

Article

A Modified MobileNetV3 Model Using an Attention Mechanism for Eight-Class Classification of Breast Cancer Pathological Images

Chang Guo, Qingjian Zhou *, Jia Jiao, Qingyang Li and Lin Zhu

College of Science, Dalian Minzu University, Dalian 116600, China; xzgc1538@163.com (C.G.); jiaojia@dlnu.edu.cn (J.J.); 202211055059@stu.dlnu.edu.cn (Q.L.); zhulin_dl原因@163.com (L.Z.)

* Correspondence: zqj68@dlnu.edu.cn

Abstract: Addressing the challenge of achieving precise subtype classification of breast cancer histopathology images with limited resources, a lightweight model incorporating multi-stage information fusion and an attention mechanism is proposed for this task. Using MobileNetV3 as the backbone, a multi-stage fusion strategy captures the rich image information in breast cancer histopathology images. Additionally, the selective kernel (SK) attention mechanism is introduced in the initial stages of feature extraction, while an improved squeeze-and-excitation coordinate attention (SCA) mechanism is integrated in the later stages to enhance the extraction of both underlying and semantic features. The final feature representations for subtype classification are determined based on the attention map weights computed at each stage. The experimental results demonstrate the model's outstanding recognition performance on the BreakHis dataset, achieving subtype classification accuracies of 96.259%, 94.763%, 95.511%, and 94.015% at four different magnifications.

Keywords: MobileNetV3; breast cancer; image classification; attention mechanism



Citation: Guo, C.; Zhou, Q.; Jiao, J.; Li, Q.; Zhu, L. A Modified MobileNetV3 Model Using an Attention Mechanism for Eight-Class Classification of Breast Cancer Pathological Images. *Appl. Sci.* **2024**, *14*, 7564. <https://doi.org/10.3390/app14177564>

Academic Editor: Douglas O'Shaughnessy

Received: 26 July 2024

Revised: 20 August 2024

Accepted: 22 August 2024

Published: 27 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rising incidence of breast cancer has posed a significant threat to human health in recent decades. Studies show that early screening and diagnosis significantly improve treatment outcomes and patient survival rates. In this context, pathologic examination plays a vital role in breast cancer diagnosis, serving as the gold standard for preoperative diagnosis. This process involves obtaining tumor tissue samples through puncture or surgery, followed by complex processing steps to create glass slides for microscopic examination and analysis by doctors [1–4]. However, this procedure is time consuming and highly reliant on the physician's expertise, with a risk of misdiagnosis and omission [5]. Due to the scarcity of pathologists and the limitations of manual diagnosis, many researchers are exploring artificial intelligence techniques to automatically analyze breast cancer histopathology images, aiding physicians in making more accurate clinical diagnoses.

Machine learning has become increasingly significant in medical image analysis. By efficiently extracting and processing features from medical images, machine learning algorithms can accurately predict and diagnose disease risks. This technique, combined with medical experts' knowledge, significantly enhances the accuracy of early diagnosis and treatment effectiveness, providing a reliable basis for prognostic assessment [6,7]. Spanhol et al. [8] created a public dataset of breast cancer histopathological images, BreakHis, extracting features using six traditional descriptors and classifying benign and malignant images with various classifiers, achieving an accuracy of 80–85%. Zhang Li et al. [9] enhanced feature extraction by fusing image features of different color gamuts and scales, using a support vector machine for classification, with accuracies of 95.31%, 94.34%, 93.07%, and 91.94% at different magnifications. Wang et al. [10] combined shape-based and texture features with a genetic algorithm and support vector machine,

achieving an accuracy of 96.19% for classifying breast cancer images as benign or malignant. However, with complex and diverse pathological tissue images, traditional feature descriptors often capture only shallow features and cannot comprehensively and accurately represent deep information and complex structures.

Advancements in technology have led to the integration of deep learning methods into medical image analysis, providing notable benefits. Spanhol et al. [11] utilized the AlexNet model as the core network to classify benign and malignant breast cancer pathology images from the BreakHis dataset, employing four different strategies for image block fusion, resulting in a 6% accuracy improvement over traditional machine learning methods. Gour et al. [12] developed a ResHist model based on residual learning, combined with data enhancement techniques, achieving an accuracy of 92.52% in classifying benign and malignant images from the BreakHis dataset. Miaolin Zhang et al. [13] introduced a DC-DenseNet model with DenseNet as the backbone network, incorporating dilation convolution in dense blocks to capture multi-scale features, reaching an accuracy of 94.10% on the BACH dataset for four-class classification. Kausar et al. [14] employed the Haar wavelet transform to decompose breast cancer images, reducing the computational load and resource usage of the deep convolutional neural network (CNN) during convolution. By merging features from each level of the VGG16 model for the multi-classification of breast cancer images in the BreakHis dataset, they achieved an accuracy of 96.85%. Srikantamurthy et al. [15] combined a CNN with a long- and short-term memory recurrent neural network to create a hybrid model, attaining different magnification accuracies of 96.5%, 92.6%, 88.94%, and 92.51% for breast cancer images from the BreakHis dataset.

The multi-layer neural network architecture utilizing convolutional neural networks (CNNs) effectively captures the deep semantic features of images, significantly enhancing the accuracy and reliability of image recognition. Despite these advancements, deep learning faces several challenges and limitations in medical image classification. Firstly, deep learning models require substantial labeled training data, which are often scarce and unevenly distributed in medical imaging. Additionally, these models have complex structures with numerous parameters, leading to long training times and high computational resource demands, thus hindering their deployment and practical application in computer-aided diagnosis.

To address these issues and enhance the accuracy of breast cancer histopathological image subtype classification while maintaining a lightweight design, this paper proposes a network model incorporating multi-layer information fusion and an attention mechanism. This approach enhances classification accuracy and ensures model efficiency, facilitating accurate breast cancer subtype classification and supporting doctors in making quick and precise diagnoses.

This article is organized as follows: Section 2 presents the overall model design. First, MobileNetV3 is used as the backbone network, and a multi-stage information fusion strategy is proposed to capture rich information in breast cancer histopathology images. Next, the selective kernel (SK) attention mechanism is introduced in the initial two stages of feature extraction, while the squeeze-and-excitation coordinate attention (SCA) mechanism, an improved attention mechanism developed in this paper, is implemented in the final two stages to target the extraction of underlying and semantic features of breast cancer histopathology images. Ultimately, the final feature representations for breast cancer subtype classification are obtained based on the attention map weights computed at each stage. Section 3 details the experimental setup and result analysis, including the dataset, implementation details, evaluation criteria, and analysis of the experimental results. Section 4 provides conclusions and final remarks.

2. Overall Design of the Model

This study introduces a lightweight model designed for the subtype classification of breast cancer pathology images, incorporating multi-stage information fusion and an attention mechanism. The model architecture, depicted in Figure 1, employs MobileNetV3

as the backbone network, leveraging its efficient and lightweight characteristics to ensure both processing speed and ease of deployment. To fully harness the rich information contained in breast cancer histopathology images, the proposed model implements a multi-stage information fusion strategy. This strategy incrementally extracts and integrates both detailed and deep semantic features of the images through various stages of processing. In the initial two stages of feature extraction, the selective kernel (SK) attention mechanism is employed to effectively capture fundamental features such as color and texture. This mechanism adapts the convolution kernel size to optimally adjust to these underlying features. In the subsequent stages, the model integrates our improved SCA module. This module enhances the focus on key spatial locations and channel information within the image, thereby capturing higher-level semantic information crucial for the accurate recognition of complex breast cancer histopathology images. This design ensures that the model captures essential underlying features in the early stages while focusing on significant semantic features in the later stages. Ultimately, the feature representation capability is further refined by computing the attention map weights at each stage.

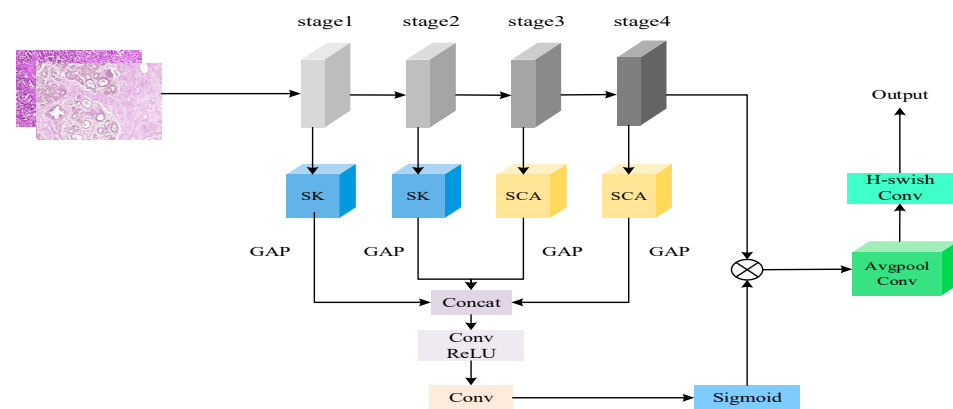


Figure 1. Overall network design.

2.1. MobileNetV3 Model

Pathology images are often characterized by significant noise and complex information, which poses a considerable challenge for classification algorithms. MobileNetV3, a lightweight deep learning model [16], is widely adopted in mobile devices and edge computing due to its compact size and high computational efficiency. It utilizes the NetAdapt algorithm [17] to determine the optimal number of convolutional kernels and channels, thereby enhancing model efficiency. MobileNetV3 retains the depthwise separable convolution from its V1 version [18], which effectively reduces the number of parameters and computations, making it well suited for mobile applications. Additionally, it incorporates the linear bottleneck residual structure from the V2 version [19], allowing the model to learn deeper network structures while maintaining low computational costs. Including the squeeze-and-excitation (SE) channel attention mechanism [20] further enhances the network's ability to concentrate on significant feature channels, thus improving its representation capability. Due to its excellent performance and lightweight nature, MobileNetV3 is widely utilized in both academia and industry.

MobileNetV3 is available in two versions: large and small. For the task of multi-categorization of breast cancer pathology images, the critical requirement is to capture subtle feature differences between images. Although MobileNetV3-Large has a slightly higher number of parameters and computational demands compared to MobileNetV3-Small, it can learn higher-level abstract features and complex patterns, aiding in differentiating breast cancer subtypes. Hence, MobileNetV3-Large was selected as the backbone network for this study.

Table 1 outlines the parameters of the MobileNetV3-Large model. In the initial stage, the H-swish activation function is introduced, which reduces computation and improves

operation speed compared to the swish activation function. In the intermediate stage, the SE attention block is incorporated to enhance classification accuracy. This process involves extracting the mean value of the input features using the average pooling layer, applying the ReLU and H-swish activation functions to compute the feature weights, and multiplying these weights with the original matrix to obtain weighted output features. This approach strengthens the model's learning capability by combining the feature channel. In the final stage, the average pooling layer is advanced, and the convolutional layer adjusts the number of feature channels to expand into a higher-dimensional space. This structural design allows the MobileNetV3-Large model to enhance computational speed and performance without sacrificing accuracy.

Table 1. MobileNetV3-Large parameter setting.

Input	Operation	SE Modular	Activation Function	Step Length
$224 \times 224 \times 3$	conv2d		HS	2
$112 \times 112 \times 16$	bneck, 3×3		RE	1
$112 \times 112 \times 16$	bneck, 3×3		RE	2
$56 \times 56 \times 24$	bneck, 3×3		RE	1
$56 \times 56 \times 24$	bneck, 5×5	✓	RE	2
$28 \times 28 \times 40$	bneck, 5×5	✓	RE	1
$28 \times 28 \times 40$	bneck, 5×5	✓	RE	1
$28 \times 28 \times 40$	bneck, 3×3		HS	2
$14 \times 14 \times 80$	bneck, 3×3		HS	1
$14 \times 14 \times 80$	bneck, 3×3		HS	1
$14 \times 14 \times 80$	bneck, 3×3		HS	1
$14 \times 14 \times 80$	bneck, 3×3	✓	HS	1
$14 \times 14 \times 112$	bneck, 3×3	✓	HS	1
$14 \times 14 \times 112$	bneck, 5×5	✓	HS	2
$7 \times 7 \times 160$	bneck, 5×5	✓	HS	1
$7 \times 7 \times 160$	bneck, 5×5	✓	HS	1
$7 \times 7 \times 160$	conv2d, 1×1	-	HS	1
$7 \times 7 \times 960$	pool, 7×7	-	-	1
$1 \times 1 \times 960$	conv2d, 1×1 , NBN	-	HS	1
$1 \times 1 \times 1280$	conv2d, 1×1 , NBN		-	1

2.2. Multi-Stage Information Fusion Strategy

MobileNetV3 serves as the backbone network, leveraging the comprehensive feature information extracted in its fourth stage. Given the complexity and diversity of pathology images, it is essential to extract both texture and color features in the early stages and deeper semantic features in the later stages. To address this, a multi-stage information fusion strategy is proposed in the present study. The MobileNetV3 network structure divides feature extraction into four stages. Figure 2 reveals that the first stage consists of one convolutional layer and three bottleneck layers (bnecks) with a 3×3 kernel size. The second stage comprises three bottleneck layers with a 5×5 kernel size. The third stage includes six bottleneck layers with a 3×3 kernel size, and the fourth stage consists of three bottleneck layers with a 5×5 kernel size and one convolutional layer. The multi-stage information fusion strategy processes the input feature maps from all four stages (x_1, x_2, x_3, x_4) and calculates an attention map, which is then multiplied with the fourth stage's feature map to produce the final output. This attention map, a matrix matching the input feature map in dimension, highlights the importance of each pixel. When multiplied pixel by pixel with the fourth feature map, it results in a new feature map that combines the original features with additional important information, thereby enhancing the model's interpretation of breast cancer histopathology images. The fusion process involves several steps, depicted in Figure 2. Initially, a global average pooling (GAP) operation reduces the spatial dimensions of each input feature map to one. These pooled features from all the four stages are concatenated along the channel dimension, and a 1×1 convolution reduces the number of channels. After applying the ReLU activation function to process

these features, another 1×1 convolution restores the original number of channels. Finally, sigmoid activation generates the weighted features, which are multiplied with the fourth input feature map to obtain the final weighted feature information. Through this multi-stage processing, the model captures distinct features at various levels: the first two stages focus on texture, color, and morphology information, while the last two stages emphasize the deeper semantic features of the breast cancer histopathology images.

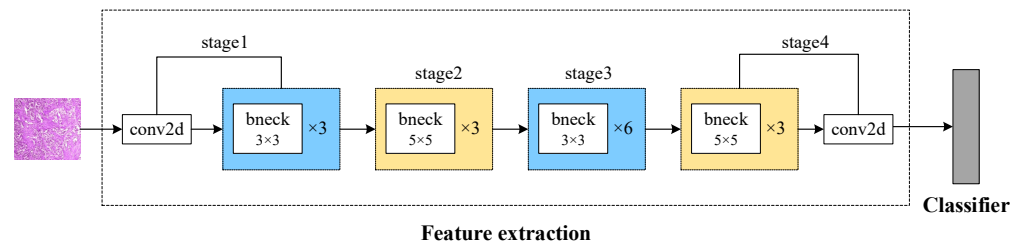


Figure 2. Multi-stage partition schematic diagram.

2.3. SK Attention Mechanism

The SK attention mechanism focuses on the receptive field size of neurons in CNNs [21]. The receptive field determines the area of input data a neuron can observe, which is crucial for processing images, as different features and objects require varying receptive field sizes. The SK attention mechanism dynamically adjusts the neuron's receptive field by introducing an attention mechanism, enhancing the model's ability to extract multi-scale features. Similar to SENet, SK attention aims to provide a lightweight gating mechanism that reduces computational burden and improves model efficiency, making it ideal for resource-constrained environments.

This article integrates the SK attention mechanism after the last bottleneck layer in the first two stages of the model's feature extraction process. The primary goal is to enhance the network's adaptability to breast cancer histopathology images by dynamically adjusting the convolution kernel size to capture features at different scales and focus on critical underlying features such as local details and textures. The designed SK attention includes 3×3 and 5×5 convolutions, as shown in Figure 3, which provide different receptive fields to capture various levels of features in breast cancer histopathology images.

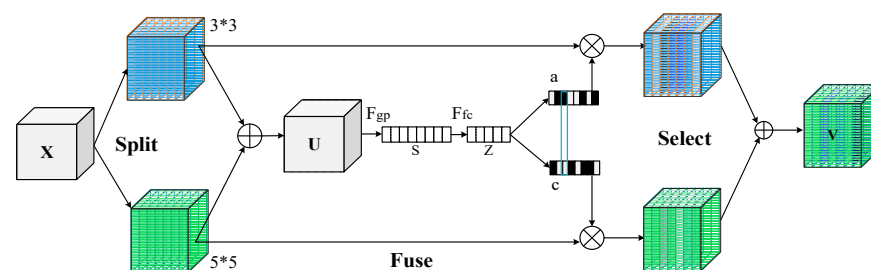


Figure 3. SK attention mechanism.

SK attention extracts features from input feature maps using convolution kernels of various sizes, and then automatically selects the most relevant information. The process involves three main operations: split, fuse, and select. Initially, convolution kernels of two different sizes are used to convolve the feature map, extracting features that are then summed element-wise. These fused feature maps undergo GAP in one spatial dimension. Next, full connectivity is employed to extract channel attention information, in which different convolution kernels are applied to the feature maps from the previous step. These are processed by the Softmax layer to obtain channel attention information specific to each convolution kernel. Finally, the channel attention information from different convolution kernels is combined with the respective feature maps and fused into the final feature map.

2.4. SCA Mechanism

Coordinate attention (CA) [22] is an attention mechanism designed for deep learning models to enhance their sensitivity and processing of spatial information in input data. This mechanism improves the model's focus on different spatial locations within an image without significantly increasing the computational load. Its lightweight design makes it particularly suitable for applications in which models must function in resource-constrained environments, such as on mobile devices.

In breast cancer histopathology images, lesion features can vary in morphology and correlation across different spatial locations. To address this, coordinate attention is introduced after the last bottleneck layer in two stages following feature extraction. This aims to enhance the model's understanding and use of spatial context information, focusing on key regions within the pathology images and thereby improving the recognition of complex pathological patterns. The coordinate attention mechanism, illustrated in Figure 4, involves two main steps: coordinate information embedding and coordinate attention generation. First, the input feature maps undergo GAP in both horizontal and vertical directions to capture positional information along the x-axis and y-axis. These pooled feature maps are then concatenated and processed through a 1×1 convolution to create attention feature maps. Subsequently, feature extraction is performed on the activation-function-processed maps along the horizontal and vertical directions. Finally, these feature maps are multiplied by weights processed in the respective directions to produce the final feature maps.

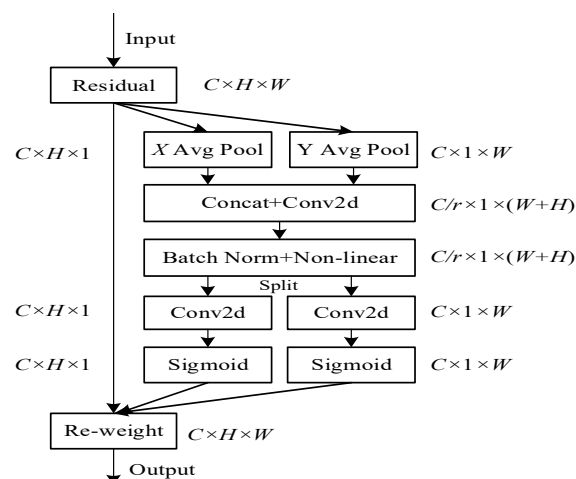


Figure 4. CA mechanism.

While classifying the subtypes of breast cancer histopathology images, model performance can be hindered by the presence of tiny lesions that are difficult to identify. To address this, an improved version of the CA mechanism, named the SCA block, is introduced. Its structure is illustrated in Figure 5. Inspired by the core principles of the SE block, the design includes two parallel fully connected paths, FC1 and FC2, which process the globally averaged pooled features separately. The outputs from these paths are concatenated in the channel dimension, followed by the generation of channel weights through another fully connected layer and a sigmoid activation function. This approach enhances the model performance to adaptively learn the importance of input feature channels. This design not only introduces nonlinear factors but also captures complex dependencies between channels, thereby improving the information identification performance of the model. The reduced-dimensional features are processed through two fully connected layers and then scaled back to the original number of channels. This allows the learning of more complex features through nonlinear transformations without losing information, facilitating the identification of various types of microscopic lesions in breast cancer histopathology images. By enhancing the input features with regional information and then applying the CA mechanism, the model's comprehension and utilization of chan-

nel information are improved. This enables the network to process spatial and channel information more effectively, leading to better recognition of subtle features across different pathology image types.

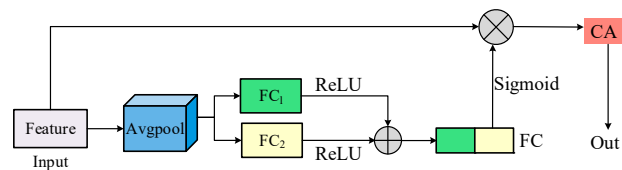


Figure 5. SCA mechanism.

3. Experimental Setup and Result Analysis

3.1. Dataset

The datasets employed in this study were sourced from the open-source BreakHis dataset, which is organized into four groups based on different magnifications: $40\times$, $100\times$, $200\times$, and $400\times$. All the datasets underwent dye normalization preprocessing before being used in the experiments. For each magnification group, 80% of the images were randomly selected as the training set, while the remaining ones were used as the test set. The input image size was set to 224×224 pixels. To mitigate overfitting, images in the training set were augmented by flipping them horizontally and vertically with a probability of 25%.

3.2. Implementation Details

This subsection focuses on developing a subtype classification model for breast cancer histopathology images and conducting a series of experiments to validate the proposed method's effectiveness. All the experiments were performed on a machine with the following setup: initial learning rate set to 0.0001, batch size of 32, and 50 training iteration rounds. The Adam optimizer was employed to update network weights, with convergence controlled by the cross-entropy loss function. Softmax was used as the output function to achieve the classification of eight pathology images. The experimental environment was based on the Ubuntu 18.04.5 operating system, equipped with an NVIDIA GeForce RTX 3080 Ti GPU and 45 GB of RAM (NVIDIA, Santa Clara, CA, USA). This setup provided ample processing power and storage space. The development was carried out using Python 3.8.10 and PyCharm, with all deep learning tasks executed under the Pytorch 1.8.0 framework. GPU acceleration was facilitated by CUDA 11.7, ensuring efficient model training and inference performance.

3.3. Evaluation Criteria

The model performance was comprehensively evaluated using accuracy, precision, recall, and the F1 value. Given that the focus is on subtype classification, the macro average was employed to calculate the final results. This approach ensures that each category is given equal weight in the calculation of the mean value. Consequently, the final result was the arithmetic mean of the corresponding metrics for each category.

3.4. Experimental Results

3.4.1. Ablation Experiments

To further investigate the impact of the various modules in the improved MobileNetV3 algorithm on the performance metrics of the base algorithm, ablation experiments were conducted on four datasets with different magnifications under the same experimental conditions. The base algorithm used was MobileNetV3. Subsequently, experiments were performed on different models by sequentially incorporating the SK, CA, and the SCA modules, and the multi-stage information fusion strategy into the feature extraction part of the backbone network.

These are shown in Tables 2–5, combining the experimental results from the four datasets with different magnifications, the improved SCA module demonstrated

superior performance compared to the CA module. Additionally, the newly incorporated mechanisms enhanced the algorithm's performance to varying degrees without conflicting with each other. The improved model, when compared to the benchmark MobileNetV3, showed enhancements in all performance metrics, including accuracy. The SK attention module, introduced in the initial two stages of the backbone network's feature extraction part, effectively extracted multi-scale features in breast cancer histopathology images. The SCA module, added in the latter two stages, focused on key features and captured finer, more detailed semantic features, enabling the recognition of subtle changes in the images. The multi-stage feature information strategy enhanced the network's ability to recognize pathological features by refining the feature extraction process at various stages. The experimental results confirm that the improved MobileNetV3 algorithm is highly effective for breast cancer pathology image classification, with the added components proving to be both scientific and effective.

Table 2. The ablation results of the dataset at 40× magnification factor (%).

Model	ACC	Precision	Recall	F1
MobileNetV3	91.272	89.485	93.397	91.168
MobileNetV3_CA	93.766	93.897	92.202	92.941
MobileNetV3_SCA	95.012	94.18	95.602	94.65
MobileNetV3_SK	92.519	93.476	89.562	91.337
MobileNetV3_Strategy	94.015	94.187	92.603	93.329
MobileNetV3_All	96.259	95.729	95.789	95.726

Table 3. The ablation results of the dataset at 100× magnification factor (%).

Model	ACC	Precision	Recall	F1
MobileNetV3	89.526	89.604	88.551	88.627
MobileNetV3_CA	91.521	92.709	87.98	90.057
MobileNetV3_SCA	92.02	89.237	90.277	89.65
MobileNetV3_SK	90.524	87.545	89.818	88.4
MobileNetV3_Strategy	94.514	93.608	92.561	92.988
MobileNetV3_All	94.763	94.091	94.457	93.95

Table 4. The ablation results of the dataset at 200× magnification factor (%).

Model	ACC	Precision	Recall	F1
MobileNetV3	90.025	89.214	87.911	88.46
MobileNetV3_CA	91.771	90.221	89.113	89.606
MobileNetV3_SCA	92.519	92.483	90.086	91.151
MobileNetV3_SK	92.269	93.945	89.99	91.673
MobileNetV3_Strategy	93.017	93.403	90.764	91.821
MobileNetV3_All	95.511	94.909	95.34	95.028

Table 5. The ablation results of the dataset at 400× magnification factor (%).

Model	ACC	Precision	Recall	F1
MobileNetV3	87.282	85.26	86.006	85.041
MobileNetV3_CA	92.768	91.834	90.946	91.298
MobileNetV3_SCA	93.267	93.226	92.011	92.5
MobileNetV3_SK	87.78	86.5	84.49	84.24
MobileNetV3_Strategy	90.773	88.89	90.05	89.177
MobileNetV3_All	94.015	92.338	93.537	92.777

The number of model parameters and the computational load (FLOPs) for the benchmark MobileNetV3 model and its various improved versions are presented in Table 6. The

model has 0.32 million more parameters than the benchmark model, but it significantly improves performance metrics such as accuracy. Specifically, the benchmark model has 233.56 million FLOPs and 4.21 million parameters. The introduction of the CA mechanism increased the number of parameters by only 0.05 million. After incorporating the improved SCA mechanism, the number of parameters increased by just 0.11 million compared to the benchmark model. The use of the multi-stage information fusion strategy added only 0.13 million parameters. Consequently, the total number of parameters for the proposed model is 4.53 million, which maintains a good balance between computational complexity and the number of parameters. The experimental results demonstrate that the various improvement mechanisms implemented in the proposed model enhanced its accuracy without significantly increasing the computational complexity or the number of parameters, thus keeping the model lightweight and efficient.

Table 6. Parameters of different improved models and FLOPs comparison.

Model	FLOPs	Parameters
MobileNetV3	233.56 M	4.21 M
MobileNetV3_CA	234.16 M	4.26 M
MobileNetV3_SCA	234.25 M	4.32 M
MobileNetV3_SK	338.49 M	4.29 M
MobileNetV3_Strategy	233.85 M	4.34 M
MobileNetV3_All	339.46 M	4.53 M

3.4.2. Contrast Experiment

To assess the effectiveness of the improved MobileNetV3 model, its performance was evaluated against that of the ResNet18 [23] and EfficientNetB0 [24] models. These models are widely used for image classification, especially in the context of breast cancer histopathology images. The evaluation was performed across four different magnification levels, providing a thorough and reliable comparison.

Table 7 presents the performance metrics for each model across datasets with different magnifications, including test accuracy, precision, and recall. The proposed network structure outperforms other models in overall classification performance for all four magnification levels. The multi-stage information fusion strategy employed here significantly enhances the model's accuracy by capturing rich image information in breast cancer histopathology images. In the initial two stages of the feature extraction process, the SK attention mechanism is introduced to extract multi-scale features adaptively, improving the learning of texture, color, morphology, and other underlying features. In the latter two stages, the improved SCA mechanism captures spatial context information more precisely. This targeted extraction of underlying and deep semantic features, followed by a weighted fusion of the rich features from different stages, enables the model to effectively learn the subtle differences among various subtypes of lesions in breast cancer histopathology images.

Table 7. Comparative experimental results of datasets at different magnification factor (%).

Magnification Factor	Model	ACC	Precision	Recall	F1
40×	MobileNetV3	91.272	89.485	93.397	91.168
	ResNet18	85.037	87.803	80.170	83.619
	EfficientNetB0	89.027	88.447	87.483	87.884
	MobileNetV3_All	96.259	95.729	95.789	95.726
100×	MobileNetV3	89.526	89.604	88.551	88.627
	ResNet18	83.226	84.108	81.286	82.759
	EfficientNetB0	86.267	86.15	85.923	85.872
	MobileNetV3_All	94.763	94.091	94.457	93.95

Table 7. Cont.

Magnification Factor	Model	ACC	Precision	Recall	F1
200×	MobileNetV3	90.025	89.214	87.911	88.46
	ResNet18	84.987	86.574	82.385	84.783
	EfficientNetB0	88.95	88.652	87.185	87.264
	MobileNetV3_All	95.511	94.909	95.34	95.028
400×	MobileNetV3	87.282	85.26	86.006	85.041
	ResNet18	80.263	82.382	78.329	79.867
	EfficientNetB0	85.012	84.322	82.982	82.389
	MobileNetV3_All	94.015	92.338	93.537	92.777

3.4.3. ROC Curve

The ROC curve is a critical tool to assess a classifier's performance, illustrating its generalization capability in classification tasks. It plots the true positive rate (TPR) on the vertical axis against the false positive rate (FPR) on the horizontal axis. The classifier generates probability values ranging from 0 to 1, indicating the likelihood that a sample belongs to a specific category. By adjusting a threshold value, samples are categorized as positive or negative; probabilities above the threshold are positive, while those below are negative. The ROC curve is constructed by gradually lowering the threshold from 1 to 0, calculating the corresponding (FPR, TPR) points, and connecting these points. The area under the ROC curve (AUC) ranges from 0 to 1, with higher values indicating better model performance.

Figure 6a shows the ROC curves for breast cancer histopathological images at 40× magnification. The AUC values for each type of breast cancer image ranged from a minimum of 0.93 to an average of 0.98. The ROC curves for adenopathy, mucinous carcinoma, and papillary carcinoma exhibited an AUC of 1. However, the model's performance was weakest for lobular carcinoma, which had the lowest AUC, indicating poorer recognition capability. At 100× magnification, as shown in Figure 6b, the minimum AUC for different breast cancer pathology types was 0.95, with an average AUC of 0.97, demonstrating balanced recognition across various pathology types. Notably, lobular carcinoma's AUC reached 0.99, indicating strong recognition at this magnification. Figure 6c displays the ROC curves at 200× magnification. Here, the AUC values for each type of breast cancer pathology image were at least 0.93, with an average AUC of 0.97. The model excelled in recognizing adenopathy and mucinous carcinoma, both achieving an AUC of 1. However, tubular adenoma had the lowest AUC, indicating poorer recognition performance at this magnification. Finally, Figure 6d presents the ROC curves at 400× magnification. The AUC values for each type of breast cancer pathology image reached a minimum of 0.95, with an average AUC of 0.96. While the improved model proposed in this chapter performed well overall, its performance slightly declined compared to other magnifications. This decline is attributed to the increased detail at 400× magnification, which may cause a loss of contextual information and consequently affect the model's recognition accuracy.

3.4.4. Confusion Matrix

The confusion matrix is a widely used evaluation metric in classification problems, illustrating the relationship between actual and predicted categories. In the context of breast cancer histopathology image classification, where the dataset includes various pathology types, the confusion matrix helps visualize how each category is accurately or inaccurately classified. Therefore, this section presents the confusion matrices corresponding to the test results of the model with different modules, using the dataset at 40× magnification as an example, as shown in Figure 7. The pathological subtypes are shown as follows: adenosis (A), fibroadenoma (F), tubular adenoma (TA), phyllodes tumor (PT), papillary carcinoma (PC), ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC).

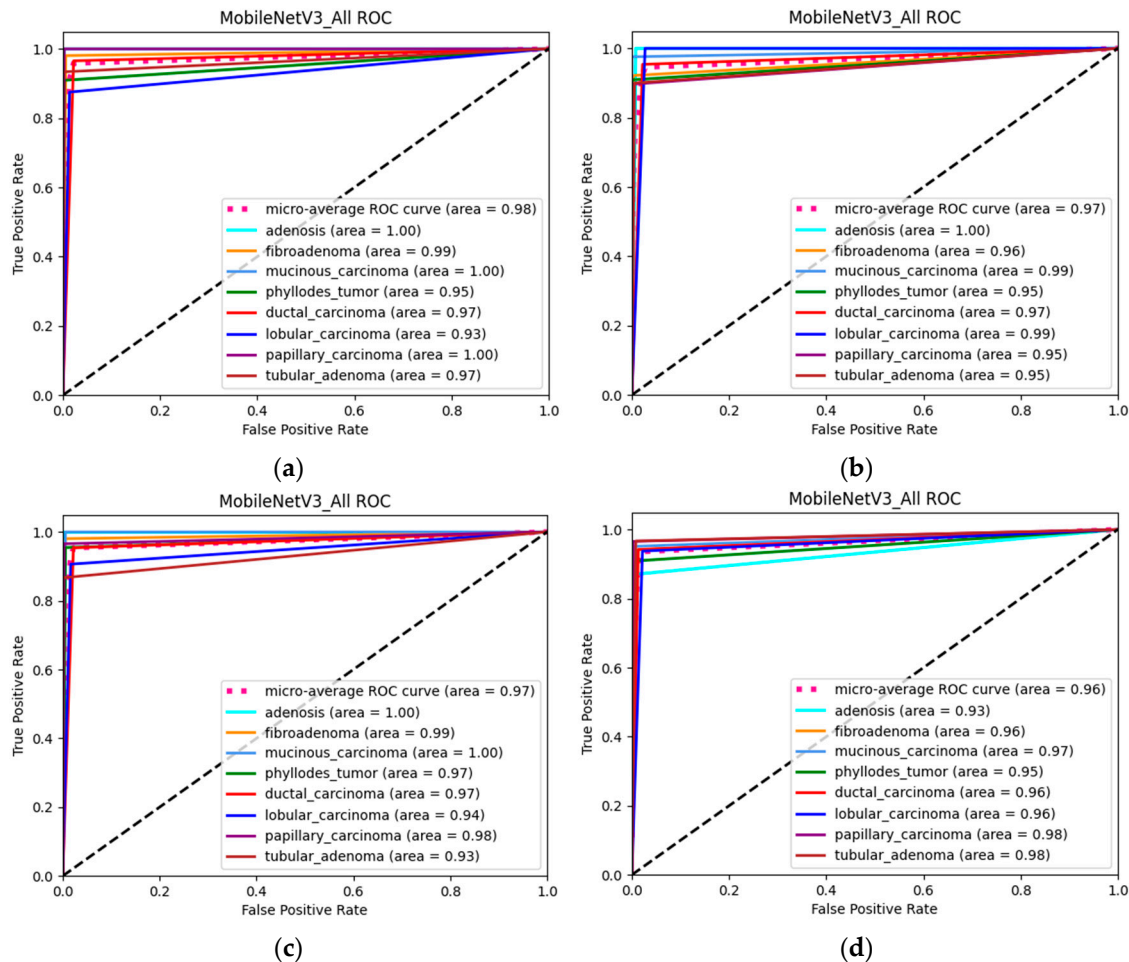


Figure 6. ROC curve at various magnification factors. (a) ROC curve at 40× magnification factor, (b) ROC curve at 100× magnification factor, (c) ROC curve at 200× magnification factor, (d) ROC curve at 400× magnification factor.

The confusion matrix for the benchmark model MobileNetV3, shown in Figure 7a, reveals that ductal carcinoma was often misclassified as lobular carcinoma. Adenopathies, on the other hand, were consistently identified correctly, likely because their cell morphology is more regular, and their textures are clearer. Other pathology types, however, were subject to various degrees of misclassification. When the coordinate attention module was introduced in the last two stages of the feature extraction part of MobileNetV3, denoted as MobileNetV3_CA, there was a notable reduction in ductal carcinoma samples misclassified as lobular carcinoma, as shown in Figure 7b. This indicates an enhanced ability of the model to correctly identify ductal carcinoma with the addition of the CA module. Further improvement was seen with MobileNetV3_SCA, which integrates the improved SCA module into the last two stages of the feature extraction part of MobileNetV3. Figure 7c demonstrates that the number of misclassified samples was significantly reduced compared to both the benchmark model and MobileNetV3_CA. This suggests that the SCA module proposed in this paper is more effective than the CA module in classifying breast cancer histopathology image subtypes. MobileNetV3_SK, which incorporates the SK attention module into the first two stages of feature extraction of the benchmark model, shows an improved recognition ability for ductal carcinoma in Figure 7d. However, the model failed to correctly recognize all adenopathy cases. This could be because adding SK attention in the early stages of feature extraction enhanced the recognition of complex underlying texture information, which benefits the identification of samples with intricate pathological features like ductal carcinoma. However, for adenopathy images with simpler textures,

this increased focus on complex textures may introduce unnecessary interference, leading to the overanalysis of straightforward textures.

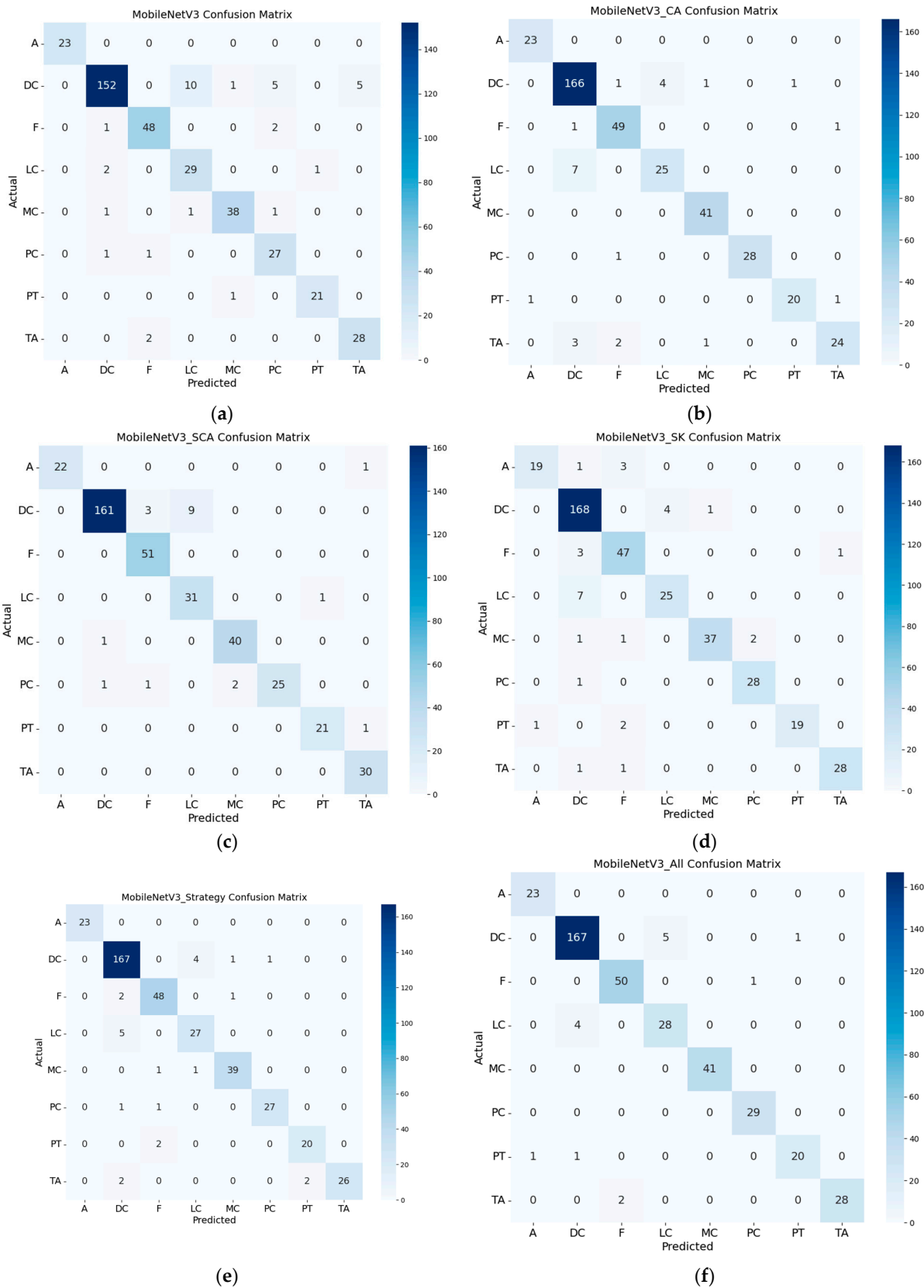


Figure 7. Confusion matrix. (a) MobileNetV3, (b) MobileNetV3_CA, (c) MobileNetV3_SCA, (d) MobileNetV3_SK, (e) MobileNetV3_Strategy, (f) MobileNetV3_All.

The confusion matrix for MobileNetV3_Strategy, which employs the multi-layer feature information fusion strategy, is shown in Figure 7e. This matrix illustrates that the model achieved a more balanced recognition across different lesion categories with this strategy. Adenopathy was correctly recognized in all instances, and the misclassification rate for ductal carcinoma was significantly reduced. This suggests that the multi-stage information fusion strategy effectively enhances feature representation, accommodating both simple and complex pathological samples.

The confusion matrix for MobileNetV3_All, the fully constructed model in this paper, is shown in Figure 7f. This model integrated various modules and strategies, leading to improved recognition performance compared to the benchmark MobileNetV3. Notably, while ductal carcinoma (DC) was often misclassified as lobular carcinoma (LC) due to the subtle differences between subtypes, the addition of attention mechanisms and multi-layer information fusion enabled the model to capture finer lesion features, thus distinguishing breast cancer subtypes more accurately. The enhanced MobileNetV3 model demonstrated superior recognition results, significantly reducing the misclassification rate. Furthermore, the model exhibited a more balanced recognition across all breast cancer histopathology image types, markedly improving its ability to identify rare subtypes.

4. Conclusions

This paper introduces a multi-classification model framework for breast cancer histopathology images, utilizing MobileNetV3 as the backbone network. A multi-stage information fusion strategy is proposed to capture the extensive image details in breast cancer histopathology images. The SK attention mechanism is integrated into the first two stages of feature extraction to harness multi-scale features, enhancing the learning of texture, color, and morphological features. In the latter two stages, the SCA mechanism, an improvement over CA, is incorporated to focus on both the underlying and semantic features of the histopathology images. This approach allows for obtaining final feature representations for subtype classification based on attention map weights calculated at each stage. The dataset, experimental environment, parameters, training methods, and evaluation metrics are described in detail. The effectiveness of the model is validated through detailed comparisons and analyses of experimental results and visualizations. By enhancing features at different stages and assigning appropriate weights, the model effectively learns subtle differences across various subtypes of lesions in breast cancer histopathology images, demonstrating improved classification performance.

Author Contributions: C.G. conceived and developed the main ideas of this study. Q.Z. provided oversight and guidance to ensure the overall quality of this study. J.J. provided insightful ideas during the discussion. Q.L. and L.Z. organized the statistics of the experiments. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is supported by the National Natural Science Foundation of China (12101106), the Natural Science Foundation of Liaoning Province of China (No. 2021-MS-146, 2022-MS-164), and the Liaoning Provincial Department of Education Basic Research Project for Higher Education Institutions—General Project (2021-20).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The instances and code used in this paper can be found at <http://github.com/karryxz/Modified-model>.

Acknowledgments: The authors gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved our work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Siegel, R.L.; Miller, K.D.; Wagle, N.S.; Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* **2023**, *73*, 17–48. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Ding, R.; Xiao, Y.; Mo, M.; Zheng, Y.; Jiang, Y.Z.; Shao, Z.M. Breast cancer screening and early diagnosis in Chinese women. *Cancer Biol. Med.* **2022**, *19*, 450–467. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Wang, W.; Li, Q.; Lan, B.; Ma, F. Interpretation of Quality Control Yearbook of Standardized Diagnosis and Treatment of Breast Cancer in China. *Electron. J. Compr. Tumor Ther.* **2023**, *9*, 23–29.
4. Williams, B.; Hanby, A.; Millican-Slater, R.; Verghese, E.; Nijhawani, A.; Wilson, I.; Besusparis, J.; Clark, D.; Snead, D.; Rakha, E.; et al. Digital pathology for primary diagnosis of screen-detected breast lesions—experimental data, validation and experience from four centres. *Histopathology* **2020**, *76*, 968–975. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Liang, Y.; Hao, M.; Guo, R.; Li, X.; Li, Y.; Yu, C.; Yang, Z. Biomarkers for early screening and diagnosis of breast cancer: A review. *Chin. J. Biotechnol.* **2023**, *39*, 1425–1444.
6. Abunasser, B.S.; AL-Hiealy, M.R.J.; Zaout, I.S.; Abu-Naser, S.S. Literature review of breast cancer detection using machine learning algorithms. *AIP Conf. Proc.* **2023**, *2808*, 040006.
7. Radak, M.; Lafta, H.Y.; Fallahi, H. Machine learning and deep learning techniques for breast cancer diagnosis and classification: A comprehensive review of medical imaging studies. *J. Cancer Res. Clin. Oncol.* **2023**, *149*, 10473–10491. [\[CrossRef\]](#)
8. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. A dataset for breast cancer histopathological image classification. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 1455–1462. [\[CrossRef\]](#)
9. Zhang, L.; Zhang, C.; Hao, Y.; Cheng, R.; Bai, Y. Breast Cancer Histopathological Image Classification Based on Multi-scale and Multi-gamut Feature Fusion. *Comput. Technol. Dev.* **2022**, *32*, 175–180+185.
10. Wang, P.; Hu, X.; Li, Y.; Liu, Q.; Zhu, X. Automatic cell nuclei segmentation and classification of breast cancer histopathology images. *Signal Process.* **2016**, *122*, 1–13. [\[CrossRef\]](#)
11. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. Breast cancer histopathological image classification using convolutional neural networks. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 2560–2567.
12. Gour, M.; Jain, S.; Sunil Kumar, T. Residual learning based CNN for breast cancer histopathological image classification. *Int. J. Imaging Syst. Technol.* **2020**, *30*, 621–635. [\[CrossRef\]](#)
13. Zhang, M.; Shuai, R. Breast cancer pathological image classification based on DC-DenseNet. *Comput. Appl. Softw.* **2023**, *4*, 116–121.
14. Kausar, T.; Wang, M.J.; Idrees, M.; Lu, Y. HWDCCNN: Multi-class recognition in breast histopathology with Haar wavelet decomposed image based convolution neural network. *Biocybern. Biomed. Eng.* **2019**, *39*, 967–982. [\[CrossRef\]](#)
15. Srikantamurthy, M.M.; Rallabandi, V.P.S.; Dudekula, D.B.; Natarajan, S.; Park, J. Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. *BMC Med. Imaging* **2023**, *23*, 19. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for mobilenetv3. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
17. Yang, T.J.; Howard, A.; Chen, B.; Zhang, X.; Go, A.; Sandler, M.; Sze, V.; Adam, H. Netadapt: Platform-aware neural network adaptation for mobile applications. In *Computer Vision—ECCV 2018: 15th European Conference, Munich, Germany, 8–14 September 2018, Proceedings, Part X*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11214.
18. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
19. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
20. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
21. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective kernel networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 510–519.
22. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 13708–13717.
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
24. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.