhattacharjee, M. Nasipuri, Patch-based system for classification of breast histology images using deep learning, Comput. Med. Imag. Graph. 71 (2019) 90–103.
[60] A. Golatkar, D. Anand, A. Sethi, Classification of breast cancer histology using deep learning, in: International Conference Image Analysis and Recognition, Springer, 2018.
[61] A. Rakhlin, A. Shvets, V. Iglovikov, A.A. Kalinin, Deep convolutional neural networks for breast cancer histology image analysis, in: Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15, Springer, 2018.
[62] F.F. Ting, Y.J. Tan, K.S. Sim, Convolutional neural network improvement for breast cancer classification, Expert Syst. Appl. 120 (2019) 103–115.