

## Research Article

# Feature Fusion Based on Convolutional Neural Network for Breast Cancer Auxiliary Diagnosis

Xiaofan Cheng ,<sup>1</sup> Liang Tan ,<sup>1,2</sup> and Fangpeng Ming ,<sup>1</sup>

<sup>1</sup>School of Computer Science, Sichuan Normal University, Cheng Du 610000, Sichuan, China

<sup>2</sup>Institute of Computer Science, Chinese Academy of Sciences, Beijing 100000, China

Correspondence should be addressed to Xiaofan Cheng; sail967642@gmail.com

Received 2 July 2021; Revised 28 August 2021; Accepted 7 September 2021; Published 15 September 2021

Academic Editor: Zhen Liu

Copyright © 2021 Xiaofan Cheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cancer is one of the leading causes of death in many countries. Breast cancer is one of the most common cancers in women. Especially in remote areas with low medical standards, the diagnosis efficiency of breast cancer is extremely low due to insufficient medical facilities and doctors. Therefore, in-depth research on how to improve the diagnosis rate of breast cancer has become a hot spot. With the development of society and science, people use artificial intelligence to improve the auxiliary diagnosis of diseases in the existing medical system, which can become a solution for detecting and accurately diagnosing breast cancer. The paper proposes an auxiliary diagnosis model that uses deep learning in view of the low rate of human diagnosis by doctors in remote areas. The model uses classic convolutional neural networks, including VGG16, InceptionV3, and ResNet50 to extract breast cancer image features, then merge these features, and finally train the model VIRNets for auxiliary diagnosis. Experimental results prove that for the recognition of benign and malignant breast cancer pathological images under different magnifications, VIRNets have a high generalization and strong robustness, and their accuracy is better than their basic network and other structures of the network. Therefore, the solution provides a certain practical value for assisting doctors in the diagnosis of breast cancer in real scenes.

## 1. Introduction

Nowadays, cancer is one of the most important diseases threatening human health. In more than 180 countries, cancer is the leading cause of premature death. According to the statistics of the World Health Organization (WHO), in 2018, there were about 2.1 million new cases of female breast cancer in the world and 1.8 million deaths, accounting for nearly a quarter of female cancer cases, which is the highest mortality of female cancer in the world [1, 2].

Breast cancer seriously endangers the physical and mental health of women all over the world, so the treatment of breast cancer is very important. Many traditional techniques, including tissue biopsy, CT imaging, magnetic resonance imaging (MRI), and magnetic resonance mammography (MRM), are important tools for cancer diagnosis [3]. As the “gold standard” of cancer detection, the pathological examination can effectively diagnose malignant

tumors, but there are still major challenges in pathological imaging and clinical diagnosis. On the one hand, the heterogeneity in the process of disease, tissue preparation, and staining leads to appearance instability, which makes it difficult to obtain accurate information in histopathological images. On the other hand, effective diagnosis methods require many professional doctors or experts to carry out long-term inspections, and some remote areas or underdeveloped areas cannot adopt pathological examination methods due to insufficient doctors and equipment. Due to the above reasons, the diagnosis efficiency of breast cancer is not high, especially for women in remote areas and underdeveloped medical areas [4].

With the rise of computer technology in recent years, technologies such as deep learning and blockchain can be applied to medical imaging and other fields [5]. In particular, some models of deep learning training have surpassed human experts. In recent years, people have proposed and

realized the application of computer vision in computer-aided diagnosis [6, 7], and the application of ultrasound and pathological images in disease-aided diagnosis has become very extensive [8]. Therefore, it is an effective method to collect histopathological images of breast tissue through microscopic imaging and then check the histopathological images through deep learning methods. Breast cancer detection methods are divided into detection, classification, and segmentation. The analysis results are used for auxiliary diagnosis, which can effectively help doctors and experts to diagnose, thereby increasing the diagnosis rate [7, 9].

There are the following problems. On the one hand, the difficulty in making breast cancer pathological images results in fewer samples, which is not conducive to the use of more complex convolutional neural networks. On the other hand, the existing convolutional neural networks are not suitable for such medical images. Feature information extraction is not sufficient. Starting from the above problems, this paper uses deep learning to train and diagnose breast cancer pathological images from the perspective of delay in breast cancer diagnosis and low diagnosis accuracy, using traditional classic convolutional neural networks.

The contributions of the paper are as follows:

- (1) An auxiliary diagnosis model based on multimodel fusion convolutional neural network (VIRNets) is proposed, which achieves an average accuracy of 98.26% for the diagnosis of breast cancer pathological images
- (2) The adopted transfer learning strategy allows the model to extract shallow features by using the parameters of traditional natural image features, reducing model training time and actual parameters

## 2. Materials and Methods

**2.1. Convolutional Neural Networks (CNNs).** Deep learning was proposed by Hinton and Salakhutdinov [10] in 2006. It is a complex machine learning that learns information from the internal laws and representation levels of data and performs tasks such as image classification and speech recognition through the learned information. Convolutional neural network (CNNs) is a multilayer neural network suitable for image classification, segmentation, or detection. In recent years, it has developed rapidly through continuous optimization of the structure. It inherits the unique feature extraction capabilities of deep learning. Because of the advantages of partial connection, weight sharing, and down-sampling, compared with fully connected neural networks, in a network structure with a large amount of data, it can effectively reduce the number of parameters, speed up the calculation, reduce the complexity of the calculation, and reduce computing memory. CNNs include convolutional layer, activation function, pooling layer, fully connected layer, and softmax. The convolutional layer is used to extract image features. It is calculated as follows:

$$g(i) = \sum_{x=1}^m \sum_{y=1}^n \sum_{z=1}^p a_{x,y,z} \times w_{x,y,z}^i + b^i, \quad i = 1, 2, \dots, q, \quad (1)$$

where  $i$  is the  $i$ -th convolution kernel,  $g(i)$  is the feature map learned by the  $i$  convolution kernel,  $a$  is the input data,  $b$  is the bias of the convolution kernel, and  $x, y, z$  are the input data dimension.

After the convolution operation, the nonlinear activation function can make the neural network better learn nonlinear features and can also capture more complex feature information models. Among them, the ReLU function is a one-sided suppression piecewise linear function. The output of the negative part is 0, and the positive remains unchanged so that the model can more effectively dig out the relevant features and fit the training data. Its function analytical formula is as follows:

$$y(i) = f(g(i)) = \max\{0, g(i)\}, \quad i = 1, 2, \dots, q. \quad (2)$$

The role of the activation function is to map all nonzero values to 0, and values greater than 0 remain unchanged. In equation (2),  $i$  represents the input data of the ReLU function and  $f(g(i))$  represents the output data of the ReLU function.

Through the feature extraction of the convolutional layer, the obtained feature map is input to the pooling layer to select features and filter information. The pooling layer performs a convolution operation similar to the convolution layer. The pooling operation is downsampling. This process can retain important pairs of feature information without changing the number of feature maps. In this way, the model reduces spatial information to obtain better computing performance and reduces the risk of model overfitting. The pooling layer is divided into maximum pooling and average pooling calculation as follows:

$$p_{\max 0}^{l(i,j)} = \max_{(j-1)w < t < jw} \{a^{l(i,t)}\}, \quad j = 1, 2, \dots, q, \quad (3)$$

$$p_{\text{avg}0}^{l(i,j)} = \frac{1}{(j-1)w} \sum_{t=(j-1)w+1}^{jw} a^{l(i,t)}, \quad j = 1, 2, \dots, q, \quad (4)$$

where  $a^{l(i,t)}$  represents the  $t$ -th neuron in the  $i$ -th feature map in the  $l$ -th layer,  $w$  is the width of the convolution kernel, and  $j$  is the  $j$ -th pooling kernel.

The fully connected layer feeds back the obtained features to the prepredicted image and converts the previous two-dimensional image into a one-dimensional vector.

Finally, the probability calculation is performed through the softmax layer, and the final category output is obtained. The softmax function converts the input into a probability value ranging from 0 to 1, and the sum of all probability values is 1. The calculation expression is as follows:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_c e^{z_c}}, \quad (5)$$

where  $z_i$  is the output value of the  $i$ -th node and  $C$  is the number of output nodes, that is, the number of classification categories.

## 2.2. Selected Networks

**2.2.1. VGG16 Model.** The VGG [11] model won the championship in the ImageNet positioning competition in 2014 and won second place in the classification task competition in the same year. The reason for the success of the VGG network is the small convolution kernel. The small convolution kernel refers to the  $3 \times 3$  convolution kernel used in VGG, and some smaller  $1 \times 1$  convolution kernels are also used. The reasons for using such a small convolution kernel stack to replace a large convolution kernel are as follows: first, the reduction of network parameters greatly simplifies the complexity and calculation of the model, and second, the increase of the convolution layer means more activation features.

**2.2.2. InceptionV3 Model.** InceptionV3 [12] is the third version of GoogLeNet. The inception module is shown in Figure 1. The input tensor extracts feature through several convolution branches of different sizes and then combines the branch features. This connection method obtains more features and restricts the expansion of characteristic dimensions. The inception structure has two characteristics. The first uses different sizes of convolution kernels, and the second uses different padding to ensure that the size of the feature map output by each branch is the same, which facilitates the direct fusion of features.

**2.2.3. ResNet50 Model.** The ResNet50 model [13] is a deep residual model, which was proposed in 2015 to solve the problem that the accuracy of the training set decreases with the deepening of the network. At the same time, it has the characteristics of easy optimization and a small amount of calculation. The ResNet model uses the shortcut connection cross-layer connection transfer method. The residual connection module is shown in Figure 2. This connection can merge the features after convolution processing with the input features without increasing the parameters and computational complexity. The  $1 \times 1$  convolution layer is used to reduce the dimensionality. In the module connection process shown in Figure 2, the identity mapping about  $x$  is added, and the function relationship is  $F(x) = H(x) - x$ , where  $H(x)$  is the output of the residual network and  $F(x)$  is the output after intermediate convolution processing.

**2.3. Transfer Learning.** Transfer learning [14] refers to the use of analogy, using existing machine learning methods and trained model parameters to solve related practical problems. Related issues refer to different fields but have certain relevance. The purpose is to use knowledge transfer to solve the situation in which there are few label samples or even real conditions in various situations that lead to the absence of label samples, such as medical data. The basic idea is to use

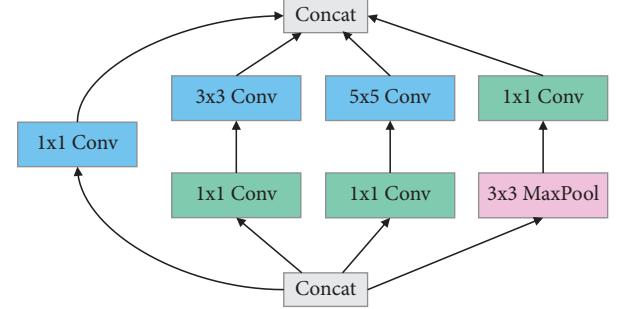


FIGURE 1: Inception module.

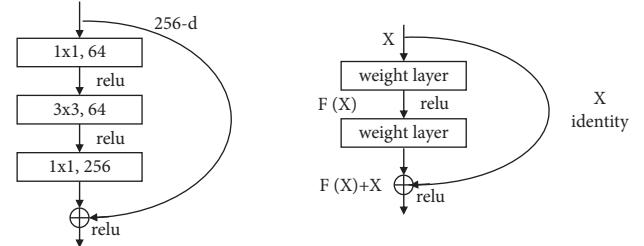


FIGURE 2: ResNet residual module and ResNet function mapping.

pretraining to use the weight values trained on a large dataset as the initial value of the network structure parameters and transfer them to the actual problem dataset for retraining and fine-tuning the network parameters. Some basic features are learned in pretraining, such as color and border features. This method can improve accuracy and save time. In this paper, the idea of transfer learning is adopted in the recognition of benign and malignant breast cancer pathological images. The main reasons are as follows: (1) although the hospital can collect many pathological images of breast cancer, most of the data is of low image quality due to noise, and the amount of data available is small. If a more complex network such as a convolutional neural network is directly used for training, it will result in fewer deep features extracted by the network, which is not conducive to the training of the network, and the diagnosis rate of the network is low. (2) We use VGG16, InceptionV3, and ResNet50 with deep network layers and many parameters. If we start training from scratch, it will make the training time very long, and the performance will be low, and using these classic network parameters trained on ImageNet to directly extract the features of the pathological image and then using the custom classifier for training not only speed up the network training speed and reduce the amount of calculation but also can improve the classification accuracy of the network. Based on the above reasons, this article uses the parameters of the pretrained model of the three basic networks on ImageNet to extract features of breast cancer pathology in the breast cancer diagnosis program.

**2.4. Dataset.** The paper uses BreakHis [15], a breast cancer dataset constructed by Spanhol in cooperation with the P&D Laboratory in Paraná, Brazil. The microscopic images of breast tumor tissues from 82 patients (including 24 benign

patients and 58 malignant patients) were collected. Around 7,909 photos were taken. Among them, 2,480 are images of benign breast tumors, and 5,429 are images of malignant breast cancer; the image size is  $700 \times 460$ ; and each image has three RGB channels. All disease types include 40 times, 100 times, 200 times, and 400 times zoom categories. Figure 3 shows images of breast tumors at each magnification, where  $B$  represents benign and  $M$  represents malignant. Table 1 lists the number of benign and malignant tumors under different magnifications.

## 2.5. Data Preprocessing

**2.5.1. Data Normalization.** The data standardization process limits the pixels in the image to a uniform standard value range. The main function is to convert to standard mode to prevent the influence of radiation transformation and reduce the influence of geometric transformation and accelerate the speed of gradient descent to find the optimal solution. The formula for normalizing the data is as follows:

$$\text{norm} = \frac{x_i - \min(x)}{\max(x) - \min(x)}, \quad (6)$$

where  $x_i$  represents the pixel value of the image and  $\min(x)$  and  $\max(x)$  represent the maximum and minimum values of image pixels, respectively.

**2.5.2. Data Enhancement.** Deep learning requires a large amount of data as the support for training the network, but it is difficult to prepare enough data in practice, and due to the imbalance of the amount of data, the network will become too strict, causing the network model to overfit. In order to solve the problem of overlearning of malignant pathological images, data enhancement is needed [16, 17]. Randomly crop the pathological image into the same pixel by random cropping, which can enhance the data image hundreds of times. However, the similarity of key parts may be too high, and the actual effect is not good. At the same time, other data enhancement methods are used to expand the size of the training samples, mainly using spatial geometric transformations, including inversion, random cropping, random rotation, random scaling, deformation, and so on, to perform an affine transformation on the image. The paper randomly crops; rotates; translates; mirrors the image horizontally; changes the brightness, contrast, and saturation of the image; and randomly adds Gaussian noise and speckle noise.

Random cropping the original image is randomly cropped based on the ratio range of  $[0.8, 1]$  of the original image, and then the cropped image is adjusted to  $(224, 224)$ . The formula is as follows:

$$\begin{aligned} x &\leftarrow \text{random}(0.8 \times W, W) - 0.8 \times W, \\ y &\leftarrow \text{random}(0.8 \times H, H) - 0.8 \times H, \\ J(x, y) &= I(x: x + 0.8 \times W, y: y + 0.8 \times H). \end{aligned} \quad (7)$$

Random rotation: the original image is randomly rotated along the center of the image; the rotation range is  $\pm 45^\circ$ ; and

the incomplete part is filled with linear interpolation. The formula is as follows:

$$\theta \leftarrow \text{random}(-45, 45),$$

$$J(x, y) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \left( I(x, y) + \begin{bmatrix} 2^{-1}W \\ 2^{-1}H \end{bmatrix} \right) - \begin{bmatrix} 2^{-1}W \\ 2^{-1}H \end{bmatrix}. \quad (8)$$

Random translation: the original image is randomly translated in the horizontal and vertical directions; the translation range is 10% of the original image size; and the incomplete part is filled with interpolation. The formula is as follows:

$$w \leftarrow \text{random}(0, 44) - 22,$$

$$h \leftarrow \text{random}(0, 44) - 22, \quad (9)$$

$$J(x, y) = I(x, y) + \begin{bmatrix} w \\ h \end{bmatrix}.$$

Random horizontal mirroring: the original image is mirrored horizontally with a probability of 0.5. The formula is as follows:

$$J(x, y) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} I(x, y) + \begin{bmatrix} W \\ 0 \end{bmatrix}. \quad (10)$$

The brightness, contrast, and saturation of the image are randomly changed by  $\pm 10\%$ . Gaussian noise and speckle noise are randomly added.

**2.6. Auxiliary Diagnosis Model Based on VIRNets.** In this paper, in order to avoid insufficient information of breast cancer pathological images extracted by a single model, network expression ability is improved, and better image classification results are achieved. We propose a new fusion convolutional neural network-VIRNets, which uses the idea of migration learning to perform feature fusion on the three basic convolutional neural network models proposed in Section 2.2. These three networks have evolved step by step according to the history of deep learning and have their own advantages. Among them, VGG16 uses a small convolution kernel instead of a large convolution kernel, which greatly reduces network parameters and the amount of model calculation; adding more convolution layers to increase the activation function and enhance the nonlinear expression ability of the network model. InceptionV3 uses different sizes of convolution kernels to extract different features for fusion and uses BatchNormalization to solve the problem of gradient disappearance. Finally, ResNet50 uses the interleaving module to solve the problem of slow network convergence and gradient disappearance caused by the deepening of the number of network layers and can effectively extract deep features to improve performance. Due to the excellent performance of these three networks on the ImageNet dataset, we can use the weight parameters trained on ImageNet to extract the features of breast cancer pathological images using the idea of migration learning, which not only can reduce the training parameters and training

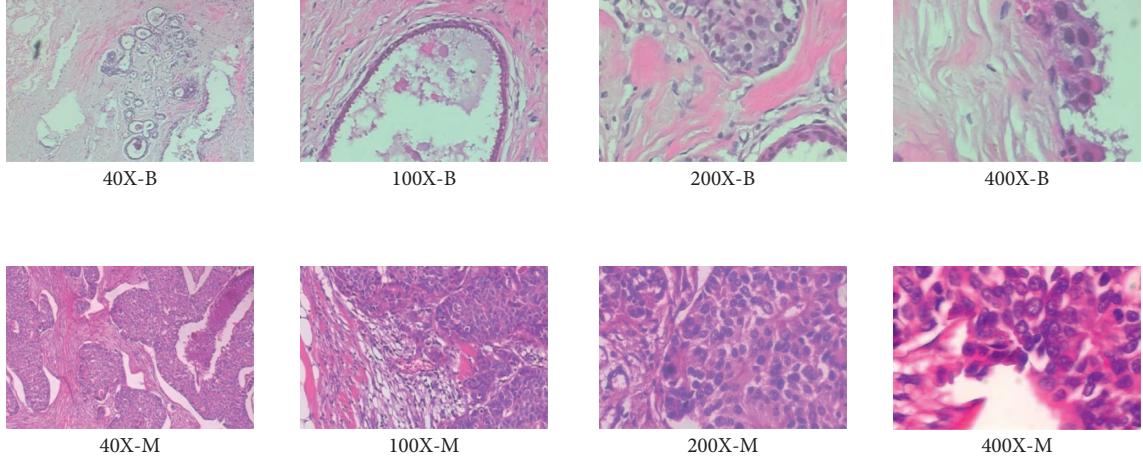


FIGURE 3: Pathological image of breast tumor.

TABLE 1: Statistics of BreakHis dataset.

Magnification	Benign	Malignant	Total
40X	652	1,370	1,995
100X	644	1,437	2,081
200X	623	1,390	2,013
400X	588	1,232	1,820
Total images	2,480	5,429	7,909

time but also can effectively improve classification performance. The core of VIRNets is to extract key features from different network types and then use feature fusion to combine the features extracted from the three parallel networks to learn more complementary features, improve network expression capabilities, and improve the accuracy of a breast cancer diagnosis.

Figure 4 is the structure diagram of the three network fusion models, from left to right are the VGG16, InceptionV3, and ResNet50 convolutional neural network models. The model inputs pathological images, and then in the feature extraction part, three pretrained models loaded with ImageNet weights are used for feature extraction. After the final convolutional layer passes global pooling, the Concat layer is merged. We add noise disturbance to the fusion feature to improve the robustness of the model. After extracting the features of the pathological image, it uses a fully connected layer for dimensionality reduction and adds Dropout and BatchNormalization to prevent overfitting. Finally, it is classified by softmax. This paper uses the cross-entropy loss function to calculate the loss value, thereby updating the weight parameters of the model. The formula of the cross-entropy loss function is as follows:

$$\text{Loss} = -1 \cdot \sum_{i=1}^{\text{outputsize}} -y_i \cdot \log \hat{y}_i, \quad (11)$$

Where  $y_i$  is the true value of the  $i$ -th data in the dataset and  $\hat{y}_i$  is the predicted value of the  $i$ -th data. The loss value is calculated for a batch of data passed through the network in each iteration.

The overall processing flow of the auxiliary diagnosis in this paper is shown in Figure 5. To obtain breast cancer pathological images from the above dataset, first the data are preprocessed, including data enhancement, data normalization, and so on [18], and then three pretraining networks are used to extract the features of the pathological images; the extracted features are fused and then sent. The custom classifier is trained for the next step, and the calculated loss value is backpropagated to make the model training converge. The final output of the classification model is a  $1 \times 2$  vector, namely,  $P_0$  and  $P_1$  in Figure 5, which represent the probabilities of benign and malignant, respectively. Finally, the two data values are compared, and the category corresponding to the number with the larger probability is selected as the final classification result.

### 3. Results and Discussion

**3.1. Experimental Setup.** The experiments in the paper are finished on Google's open-source TensorFlow deep learning framework. The experimental hardware environment is that the operating system is Ubuntu 18.04, and the NVIDIA GTX 1080TI GPU is used for accelerated processing. The BreakHis dataset is divided into benign and malignant binary data under 40X, 100X, 200X, and 400X magnifications to conduct experiments. The data are divided into training set, validation set, and testing set according to the data ratio of 6:2:2. Sixty percent of the data is used for training; 20% of the data is used for model optimization and verification, and the remaining 20% is used for testing model performance. The number of training data iterations is set to

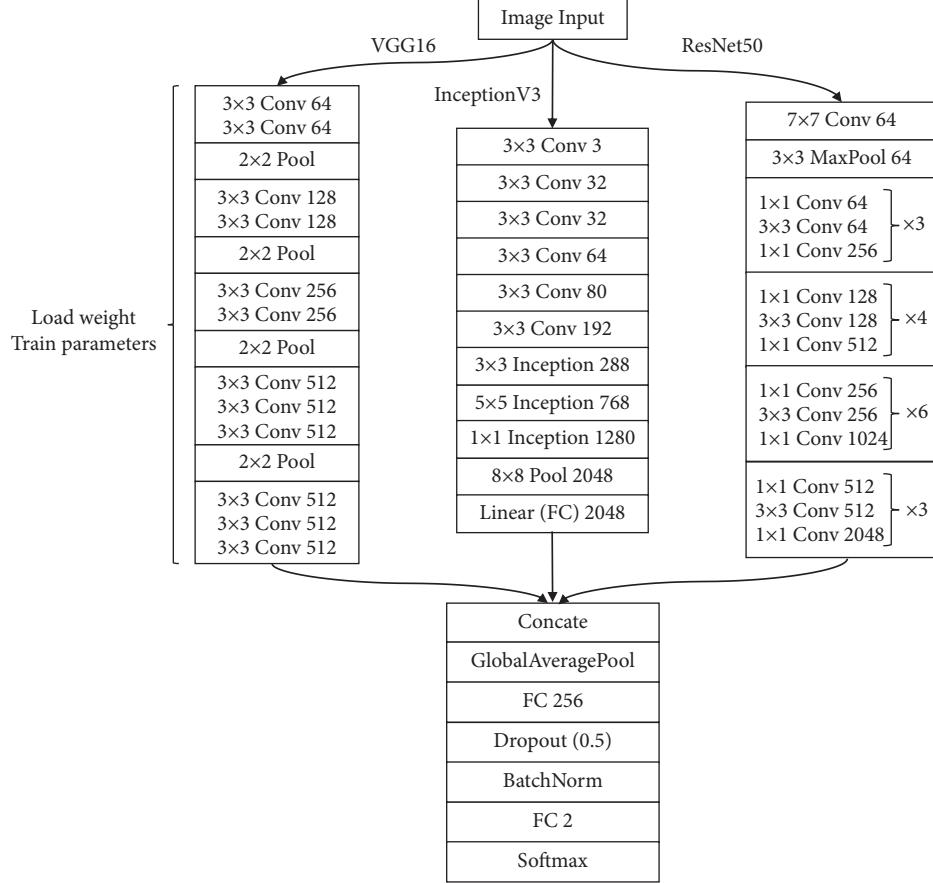


FIGURE 4: VGG16, InceptionV3, and ResNet50 fusion model.

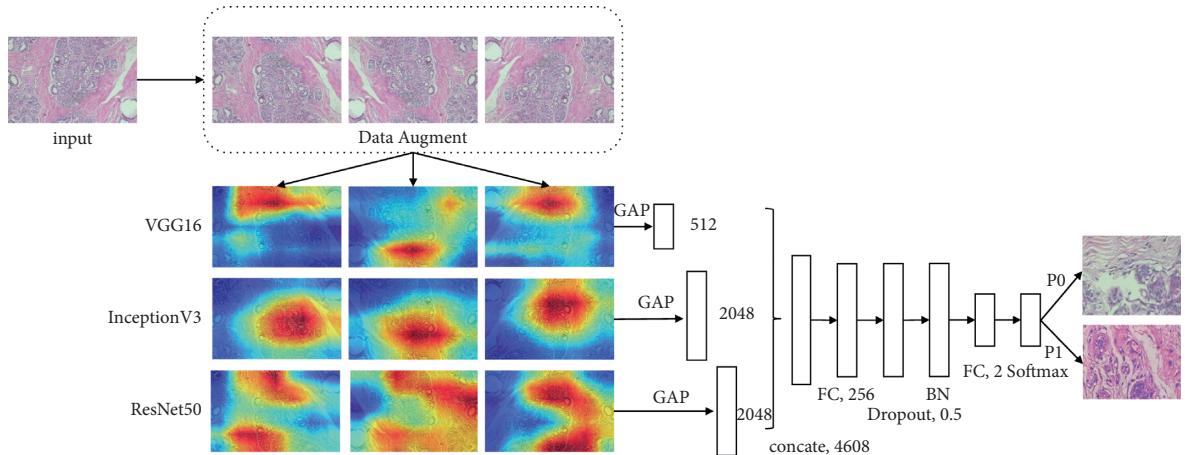


FIGURE 5: Process flow of auxiliary diagnosis model.

100 rounds, and the batch size is set to 32. The Adam optimizer is used to optimize the fusion convolutional neural network, in which beta\_1 is set to 0.9 and beta\_2 is set to 0.999. Using the operation of learning rate attenuation, the learning rate is initialized to 1e-4, and the accuracy of the verification set is observed. If the accuracy of the verification set is not improved for five times, the learning rate is

reduced, and the minimum is 1e-7 to ensure that the initial stage of the model can approach the global optimal value with not too large fluctuations.

**3.2. Evaluation Methods.** The goal of this paper is to classify breast cancer pathological image data. After training the

TABLE 2: Comparison of the results of various indicators under different magnifications.

Enlarge size	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 – score (%)
40x	VGG16	94.39	92.99	94.22	93.56
	InceptionV3	92.20	97.73	89.85	90.71
	ResNet50	84.15	81.76	85.70	82.79
	VIRNets (ours)	<b>99.02</b>	<b>98.86</b>	<b>98.97</b>	<b>98.90</b>
100X	VGG16	86.19	83.51	86.61	84.63
	InceptionV3	96.67	95.78	97.93	96.74
	ResNet50	90.95	88.64	91.75	89.86
	VIRNets (ours)	<b>97.62</b>	<b>96.43</b>	<b>98.28</b>	<b>97.27</b>
200X	VGG16	83.66	80.85	81.62	81.21
	InceptionV3	97.03	98.12	97.26	97.67
	ResNet50	91.58	91.43	88.68	89.88
	VIRNets (ours)	<b>98.02</b>	<b>97.02</b>	<b>98.56</b>	<b>97.73</b>
400X	VGG16	86.41	84.84	83.9	84.34
	InceptionV3	95.56	94.27	96.34	95.17
	ResNet50	83.15	80.73	82.77	81.50
	VIRNets (ours)	<b>98.37</b>	<b>97.95</b>	<b>98.36</b>	<b>98.15</b>

model, the test set is tested by loading the trained model weight, and the number of correct predictions for each category is counted according to common classification indicators, including precision, recall, F1-score, and accuracy.

First of all, the accuracy rate is based on the percentage of the number of samples correctly predicted to the total sample size. The specific formula is as follows:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}. \quad (12)$$

The second is the accuracy rate, also called the precision rate, the percentage of the true positive samples among the results predicted by the model to be positive samples. The formula is calculated as follows:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (13)$$

The recall rate is different from the precision rate. Percentage of predicted positive samples in all positive samples. The calculation formula is as follows:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (14)$$

The last indicator is F1-score, which is mainly a method proposed for the shortcomings of precision rate and recall rate, which is a compromise method. The

$$F1 = \frac{2 \times P \times R}{P + R}. \quad (15)$$

In these definitions, the TP, TN, FP, and FN scores represent the number of true positives, true negatives, false positives, and false negatives. In addition, we use the ROC curve and the area under the curve AUC to evaluate the performance of the classifier.

**3.3. Result Analysis.** This paper designs experiments for the categories of benign and malignant breast cancer, and the basic models used are VGG16, InceptionV3, and

ResNet50. The experimental results consist of two parts. First, we compare the results of the basic network model and the VIRNets network using the various evaluation indicators described in Section 2.4 at different magnifications (40X, 100X, 200X, and 400X). Second, we compare VIRNets with models that have used the same dataset in recent years.

It can be seen from Table 2 that in the classification of benign and malignant breast cancer under different magnifications, the classification accuracy of a single model is low, which is due to the fact that effective features cannot be extracted with a small amount of data. The classification accuracy, precision, recall, and F1-score of our proposed fusion convolutional neural network model VIRNets under different magnifications are much higher than the basic model. Among them, the highest accuracy rate is 99.02 at 40X.

Figure 6 shows the comparison of ROC curves of VIRNets and its basic network at different multiples. We can observe from the figure that the AUC values corresponding to VIRNets at different multiples exceed the AUC value of the basic network model to varying degrees. Among them, the AUC value below 40x is the highest, reaching 0.9965, which is larger than other basic models. This shows that our fusion model network can effectively extract feature information, and the network has high robustness, which can effectively improve the accuracy of breast cancer classification. Relatively speaking, the AUC value of the VGG16 network is significantly lower, indicating that the feature proposal for breast cancer images is not sufficient, and there is a certain gap between the feature fusion model.

Figure 7 shows the confusion matrix of our method for the two classifications of breast cancer under four magnifications. It can be seen from the figure that our model is generally more accurate for the classification of benign and malignant breast cancer, and the number of classification errors under each magnification is not more than 5, which fully shows that the method we proposed is effective and can be two classification problems applied to breast cancer.

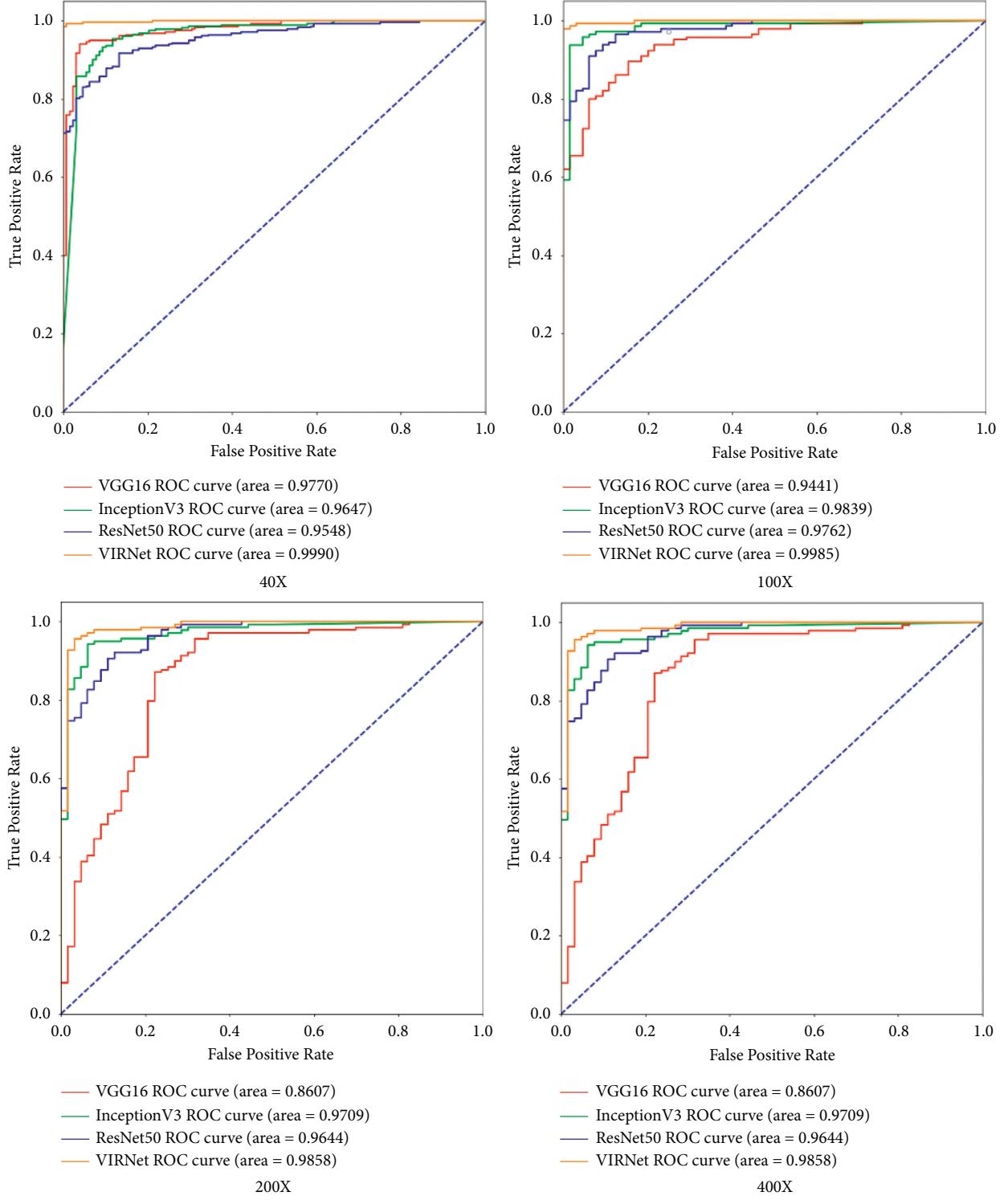


FIGURE 6: ROC curves of four types of networks under different magnifications.

Table 3 shows the comparison of different architectures for breast cancer classification under different magnifications using the BreakHis dataset and the VIRNets model in this paper. It can be seen from the table that the correct rate of VIRNets in this article is very high under different multiples. The correct rate under 40X and 400X is higher than that of other structures, especially under 40X, the

accuracy rate reaches 99.02%. And under 100X and 200X, there is not much difference in accuracy with other structures. This shows that our model can be effectively applied to actual scenarios to assist doctors in diagnosis.

Compared with the single-model network, the fusion network model we proposed has greatly improved its accuracy. Regarding the amount of parameters and running

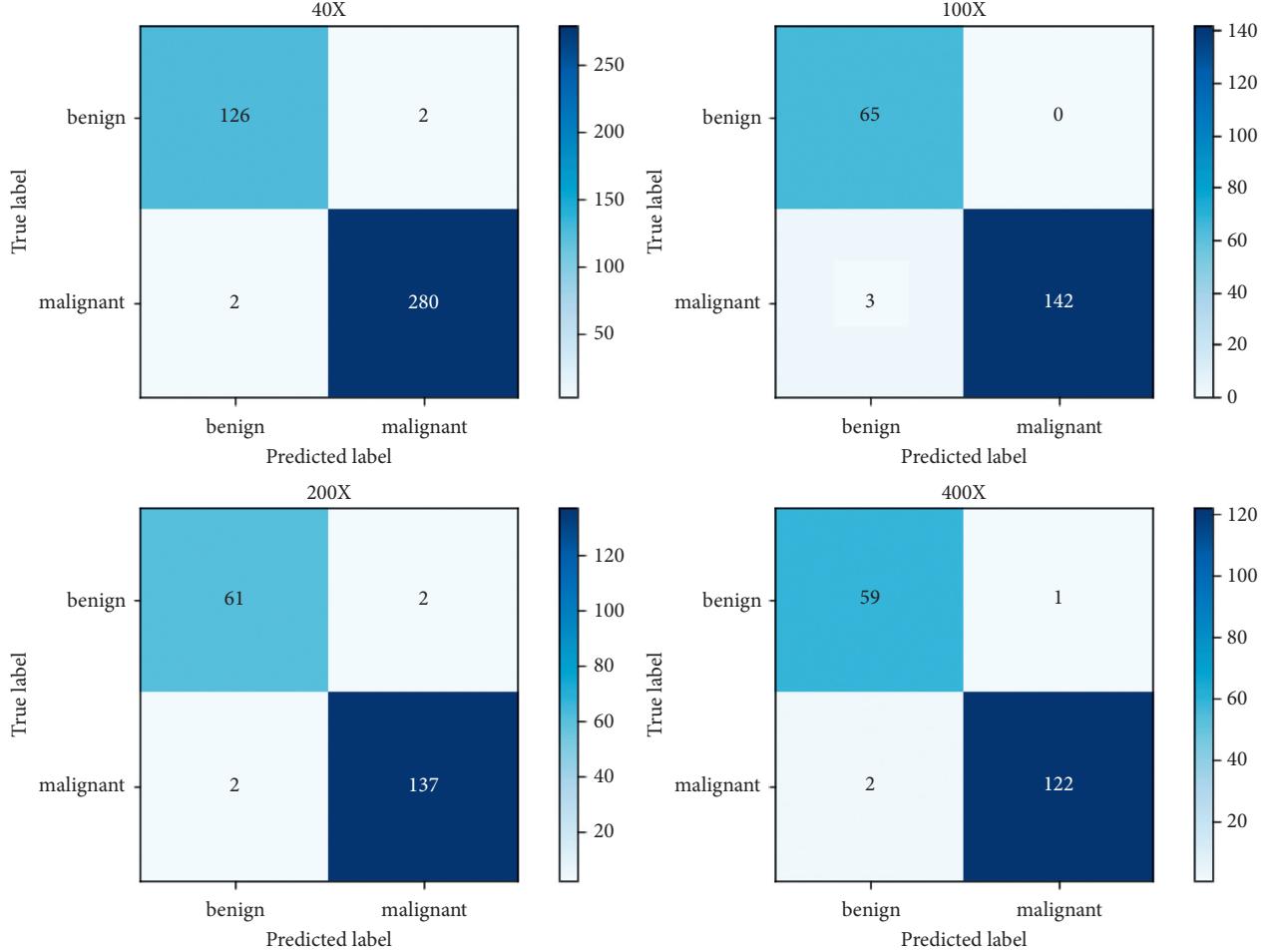


FIGURE 7: Confusion matrix of VIRNets at different magnifications.

TABLE 3: Comparison of the correct rate of the VIRNets model using the BreaKHis test set and different architectures.

Structure	Accuracy under different magnifications				Average accuracy (%)
	40X (%)	100X (%)	200X (%)	400X (%)	
CSDCNN [19]	95.9	96.9	96.7	94.9	96.1
IRRCNN [20]	97.95	97.57	97.32	97.36	97.55
Bi-LSTM [21]	96.2	97.2	97.1	95.4	96.48
BreastNet [22]	97.99	97.84	98.51	95.88	97.55
Sharma et al. [23]	97.4	98.6	97.7	96.8	97.63
VIRNets (ours)	<b>99.02</b>	<b>97.62</b>	<b>98.02</b>	<b>98.37</b>	<b>98.26</b>

time, as the network increases, the parameter amount and running time of our proposed model are more than that of a single network, but we use the idea of transfer learning in its process and choose the one that performs well in ImageNet. The initial weight of the network training, using it to extract the features of breast cancer pathological images, helps us reduce a certain amount of training time and parameters, making the training of the model relatively simple and fast. Therefore, it is worthwhile to replace the increase in classification accuracy with the increased parameter amount and running time.

#### 4. Conclusions

This paper starts from the problem of the low accuracy of traditional medical equipment for breast cancer diagnosis, combined with the current popular deep learning methods for research. A variety of convolutional neural network model fusion methods are used to fuse breast cancer features extracted from different networks, which effectively improve the accuracy of the diagnosis of benign and malignant breast cancer. In the end, the classification result reached an accuracy of 98.26%. It shows that our model can effectively

solve the problems of doctors and incomplete facilities in some remote areas and can use computer-assisted diagnosis to help breast cancer patients get timely treatment. However, we can find that regardless of the public dataset and the real scene, there are far more malignant breast pathological images than benign breast images, which leads to the tendency of the network model to learn the characteristics of malignant breast pathological images. Our next plan is to use generative adversarial networks (GAN) to expand the dataset with more image samples or use undersampling methods (such as short-range undersampling) and use it to train better models to improve performance.

## Data Availability

The data can be obtained through an official application because the authors have no right to publicly provide the data. Visit <https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] F. Bray, J. Ferlay, and I. Soerjomataram, “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[,” *J. CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] C. P. Wild, E. Weiderpass, and B. W. Stewart, *World Cancer Report: Cancer Research for Cancer Prevention*, pp. 23–33, International Agency for Research on Cancer, Lyon, France, 2020.
- [3] B. Prabhakar, P. Shende, and S. Augustine, “Current trends and emerging diagnostic techniques for lung cancer,” *Bio-medicine & Pharmacotherapy*, vol. 106, pp. 1586–1599, 2018.
- [4] B. M. C. Silva, J. J. P. C. Rodrigues, and A. Ramos, “A mobile health system to empower healthcare services in remote regions,” in *Proceedings of the 2019 IEEE international conference on E-health networking, application & services (HealthCom)*, pp. 1–6, IEEE, Dalian, China, October 2019.
- [5] Y. Xu, C. Zhang, Q. Zeng, G. Wang, J. Ren, and Y. Zhang, “Blockchain-enabled accountability mechanism against information leakage in vertical industry services,” *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1202–1213, 2021.
- [6] P. Guillén-Rondon, M. Robinson, and J. Ebalunode, *Breast Cancer Classification: A Deep Learning Approach for Digital pathology*, pp. 33–40, Springer, Berlin, Germany, 2018.
- [7] Z. Hu, J. Tang, and Z. Wang, “Deep learning for image-based cancer detection and diagnosis— A survey,” *Pattern Recognition*, vol. 83, pp. 134–149, 2018.
- [8] D. R. Martin, J. A. Hanson, and R. R. Gullapalli, “A deep learning convolutional neural network can recognize common patterns of injury in gastric pathology[J],” *Archives of pathology & laboratory medicine*, vol. 144, no. 3, pp. 370–378, 2020.
- [9] D. Komura and S. Ishikawa, “Machine learning approaches for pathologic diagnosis,” *Virchows Archiv*, vol. 475, no. 2, pp. 131–138, 2019.
- [10] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [11] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, <http://arxiv.org/abs/1404.3978>.
- [12] C. Szegedy, V. Vanhoucke, and S. Ioffe, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, Las Vegas, NV, USA, June 2016.
- [13] K He, X Zhang, and S Ren, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [14] F. Zhuang, Z. Qi, and K. Duan, “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.
- [15] F. A. Spanhol, L. S. Oliveira, and C. Petitjean, “A dataset for breast cancer histopathological image classification,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2016.
- [16] G. Litjens, T. Kooi, and B. E. Bejnordi, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, no. 9, pp. 60–88, 2017.
- [17] F. Shahidi, S. M. Daud, and H. Abas, “Breast cancer classification using deep learning approaches and histopathology image: a comparison study,” *IEEE Access*, vol. 8, pp. 187531–187552, 2020.
- [18] Y. Xu, J. Ren, Y. Zhang, C. Zhang, B. Shen, and Y. Zhang, “Blockchain empowered arbitrable data auditing scheme for network storage as a service,” *IEEE Transactions on Services Computing*, vol. 13, no. 2, pp. 289–300, 2020.
- [19] Z. Han, B. Wei, and Y. Zheng, “Breast cancer multi-classification from histopathological images with structured deep learning model[J],” *Scientific Reports*, vol. 7, no. 1, p. 4172, 2017.
- [20] M. Z. Alom, C. Yakopcic, and M. S. Nasrin, “Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network,” *Journal of Digital Imaging*, vol. 32, no. 5, 2019.
- [21] S. Pratiher, S. Chattoraj, and S. Agarwal, “Grading tumor malignancy via deep bidirectional LSTM on graph manifold encoded histopathological image,” in *Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, IEEE, Singapore, November 2018.
- [22] M. Toğçaçar, K. B. Özkurt, and B. Ergen, “BreastNet: a novel convolutional neural network model through histopathological images for the diagnosis of breast cancer,” *Physica A: Statistical Mechanics and Its Applications*, vol. 545, p. 123592, 2020.
- [23] M. Sharma, R. Verma, and A. Mishra, “A novel approach to classify breast cancer tumors using deep learning approach and resulting most accurate magnification factor,” *Performance Vision Intelligence*, pp. 185–201, Springer, Singapore, 2020.