

Classification of Breast Cancer Histopathological Images Using Discriminative Patches Screened by Generative Adversarial Networks

RUI MAN¹, PING YANG¹, AND BOWEN XU²

¹Smart City College, Beijing Union University, Beijing 100101, China

²Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

Corresponding author: Ping Yang (xtyangping@buu.edu.cn)

ABSTRACT Computer-aided diagnosis (CAD) systems of breast cancer histopathological images automated classification can help reduce the manual observation workload of pathologists. In the classification of breast cancer histopathology images, due to the small number and high-resolution of the training samples, the patch-based image classification methods have become very necessary. However, adopting a patches-based classification method is very challenging, since the patch-level datasets extracted from whole slide images (WSIs) contain many mislabeled patches. Existing patch-based classification methods have paid little attention to addressing the mislabeled patches for improving the performance of classification. To solve this problem, we propose a novel approach, named DenseNet121-AnoGAN, for classifying breast histopathological images into benign and malignant classes. The proposed approach consists of two major parts: using an unsupervised anomaly detection with generative adversarial networks (AnoGAN) to screen mislabeled patches and using densely connected convolutional network (DenseNet) to extract multi-layered features of the discriminative patches. The performance of the proposed approach is evaluated on the publicly available BreakHis dataset using 5-fold cross validation. The proposed DenseNet121-AnoGAN can be better suited to coarse-grained high-resolution images and achieved satisfactory classification performance in 40X and 100X images. The best accuracy of 99.13% and the best $F1_{score}$ of 99.38% have been obtained at the image level for the 40X magnification factor. We have also investigated the performance of AnoGAN on the other classification networks, including AlexNet, VGG16, VGG19, and ResNet50. Our experiments show that whether it is at the patient-level accuracy or at the image-level accuracy, the classification networks with AnoGAN have provided better performance than the classification networks without AnoGAN.

INDEX TERMS Breast cancer histopathological images, densely connected convolutional networks, discriminative patches, generative adversarial networks, image classification.

I. INTRODUCTION

Breast cancer is the top cancer in women, impacting 2.1 million women each year, and also causes the greatest number of cancer-related deaths among women. Breast cancer is a serious disease that can start in almost any organ or tissue of the body when abnormal cells grow uncontrollably, go beyond their usual boundaries to invade adjoining parts of the body or spread to other organs [1]. According to the data provided by the American Cancer Society, in 2020 in U.S., there will be an estimated 276,480 new cases of invasive breast cancer and 48,530 new cases of non-invasive breast cancer expected to be diagnosed in women.

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar¹.

About 42,170 women in U.S. are expected to die in 2020 from breast cancer [2].

Due to the high death rate of breast cancer, women are suggested to do regular screenings via mammograms and computerized tomography (CT) [3]. If abnormal cells are found, biopsy procedure is performed to diagnose the abnormality in breast. Usually, the collected sample is stained with hematoxylin and eosin (H&E). Hematoxylin reacts to Deoxyribonucleic Acid (DNA) and it stains purple or blue color to the nuclei, while Eosin reacts to proteins and it stains pink color to other structures [4].

Diagnosis from a histopathological image is considered as the gold standard in diagnosing all kinds of cancer, including breast cancer [5]–[7]. However, histopathological analysis is a very time-consuming professional task that depends on

the experience of the pathologist, and the diagnosis can be influenced by factors such as the pathologist's fatigue and decreased attention [7], [8]. Therefore, there is an urgent need for computer-aided diagnosis (CAD) systems to provide an objective assessment to pathologists and improve the diagnostic efficiency [9], [10].

With the advancements in medical image processing and deep learning, classification of breast histopathological images has become an important area for research [11], [12]. Due to the high-resolution breast cancer histopathological images, the exiting traditional machine learning methods and the deep neural network models used to directly analyze the whole slide images (WSIs) have caused a very complex architecture that hard to train [13]. During the past few decades, some researchers proposed the strategies that relied on the segmentation of nucleus, and then used the extracted handcrafted features to train a classifier [12], [14]–[16]. Kowal *et al.* [14] segmented the nucleus by color-based clustering, and George *et al.* [15] used the circular Hough transform to detect the location of the nucleus, then refined feature-based candidates via watersheds algorithm [17]. These studies allowed to extract features that are usually related to morphology, topology, and texture. The calculated features can then be used to train one or more classifiers. Kowal *et al.* [14] achieved an accuracy rate of 84%–93% on 500 images from 50 patients and George *et al.* [15] achieved an accuracy between 72% and 97% on 92 images. In addition to the nuclei-related information, Belsare *et al.* [16] also considered to segment the epithelial layer around the cell cavity by using the spatio-color-texture graph, and statistical texture features were used to train the final classifier. Belsare *et al.* [16] reported the accuracy rates between 70% and 100% of 70 breast histology H&E datasets from 40X magnification level. Spanhol *et al.* [18] constructed a public dataset called BreaKHis and explored the effectiveness of six state-of-the-art handcrafted features descriptors, *i.e.*, Local Binary Pattern (LBP) [19], Completes Local Binary Pattern (CLBP) [20], Local Phase Quantization (LPQ) [21], Gray-Level Co-Occurrence Matrix (GLCM) [22], Parameter-Free Threshold Adjacency Statistics (PFTAS) [23], and Oriented FAST and Rotated BRIEF (ORB) [24]. Then they made experiment on four different classifiers and reported the accuracy between 80% and 85%.

The results obtained from different handcrafted features given above were considered to be relatively acceptable results, but highly unstable. As a matter of fact, the main limitation of these traditional methods is that the quality of the model depends on the extracted features, however, obtaining highly representative features is a very complicated task. Even if we choose the most appropriate descriptor, or combine various descriptors together to improve their recognition ability, the results obtained are still relatively low and unstable between different magnification levels [25].

Recently, the convolutional neural network (CNN) has been employed in visual classification system [26]–[29]. In the classification of breast cancer histopathology images,

the number of samples is small and the size of images is large, which makes it difficult or even impossible to train a deep learning model based on CNN. In addition, directly resizing the whole histopathology images to the input size for the deep learning model will lose a host of detailed feature information. Consequently, some researchers proposed the patch-based image classification methods to solve this problem. Spanhol *et al.* [30] adopted the random extracting patches strategy and the sliding window strategy to extract the image patches of the BreakHis dataset. They trained AlexNet [31] based on the extracted image patches as input and combined the patch-level classification results with three fusion rules for final classification. Araújo *et al.* [32] proposed a convolutional neural network (CNN) architecture, which is designed to extract features from patch-level dataset of 512×512 pixels. By training the network, images were classified into four classes, normal, benign, in situ carcinoma, and invasive carcinoma, and into two classes, carcinoma, and non-carcinoma. The image patches extraction strategy enabled CNN to train the WSIs with a certain resolution. Hou *et al.* [33] proposed a patch-level convolutional neural network (CNN) for high-resolution WSIs classification which has two-level model. The first level (patch-level) model is an Expectation Maximization (EM) that can automatically identify patches for patch-level CNN training, and the second level (image-level) model is multiclass logistic regression or support vector machine (SVM). Alom *et al.* [34] proposed a method to classify breast cancer histopathology images using the Inception Recurrent Residual Convolutional Neural Network (IRRCNN) model. Random patches were cropped to create a patch dataset for training and testing the IRRCNN model, then used the Winner Take ALL (WTA) method [35] to generate the final classification results.

Although the above researches show that patch-based image classification methods have been widely used in various breast cancer histopathology datasets. Adopting a patch-based classification method is very challenging. This is because labeled data is critical to the performance of the deep learning approaches. Automated image classification tasks require large amounts of annotated data. Because of the complexity of breast cancer histopathology images, the annotation process is laborious and costly. As only the image-level label is given in the datasets, the label of the whole input histopathological images is assigned to the corresponding generated patches. However, there are benign areas in the malignant WSIs, which makes the patch-level label maybe not consistent with the image-level label, and only a small part of extracted image patches is correctly labeled. This can result in training with mislabeled patches. When the training model receives the incorrect label information, the classification performance will be reduced.

To address these mislabeled patches and further improve the accuracy of classification. We propose a novel approach, named DenseNet121-AnoGAN, for classifying histopathological images into benign and malignant classes.

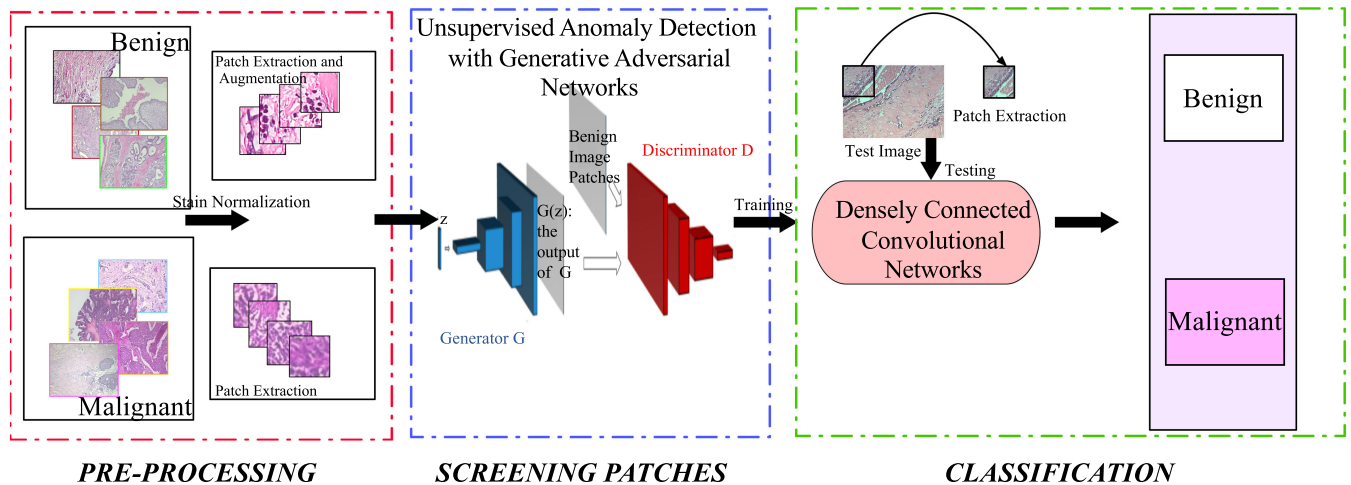


FIGURE 1. Overview of the proposed framework used for classification of breast cancer histopathological images.

The proposed approach consists of two major parts: using an unsupervised anomaly detection with generative adversarial networks (AnoGAN) [36] to screen mislabeled patches as well as using densely connected convolutional network (DenseNet) [37] to extract multi-layered features of the discriminative patches. The main contributions of our work can be summarized as follow:

- 1) We propose a screening patches method based on an unsupervised anomaly detection with generative adversarial networks (AnoGAN). We use benign patches to train AnoGAN. The data distribution of benign patches can be obtained by AnoGAN and it will generate a fake patch with a probability distribution similar to that of the benign patch. By defining the threshold of residual loss and discrimination loss between the malignant patch to be tested and fake patch, this well-trained AnoGAN can yield the high anomaly score of the malignant patches. However, the anomaly score of mislabeled patches in malignant patches is low, which can screen the most discriminative histopathological image patches and improve the classification performance of the subsequent network.
- 2) We design a breast cancer histopathological images classification method based on DenseNet121. We note that the presented research rarely involves state-of-the-art network architecture, e.g. DenseNet. DenseNet achieves multi-scale feature extraction by integrating convolutional neural networks into dense blocks.
- 3) Experiments were conducted on the BreakHis dataset using 5-fold cross validation. The results demonstrate that the proposed approach for breast cancer histopathology image classification has an excellent performance in both image-level and patient-level classification. The best accuracy of 99.13% and the best $F1_{score}$ of 99.38% have been obtained at the image level for the 40X magnification factor.

The rest of this paper is organized as follows: in Section II, we give the information about the dataset and describe the proposed method. Section III provides the experiments and results. Discussions are shown in Section IV. In Section V, we summarize the conclusion of this paper.

II. METHODOLOGY

As shown in FIGURE 1, the proposed approach mainly includes three main steps, described as below.

- 1) *Pre-processing*: To solve the stain variability of BreakHis dataset, the stain normalization pre-processing of histopathological images is firstly carried out. Secondly, to increase the number of training samples, we use patch extraction and data augmentation algorithm on benign images and use patch extraction algorithm on malignant images.
- 2) *Screening patches*: We use benign patches to train AnoGAN, which will generate a fake patch $G(z)$ of random sample z with a probability distribution similar to that of the benign patch. The trained parameters of the generator and discriminator are kept fixed. We calculate the anomaly score between the malignant patch to be tested and fake patch. This well-trained AnoGAN can yield the high anomaly score of the malignant patches and yield the low anomaly score of mislabeled patches in malignant patches. The patches labeled correctly in malignant are processed with data augmentation.
- 3) *Classification*: We use the discriminative patches to train DenseNet121. During testing, 100 random patches with the size of 224×224 pixels are cropped from each image in the testing set. These patches are passed to the well-trained DenseNet121, and we use majority voting for obtaining the final image label from the individual patch classifications.

In this section, we introduce the detail of the main technologies for the classification of breast cancer histopathological images used in the overall framework.

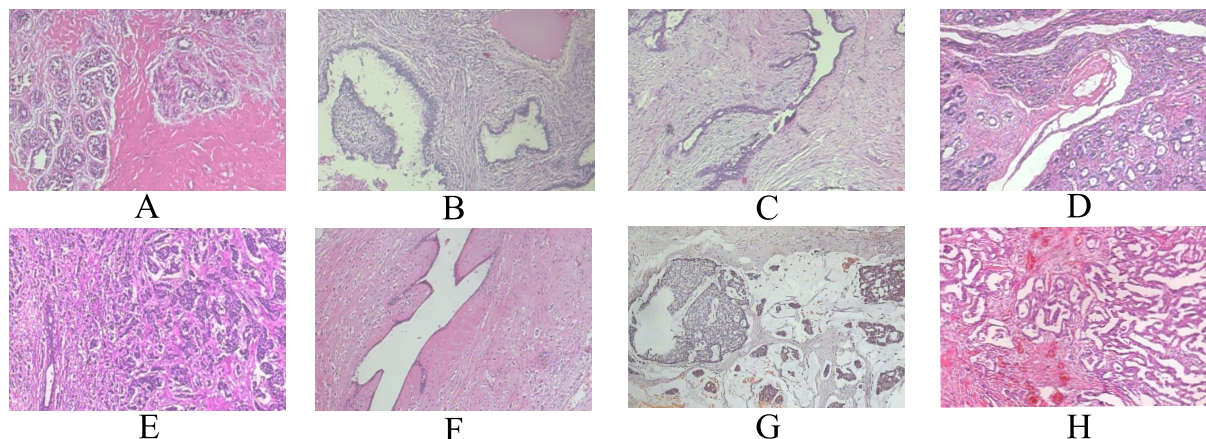


FIGURE 2. H&E stained images from BreKHis, (A): Adenosis, (B): Fibroadenoma, (C): Phyllodes Tumor, (D): Tubular adenoma, (E): ductal carcinoma, (F): lobular carcinoma, (G): Mucinous carcinoma, (H): Papillary carcinoma.

A. DATASET

The dataset used in this work is BreKHis, the latest public breast cancer histopathological images dataset, which was collected through a clinical study in 2014. During this period, all patients referred to the P&D Laboratory (Brazil) with a clinical indication of breast cancer were invited to participate this study [18]. The institutional review board approved the study and all patients signed written informed consent. All data were anonymized. Samples were generated from breast tissue biopsy slides and stained with hematoxylin and eosin (H&E). The samples were collected by a surgical open biopsy (SOB), prepared for histological research and labeled by pathologists of the P&D laboratory. Each case was diagnosed by an experienced pathologist and confirmed by immunohistochemical analysis and other complementary exams [38].

To date, BreKHis dataset is composed of 7909 histopathological biopsy images collected from 82 patients. Images were acquired in three-channel RGB color space, with a dimension of 700×460 using four magnification factors (40X, 100X, 200X, and 400X). Each images is labeled as either benign or malignant categories, and also distributed into eight sub-categories: Adenosis (A), Fibroadenoma (F), Phyllodes Tumor (PT), and Tubular Adenoma (TA) for benign images and Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC) and Papillary Carcinoma (PC) for malignant ones. The distribution of BreakHis images and patients into four magnification levels for both main tumor categories and each sub-category is provided in TABLE 1. FIGURE 2 shows samples from eight sub-categories breast tumors in 40X magnification factor.

B. STAIN NORMALIZATION PRE-PROCESSING

A deep learning-based method for the classification of breast cancer histopathology images, which relies on training set to capture a wide range of changes to distinguish the differences between intra-class and inter-class. Due to the color response

TABLE 1. Image and patient distribution among the main categories and each sub-category.

Main category	Sub-category	Magnification				Patients
		40X	100X	200X	400X	
Benign	Adenosis	114	113	111	106	4
	Fibroadenoma	253	260	264	237	10
	Phyllodes Tumor	149	150	140	130	3
	Tubular Adenoma	109	121	108	115	7
Malignant	Ductal Carcinoma	864	903	896	788	38
	Lobular Carcinoma	156	170	163	137	5
	Mucinous Carcinoma	205	222	196	169	9
	Papillary Carcinoma	145	142	135	138	6

of the digital scanners, the material and manufacturing technology of the staining supplier, and the different staining protocol in the different labs, it may cause large color differences in the histopathological images. Therefore, stain normalization is a fundamental and necessary step in the pre-processing of H&E stained breast cancer histopathology images.

Many methods have been proposed for stain normalization [39]–[41]. In this paper, we use a stain normalization method proposed by Vahadane *et al.* [41] on BreKHis dataset. This method adopts a novel structure-preserving color normalization (SPCN) scheme. It transforms the stain separation problem into a non-negative matrix factorization (NMF) [42] to which we add a sparseness constraint, which is called sparse non-negative matrix factorization (SNMF). One advantage of this method is that the color basis is

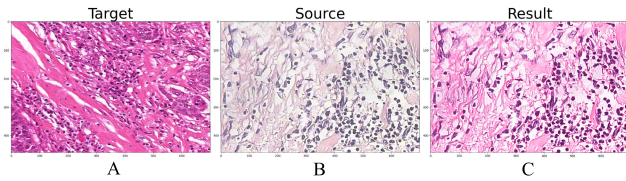


FIGURE 3. H&E stained images normalization, (A): The target image, (B): original image, (C): image after stain normalization.

determined in an unsupervised manner, and there is no need to manually label the pure stains in different areas. The working principle of SPCN is to replace the color basis of a source image with the color of a pathologist-preferred target image while reliably keeping the source image structural information intact, and still maintaining its original staining concentration. FIGURE 3 shows images before and after stain normalization.

C. PATCH EXTRACTION AND DATA AUGMENTATION

The performance of deep learning model depends on the large number of samples used for training. The number of training samples for each category in the BreaKHis dataset is limited. So we must increase the number of training samples through patch extraction and data augmentation algorithm to overcome the overfitting problem in the network.

Because of the high-resolution of breast cancer histopathology images, direct training will lead to excessive memory consumption and long training time. Inspired by Spanhol *et al.* [30] and Krizhevsky *et al.* [43], we apply patch extraction and data augmentation algorithm to increase the number of training samples and use these training samples to train the proposed model. Then we use majority voting for obtaining the final image label from the individual classifications. It is worth mentioning that we avoid using smaller image patches of size 32×32 or 64×64 [30]. This is because in the BreaKHis dataset, the label has been assigned to the whole input breast cancer histopathological image with a size of 700×460 , and there is no guarantee that a smaller image patch with a size of 32×32 or 64×64 will carry sufficient diagnostic information. Therefore, we divide the images of size 700×460 into the patches of size 224×224 that would provide a larger field of view and carry more local discrimination features in contrast to the smaller patches [44]–[47].

As can be seen from TABLE 1, BreaKHis dataset has a problem with data imbalance. The imbalance ratio between malignant and benign classes is 0.45 at the image level and 0.41 at the patient level. In classification tasks, the data imbalance problem may cause the discrimination ability of the computer-aided diagnosis (CAD) systems to be biased towards the majority class. To minimize the influence of data imbalance on the model performance, we adopt a random patches extraction strategy. In the j^{th} category, the number of

patches generated from per image is defined by Equation (1):

$$N_j = \left\lfloor \frac{\sum_{i=1}^n x_i}{\frac{n}{x_j}} \times \alpha \right\rfloor, \quad (1)$$

where N_j is the number of patches generated from per image in the j^{th} category, x_i refers the number of the i^{th} category, x_j refers the number of the j^{th} category, and n is the number of categories. In our experiment, we set the fixed parameter α to 64. Then all classes have the roughly equal number of patches.

The main advantage of using image patches in training for each category is that it retains the local discrimination information of the histopathological image, which helps the model to learn local features [48]. The random image patches generation strategy can also reduce the size of the training images and increase the number of training samples at the same time.

Data augmentation is an integral part of deep learning since it helps to overcome overfitting on models by increasing the number of training samples [43]. For breast cancer histopathological images, pathologists can examine a tissue slide from different angles without tampering the diagnostic results. So these images are rotation-invariant. We use the data augmentation algorithm to increase the prediction accuracy of the CAD systems, while increasing the number of training samples without changing the tissue morphology and cell structure of the image. The data augmentation algorithm is given in **Algorithm 1**.

D. SCREENING PATCHES

The strategy for sampling patches from breast cancer histopathological images is described in Section II-C. As only image-level labels are given in the breast cancer histopathological image classification task, the label of the whole image is usually assigned to the corresponding generated image patches. Therefore, the patch-level labels may not be consistent with the image-level labels. These mislabeled patches may affect the training of subsequent network and reduce the classification performance. To avoid mislabeled image patches when we use patch-based classification method, inspired by Schlegl *et al.* [36], we propose a screening mislabeled patches method based on an unsupervised anomaly detection with generative adversarial networks (AnoGAN). FIGURE 4 shows the framework of screening patches using AnoGAN.

1) GENERATIVE ADVERSARIAL NETWORK

The generative adversarial network (GAN) [49] consists of two adversarial models, a generator G and a discriminator D . The generator network captures the data distribution and maps $G(z)$ of random samples z , 1D vectors of uniformly distributed input noise sampled from the latent space Z , to data space. The discriminator network estimates the

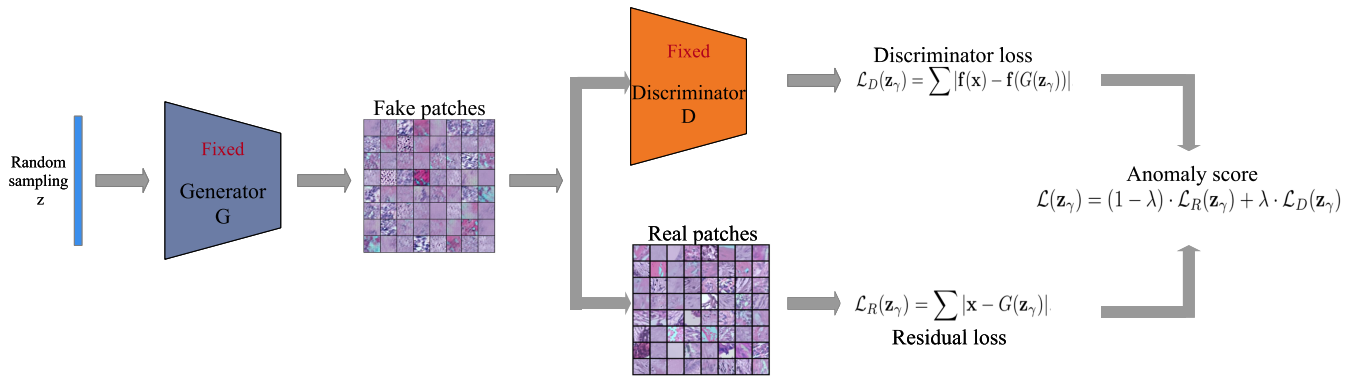


FIGURE 4. The framework of screening patches from Anomaly Detection with Generative Adversarial Networks (AnoGAN). Generative adversarial training is performed on benign data and testing is performed on both, unseen healthy cases and anomalous data.

Algorithm 1 Data Augmentation of BreakHis Histopathological Images

Input: Breast cancer histopathological image I_k from BreakHis after stain normalization pre-processing.

Output: Augmented images $\{I_{a1}, I_{a2}, \dots, I_{an}\}$.

Functions:

RanPatchGen() represents the methods of random patches extraction with Eq.(1);

Rotation() represents the methods of rotation with $\frac{\pi Q}{4}$ variations with Q in $\{0, 1, \dots, 7\}$;

Filp() represents the methods of horizontal and vertical reflection.

Step 1:

Take histopathological image I_k after stain normalization from training set;

Step 2:

Apply random patches extraction algorithm on image I_k :

$$RanPatchGen() = \{I_{k1}, I_{k2}, \dots, I_{kn}\}$$

Step 3:

Apply affine transformations on image patches $\{I_{k1}, I_{k2}, \dots, I_{kn}\}$:

for $I_{ki} = I_{k1} : I_{kn}$ **do**

Rotation()

Filp()

end for

Break;

probability that a sample comes from the real data rather than the generator network. During the training process, the generator network is optimized by the results of the discriminator network to improve the generating ability, and generates the image as much closer to x as possible to “fool” the discriminator network. At the same time, the discriminator network also optimizes itself to be better at flagging the generated samples. Goodfellow et al. [49] compared the generative adversarial network (GAN) to the minimax two-player game between the generator G and the discriminator D .

Let x be data representing an image. For the generator network, let z be a latent space vector sampled from uniform distribution. $G(z)$ refers the generator function which maps z to data space. The generator can generate fake samples from the estimated distribution p_g by estimating the training data comes from p_{data} . $D(x)$ represents the probability that x came from the training data rather than the generator. $D(G(z))$ is the probability that the output of the generator G is a real image. As Goodfellow et al. [49] described, a discriminator D tries to maximize the probability of classifying reals and fakes ($\log D(x)$), and a generator G simultaneously tried to fool a discriminator D via minimizing $\log(1 - D(G(z)))$. Therefore, we can find D and G through the following two-player minimax game, with the value function $V(G, D)$ [49]:

$$\min_{G, D} V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(x)} [\log(1 - D(G(z)))] \quad (2)$$

In order to screen the most discriminative breast cancer histopathological image patches in malignant images, the data distribution of benign image patches can be obtained by GAN, and when we use this GAN to learn the data distribution of malignant image patches, there are obvious differences, which provides the possibility to screen the most discriminative histopathological image patches and improve the classification performance of the subsequent network.

2) MAPPING NEW IMAGES TO THE LATENT SPACE

When adversarial training is completed, the generator has learned the mapping $G(z) = z \rightarrow x$ from the latent space representations z to the benign image patches x . However, GAN does not automatically generate the inverse mapping $\mu(x) = x \rightarrow z$ from the test image patches x to the latent space representations z for free, and it needs to find z iteratively [36]. The transition of the latent space is smooth. In other words, the images generated from two points at close distances in the latent space are very similar [50]. Given a malignant image patch x , we aim to find a point z in the latent space, which corresponds to the image $G(z)$ that is visually most similar to the malignant image patch x and that

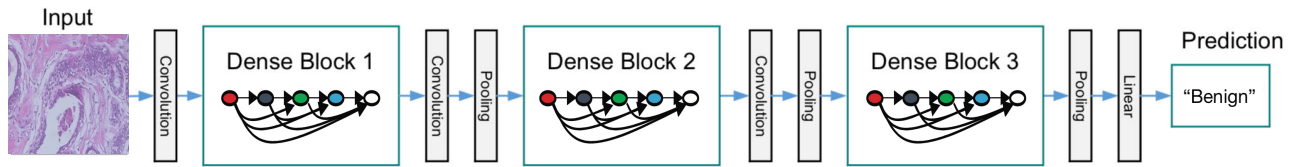


FIGURE 5. Structure diagram of breast cancer histopathology image classification based on DenseNet.

is located on the data distribution of benign image patches. Inspired by the feature matching [51], in order to find the best z , the following steps are used:

Step1: Define loss function, which represents the loss of latent space vector mapping to the image patches.

Step2: Randomly sample z_1 from the latent space distribution Z and feed z_1 into a well-trained generator to obtain the generated image $G(z_1)$. Use the loss function to calculate the loss.

Step3: Calculate the gradient of the loss function about z_1 , and use the gradient descent method to continuously update the coefficients of z_1 . During the iteration process via $\gamma = 1, 2, \dots, \Gamma$ backpropagation steps to optimize the position of z in the latent space Z . Until finding the most similar image $G(z_\Gamma)$.

3) LOSS FUNCTION

We use a loss function that maps malignant image patches to the latent space. This loss function includes two components, a residual loss and a discrimination loss.

Residual Loss The residual loss is used to measure the dissimilarity between the generated image $G(z_\gamma)$ and the malignant image patches x .

$$L_{Residual}(z_\gamma) = \sum |x - G(z_\gamma)| \tag{3}$$

For an ideal normal query situation, the image patches x and $G(z_\gamma)$ are the same. In this case, the residual loss is zero.

Discrimination Loss Inspired by the proposed feature matching technique, we regard the discriminator as a feature extractor, and the output of a certain layer of the discriminator is used as the function $f(\cdot)$ to specify the statistics of an input image. The discriminator loss reflects the difference of the extracted features by the discriminator on the two feature maps.

$$L_{Discriminator}(z_\gamma) = \sum |f(x) - f(G(z_\gamma))| \tag{4}$$

To map to the latent space, we define the total loss as the weighted sum of residual loss and discrimination loss:

$$L(z_\gamma) = (1 - \lambda) \cdot L_{Residual}(z_\gamma) + (\lambda) \cdot L_{Discriminator}(z_\gamma) \tag{5}$$

Thus, an anomaly score, which expresses the fit of a query image x to the model of benign image patches, can be directly obtained from the total loss function in Eq.(5). This model yields a large anomaly score for malignant image patches whereas a small anomaly score for benign image patches.

E. DENSELY CONNECTED CONVOLUTIONAL NETWORK TOPOLOGY

Densely connected convolutional network (DenseNet) [37] combines the advantages of ResNet [52] and Highway [53] to alleviate the vanishing-gradient problem in deep neural networks. The idea of DenseNet is to ensure maximum information flow between layers in the network. so we directly connect all layers (with matching feature-map sizes). The patch-level breast cancer histopathology images classification algorithm includes: input the most discriminative patches screened by AnoGAN, use DenseNet to extract features, and softmax classifier. First, the preprocessed image patches are used as the input of the model. During the training, DenseNet can extract the features of the patches. Finally, the extracted feature vector is sent to the softmax classifier to complete the classification of breast cancer histopathology images. The structure of breast cancer histopathology image classification model used DenseNet is shown in FIGURE 5.

The dense block is the main part of DenseNet. The main characteristic is that each layer connects to every other layer in a feed-forward fashion and passes its own feature maps to all subsequent layers. It can promote better information and gradient flow, alleviate the vanishing-gradient problem, and the network can converge better [37]. Assuming that an image patch x_0 passes through the DenseNet, the network comprises L layers, each layer implements a non-linear transformation $H(\cdot)$, and x_l is the output of the l^{th} layer. The output of the l^{th} layer is given in Eq.(6):

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]), \tag{6}$$

where $[x_0, x_1, \dots, x_{l-1}]$ refers to the concatenation of the feature-maps produced in layer $0, 1, \dots, l - 1$. $H(\cdot)$ includes three consecutive operations: batch normalization (BN) [54], rectified linear unit (ReLU) [55] and convolution (Conv). If each function $H_l(\cdot)$ produces k feature maps, the l^{th} layer consequently has $k_0 + k \times (l - 1)$ input feature maps, where k_0 is the number of channels in the input layer. The hyperparameter k is also called growth rate of the DenseNet. FIGURE 6 illustrates the structure of dense block schematically.

DenseNet is divided into multiple dense blocks. These layers between dense blocks are called transition layers, which take care of down-sampling applying a batch normalization, a 1×1 convolution, and a 2×2 average pooling.

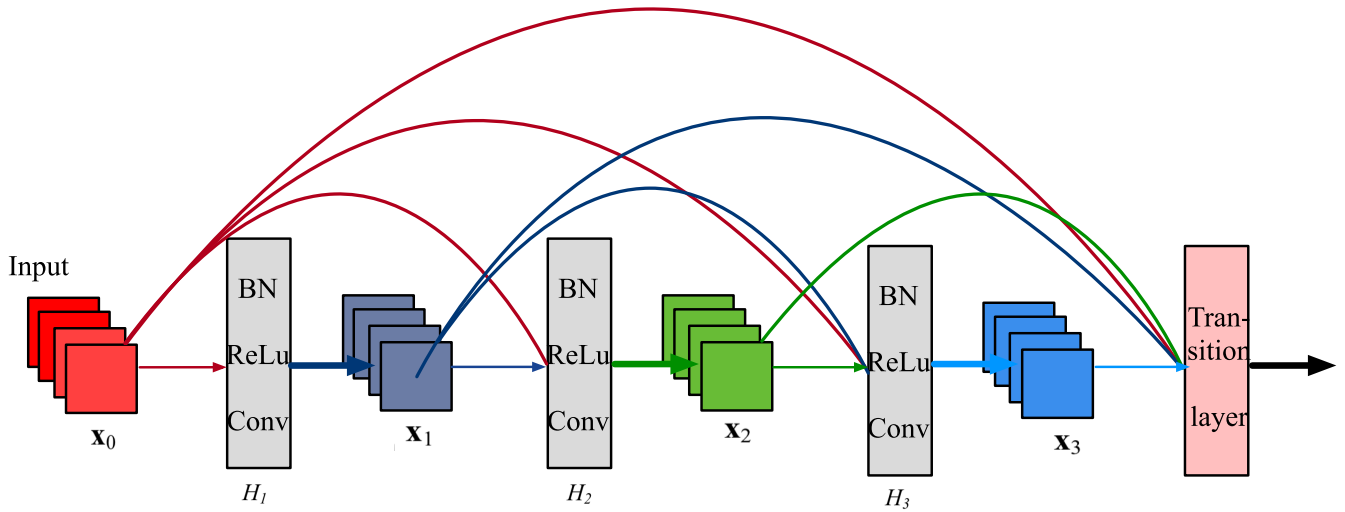


FIGURE 6. A 4-layer dense block with a growth rate of $k = 4$. Each layer takes all preceding feature-maps as input.

We define the i^{th} input image patch x_i with the label y_i . The DenseNet optimization is supervised by the softmax loss(L) [56] which can be written as

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i -\log \left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \right), \quad (7)$$

where f_j denotes the j^{th} element ($j \in [1, K]$, K is the number of classes) of the vector of class scores f , and N is the number of training image patches.

Compared with the traditional convolutional neural network, dense connectivity strengthens the feature propagation of breast cancer histopathological images, improves the information flow between the various layers, and greatly enhances the feature reuse. Therefore, the DenseNet can automatically learn the discriminative features in breast cancer histopathological images and increase the accuracy of classification.

III. EXPERIMENTS AND RESULTS

A. PERFORMANCE EVALUATION

The purpose of the proposed BreKHis dataset is to form the benchmark of breast cancer CAD systems. For this reason, BreKHis authors proposed two classification level evaluation metrics [18].

The first one is patient level accuracy that reflects the achieved performance in a patient-wise level. Let N_{np} be the number of pathological images of each patient, N_{rp} be the number of correctly classified images of each patient, and N_p be the total number of patients. The patient score for each patient is as follows:

$$PatientScore = \frac{N_{rp}}{N_{np}} \quad (8)$$

The global patient level accuracy as:

$$PatientLevelAccuracy = \frac{\sum PatientScore}{N_p} \quad (9)$$

In the second case, the evaluation metric is image level accuracy. Let N_{all} be the number of breast cancer images of testing set. If the CAD systems classify correctly N_r breast cancer images, the image level accuracy is:

$$ImageLevelAccuracy = \frac{N_r}{N_{all}} \quad (10)$$

Conventionally, during cancer diagnosis, malignant case is considered to be positive while benign case is considered to be negative. The sensitivity (also called recall) of the CAD systems is more important in clinically diagnosis. Therefore, we not only use the first two evaluation metrics but also other evaluation metrics such as precision, recall, $F1_{score}$ are used to evaluate the performance of breast cancer classification. The metrics are calculated respectively as follows:

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (11)$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (12)$$

$$F1_{score} = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (13)$$

In order to visualize the classification performance, we also use the confusion matrix that is a specific contingency table.

B. EXPERIMENTAL PROTOCOL

Following the standard labeling conventions used in medical research, the label “positive” refers to malignant images, and “negative” refers to benign images [38]. For further reducing the color inconsistency and improving efficiency in learning high-level features, stain normalization pre-processing method described in Section II-B

TABLE 2. AnoGAN architecture for screening the discriminative patches.

Modules	Layers	Output Size	Layer configurations of AnoGAN
Generator	Input: $100 \times 1 \times 1$	$100 \times 1 \times 1$	
	Transposed Convolution	$512 \times 7 \times 7$	7×7 Trans Conv, stride=1, BN, ReLU
	Transposed Convolution	$256 \times 14 \times 14$	4×4 Trans Conv, stride=2, padding = 1, BN, ReLU
	Transposed Convolution	$128 \times 28 \times 28$	4×4 Trans Conv, stride=2, padding = 1, BN, ReLU
	Transposed Convolution	$64 \times 56 \times 56$	4×4 Trans Conv, stride=2, padding = 1, BN, ReLU
	Transposed Convolution	$3 \times 112 \times 112$	4×4 Trans Conv, stride=2, padding = 1, BN, ReLU
	Transposed Convolution	$3 \times 224 \times 224$	4×4 Trans Conv, stride=2, padding = 1, Tanh
Discriminator	Input: $3 \times 224 \times 224$	$3 \times 224 \times 224$	
	Convolution	$64 \times 112 \times 112$	4×4 Conv, stride=2, padding = 1, LeakyReLU
	Convolution	$128 \times 56 \times 56$	4×4 Conv, stride=2, padding = 1, BN, LeakyReLU
	Convolution	$256 \times 28 \times 28$	4×4 Conv, stride=2, padding = 1, BN, LeakyReLU
	Convolution	$512 \times 14 \times 14$	4×4 Conv, stride=2, padding = 1, BN, LeakyReLU
	Convolution	$1024 \times 7 \times 7$	4×4 Conv, stride=2, padding = 1, BN, LeakyReLU
	Convolution	$2048 \times 1 \times 1$	7×7 Conv, stride=1
	Fully Connected	1×1	sigmoid

was employed on BreKHis datasets. In order to prevent generating random results, we used 5-fold cross validation to evaluate the proposed method for each magnification factor. We divided the BreKHis dataset into five folds and each fold contained 20% of the overall samples. During training four of the folds were used as the training set whereas the remaining set was used for testing.

We applied the random patches extraction strategy mentioned in Section II-C on the training set, so a roughly equal number of patches were generated for each category could we get. The size of the patches is 224×224 pixels because it has been shown to be particularly relevant to CNN-based classification [44]–[47].

For screening the most discriminative patches, we applied the data augmentation algorithm described in Section II-C on benign patches for training AnoGAN. The malignant patches were sent to the well-trained AnoGAN for testing, and the discriminative malignant patches were screened by the anomaly score. Then we used affine transformations mentioned in

TABLE 3. Densely Connected Convolutional Networks121 (DenseNet121) architecture for BreakHis patches. The growth rate for DenseNet121 is $k = 32$. Note that each “conv” layer shown in the table corresponds the sequence BN-ReLU-Conv.

Layers	Output size	Layer configurations of DenseNet-121(k = 32)
Input: $224 \times 224 \times 3$	$224 \times 224 \times 3$	
Convolution	$112 \times 112 \times 64$	7×7 conv, stride = 2
Max Pooling	$56 \times 56 \times 64$	3×3 max pool, stride = 2
Dense Block(1)	$56 \times 56 \times 256$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer(1)	$56 \times 56 \times 128$	1×1 conv, stride = 1
	$28 \times 28 \times 128$	2×2 average pool, stride = 2
Dense Block(2)	$28 \times 28 \times 512$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer(2)	$28 \times 28 \times 256$	1×1 conv, stride = 1
	$14 \times 14 \times 256$	2×2 average pool, stride = 2
Dense Block(3)	$14 \times 14 \times 1024$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$
Transition Layer(3)	$14 \times 14 \times 512$	1×1 conv, stride = 1
	$7 \times 7 \times 512$	2×2 average pool, stride = 2
Dense Block(4)	$7 \times 7 \times 1024$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$
Classification Layer	$1 \times 1 \times 1024$	7×7 global average pool, stride = 1
	1×2	fully connected
	1×2	softmax

Algorithm 1 to increase the number of malignant patches. Finally, we used the discriminative patches for training DenseNet121. During testing, 100 random patches with the size of 224×224 pixels were cropped from each image in the testing set. These patches were passed to the well-trained DenseNet121 and the class label of the image was obtained by majority voting from the individual patch classifications. TABLE 2 shows the details of AnoGAN architectures and TABLE 3 shows the details of DenseNet121 architectures.

Firstly, we performed 200 epochs utilizing Adam optimizer with the learning rate 0.001 for training AnoGAN. The trained parameters of the generator and discriminator were kept fixed. We ran 500 backpropagation steps for mapping malignant patches to the latent space. We set $\lambda = 0.1$ in Equations (5) (0.1 is an empirical value found in the original

TABLE 4. The mean and standard deviations of classification accuracy of DenseNet121 on BreakHis (without and with data balance). The best mean results are in bold.

Accuracy (in %)	Patch extraction strategy	Method	Magnification factors			
			40X	100X	200X	400X
Patient level	Random patches strategy with data imbalance	DenseNet121	72.43 ± 2.5	72.09 ± 1.2	68.65 ± 1.8	67.98 ± 4.3
	Random patches strategy with data balance	DenseNet121	92.02 ± 0.9	90.21 ± 2.4	81.94 ± 1.7	80.09 ± 0.7
Image level	Random patches strategy with data imbalance	DenseNet121	74.43 ± 3.9	70.09 ± 0.5	69.65 ± 2.8	64.98 ± 2.3
	Random patches strategy with data balance	DenseNet121	94.26 ± 3.2	92.71 ± 0.4	83.90 ± 2.8	82.74 ± 1.5

TABLE 5. The mean and standard deviations of classification accuracy of DenseNet121 on BreakHis (without and with AnoGAN). The best mean results are in bold.

Accuracy (in %)	Method	Magnification factors			
		40X	100X	200X	400X
Patient level	DenseNet121	92.02 ± 0.9	90.21 ± 2.4	81.94 ± 1.7	80.09 ± 0.7
	DenseNet121-AnoGAN	96.32 ± 1.3	95.98 ± 0.9	86.91 ± 2.0	85.16 ± 1.3
Image level	DenseNet121	94.26 ± 3.2	92.71 ± 0.4	83.90 ± 2.8	82.74 ± 1.5
	DenseNet121-AnoGAN	99.13 ± 0.2	96.39 ± 0.7	86.38 ± 1.2	85.20 ± 2.1

TABLE 6. The mean and standard deviations of precision, recall, and $F1_{score}$ computed from DenseNet121 on BreakHis (without and with AnoGAN). The best mean results are in bold.

Magnification factors	Method	Precision (in %)	Recall/Sensitivity (in %)	$F1_{score}$ (in %)
40X	DenseNet121	96.81 ± 0.4	95.39 ± 0.7	96.08 ± 0.3
	DenseNet121-AnoGAN	99.53 ± 0.1	99.16 ± 0.3	99.38 ± 0.4
100X	DenseNet121	95.11 ± 0.5	96.39 ± 0.2	96.48 ± 0.3
	DenseNet121-AnoGAN	97.87 ± 0.2	97.89 ± 0.6	98.02 ± 0.2
200X	DenseNet121	87.98 ± 0.3	88.66 ± 0.7	87.92 ± 0.4
	DenseNet121-AnoGAN	87.09 ± 1.1	92.22 ± 0.6	90.12 ± 0.5
400X	DenseNet121	80.97 ± 0.4	79.83 ± 0.6	80.39 ± 0.2
	DenseNet121-AnoGAN	81.37 ± 1.4	91.55 ± 0.5	87.63 ± 0.6

paper [36]). Secondly, we used Adam optimizer with a batch size of 64 to train the classification model. The learning rate was set as 0.001. Our experiments were implemented in Python using Pytorch as deep learning framework backend and conducted on three NVIDIA GeForce GTX 1080 Ti GPUs with 24GB RAM.

C. EXPERIMENTAL RESULTS

This section presents the experimental results of the proposed approach evaluated on the BreakHis dataset. In Section III-C-1, we verify the classification of BreakHis dataset affected by data imbalance. Section III-C-2 presents the performance of the proposed model (DenseNet121-AnoGAN). Additionally, the performance of AnoGAN on the existing classification networks is evaluated and presented in Section III-C-3.

1) THE IMPACT OF DATA IMBALANCE ON THE PERFORMANCE OF THE DENSENET

In this section, we experimentally evaluate the impact of data imbalance on the DenseNet121 performance in the

breast cancer histopathological images classification task. First, the random patches extraction strategy and data augmentation algorithm described in Section II-C were performed on the training set to obtain roughly equal numbers of patches in both classes (benign and malignant) of size 224×224 . These patches were used to train DenseNet121. Second, we randomly extracted 64 patches with the size of 224×224 from each image in the training set and used the affine transformations mentioned in Algorithm 1 to increase the number of training samples. DenseNet121 was trained with the training patches unevenly distributed to the two classes (benign and malignant). TABLE 4 shows the accuracy performance of the two experiments. In the case of the BreakHis dataset, especially the majority class consists of the images of malignant tissue. It can be seen that high data imbalance significantly affects the performance of DenseNet121, slight data imbalance is actually beneficial for the performance of DenseNet121. This conclusion is consistent with Koziarski [57]. Therefore, we also applied the random patches extraction strategy and data augmentation algorithm described in Section II-C in subsequent experi-

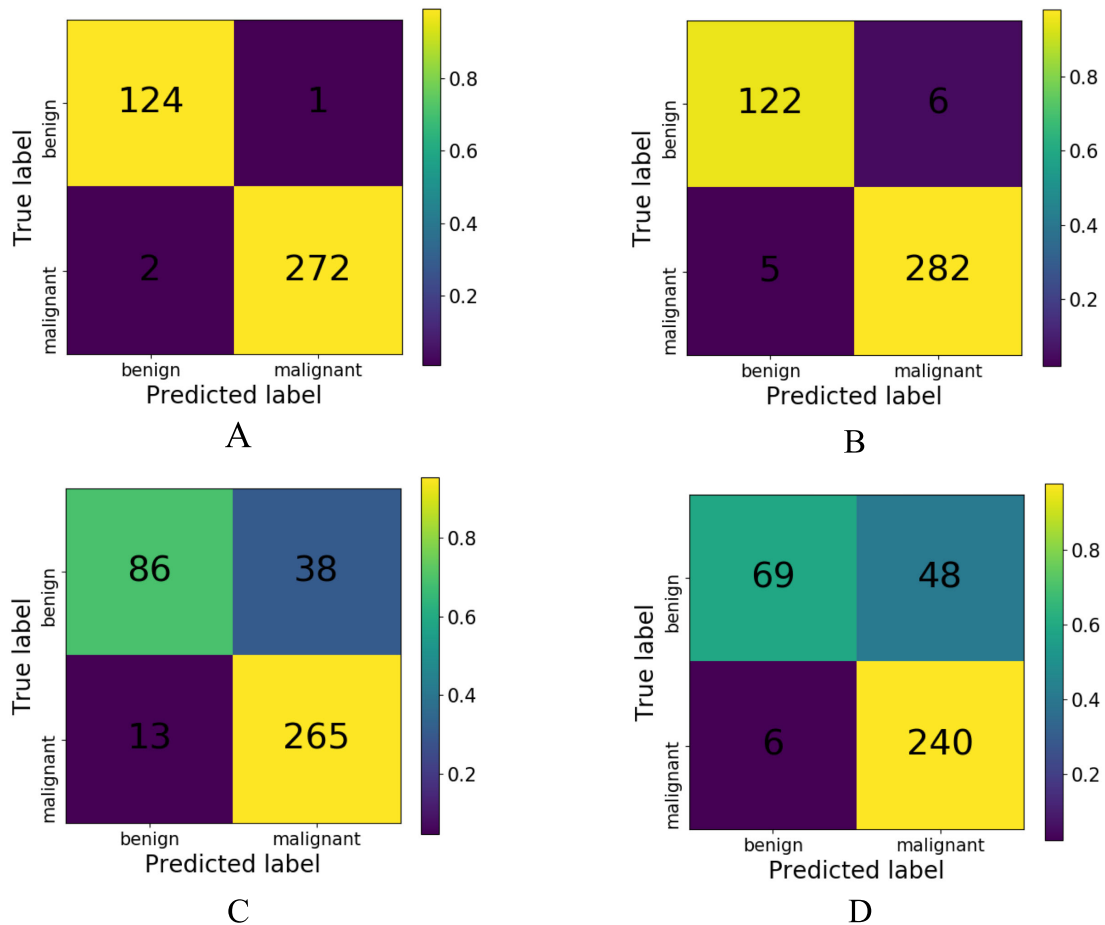


FIGURE 7. Confusion matrices of DenseNet121-AnoGAN model which has the best score in the BreakHis testing set among 5-fold cross validation, (A): Confusion matrix of 40X magnification, (B): Confusion matrix of 100X magnification, (C): Confusion matrix of 200X magnification, (D): Confusion matrix of 400X magnification.

ments. It can solve the problem of high data imbalance and increase the number of training samples.

2) THE PROPOSED MODEL RESULTS

DenseNet121-AnoGAN is a novel network for breast cancer histopathological images classification which can screen the discriminative patches and improve the performance of classification. In order to verify the effect of using the proposed approach, we conducted two sets of experiments. In the first set of experiments, we did not use AnoGAN for patches screening, and all patches were trained based on the DenseNet121. In the second set of experiments, the discriminative patches were screened by AnoGAN. Then we used these discriminative patches to train the DenseNet121. TABLE 5 provides the accuracy performance of two sets of experiments at the corresponding magnification factors. It can be noticed that there is a significant improvement in the performance of DenseNet121 when AnoGAN is employed. For breast cancer histopathological images with magnification factors of 40X, 100X, 200X, and 400X, whether it is at the patient-level accuracy or at the image-level

accuracy, the classification network with AnoGAN screening the discriminative patches has further improved the accuracy of the classification network without using AnoGAN for patches screening. The best accuracy of 99.13% has been obtained at the image level for the 40X magnification factor. In TABLE 6, the assessment of the proposed model based on evaluation metrics like precision, recall, and $F1_{score}$ are further presented. At 40X magnification factor, we achieved the best precision of 99.53%, the best recall of 99.16%, and the best $F1_{score}$ of 99.38%.

A false negative means that a subject with breast cancer is misclassified as not having the disease on the basis of the classification model. The subject is given a misleading result that she is free of breast cancer and thus does not undergo more suitable diagnostic tests. FIGURE 7 shows the confusion matrices of the DenseNet121-AnoGAN model which has the best score in testing set among 5-fold cross validation. In the confusion matrices, we can see that the proposed model produces few false negatives at all magnification factors, which proves that the proposed model can further improve the performance of the computer-aided

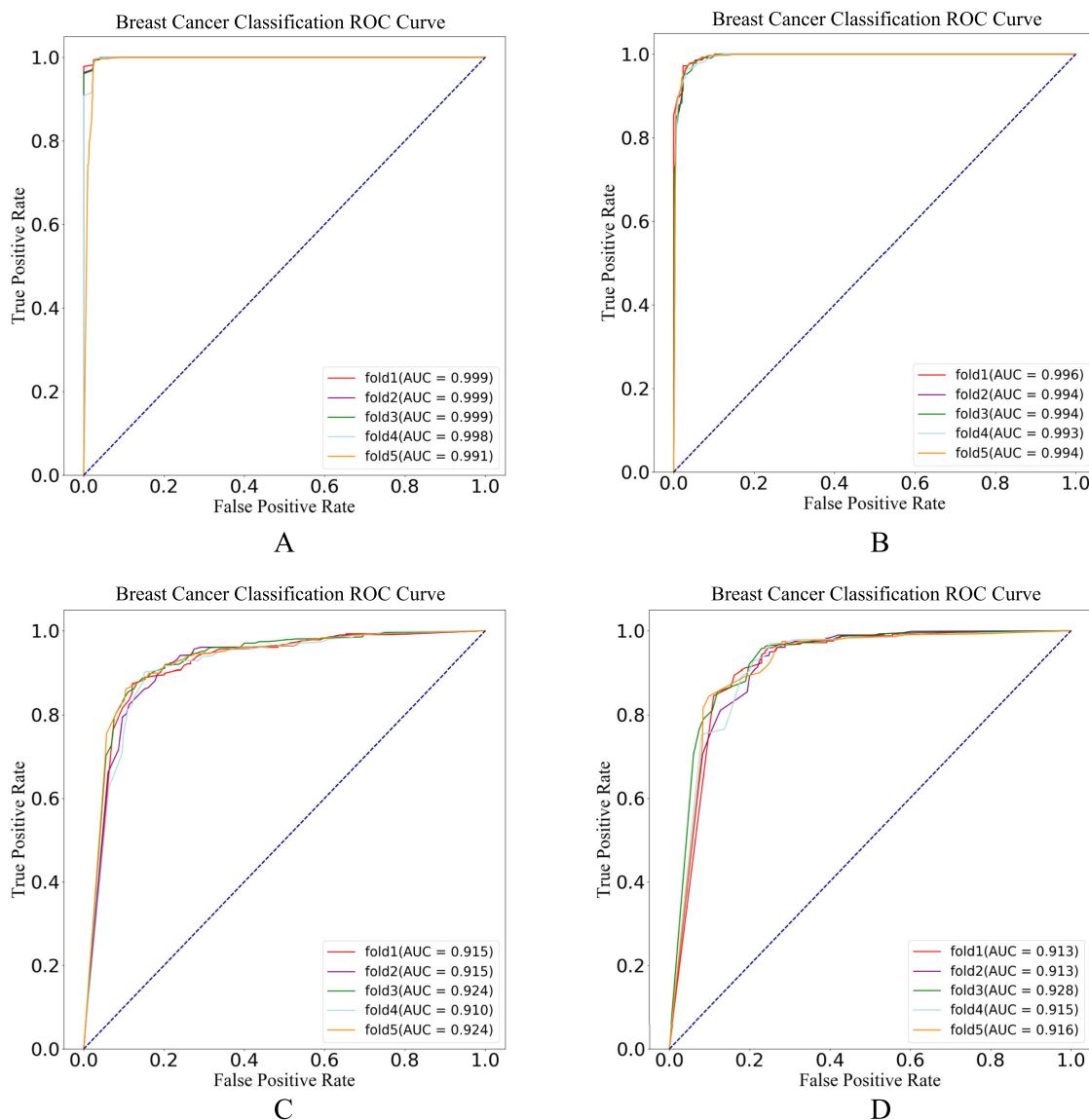


FIGURE 8. The receiver operating characteristic (ROC) curves of DenseNet121-AnoGAN models based on BreakHis dataset, (A): ROC curve of 40X magnification, (B): ROC curve of 100X magnification, (C): ROC curve of 200X magnification, (D): ROC curve of 400X magnification.

diagnosis (CAD) systems of breast cancer. The performance of DenseNet121-AnoGAN is further analyzed by using receiver operating characteristic (ROC) curves corresponding to each magnification factor (see FIGURE 8).

3) ANOGAN ON THE EXISTING CLASSIFICATION NETWORKS RESULTS

In this section, we presented the performance of the method of screening patches by AnoGAN on the other classification networks. We tested AlexNet [31], VGG16 [58], VGG19 [58], and ResNet50 [52] on the original patches as well as the discriminative patches screening by AnoGAN. The experimental results are shown in TABLE 7 (the best mean results are in bold). From the experimental results, it can

be seen that all the classification networks with AnoGAN to screen patches achieve better performance than the classification networks without AnoGAN. While comparing the performance of the classification network, it can be observed that the ResNet50-AnoGAN achieves the overall best accuracy of 86.72% at the patient level and ResNet50-AnoGAN achieves the overall best accuracy of 87.02% at the image level. In the task of breast cancer histopathological images classification, these classification networks only learn low-level features, such as colors, textures, and edges. However, DenseNet121 can concatenate features from different layers, strengthen features propagation, encourage feature reuse, and also has the narrow layers which means that the model has fewer parameters to train, so DenseNet121 makes

TABLE 7. The mean and standard deviations of classification accuracy of the existing classification networks on BreakHis (without and with AnoGAN). The best mean results are in bold.

Accuracy (in %)	Method	Magnification factors			
		40X	100X	200X	400X
Patient level	AlexNet	76.04 ± 2.3	74.19 ± 3.1	72.65 ± 1.0	70.83 ± 2.5
	AlexNet-AnoGAN	80.83 ± 3.2	81.19 ± 2.7	76.32 ± 2.9	73.71 ± 1.2
	VGG16	81.24 ± 2.8	80.94 ± 3.1	78.62 ± 3.1	76.79 ± 4.0
	VGG16-AnoGAN	83.92 ± 1.2	82.87 ± 2.6	83.39 ± 0.7	78.19 ± 3.5
	VGG19	78.46 ± 3.1	76.79 ± 1.9	75.05 ± 2.2	71.84 ± 1.7
	VGG19-AnoGAN	83.52 ± 2.7	82.49 ± 1.9	76.39 ± 1.6	75.61 ± 2.3
	ResNet50	85.74 ± 1.9	82.09 ± 0.9	78.65 ± 2.4	79.08 ± 1.2
	ResNet50-AnoGAN	86.72 ± 2.0	84.61 ± 1.1	82.09 ± 2.9	82.11 ± 2.4
Image level	AlexNet	74.34 ± 1.3	71.85 ± 2.8	63.68 ± 2.3	62.08 ± 2.5
	AlexNet-AnoGAN	75.62 ± 2.1	72.91 ± 1.7	68.32 ± 2.4	69.89 ± 3.2
	VGG16	82.64 ± 1.9	83.09 ± 2.0	81.05 ± 2.1	80.18 ± 2.3
	VGG16-AnoGAN	84.22 ± 0.9	84.21 ± 1.2	82.65 ± 2.3	80.13 ± 1.4
	VGG19	79.63 ± 2.3	81.01 ± 1.4	72.57 ± 2.5	73.82 ± 3.1
	VGG19-AnoGAN	83.64 ± 0.7	84.01 ± 1.9	78.07 ± 1.5	74.89 ± 2.0
	ResNet50	84.64 ± 0.8	83.09 ± 1.8	84.06 ± 2.8	80.71 ± 3.4
	ResNet50-AnoGAN	87.02 ± 1.2	84.12 ± 0.9	82.32 ± 1.4	82.19 ± 1.6

the classification task easier and more efficient to train than any other network.

IV. DISCUSSION

Breast cancer is one of the common types among hundreds of cancer diseases. The incidence of this disease is increasing day by day, especially among women. If the disease is not diagnosed in time, the mortality rate will be fairly high. In this work, we propose a novel approach for the classification of breast cancer histopathology images, named DenseNet121-AnoGAN. Many researchers have conducted studies on the BreakHis dataset. The performance comparison of the proposed model with the existing studies using the BreakHis dataset is shown in TABLE 8.

Compared with all the studies given in TABLE 8, our proposed model obtained the best performance for 40X and 100X histopathology images. In particular, the accuracy at the image level of our proposed model for the 40X magnification factor is 99.13%, the best precision is 99.53%, the best recall is 99.16%, and the best $F1_{score}$ is 99.38%. The classification performance of our proposed model has clearly outperformed the methodologies of Spanhol *et al.* [18], Spanhol *et al.* [30], Spanhol *et al.* [59], and Kumar and Rao [60]. As can be seen from TABLE 8, our proposed model has the best performance in low-level magnifications *i.e.* 40X and 100X compared with the methodologies of Gupta and Bhavsar [61], Sudharshan *et al.* [38], and Gour *et al.* [62].

This is the first attempt that we use AnoGAN for screening discriminative patches to deal with the mislabeled patches. We use benign patches to train AnoGAN. The data distribution of benign patches can be obtained by

AnoGAN and it will generate a fake patch with a probability distribution similar to that of the benign patch. By defining the threshold of residual loss and discrimination loss between the malignant patch to be tested and fake patch, this well-trained AnoGAN can yield the high anomaly score of the malignant patches. However, the anomaly score of mislabeled patches in malignant patches is low. Therefore, we can use the obvious differences produced by AnoGAN to screen the discriminative patches in malignant patches and improve the classification performance of the subsequent network. Compared with the patch-based image classification methods, the proposed approach named DenseNet121-AnoGAN can effectively solve the problem of mislabeled patches in malignant patches and improve the classification performance. Nuclei and tissue organization are related to the diagnosing process [32]. As the magnification increases, the number of nuclei you are able to see in the patches will decrease, resulting in incomplete nuclei edge-related features extracted. The classification network is based on the extracted features from different scales, including nuclei and tissues organization. If the classification network cannot extract these relevant features at high-level magnifications, the accuracy of classification network will decline. Although compared with the existing studies, the proposed model is insufficient in the classification performance of high-level magnifications, our proposed DenseNet121-AnoGAN can be better suited coarse-grained high-resolution images from breast tissue biopsy slides stained with hematoxylin and eosin (H&E) and achieved satisfactory classification performance at 40X and 100X magnifications. It has laid a foundation for helping pathologists to diagnose diseases in the future.

TABLE 8. Performance comparison with existing studies using the BreakHis dataset in terms of patient level accuracy and image level accuracy.

Accuracy(in %)	Authors	Year	Preprocessing	Model	Magnification factors			
					40X	100X	200X	400X
Patient level	Spanhol <i>et al.</i> [18]	2016	None	Descriptor: PFTAS Classifier: QDA	83.8 ± 4.1	82.1 ± 4.9	85.1 ± 3.1	82.3 ± 3.8
	Spanhol <i>et al.</i> [30]	2016	Res(350 × 230)	Pre-trained AlexNet Model	88.6 ± 5.6	84.5 ± 2.4	85.3 ± 3.8	81.7 ± 4.9
	Spanhol <i>et al.</i> [59]	2017	None	DeCAF Model	84.0 ± 6.9	83.9 ± 5.9	86.3 ± 3.5	82.1 ± 2.4
	Kumar <i>et al.</i> [60]	2018	None	New Designed CNN Model	83 ± 3.2	81 ± 4.2	84.2 ± 3.4	81 ± 2.4
	Gupta <i>et al.</i> [61]	2018	Res	Sequential framework with deep multi-layered features	94.71 ± 0.88	95.9 ± 4.2	96.76 ± 1.09	89.11 ± 0.12
	Sudharshan <i>et al.</i> [38]	2019	None	Descriptor: PFTAS Classifier: NPMIL	92.1 ± 5.9	89.1 ± 5.2	87.2 ± 4.3	82.7 ± 3.0
	Gour <i>et al.</i> [62]	2020	DA(Rot, Flip)	ResHist Model	87.4 ± 3.3	88.1 ± 2.9	92.5 ± 2.8	87.7 ± 2.4
	Proposed		DA(Rot, Flip)	DenseNet121-AnoGAN Model	96.32 ± 1.3	95.98 ± 0.9	86.91 ± 2.0	85.16 ± 1.3
Image level	Spanholet <i>et al.</i> [30]	2016	Res(350 × 230)	Pre-trained AlexNet Model	89.6 ± 6.5	85.0 ± 4.8	84.0 ± 3.2	80.8 ± 3.1
	Spanhol <i>et al.</i> [59]	2017	None	DeCAF Model	84.6 ± 2.9	84.8 ± 4.2	84.2 ± 1.7	81.6 ± 3.7
	Kumar <i>et al.</i> [60]	2018	None	New Designed CNN Model	82 ± 2.8	86.2 ± 4.6	84.6 ± 3	84 ± 4
	Sudharshan <i>et al.</i> [38]	2019	None	Descriptor: PFTAS Classifier: NPMIL	87.8 ± 5.6	85.6 ± 4.3	80.8 ± 2.8	82.9 ± 4.1
	Gour <i>et al.</i> [62]	2020	SN + DA(Rot, Flip)	ResHist Model	87.4 ± 3.0	87.2 ± 3.5	91.1 ± 2.3	86.2 ± 2.1
		Proposed		SN+DA(Rot, Flip)	DenseNet121-AnoGAN Model	99.13 ± 0.2	96.39 ± 0.7	86.38 ± 1.2

Note that "Res" refers to resizing original images to a new size, "DA" refers to data augmentation, "Rot" refers to rotation, "Flip" refers to flipping, "SN" refers to stain normalization.

Note that "PFTAS" refers to Parameter-Free Threshold Adjacency Statistics, "QDA" refers to Quadratic Linear Analysis, "NPMIL" refers to Non-parametric Multi-instance learning.

V. CONCLUSION

In this paper, we propose a novel approach, named DenseNet121-AnoGAN, for the classification of breast cancer histopathology images, which screens patches based on an unsupervised anomaly detection with generative adversarial networks (AnoGAN). The proposed model can effectively solve the problem of mislabeled patches when we adopt a patch-based classification method and improve the performance of classification. We have experimentally evaluated the proposed model for binary classification using 5-fold cross validation in the BreakHis dataset at four different magnification factors (40X, 100X, 200X, 400X). The best accuracy of 99.13% and the best $F1_{score}$ of 99.38% have been obtained at the image level for the 40X magnification factor. In addition, we have also preliminarily explored the impact of data imbalance on the classification network and investigated the performance of the method of screening patches by AnoGAN on the other classification networks, including AlexNet, VGG16, VGG19, and ResNet50.

For breast cancer histopathological images with the magnification factors of 40X, 100X, 200X, and 400X, our experiments show that whether it is at the patient-level accuracy or the image-level accuracy, the method of screening the discriminative patches by AnoGAN has further improved the accuracy of the method without using AnoGAN for patches screening.

Although the proposed model is very effective for breast cancer diagnosis in low-level magnifications *i.e.* 40X and 100X, future work can explore different activation functions in the final layer of CNN architectures, the optimization of the hyperparameters, and the size of patches to improve the accuracy at high-level magnification. Furthermore, the problem of data imbalance is ubiquitous in the medical domain, we should explore some approaches for dealing with data imbalance in the future. Before being used for clinical diagnosing, we need to validated on the other breast cancer histopathological image datasets.

REFERENCES

- [1] *Latest Global Cancer Data: Cancer Burden Rises to 18.1 Million New Cases and 9.6 Million Cancer Deaths in 2018*, Int. Agency Res. Cancer, Lyon, France, 2018.
- [2] C. Wild, E. Weiderpass, and B. Stewart, *World Cancer Report: Cancer Research for Cancer Prevention*. Lyon, France: The International Agency for Research on Cancer (IARC), 2020. [Online]. Available: <https://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-Cancer-Research-For-Cancer-Prevention-2020>
- [3] S. C. DeWit, H. Stromberg, and C. Dallred, *Medical-Surgical Nursing: Concepts & Practice*. Amsterdam, The Netherlands: Elsevier, 2016.
- [4] D. Bardou, K. Zhang, and S. M. Ahmad, "Classification of breast cancer based on histology images using convolutional neural networks," *IEEE Access*, vol. 6, pp. 24680–24693, 2018.
- [5] Institute of Medicine (US) and National Research Council, *Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer*. Washington, DC, USA: National Academies Press, 2001.
- [6] J. R. Won, D. Gao, C. Chow, J. Cheng, S. Y. Lau, M. J. Ellis, C. M. Perou, P. S. Bernard, and T. O. Nielsen, "A survey of immunohistochemical biomarkers for basal-like breast cancer against a gene expression profile gold standard," *Modern Pathol.*, vol. 26, no. 11, pp. 1438–1450, Nov. 2013.
- [7] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.
- [8] G. Aresta, T. Araújo, S. Kwok, S. S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Prastawa, M. Chan, M. Donovan, G. Fernandez, J. Zeineh, M. Kohl, C. Walz, F. Ludwig, S. Braunewell, M. Baust, Q. D. Vu, and P. Aguiar, "BACH: Grand challenge on breast cancer histology images," *Med. Image Anal.*, vol. 56, pp. 122–139, Aug. 2019.
- [9] F. Oloumi, R. M. Rangayyan, P. Casti, and A. L. Ells, "Computer-aided diagnosis of plus disease via measurement of vessel thickness in retinal fundus images of preterm infants," *Comput. Biol. Med.*, vol. 66, pp. 316–329, Nov. 2015.
- [10] Z. Gandomkar, P. Brennan, and C. Mello-Thoms, "Computer-based image analysis in breast pathology," *J. Pathol. Informat.*, vol. 7, no. 1, p. 43, 2016.
- [11] J. Xie, R. Liu, J. Luttrell, and C. Zhang, "Deep learning based analysis of histopathological images of breast cancer," *Frontiers Genet.*, vol. 10, p. 80, Feb. 2019.
- [12] C. Kaushal, S. Bhat, D. Koundal, and A. Singla, "Recent trends in computer assisted diagnosis (CAD) system for breast cancer diagnosis using histopathological images," *IRBM*, vol. 40, no. 4, pp. 211–227, Aug. 2019.
- [13] A. Das, M. S. Nair, and S. D. Peter, "Computer-aided histopathological image analysis techniques for automated nuclear atypia scoring of breast cancer: A review," *J. Digit. Imag.*, pp. 1–31, Jan. 2020, doi: [10.1007/s10278-019-00295-z](https://doi.org/10.1007/s10278-019-00295-z).
- [14] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, and R. Monczak, "Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images," *Comput. Biol. Med.*, vol. 43, no. 10, pp. 1563–1572, Oct. 2013.
- [15] Y. M. George, H. H. Zayed, M. I. Roushdy, and B. M. Elbagoury, "Remote computer-aided breast cancer detection and diagnosis system based on cytological images," *IEEE Syst. J.*, vol. 8, no. 3, pp. 949–964, Sep. 2014.
- [16] A. D. Belsare, M. M. Mushrif, M. A. Pangarkar, and N. Meshram, "Classification of breast cancer histopathology images using texture feature analysis," in *Proc. TENCON IEEE Region 10 Conf.*, Nov. 2015, pp. 1–5.
- [17] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.
- [18] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016.
- [19] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [20] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jun. 2010.
- [21] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Proc. Int. Conf. Image Signal Process.* Cham, Switzerland: Springer, 2008, pp. 236–243.
- [22] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
- [23] L. P. Coelho, A. Ahmed, A. Arnold, J. Kangas, A.-S. Sheikh, E. P. Xing, W. W. Cohen, and R. F. Murphy, "Structured literature image finder: Extracting information from text and images in biomedical literature," in *Linking Literature, Information, and Knowledge for Biology*. Cham, Switzerland: Springer, 2010, pp. 23–32.
- [24] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571.
- [25] Y. Benhammou, B. Achchab, F. Herrera, and S. Tabik, "BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights," *Neurocomputing*, vol. 375, pp. 9–24, Jan. 2020.
- [26] M. Hammad, P. Pławiak, K. Wang, and U. R. Acharya, "ResNet—Attention model for human authentication using ECG signals," *Expert Syst.*, p. e12547, Mar. 2020, doi: [10.1111/exsy.12547](https://doi.org/10.1111/exsy.12547).
- [27] M. Hammad, Y. Liu, and K. Wang, "Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint," *IEEE Access*, vol. 7, pp. 26527–26542, 2018.
- [28] A. Alghamdi, M. Hammad, H. Ugail, A. Abdel-Raheem, K. Muhammad, H. S. Khalifa, and A. A. El-Latif, "Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities," *Multimedia Tools Appl.*, pp. 1–22, Mar. 2020, doi: [10.1007/s11042-020-08769-x](https://doi.org/10.1007/s11042-020-08769-x).
- [29] M. Hammad, S. Zhang, and K. Wang, "A novel two-dimensional ECG feature extraction and classification algorithm based on convolution neural network for human authentication," *Future Gener. Comput. Syst.*, vol. 101, pp. 180–196, Dec. 2019.
- [30] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2560–2567.
- [31] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [32] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, "Classification of breast cancer histology images using convolutional neural networks," *PLoS ONE*, vol. 12, no. 6, Jun. 2017, Art. no. e0177544.
- [33] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2424–2433.
- [34] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," *J. Digit. Imag.*, vol. 32, no. 4, pp. 605–617, Aug. 2019.
- [35] J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead, "Winner-take-all networks of O(n) complexity," in *Proc. Adv. Neural Inf. Process. Syst.*, 1989, pp. 703–711.
- [36] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2017, pp. 146–157.
- [37] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4700–4708.
- [38] P. J. Sudharshan, C. Petitjean, F. Spanhol, L. E. Oliveira, L. Heutte, and P. Honeine, "Multiple instance learning for histopathological breast cancer image classification," *Expert Syst. Appl.*, vol. 117, pp. 103–111, Mar. 2019.
- [39] E. Reinhard, M. Adhikmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 4, pp. 34–41, Jul. 2001.
- [40] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symp. Biomed. Imag. From Nano to Macro*, Jun. 2009, pp. 1107–1110.
- [41] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Trans. Med. Imag.*, vol. 35, no. 8, pp. 1962–1971, Aug. 2016.

- [42] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.
- [43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [44] K. Das, S. P. K. Karri, A. Guha Roy, J. Chatterjee, and D. Sheet, "Classifying histopathology whole-slides using fusion of decisions from deep convolutional network on a collection of random multi-views at multi-magnification," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 1024–1027.
- [45] W. Zhi, H. W. F. Yueng, Z. Chen, S. M. Zandavi, Z. Lu, and Y. Y. Chung, "Using transfer learning with convolutional neural networks to diagnose breast cancer from histopathological images," in *Proc. Int. Conf. Neural Inf. Process. Cham, Switzerland: Springer*, 2017, pp. 669–676.
- [46] K. Das, S. Conjeti, A. G. Roy, J. Chatterjee, and D. Sheet, "Multiple instance learning of deep convolutional neural networks for breast histopathology whole slide classification," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 578–581.
- [47] Z. Gandomkar, P. C. Brennan, and C. Mello-Thoms, "MuDeRN: Multi-category classification of breast histopathological image using deep residual networks," *Artif. Intell. Med.*, vol. 88, pp. 14–24, Jun. 2018.
- [48] Y. Ono, E. Trulls, P. Fua, and K. M. Yi, "LF-Net: Learning local features from images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 6234–6244.
- [49] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [50] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," 2016, *arXiv:1607.07539*. [Online]. Available: <http://arxiv.org/abs/1607.07539>
- [51] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2234–2242.
- [52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [53] R. Kumar Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," 2015, *arXiv:1505.00387*. [Online]. Available: <http://arxiv.org/abs/1505.00387>
- [54] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: <http://arxiv.org/abs/1502.03167>
- [55] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
- [56] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," in *Proc. ICML*, 2016, vol. 2, no. 3, p. 7.
- [57] M. Koziarski, "Radial-based undersampling algorithm for classification of breast cancer histopathological images affected by data imbalance," in *Proc. 12th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2019, pp. 1–5.
- [58] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [59] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, "Deep features for breast cancer histopathological image classification," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2017, pp. 1868–1873.
- [60] K. Kumar and A. C. S. Rao, "Breast cancer classification of image using convolutional neural network," in *Proc. 4th Int. Conf. Recent Adv. Inf. Technol. (RAIT)*, 2018, pp. 1–6.
- [61] V. Gupta and A. Bhavsar, "Sequential modeling of deep features for breast cancer histopathological image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 2254–2261.
- [62] M. Gour, S. Jain, and T. Sunil Kumar, "Residual learning based CNN for breast cancer histopathological image classification," *Int. J. Imag. Syst. Technol.*, vol. 30, no. 3, pp. 621–635, Sep. 2020.



RUI MAN was born in Dezhou, China, in 1996. She received the bachelor's degree from Qufu Normal University, Qufu, China, in 2018. She is currently pursuing the master's degree with Beijing Union University. Her research interests include medical image processing and deep learning.



PING YANG received the Ph.D. degree in control theory and control engineering from the China University of Mining and Technology, Beijing, in 2006. She is currently an Associate Professor with the Smart City College, Beijing Union University. Her works have been published in several international conferences and journals. Her research interests include information acquisition and processing, image processing, and the Internet of Things.



BOWEN XU was born in Dezhou, China, in 1996. He received the bachelor's degree from the Shandong University of Science and Technology, Qingdao, China, in 2018. He is currently pursuing the master's degree with the Beijing University of Technology. His research interests include anomaly detection and water quality time series prediction.

...