

## Article

# Breast Histopathological Image Classification Method Based on Autoencoder and Siamese Framework

Min Liu, Yu He, Minghu Wu \* and Chunyan Zeng

School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China; liu\_min@hbut.edu.cn (M.L.); heyu@hbut.edu.cn (Y.H.); cyzeng@hbut.edu.cn (C.Z.)

\* Correspondence: wuxx1005@hbut.edu.cn

**Abstract:** The automated classification of breast cancer histopathological images is one of the important tasks in computer-aided diagnosis systems (CADs). Due to the characteristics of small inter-class and large intra-class variances in breast cancer histopathological images, extracting features for breast cancer classification is difficult. To address this problem, an improved autoencoder (AE) network using a Siamese framework that can learn the effective features from histopathological images for CAD breast cancer classification tasks was designed. First, the inputted image is processed at multiple scales using a Gaussian pyramid to obtain multi-scale features. Second, in the feature extraction stage, a Siamese framework is used to constrain the pre-trained AE so that the extracted features have smaller intra-class variance and larger inter-class variance. Experimental results show that the proposed method classification accuracy was as high as 97.8% on the BreakHis dataset. Compared with commonly used algorithms in breast cancer histopathological classification, this method has superior, faster performance.

**Keywords:** histopathological image classification; autoencoder; Siamese network; multi-scale input



**Citation:** Liu, M.; He, Y.; Wu, M.; Zeng, C. Breast Histopathological Image Classification Method Based on Autoencoder and Siamese Framework. *Information* **2022**, *13*, 107. <https://doi.org/10.3390/info13030107>

Academic Editor: Vincenzo Moscato

Received: 21 January 2022

Accepted: 21 February 2022

Published: 24 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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

Cancer is a major public health problem worldwide. In 2020, the Global Center for Cancer Research released a statistical report on the incidence and mortality of 36 types of cancers in 185 countries. The report states that breast cancer is the most commonly diagnosed cancer, accounting for 11.7% of total cancer cases [1]. Early detection of breast cancer is paramount in reducing the mortality rate. With the development of computer and artificial intelligence technology, CADs can help doctors diagnose breast cancer while improving diagnostic efficiency and accuracy [2–4]. Medical imaging is the most effective method of primary screening for breast cancer. Depending on the imaging modality, the most common types of medical imaging include MRI, low-dose CT, mammography and ultrasound [5,6]. The above methods determine whether a breast is cancerous by analyzing changes in the shape of the breast. Unlike the above methods, pathological diagnosis is mainly based on changes in tissue structure and cell morphology and is the gold standard for the final diagnosis of malignancy. However, digital histopathology image analysis requires an experienced radiologist, and is time-consuming and laborious, while the results are subject to subjective physician interpretation. The prognostic model automatically extracts useful information from the images, without the need for specialist judgement [7]. Thus, the classification of breast pathology images has become a popular research topic in recent years.

Accurate classification of histopathological images of breast cancer can help doctors to develop more rational treatment plans and can have a significant impact on patient prognosis. A great deal of research has been done to achieve accurate histopathological classification of breast cancer [8]. These methods can basically be divided into two types; machine learning methods based on hand-crafted features [9–11], and end-to-end deep learning methods [12–17].

In breast cancer histopathological image classification by machine learning, feature extraction is usually conducted to reduce the amount of data and obtain more easily classified features. For instance, in [18] Wan et al. used Local Binary Pattern (LBP) to construct a method to measure the relationship between a pixel and its surrounding pixels to describe image local texture features, and then used Support Vector Machine (SVM) as a classifier to complete classification. In [19] Nanni et al. obtained texture descriptors using Local Phase Quantization (LPQ). Each descriptor was obtained by changing the LPQ parameters and a SVM was trained for each descriptor. The SVMs were then merged using summation rules. In [20] Sharma et al. presented a CADs based on texture analysis, using Parameter-Free Threshold Adjacency Statistic (PFTAS). After pre-treating the pathology image, the nucleus area was used to extract features and provide texture descriptors, and Random Forest (RF) classifier was then used to obtain the features vector classification for benign and malignant tumor characteristics. A method for classifying breast histopathology images using joint color texture features and a series of combined classifiers was proposed by G et al. The method concatenates six color texture features extracted separately and classifies the concatenated features using multiple classifiers such as SVMs, decision trees (DTs), Nearest Neighbors classifiers (NNCs), discriminant analysis (DA) and Ensemble classifiers, and the final results are obtained after pooled voting with an average recognition rate of 87.53% [21]. However, the automatic classification of breast histopathology images based on hand-crafted features are often subjective. Classifier performance depends on the feature extractor sensitivity, which leads to low final diagnostic results reliability due to the defects of the feature extractor itself.

Compared with traditional methods, deep learning methods adopt end-to-end training mechanisms and have enhanced accuracy while avoiding complex feature extraction. Due to the ultra-high resolution of histopathological images, directly inputting them into the deep neural network is expensive. The usual method is to use sliding window to cut them into many input-image patches. Feng [22] et al. randomly separated patches and assigned original image labels to them. After classifying all the patches, the prediction labels of the separate patches from each image were selected to obtain the image prediction labels. Alom et al. [23] used two methods to select image patches: one was to randomly cut patches from samples, the other was to select non-overlapping and sequential patches. After image augmentation using rotation, flipping, shearing, and translation, the patches were input into an Inception Recurrent Residual Convolutional Neural Network (IRRCNN) model for classification. In [24], Sudharshan et al. proposed a method that combined non-parametric multi-instance learning and Convolutional Neural Networks (CNN), which considered an image patch as an instance and the whole image as a bag. In contrast to other strategies that assign whole image labels to patches, this method chooses to label the lesion areas in the image. Naturally, an image is labeled positive if at least one patch is labeled malignant, and an image is labeled benign if no part of it is labeled malignant.

However, such patch-based methods only consider the information within patches and ignore between patch, or global, information. In [25], Yan et al. proposed a hybrid deep neural network for classification, using the Bi-directional Long Short-Term Memory Network (BLSTM), following feature extraction of input-image patches using the Inception\_v3 network. The features of image patches are time-shared into BLSTM, and the BLSTM long-term dependence is used to learn the information between, and within, patches simultaneously. In [26], a Multi-Scale Input and Multi-Feature Network (MSI-MFNET) model was proposed by Sheikh et al., which extracted different scale features and learned global and texture features at the same time. To evaluate complex biological processes such as multi-cell-type migration and cell-cell interactions, Comes et al. proposed a Multi-scale Generative Adversarial Network that performs image reconstruction at multiple scales to obtain information at different scales for predicting high-quality interlaced video frames. The relationship between frames in such videos is similar to the relationship between image patches [27]. However, histopathological images have characteristics of both large intra-class and small inter-class variances, which makes classification difficult. Moreover,

all the deep learning methods mentioned above require significant time to train the model. Although computer hardware has improved greatly, some hospitals in remote areas still lack decent computers to assist in diagnosis.

In this paper, we propose a novel AE, constrained by the Siamese framework, as a feature extractor and connect the AE layers with a SoftMax classifier to construct a deep neural network for breast cancer histopathological image classification. The contributions of this study are as follows:

1. As the neural network receptive field is fixed, it cannot use global and local information simultaneously during the training process, therefore multi-scale images are used as the input images. In this way, the model can learn both coarse, and fine, scale features simultaneously, and attain better classification ability.

2. We design the AE + SoftMax model structure, in which AE extracts the effective inputted features by copying the input to the output through learning. Consequently, effective feature extraction can be conducted, and feature selection subjectivity is also avoided. Moreover, the model training time and memory consumption is far less than that of CNN.

3. To overcome the shortcomings of histopathological image small inter-class, and large intra-class, variances, a Siamese framework was used to constrain the pre-trained AE. The improved contrasting function was used as a loss function to ensure the feature similarity of same class images was high, and that of different class images was low.

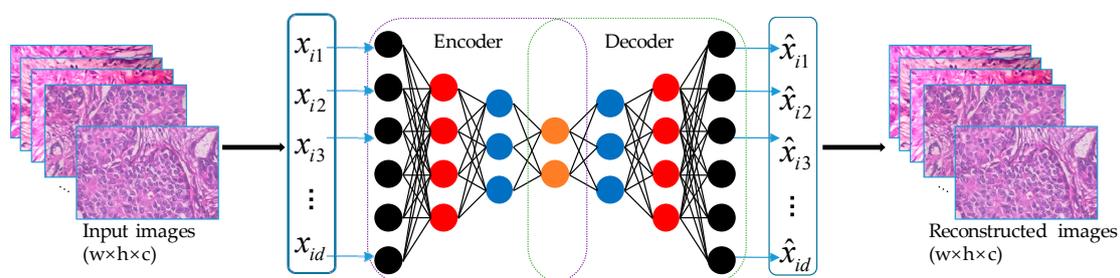
## 2. Materials and Methods

### 2.1. Preliminaries

#### 2.1.1. Autoencoder

An autoencoder is a symmetric neural network that trains the network in an unsupervised way. Specifically, it learns the effective features of the input images by forcing the inputted image to be the same as the reconstructed image. As is shown in Figure 1, the size of each inputted image is  $w \times h \times s$ , and each inputted image is re-shaped into one-dimensional vectors  $x(i)$ .  $x(i) = (x_{i1}, x_{i2}, \dots, x_{id})$ ,  $d = w \times h \times c$ , and let  $X = \{x(1), x(2), \dots, x(n)\}$  be a given training set, where  $x(i)$  is the input vector for the  $i$ -th image. The autoencoder encodes the input as a hidden representation of the latent space and then decodes it back to its original space. The effective features of training set  $X$  can be learned by minimizing the error between input data  $x(i)$  and reconstructed data  $\hat{x}(i)$ . The mean square error provides a metric to measure the similarity between the input and reconstructed data. The loss function is expressed as:

$$L(x(i), \hat{x}(i)) = \frac{1}{2n} \sum_{i=1}^n \|x(i) - \hat{x}(i)\|^2 \tag{1}$$



**Figure 1.** The architecture of Autoencoder. Autoencoder learns the effective features of data by minimizing the error between the input data and the reconstructed data.

After training, the decoder part of the autoencoder is removed, leaving only the encoder part as the feature extractor.

### 2.1.2. Siamese Network

A Siamese neural network system is a neural network consisting of two identical neural networks. The two identical networks share parameters such as weights and biases, but the inputted data varies [28]. Through the forward propagation of the neural network, data which is difficult to distinguish in its original space can be dimensionally constrained to improve classification performance.

The architecture of the Siamese network is shown in Figure 2, where  $G_w(\cdot)$  represents a feature extraction neural network,  $f_i$  and  $f_j$  are feature vectors extracted by  $G_w(\cdot)$  from input images  $x(i)$  and  $x(j)$ , respectively. Following feature extraction, the input-image similarity is obtained by calculating the feature vector similarity.

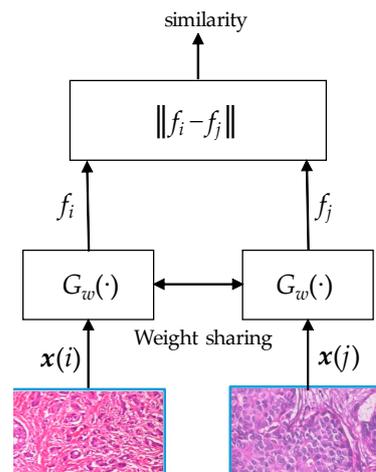


Figure 2. The architecture of Siamese neural network.

### 2.2. Proposed Approach

The proposed model block diagram is shown schematically in Figure 3, and includes the sub-block diagram of the image preprocessing using multi-scale transformation. The autoencoder is pre-trained with multi-scale inputted images, and is then used as the benchmark to build the Siamese networks. Images are inputted into the Siamese network in pairs, and contrastive loss function is used to perform distance constraints on the extracted features. Finally, the extracted features are inputted into SoftMax classifier to obtain the histopathological image benign and malignant classification designations.

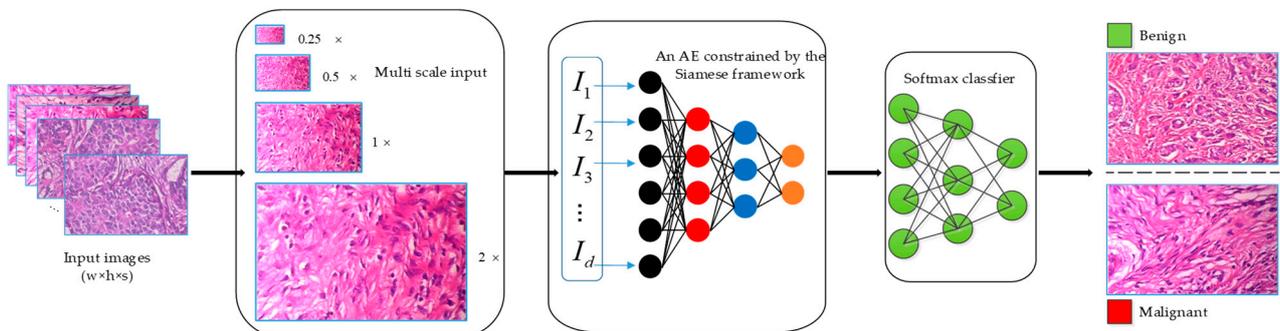


Figure 3. The procedure of the proposed method for breast histopathology image classification.

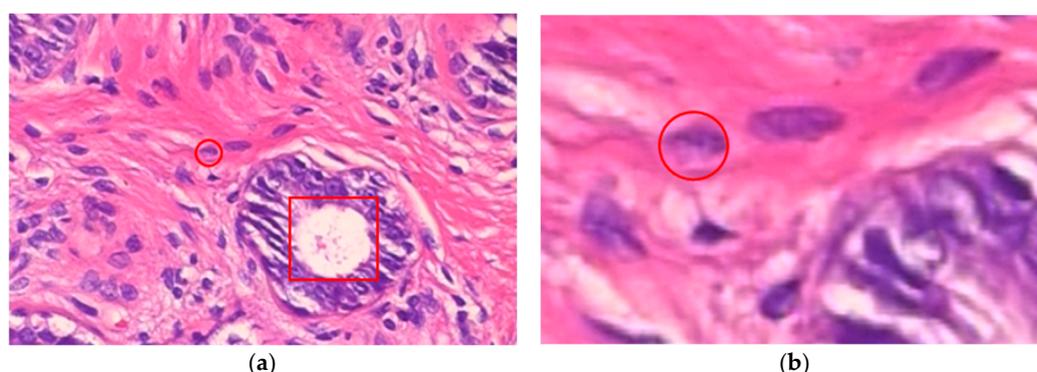
Multi-scale input is used to simultaneously learn the global features and local texture information of histopathology images at different scales. The image is transformed into five different scales ( $0.125\times$ ,  $0.25\times$ ,  $0.5\times$ ,  $1\times$ , and  $2\times$ ), thus enhancing the recognition of important areas in the image and overcoming the limitation of fixed receptive fields in spatial regions at different scales in spatial areas.

### 2.2.1. Multi-Scale Input

Image scale does not refer to the size of the image, but to the degree of blur in the image. For example, the degree of blur of an object viewed from a close distance is different from that of an object viewed from a longer distance. An image moving from a close distance to a long distance also involves the scale of the image becoming larger and larger. When analyzing an unknown scenario with a machine vision system, the computer does not know the size of the object in the image. Therefore, to better classify a particular object of interest, it is necessary to find an image scale description suitable for the object classification.

When extracting image features, if fixed scale feature detection is employed, the detection results will be biased towards this size, and several other scale features will be missed. To detect the same features in images at different scales, detecting and matching images at multiple scales is necessary.

As shown in Figure 4, it is easier to identify rectangular frame features at a coarse-scale, and easier to identify circular frame features at a finer scale. Therefore, multi-scale transformations are used with the input images.



**Figure 4.** (a) show histopathological images at coarse-scale; (b) show histopathological images at finer scale.

In this paper, a Gaussian pyramid is used to process and transform the dataset. The scale space  $L(x,y,\sigma)$  of an image is defined as the convolution operation between the original image  $I(x,y)$  and a 2-dimensional variable scale Gaussian function  $G(x,y,\sigma)$  [29]. Specifically, the scale space form is expressed as:

$$G(x_i, y_i, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma^2}\right) \quad (2)$$

The construction of the Gaussian pyramid involves two steps. First, the image is Gaussian smoothed and second, it is sampled. An example of a Gaussian pyramid is shown in Figure 5.

### 2.2.2. Siamese Framework Based on AE

AE model performs well in feature extraction. However, if  $x(i)$  and  $x(j)$  are neighbors in the input space, the distance between them in the feature space should be small following AE feature extraction. After the pre-training of AE, to reduce the distance between intra-class images and increase the distance between inter-class images, a Siamese framework was built, based on AE, to optimize it.

We used two AEs, with shared weights and bias parameters and different inputs, as the benchmark for the Siamese network. After AE processing the inputted image, its feature vectors were extracted, and feature similarity calculation was conducted at the backend of the Siamese network. The feature similarity extracted from different images is taken as the optimization objective to train the model feature extraction ability, and the

feature distance extracted from the image should be as small as possible within classes, and as large as possible between classes, to enhance classifier performance.

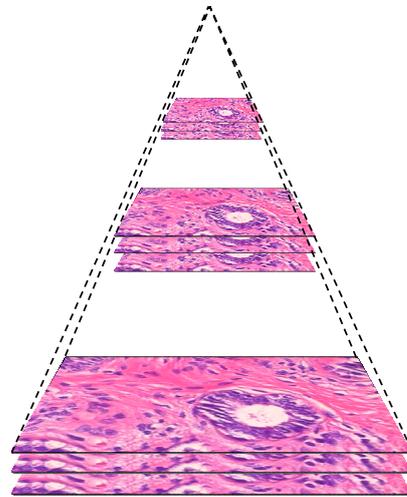


Figure 5. The image pyramid of inputting image.

The architecture of the Siamese framework based on an AE model is shown in Figure 6. After training the AE, encoder layers of the trained AE are employed to obtain representations for the breast histopathology images.

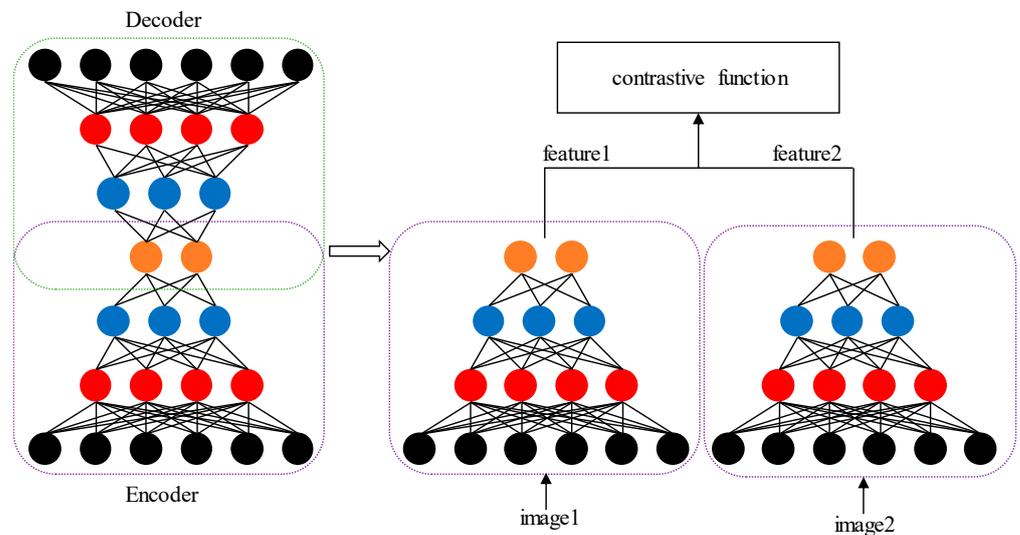


Figure 6. The architecture of AE with five hidden layers at the left side and the architecture of the Siamese framework based on AE model at the right side.

The image is selected to create an image pair  $(x(i), x(j), y(i), y(j), y(ij))$ , where  $x(i)$  and  $x(j)$  represent the different input images,  $y(i)$  and  $y(j)$  are their corresponding labels, respectively.

If  $x(i)$  and  $x(j)$  belong to the same class, the attribute value of  $y(ij) = 1$ , if  $x(i)$  and  $x(j)$  do not belong to the same class,  $y(ij) = 0$ . The Euclidean distance between  $f_i$  and  $f_j$ , where the extracted features of networks with  $x(i)$  and  $x(j)$  as inputs, is then calculated.

The following loss function is adopted to constrain the features extraction, and is expressed as:

$$L(f_i, f_j, y(ij), m) = y_{ij}D^2 + (1 - y(ij))(max(0, m - D))^2 \tag{3}$$

where  $D$  represents the feature distance of image pairs after feature extraction,  $f_i, f_j$ , and  $y(ij)$  have been described above, and  $m$  is a margin parameter, which is used to limit the minimum inter-class distance.

### 2.2.3. Loss Function

To better constrain the feature distance, we improved the loss function. According to the contrastive loss formula, when the input-image pairs belong to the same class, positive pair ( $y(ij) = 1$ ), the loss function can be simplified as:

$$L_p = D^2 \quad (4)$$

when the inputted image pairs belong to different classes, negative pair ( $y(ij) = 0$ ), the loss function can be simplified as:

$$L_n = (\max(0, m - D))^2 \quad (5)$$

According to Formula (5), no loss is caused by inputted image pairs belonging to different classes with a distance higher than  $m$ . This means the distance between different classes needs to be greater than  $m$  during the training process. A key question is how to choose the margin value  $m$ , which depends on the range of the image pair distance value. However, the distance between feature pairs is variable and uncertain. Since the range of  $D$  is  $[0, +\infty)$ , sigmoid function is used to normalize it. The specific normalization function is defined as follows:

$$D_n = \frac{2}{1 + e^{-D}} - 1 \quad (6)$$

And the final loss function is defined as

$$L(f_i, f_j, y(ij), m) = y_{ij}D_n^2 + (1 - y(ij))(\max(0, m - D_n))^2 \quad (7)$$

In subsequent experiments, the value of  $m$  is set to 0.7. The loss can be obtained and used to train the proposed model. Finally, Adam optimization is used for training.

The proposed model is a kind of supervised learning which learns the classification features of histopathological images by making the distance between similar images features as small as possible, and the distance between features from different types of images to be as large as possible. This can effectively reduce the extracted feature variance of similar histopathological images, and preserve the variance of different histopathological image extracted features, therefore making the model more discriminative.

In this paper, we pre-train AE using the training set. The Siamese network is then formed using the pre-trained AE and the network parameters are optimized by re-training. The final algorithm flow is shown in Algorithm 1.

To divide the histopathological images into two categories, a three-layer SoftMax classifier was built to obtain the probability distribution of the two categories after feature extraction from an AE network constrained by the Siamese framework. Cross entropy loss was used for the network training.

---

**Algorithm 1.** The training procedure for AE + Siamese Network.

---

**Input:**

The training set:  $X = \{x(1), x(2), \dots, x(n)\}$ ; learning rate:  $\alpha$ ; and iterative number:  $I_t$ .

**Output:**

The weights and biases:  $\{\mathbf{W}^{(i)}, \mathbf{b}^{(i)}\}_{i=1}^L$

1. Initialize  $\{\mathbf{W}^{(i)}, \mathbf{b}^{(i)}\}_{i=1}^L$  according to the trained AE.
  2. Build a Siamese network with two AE with shared weights.
  3. **for** each  $t \in [1, I_t]$  **do**
    - for** each  $i \in [1, L]$  **do**

Do forward propagation.
    - End for**
    - for** each  $i \in [1, L]$  **do**

Fine-tune  $\mathbf{W}^{(i)}, \mathbf{b}^{(i)}$  by minimizing loss function of Siamese network.
    - End for**
  - End for**
  4. Return  $\{\mathbf{W}^{(i)}, \mathbf{b}^{(i)}\}_{i=1}^L$
- 

### 3. Results

We conducted a series of experiments on a common breast cancer histopathology dataset to evaluate the performance of the improved AE model. PFTAS + QDA, PFTAS + SVM, PFTAS + RF, Inception\_v3, Resnet50, Inception\_resnet\_v2, and Xception were used as control models. The first three use the PFTAS algorithm to extract features from images, and three classification methods, namely quadratic discriminant analysis (QDA), SVM, and RF, are then used for classification. These three algorithms belong to traditional machine learning methods. The last four algorithms belong to CNN. We have also compared the performance of the proposed method with other state-of-the-art methods, such as IDNet [30], FE-VGGNET16-SVM(POLY) [31] and FCN-Bi-LSTM [32].

#### 3.1. BreakHis Dataset

The BreakHis dataset is the largest publicly available breast cancer histopathology dataset and is commonly used to classify breast cancer. It was generated from breast tissue biopsy slides, stained with hematoxylin and eosin (H&E), and labeled by pathologists at the P&D medical laboratory in Brazil [33]. The BreakHis dataset is composed of 7909 microscopic images ( $700 \times 460$  pixels, 3-channel RGB, 8-bit depth in each channel, and a PNG format) of breast tumor tissues collected from 82 patients using magnifying factors of  $40\times$ ,  $100\times$ ,  $200\times$ , and  $400\times$ , corresponding to objective lens of  $4\times$ ,  $10\times$ ,  $20\times$ , and  $40\times$ . The subsets of these images include 2480 benign and 5429 malignant images.

Both benign and malignant breast tumors can be sorted into different types based on the way the tumoral cells appear under the microscope. Various subtypes of breast tumors have different prognoses and treatment implications. Table 1 summarizes the image distribution in the BreakHis database.

**Table 1.** Image distribution of different magnifications in BreakHis dataset.

Magnification	Benign	Malignant	Total
$40\times$	625	1370	1995
$100\times$	644	1437	2081
$200\times$	623	1390	2013
$400\times$	588	1232	1820
total	2480	5429	7909
patient	24	58	82

To evaluate the performance of our model for clinical situations, we use 80% of the samples for training and 20% of the samples for testing.

### 3.2. Performance Metrics

We selected five indicators to evaluate our model performance, including accuracy, precision, specificity, recall and F1-score.  $N_{all}$  represents the total number of histopathological images in the dataset and  $N_r$  is the number of correctly classified images. Therefore, the image level recognition rate can be expressed as:

$$\text{Accuracy} = \frac{N_r}{N_{all}} = \frac{TP + FN}{TP + FP + TN + FN} \quad (8)$$

True positives (TP) are those who are diagnosed with cancer (positive) and actually have cancer. True negatives (TN) are those diagnosed as cancer-free (negative) who are actually cancer-free. False positives (FP) are those people diagnosed with cancer who do not actually have cancer. False negatives (FN) are those people diagnosed as cancer-free who actually have cancer.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (10)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

### 3.3. Experimental Settings

In this paper, we conducted all experiments in the following environments: the hardware environment was Inter Xeon Gold 5218 CPU @2.30 Ghz, 32 GB Nvidia TITAN RTX GPU, and 32 G running memory. The software environment included Ubuntu 16.04 with a 64-bit operating system. The whole programming environment was based on Python compilation language, and the model construction was based on the Keras framework. The parameters of  $\alpha$ ,  $m$  and iterative number for our method were setting for 0.001, 0.7, 100, respectively. All the above parameters were selected using a grid search strategy and comparison results are given.

### 3.4. Results and Analysis

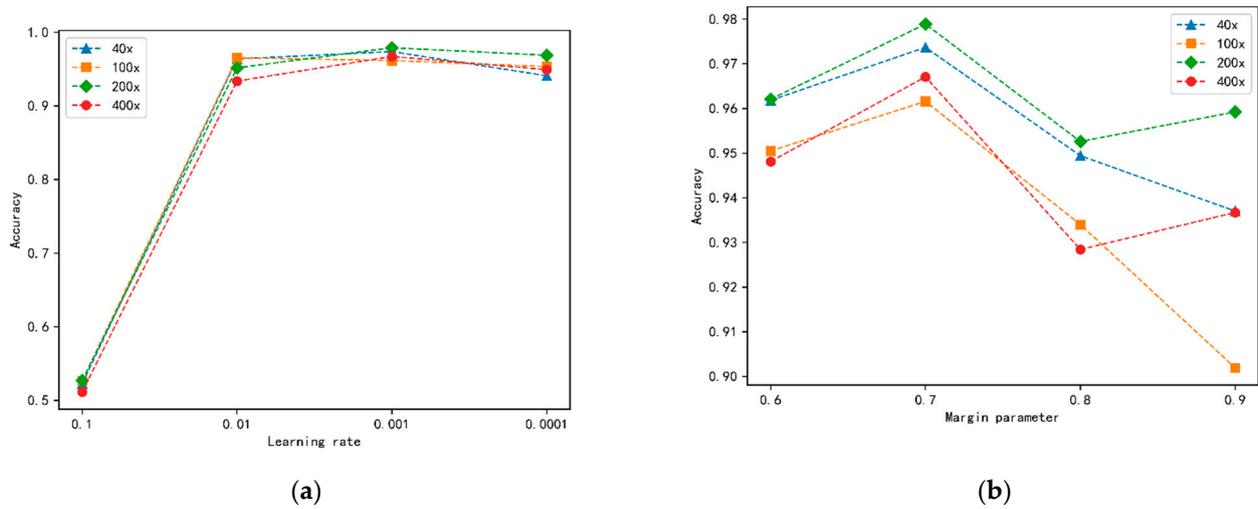
In a first experiment, we investigated the effect of different learning rates and different margin parameters on the performance of the proposed method. For this purpose, we evaluated the performance of the proposed method by comparing it for different learning rates (0.1, 0.01, 0.001 and 0.001) and different margin parameters (0.6, 0.7, 0.8 and 0.9), separately. The aim of this experiment is to find good parameters for the classification of histopathological images of breast cancer. The results are shown in Figure 7, from which we can observe the following:

The method achieves best results when the learning rate is 0.001. When  $\alpha = 0.1$ , the learning rate is too large causing the model to fail to converge. When the learning rate is greater than 0.01, the accuracy does not improve further as the learning rate decreases, due to the fact that the network training becomes very slow at this point and requires more training time. On balance, we chose a learning rate  $\alpha = 0.001$ .

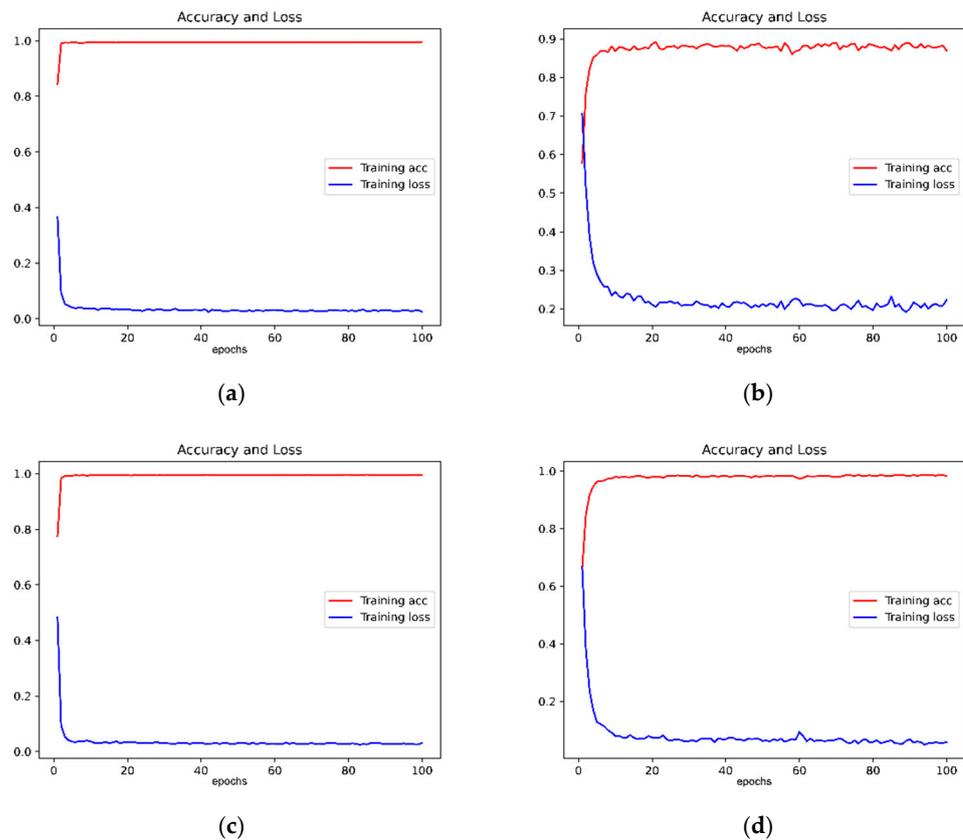
The margin parameter  $m = 0.7$  gives the best results at magnifications of  $40\times$ ,  $100\times$ ,  $200\times$  and  $400\times$ . Margin parameter  $m$  is the maximum distance threshold for the feature pairs involved in the training, and when  $m$  is too small, the model can easily fall into underfitting. When  $m$  is too large, the model can be over-trained and over-fitted. Therefore, in the following experiment, the value of  $m$  is 0.7.

Meanwhile, by observing the learning curve of the model in the first experiment, we found that the model had reached convergence when the number of iterations was at 100,

and the accuracy and loss basically stopped changing. The learning curve is illustrated in Figure 8.



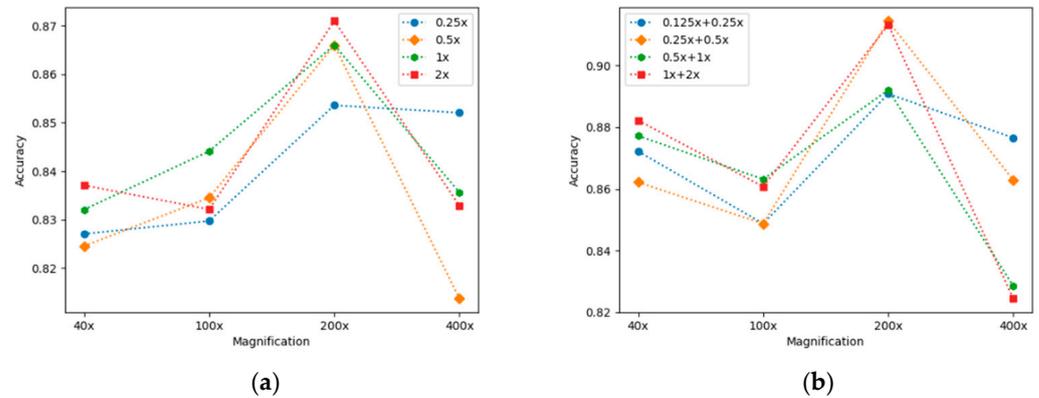
**Figure 7.** (a) A comparison of the classification accuracy of our proposed method at different learning rates at four different magnifications database. (b) A comparison of the classification accuracy of our proposed method at different learning rates at four different magnifications database.



**Figure 8.** This figure is learning curves of our model at different magnification images: (a) Training progress of the proposed model on 40× magnification images; (b) Training progress of the proposed model on 100× magnification images; (c) Training progress of the proposed model on 200× magnification images; (d) Training progress of the proposed model on 400× magnification images.

In the second experiment, we used AE + SoftMax as the benchmark network to compare the influence of multi-scale processing on the inputted images to extract feature

separability. After AE acquired the image features, SoftMax classifier was used to verify the validity of the features extracted at different scales. Furthermore, we compared classification performance of the inputted images with different magnification at different scales. The aim of this experiment was to determine a suitable input scale for the classification of breast cancer histopathological images. The results are shown in Figure 9, from which we make the following observations:



**Figure 9.** (a) Single-scale input-image accuracy; (b) Multi-scale input-image accuracy. Comparison of the classification accuracy of multi-scale input-image method with single-scale input-image on four different magnifications database.

Both single-scale and multi-scale input can achieve a high recognition rate under  $200\times$  magnification.

Compared with single-scale input, the multi-scale input recognition rate at different magnifying factors was improved by varying degrees.

At  $40\times$ ,  $100\times$ ,  $200\times$ , and  $400\times$  magnification, the input of  $1\times + 2\times$ ,  $0.5\times + 1\times$ ,  $0.25\times + 0.5\times$ , and  $0.125\times + 0.25\times$  achieved a higher recognition effect, respectively. Therefore, the subsequent experiment used those settings.

Owing to the different image magnifications, their original scale was also different, and consequently the best input scale combination also varies. As seen in Figure 9, the optimal input scale is continuous. This indicates that there is a universal optimal image scale for breast cancer histological image classification.

In the third experiment, we compared AE + SoftMax and AE + Siamese network, the latter improving on the AE feature extraction ability by making the differences in features within the same class smaller and the differences in features of different classes greater. The results are shown in Table 2. In addition, the ROC curve of the AE + Siamese network is shown in Figure 10. From these results, we can draw the following conclusion:

The AE + Siamese network achieved a better classification effect than AE + SoftMax under all magnification factors. The data shows that features with distance constraint are more favorable to classification.

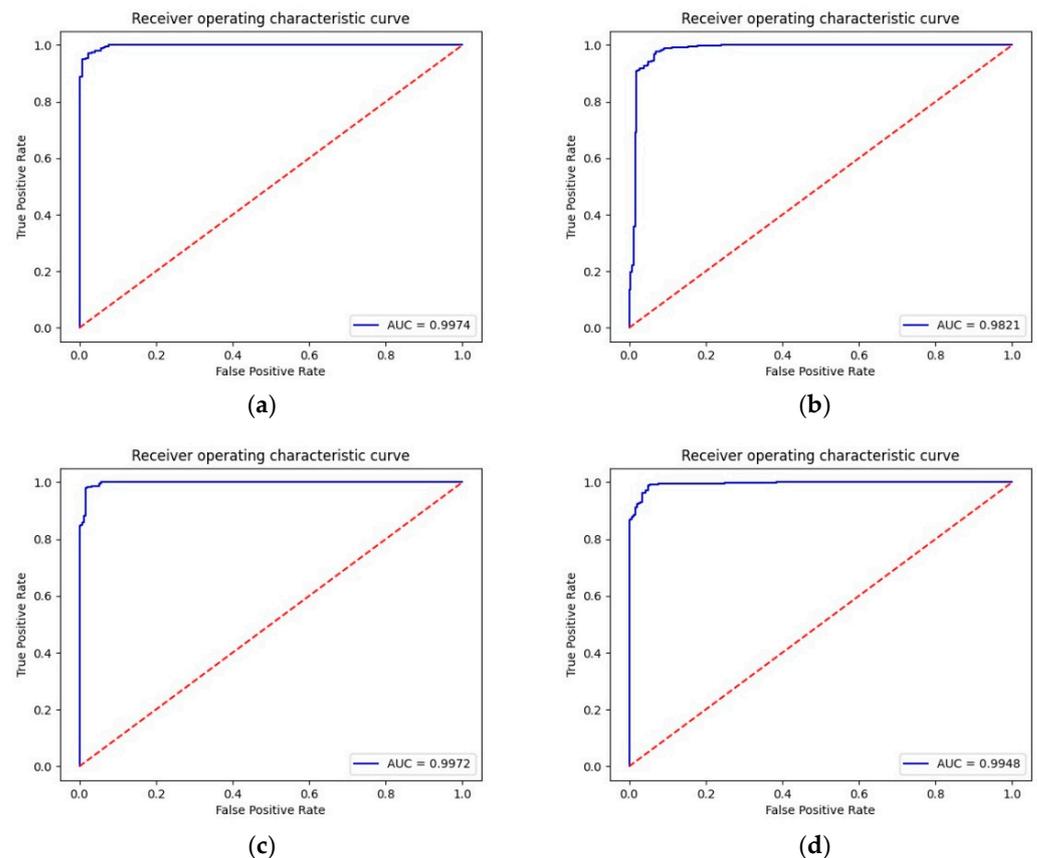
The model has the best recognition rate at  $200\times$  magnification, which is the closest to the optimal scale of the four magnifying factors.

Among the selected model performance indexes, the specificity was the worst. This suggests that the model is more likely to predict a positive outcome.

In the third experiment, we compared the performance of our model with ten other models. The first three being traditional models and the last seven being deep learning models. The results are shown in Table 3.

**Table 2.** Comparison of the performance (%) of AE + SoftMax with AE + Siamese Network.

(a). 40× magnification.				
Methods	Accuracy	Precision	Recall	Specificity
AE + SoftMax	88.2	88.6	95.0	73.2
AE + Siamese Network	97.3	96.9	99.2	93.2
(b). 100× magnification				
Methods	Accuracy	Precision	Recall	Specificity
AE + SoftMax	86.3	86.8	94.4	68.2
AE + Siamese Network	96.1	95.7	98.7	90.3
(c). 200× magnification.				
Methods	Accuracy	Precision	Recall	Specificity
AE + SoftMax	91.4	92.3	95.5	82.4
AE + Siamese Network	97.8	97.6	99.2	94.8
(d). 400× magnification.				
Methods	Accuracy	Precision	Recall	Specificity
AE + SoftMax	87.6	89.1	93.1	76.2
AE + Siamese Network	96.7	95.7	99.5	90.6



**Figure 10.** This figure is ROC curves of our model at different magnification images: (a) ROC of 40× magnification images; (b) ROC of 100× magnification images; (c) ROC of 200× magnification images; (d) ROC of 400× magnification images. The ROC curve analysis showed that the AUCs of the proposed model ranged from 0.982~0.997 for different magnification factors, indicating the effectiveness of the framework.

**Table 3.** Comparison of the performance (%) of AE + Siamese Network with other methods.

(a). 40× magnification.					
Methods	Accuracy	Precision	Recall	F1-score	Time (s)
PFTAS + QDA [33]	83.8	-	-	-	-
PFTAS + SVM [33]	81.6	-	-	-	-
PFTAS + RF [33]	81.8	-	-	-	-
Inception_v3	73.4	79.5	82.4	81.0	1003.0
Resnet50	79.1	77.5	98.1	86.6	3264.1
Inception_resnet_v2	77.9	81.6	87.5	84.4	2123.5
Xception	79.9	79.9	94.8	86.7	2346.8
IDSNet [30]	89.1	-	-	-	-
FE-VGGNET16-SVM(POLY) [31]	94.1	-	-	-	-
FCN-Bi-LSTM [32]	95.6	-	-	-	-
AE + Siamese Network	97.3	96.9	99.2	98.1	320.6
(b). 100× magnification.					
Methods	Accuracy	Precision	Recall	F1-score	Time (s)
PFTAS + QDA [33]	82.1	-	-	-	-
PFTAS + SVM [33]	79.9	-	-	-	-
PFTAS + RF [33]	81.3	-	-	-	-
Inception_v3	76.4	94.8	69.7	80.4	1063.7
Resnet50	71.2	72.8	93.0	81.7	3422.8
Inception_resnet_v2	70.0	90.9	62.8	74.2	2215.4
Xception	82.4	89.6	84.3	86.9	2445.8
IDSNet [30]	85.0	-	-	-	-
FE-VGGNET16-SVM(POLY) [31]	95.1	-	-	-	-
FCN-Bi-LSTM [32]	93.6	-	-	-	-
AE + Siamese Network	96.1	95.7	98.7	97.2	353.6
(c). 200× magnification.					
Methods	Accuracy	Precision	Recall	F1-score	Time (s)
PFTAS + QDA [33]	84.2	-	-	-	-
PFTAS + SVM [33]	85.1	-	-	-	-
PFTAS + RF [33]	83.5	-	-	-	-
Inception_v3	86.6	95.9	84.1	89.6	1083.7
Resnet50	89.3	91.2	93.5	92.3	3307.6
Inception_resnet_v2	80.8	92.7	78.4	84.9	2167.4
Xception	92.3	90.7	98.9	94.6	2377.2
IDSNet [30]	87.0	-	-	-	-
FE-VGGNET16-SVM(POLY) [31]	97.0	-	-	-	-
FCN-Bi-LSTM [32]	96.3	-	-	-	-
AE + Siamese Network	97.8	97.6	99.2	98.4	347.4
(d). 400× magnification.					
Methods	Accuracy	Precision	Recall	F1-score	Time (s)
PFTAS + QDA [33]	82.0	-	-	-	-
PFTAS + SVM [33]	82.3	-	-	-	-
PFTAS + RF [33]	81.0	-	-	-	-
Inception_v3	91.5	92.5	95.1	93.8	911.9
Resnet50	72.6	71.6	98.3	82.9	2997.4
Inception_resnet_v2	83.8	85.6	91.4	88.4	3818.8
Xception	86.8	89.9	90.6	90.3	1967.9
IDSNet [30]	84.5	-	-	-	-
FE-VGGNET16-SVM(POLY) [31]	93.4	-	-	-	-
FCN-Bi-LSTM [32]	94.2	-	-	-	-
AE + Siamese Network	96.7	95.7	99.5	97.6	318.1

Since the Siamese network is a characteristic pair input, the cost of training time actually increases compared to AE model that does not use the Siamese framework. However,

our model is still less time-consuming and shows a greatly improvement in performance compared to other models.

#### 4. Discussion

This study has demonstrated the superior performance of our model by comparing it with ten other methods and evaluating several their metrics. The second experiment and the third experiment respectively demonstrate that our proposed model possesses the following two superiorities:

(1) Suitable scale feature helps to classify the breast tumors in the histopathology images. Multi-scale input is used in our work to learn both the global and local features information at different scales. Extracting image features at the finer scale can solve the problem of fixed perceptual fields in neural networks.

This consideration of multi-scale input is further validated by the experimental results. The histopathological recognition of breast cancer was improved to varying degrees with the combination of multiple inputs at different scales. Furthermore, our results suggest that histopathological images of breast cancer imaged at different magnifications have the greatest improvement in recognition over a continuous scale change. However, finding the optimal combination of scales is still a question worth considering.

(2) If distance between the extracted features could be as small as possible within classes, and as large as possible between classes, it is good to enhance classifier performance. To further improve the performance of the breast pathology images, two AEs with shared weights and bias parameters are used as the benchmark of the proposed Siamese network. This allows the classifier to better distinguish between benign and malignant histopathological images of the breast.

In addition, due to the similarity of histopathological images of different cancers, the model proposed in this study can also be considered for extension to predict other pathological image classifications beyond breast cancer.

#### 5. Conclusions

In this paper, we proposed an AE using Siamese frame constraints as a feature extractor to construct an end-to-end network model for breast cancer histopathological image classification. The proposed model aims to learn useful image features by minimizing the distance between the input and output images, while keeping the feature extraction inter-class distance as large as possible and the intra-class distance as small as possible. This model combines the powerful feature extraction capability of an AE network with the feature distance constraints of a Siamese framework, to obtain more easily discriminative image features. In addition, the multi-scale transformation of the inputted image enables the network model to learn both global and local information. The experimental results show that the features extracted from multi-scale input can significantly improve classification performance, particularly when Siamese frameworks is used for feature distance constraint. Our model not only has high accuracy, but also shows good specificity, sensitivity and less training time in different cases, which indicates that it is especially useful for pathologists and researchers in breast cancer diagnosis. CADs composed with our models do not require high hardware costs. With greater expectation, the trained model could be applied to the breast cancer detection interface and could even be considered in combination with information from the patient clinical or family history to form an expert system to improve the diagnostic accuracy made by the expert.

**Author Contributions:** Conceptualization, M.L. and Y.H.; methodology, M.L., Y.H. and M.W.; validation, M.W. and C.Z.; writing—review and editing, Y.H.; visualization, Y.H. and C.Z.; supervision, M.W.; funding acquisition, C.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by National Natural Science Foundation of China, grant number 61,901,165.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The BreakHis dataset is publicly available at <https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-BreakHis/> (20 February 2022) and the only condition for using it is citing [33] (in our reference list).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Al-antari, M.A.; Al-masni, M.A.; Park, S.U.; Park, J.; Metwally, M.K.; Kadah, Y.M.; Han, S.M.; Kim, T.S. An Automatic Computer-Aided Diagnosis System for Breast Cancer in Digital Mammograms via Deep Belief Network. *J. Med. Biol. Eng.* **2018**, *38*, 443–456. [[CrossRef](#)]
3. Celaya-Padilla, J.M.; Guzmán-Valdivia, C.H.; Galván-Tejada, C.E.; Galván-Tejada, J.I.; Gamboa-Rosales, H.; Garza-Veloz, I.; Martínez-Fierro, M.L.; Cid-Báez, M.A.; Martínez-Torteya, A.; Martínez-Ruiz, F.J. Contralateral asymmetry for breast cancer detection: A CADx approach. *J. Med. Biol. Eng.* **2018**, *38*, 115–125. [[CrossRef](#)]
4. Chan, H.P.; Samala, R.K.; Hadjiiski, L.M. CAD and AI for breast cancer—Recent Development and Challenges. *Br. J. Radiol.* **2019**, *93*, 1108. [[CrossRef](#)]
5. Avanzo, M.; Porzio, M.; Lorenzon, L.; Milan, L.; Sghedoni, R.; Russo, G.; Massafra, R.; Fanizzi, A.; Barucci, A.; Ardu, V.; et al. Artificial intelligence applications in medical imaging: A review of the medical physics research in Italy. *Phys. Med.* **2021**, *83*, 221–241. [[CrossRef](#)] [[PubMed](#)]
6. Massafra, R.; Bove, S.; Lorusso, V.; Biafora, A.; Comes, M.C.; Didonna, V.; Diotaiuti, S.; Fanizzi, A.; Nardone, A.; Nolasco, A.; et al. Radiomic Feature Reduction Approach to Predict Breast Cancer by Contrast-Enhanced Spectral Mammography Images. *Diagnostics* **2021**, *11*, 684. [[CrossRef](#)]
7. Comes, M.C.; La Forgia, D.; Didonna, V.; Fanizzi, A.; Giotta, F.; Latorre, A.; Martinelli, E.; Mencattini, A.; Paradiso, A.V.; Tamborra, P.; et al. Early Prediction of Breast Cancer Recurrence for Patients Treated with Neoadjuvant Chemotherapy: A Transfer Learning Approach on DCE-MRIs. *Cancers* **2021**, *13*, 2298. [[CrossRef](#)]
8. Krithiga, R.; Geetha, P. Breast Cancer Detection, Segmentation and Classification on Histopathology Images Analysis: A Systematic Review. *Arch. Comput. Methods Eng.* **2020**, *10*, 2607–2619. [[CrossRef](#)]
9. Kowal, M.; Filipczuk, P.; Obuchowicz, A.; Korbicz, J.; Monczak, R. Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. *Comput. Biol. Med.* **2013**, *43*, 1563–1572. [[CrossRef](#)]
10. George, Y.M.; Zayed, H.H.; Roushdy, M.I.; Elbagoury, B.M. Remote Computer-Aided Breast Cancer Detection and Diagnosis System Based on Cytological Images. *IEEE Syst. J.* **2014**, *8*, 949–964. [[CrossRef](#)]
11. Filipczuk, P.; Fevens, T.; Krzyżak, A.; Monczak, R. Computer-Aided Breast Cancer Diagnosis Based on the Analysis of Cytological Images of Fine Needle Biopsies. *IEEE Trans. Med. Imaging* **2013**, *32*, 2169–2178. [[CrossRef](#)] [[PubMed](#)]
12. Bayramoglu, N.; Kannala, J.; Heikkila, J. Deep learning for magnification independent breast cancer histopathology image classification. In Proceedings of the International Conference on Pattern Recognition (ICCV), Cancun, Mexico, 4–8 December 2016.
13. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. Breast Cancer Histopathological Image Classification using Convolutional Neural Networks. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016.
14. Spanhol, F.A.; Oliveira, L.S.; Cavalin, P.R.; Petitjean, C.; Heutte, L. Deep features for breast cancer histopathological image classification. In Proceedings of the IEEE International Conference on Systems, Rome, Italy, 9–11 October 2017.
15. Wang, Z.; Dong, N.; Dai, W.; Rosario, S.D.; Xing, E.P. Classification of breast cancer histopathological images using convolutional neural networks with hierarchical loss and global pooling. In Proceedings of the International Conference Image Analysis and Recognition (ICIAR), Póvoa de Varzim, Portugal, 27–29 June 2018.
16. Brancati, N.; Frucci, M.; Riccio, D. Multi-classification of breast cancer histology images by using a fine-tuning strategy. In Proceedings of the International Conference Image Analysis and Recognition (ICIAR), Póvoa de Varzim, Portugal, 27–29 June 2018.
17. Pimkin, A.; Makarchuk, G.; Kondratenko, V.; Pisov, M.; Krivov, E.; Belyaev, M. Ensembling neural networks for digital pathology images classification and segmentation. In Proceedings of the International Conference Image Analysis and Recognition (ICIAR), Póvoa de Varzim, Portugal, 27–29 June 2018.
18. Wan, S.; Lee, H.C.; Huang, X.; Xu, T.; Xu, T.; Zeng, X.; Zhang, Z.; Sheikine, Y.; Connolly, J.L.; Fujimoto, J.G. Integrated local binary pattern texture features for classification of breast tissue imaged by optical coherence microscopy. *Med. Image Anal.* **2017**, *38*, 104–116. [[CrossRef](#)] [[PubMed](#)]
19. Nanni, L.; Brahnam, S.; Lumini, A. A very high performing system to discriminate tissues in mammograms as benign and malignant. *Expert Syst. Appl.* **2012**, *39*, 1968–1971. [[CrossRef](#)]

20. Sharma, M.; Singh, R.; Bhattacharya, M. Classification of breast tumors as benign and malignant using textural feature descriptor. In Proceedings of the IEEE International Conference on Bioinformatics & Biomedicine, Kansas City, MO, USA, 13–16 November 2017.
21. Gupta, V.; Bhavsar, A. Breast Cancer Histopathological Image Classification: Is Magnification Important? In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017.
22. Feng, Y.; Lei, Z.; Mo, J. Deep Manifold Preserving Autoencoder for Classifying Breast Cancer Histopathological Images. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *17*, 91–101. [[CrossRef](#)]
23. Alom, M.Z.; Yakopcic, C.; Nasrin, M.S.; Taha, T.M.; Asari, V.K. Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network. *J. Digit. Imaging* **2019**, *32*, 605–617. [[CrossRef](#)]
24. Sudharshan, P.J.; Petitjean, C.; Spanhol, F.; Oliveira, L.E.; Heutte, L.; Honeine, P. Multiple instance learning for histopathological breast cancer image classification. *Expert Syst. Appl.* **2019**, *117*, 103–111. [[CrossRef](#)]
25. Yan, R.; Ren, F.; Wang, Z.; Wang, L.; Zhang, T.; Liu, Y.; Rao, X.; Zheng, C.; Zhang, F. Breast cancer histopathological image classification using a hybrid deep neural network. *Methods* **2020**, *173*, 52–60. [[CrossRef](#)]
26. Sheikh, T.S.; Lee, Y.; Cho, M. Histopathological Classification of Breast Cancer Images Using a Multi-Scale Input and Multi-Feature Network. *Cancers* **2020**, *12*, 2031. [[CrossRef](#)]
27. Comes, M.C.; Filippi, J.; Mencattini, A.; Casti, P.; Cerrato, G.; Sauvat, A.; Vacchelli, E.; De Ninno, A.; Di Giuseppe, D.; D’Orazio, M.; et al. Multi-scale generative adversarial network for improved evaluation of cell–cell interactions served in organ-on-chip experiments. *Neural. Comput. Appl.* **2021**, *8*, 3671–3689. [[CrossRef](#)]
28. Bromley, J.; Bentz, J.W.; Bottou, L.; Guyon, I.; LeCun, Y.; Moore, C.; Sackinger, E.; Shah, R. Signature verification using a “siamese” time delay neural network. *Int. J. Pattern Recognit. Artif.* **1993**, *7*, 669–688. [[CrossRef](#)]
29. Kallenberg, M.; Petersen, K.; Nielsen, M.; Ng, A.Y.; Diao, P.; Igel, C.; Vachon, C.M.; Holland, K.; Winkel, R.R.; Karssemeijer, N. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE Trans. Med. Imaging* **2016**, *35*, 1322–1331. [[CrossRef](#)] [[PubMed](#)]
30. Li, X.; Shen, X.; Zhou, Y.; Wang, X.; Li, T.Q. Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet). *PLoS ONE* **2020**, *15*, e0232127. [[CrossRef](#)] [[PubMed](#)]
31. Kumar, A.; Singh, S.K.; Saxena, S.; Lakshmanan, K.; Sangaiah, A.K.; Chauhan, H.; Shrivastava, S.; Singh, R.K. Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer. *Inf. Sci.* **2020**, *508*, 405–421. [[CrossRef](#)]
32. Budak, Ü.; Cömert, Z.; Rashid, Z.N.; Şengür, A.; Çıbuk, M. Computer-aided diagnosis system combining FCN and Bi-LSTM model for efficient breast cancer detection from histopathological images. *Appl. Soft Comput.* **2019**, *85*, 105765. [[CrossRef](#)]
33. 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](#)]