IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DEEP LEARNING FOR COMPUTER-AIDED MEDICAL DIAGNOSIS

Received January 6, 2019, accepted January 28, 2019, date of publication February 7, 2019, date of current version February 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2898044

Classification of Breast Cancer Histology Images Using Multi-Size and Discriminative Patches Based on Deep Learning

YUQIAN LI[1], JUNMIN WU[1,2], AND QISONG WU[3]

[1]College of Computer Science and Technology, University of Science and Technology of China, Hefei 230022, China
[2]College of Software, University of Science and Technology of China, Suzhou 215123, China
[3]Pathology Department, Suzhou Kowloon Hospital Institute, Suzhou 215123, China

Corresponding author: Junmin Wu (jmwu@ustc.edu.cn)

This work was supported by the National Key Research and Development Program of China under Grant 2016YFB1000403.

ABSTRACT The diagnosis of breast cancer histology images with hematoxylin and eosin stained is non-trivial, labor-intensive and often leads to a disagreement between pathologists. Computer-assisted diagnosis systems contribute to help pathologists improve diagnostic consistency and efficiency. With the recent advances in deep learning, convolutional neural networks (CNNs) have been successfully used for histology images analysis. The classification of breast cancer histology images into normal, benign, and malignant sub-classes is related to cells' density, variability, and organization along with overall tissue structure and morphology. Based on this, we extract both smaller and larger size patches from histology images, including cell-level and tissue-level features, respectively. However, there are some sampled cell-level patches that do not contain enough information that matches the image tag. Therefore, we propose a patches' screening method based on the clustering algorithm and CNN to select more discriminative patches. The approach proposed in this paper is applied to the 4-class classification of breast cancer histology images and achieves 95% accuracy on the initial test set and 88.89% accuracy on the overall test set. The results are competitive compared to the results of other state-of-the-art methods.

INDEX TERMS Breast cancer histology images, multi-size patches, discriminating patches, CNN, image classification.

I. INTRODUCTION

Breast cancer is the most common cancer and the second main cause of cancer death in women, after lung cancer. The chance of any woman dying from breast cancer is around 1 in 37, or 2.7 percent [1]. The diagnosis and treatment of breast cancer in the early stage is crucial to reduce the morbidity rates and prevent the progression of the disease.

The diagnosis from a histology image is the gold standard in diagnosing considerable types of cancer. Pathologists analyze the regularities of cell shapes, density, and tissue structures by examining a thin slice of tissue under an optical microscope and determine cancerous regions and malignancy degree. Due to the complexity and diversity of histology images, the manual examination requires abundant knowledge and experience of the pathologists and is fairly time-consuming and error-prone [2]–[4]. The subjective of the application of morphological criteria in histology images classification leads to the result that the average diagnostic concordance between specialists is approximately 75% [5]. Researchers in the pathology fields have recognized the necessity of quantitative analysis of histology images. Furthermore, tissue histology slides can be digitized by using whole slide digital scanners and stored in digital image form. Consequently, computer-assisted diagnosis (CAD) algorithms have begun to be developed for disease detection, diagnosis, and prognosis prediction. CAD systems contribute to complement the diagnosis of the pathologist, eliminate inter-pathologist variations in diagnosis, and improve the efficiency [6].

Historically, analyses of digital histology images have focused on low-level image analysis tasks primarily, such as nuclei segmentation and feature engineering, followed by classical classification models, including random forests (RF)

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

and support vector machines (SVM). Considerable numbers of works focused on the analysis of nuclei morphology and tissue structure for breast cancer histology images classification. Kowal *et al.* [7] tested and compared four different clustering algorithms for nuclei segmentation on 500 breast microscopic images primarily, followed by extracting 42 morphological, topological and texture features used in a classification procedure with three different classifiers. George *et al.* [8] and Filipczuk *et al.* [9] detected the locations of the nuclei with circular Hough transform, followed by false-positive elimination using Otsu's thresholding and other methods. After accomplishing the segmentation of the nuclei, shape-based features and textural features were extracted for classical classification models used on 92 and 737 breast cytological images respectively. Wang *et al.* [10] focused on the regions of interest (ROIs) primarily, then split overlapped cells. Similarly, 4 shape-based features and 138 textural features based on color spaces were extracted on 68 images for support vector machine. Aforementioned works focus on nuclei segmentation methods. There are also some works focused on the features extracted from the whole image additionally. For example, Naik *et al.* [11] proposed a methodology that integrated information from low-level information based on pixel values, high-level information based on relationships between pixels, and domain-specific information based on relationships between histological structures for detection and segmentation of structures of interest. Morphological and nuclear features were extracted for SVM after using the segmentation algorithm.

The performance of the aforementioned works relies on appropriate data representation. Much of the efforts are dedicated to feature engineering which is a labor-intensive process that uses abundant expert domain knowledge to extract useful features. Moreover, these works focused on the classification of low resolution breast cancer histology images in small datasets into benign or malignant.

Over the past decade, dramatic increases in computational power and improvement in deep learning [12], especially Convolutional Neural Networks (CNN) [13], have allowed the development of computer-assisted analytical approaches to image analysis field [14]–[17], including medical histology images. Contrarily to the hand-crafted feature extraction methods, CNNs learn features directly from the histology images. Nowadays, more and more institutions provided datasets containing multi-class and high resolution histology images. It is unrealistic to train a CNN with the extremely large size of a histology image. Moreover, rescaling the entire histology image to the input size for CNN directly will lose vast of detail information. Consequently, lost of works adopted a patch sampling method in order to extract CNN activation features of sampled patches and preserved essential information for multi-class classification. In study, Spanhol *et al.* [18], [19] constructed a dataset of 7909 breast cancer histology images named BreakHis acquired on 82 patients. They trained AlexNet based on the extraction of patches obtained randomly or by a sliding window mechanism from breast cancer images with multiple magnifications and combined the patch probabilities with three fusion rules for final classification. Araújo *et al.* [20] proposed a CNN architecture designed for extracting features from patches of 512 × 512 pixels and performed 4-class classification based on 249 high resolution images released for the bioimaging 2015 breast cancer histology classification challenge [21]. The patch extraction strategy allowed CNNs to process the Whole Slide Histopathological Images (WSIs) with extremely-high resolution. Wang *et al.* [22] used sampling patches to train a CNN to make patch-level predictions, then aggregated the results to create tumor probability heatmaps and made slide-level predictions. The methodology was tested on the Camelyon16 dataset including 400 WSIs [23]. In [24], context-aware stacked convolutional neural networks for 3-class classification of breast WSIs were presented. Bejnordi *et al.* used a CNN trained by high pixel resolution patches to extract cell-level features primarily, followed by a second CNN. Then, large input patches were used to train the stacked CNNs to learn both cellular information and global tissue structures.

In the work herein described, histology image classification was performed by processing several patches with fixed size. Microscopically, cancer cells have distinguishing histological features. The nucleus is often large and irregular, and the cytoplasm may also displays atypia. Furthermore, there are clear structural differences between diseased tissues and normal tissues. In addition, different from *in situ* carcinoma within a particular tissue compartment, *invasive* carcinoma refers to malignant abnormal proliferation of neoplastic cells in the breast tissue, which has penetrated into stroma [25]. Therefore, referring to the pathologists' diagnostic process, features related to cells and global tissue structures extracted from two kinds of patches with different sizes will improve the performance of the classification of breast cancer histology images into one of the 4 target classes. The labels of histology images for the classification task given by the pathologists are based on the whole images. Larger size patches sampled from a histology image contain sufficient information so that the image label can be used for the patches. However, cell-level patches extracted from high resolution histology images, especially ultra-high resolution WSIs, may not contain sufficient diagnostic information. There exists some patches with large areas of fat cells and stroma, sparse breast cells, and normal patches extracted from malignant histology images. CNNs trained by these patches can't extract discriminative features. Consequently, we present a methodology to automatically screen more discriminative patches based on clustering algorithm and convolutional neural network. Based on the above two aspects, the main objective of this paper is to propose a comprehensive and effective scheme for the multi-classification of breast histology images in order to improve the diagnostic performance.

To achieve this, the main contributions of our work can be summarized as follow: (i) We propose a patch sampling

strategy to extract two kinds of patches with different sizes to preserve essential information and contain cell-level and tissue-level features respectively., (ii) We design a patch selecting method to select more discriminative patches based on CNN and K-means., (iii) We design a classification framework which extracts features from the patches using the feature extractors and compute the final feature of each whole image for classification through a classifier.

The remainder of this paper is organized in the following manner. In Section II, we introduce the information about the data, while Section III is dedicated to introducing the framework of the proposed scheme and the implementation of main methods. The experiments and performance comparison are discussed in Section IV. In Section VI, we summarize the conclusion of this paper.

II. DATASET

This section is dedicated to introducing the dataset used in our work and pre-processing of images. The dataset is from the bioimaging 2015 breast histology classification challenge [21], composed of high-resolution (2048×1536 pixels) and H&E stained breast cancer histology images. The images were digitized with a magnification of 200x and pixel size of $0.42\mu m \times 0.42\mu m$. Two pathologists labeled images as normal, benign, *in situ* carcinoma or *invasive* carcinoma according to the predominant cancer type in each image, without specifying the area of interest. Fig. 1 illustrates images from each class mentioned in the dataset.

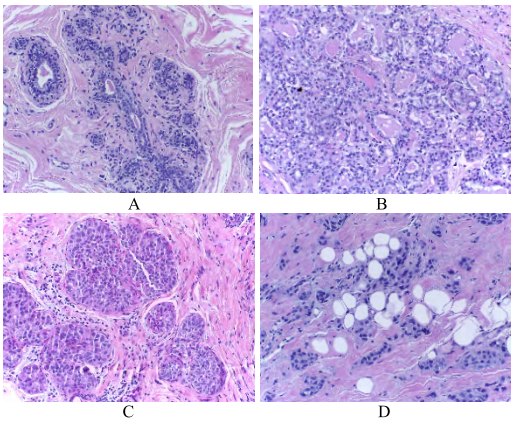


FIGURE 1. H&E stained images from each type, (A): normal tissue, (B): benign abnormality, (C): *in situ* carcinoma, and (D): *invasive* carcinoma.

This dataset composed of a training set of 249 images, an initial test set of 20 images and an extended test set of 16 images with increased ambiguity is publicly available at https://rdm.inesctec.pt/dataset/nis-2017-003. The numbers of images in the four categories are shown in TABLE 1.

The main goal of this paper is to propose an effective scheme for the 4-class breast histology images classification.

TABLE 1. The number of breast histology images per class.

Class	Training set	Test Set	
		Initial	Extended
Normal	55	5	4
Benign	69	5	4
In Situ	63	5	4
Invasive	62	5	4
Overall	249	20	16

A. PRE-PROCESSING

Stain inconsistency of histology images, due to differences in color responses of slide digital scanners, will affect the performance of image analysis. As can be seen from Fig. 1, the images in the dataset have large stain variation. To this end, stain normalization is essential prior to other processes. There are various research for stain normalization in histology images [26], [27].

In this paper, we use a method proposed by Reinhard *et al.* [28], which transforms the RGB images to the decorrelated $l\alpha\beta$ color space, followed by computing the means and standard deviations for each channel separately in $l\alpha\beta$ space and a set of linear transforms in order to match the color distribution of the source and target images, finally, converts the results back to RGB. Fig. 2 illustrates the effect of the method on a breast histology image.

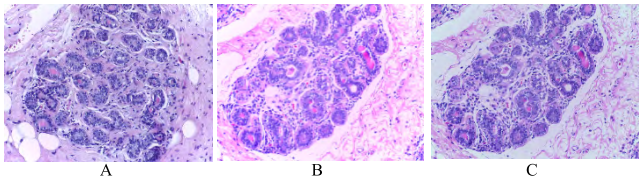


FIGURE 2. Image normalization. (A): The target image, (B): original image, (C): image after normalization.

III. MATHODOLOGY

The multi-classification scheme of breast histology images is presented in this section. We introduce the overall framework at first, and then describe each process in detail.

A. THE FRAMEWORK

Fig. 3 illustrates the framework of our approach used for multi-class classification of breast histology images. The main processes can be summarized as follow: (i) We extract two kinds of patches with different sizes by a sliding window mechanism from breast cancer histology images to preserve essential information and contain cell-level and tissue-level features, and then train two CNNs as feature extractors respectively. (ii) We split the small patches into multiple clusters using k-means clustering algorithm and select more discriminative patches based on the network trained by small patches to retrain the network. (iii) We extract features from

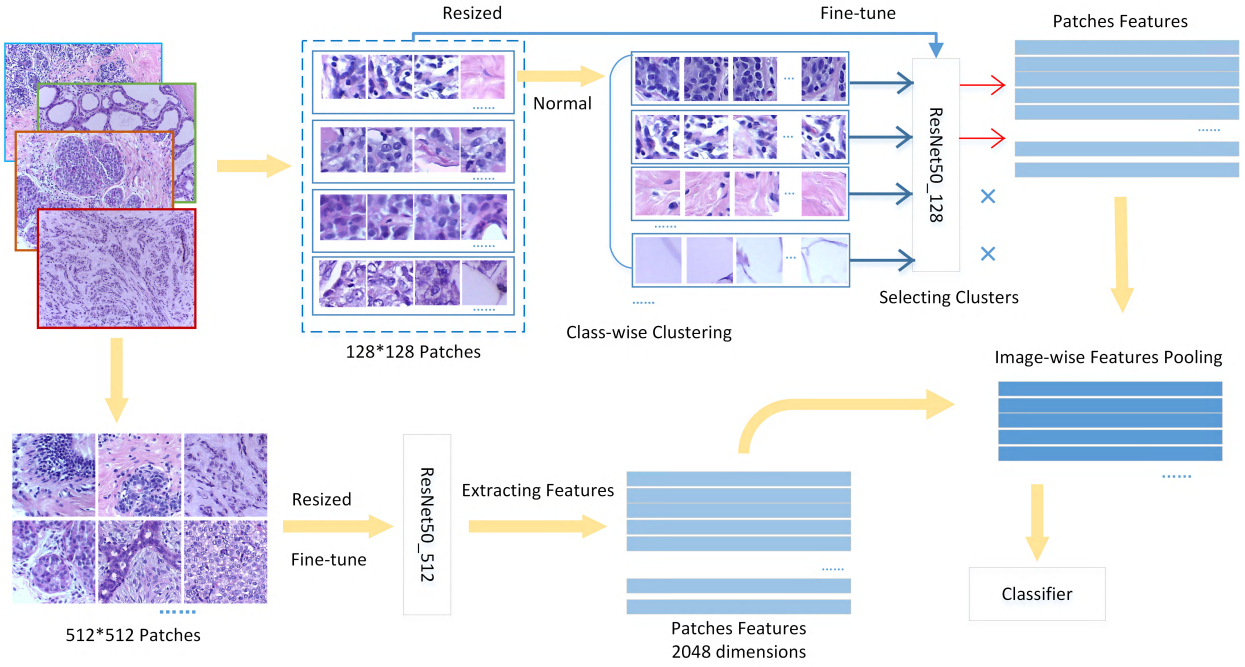


FIGURE 3. A schematic illustration of the proposed framework.

the select smaller patches and larger patches using the feature extractors and compute the final feature of each whole image to train a classifier for classification.

B. SAMPLING PATCHES

Our goal is to classify the breast histology image into four classes: normal tissue, benign tissue, *in situ* carcinoma and *invasive* carcinoma. The performance of classification is highly dependent on the information extracted from the images. We use features related to breast cells and global tissue structures to represent each whole image. Firstly, because the arrangement of cancer cells is extremely disordered and the cancerous cells have atypia such as larger nuclei and inconsistent morphology, therefore, cell-level features including the nuclei information, such as shape and variability, as well as cells organization features like density and morphology, are used to diagnose whether cells are cancerous. The pixel size of the breast histology images in the dataset is $0.42\mu m \times 0.42\mu m$, and the radius of cells is between 3 and 11 pixels approximately. Consequently, we extract small patches of 128 × 128 pixels to contain cell-level features. Secondly, the structure of the diseased tissue may be atypical. *In situ* carcinoma is growth of low-grade cancerous or precancerous cells within a particular tissue compartment such as the mammary duct without invasion of the surrounding tissue. In contrast, *invasive* carcinoma does not confine itself to the initial tissue compartment [29]. Therefore, tissue structures information is essential to differentiate between *in situ* and *invasive* carcinomas. It is unpractical for CNNs to extract features from a histology image with a large size directly. According to the size of images in the provided dataset, we extract patches of 512 × 512 pixels to contain the global tissue structures information.

We extract patches by a sliding window mechanism from breast cancer histology images. The patches of 128×128 pixels are small and focus on cell-related characteristics, therefore, we extract contiguous non-overlapping patches from the breast histology images. In addition, we extract overlapping 512 × 512 pixels patches with a 50% overlap to contain continuous tissue morphology and structures information. All extracted patches are given the same label as the corresponding histology image.

C. FEATURE EXTRACTOR

The histology images have different cell morphology, texture, tissue structures, and so on. The representation of complex features is significant for the classification task. The handcrafted feature extraction method needs abundant expert domain knowledge, and it is labor-intensive and difficult to extract discriminative features. CNNs can directly extract representative features from images, and have achieved remarkable results in various fields. ResNet50 [30] is used as feature extractor in this paper because it is a classical CNN and easy to train compared to other deeper models under the premise of ensuring the extraction of usability features.

The *deep residual learning* framework (ResNet) is proposed by He and Sun [31] to address the *degradation* of deep networks. Formally, the desired underlying mapping is denoted as $H(x)$, then the stacked nonlinear layers are fitted to another mapping of $F(x) := H(x) - x$ and the

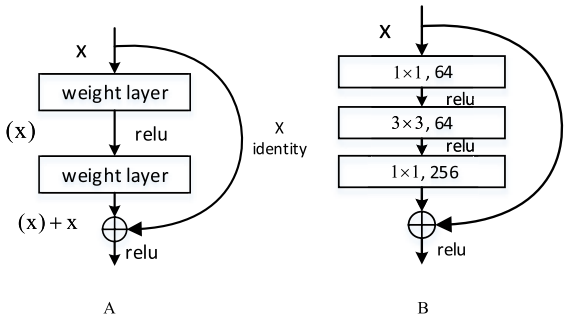


FIGURE 4. (A): A building block, (B): a "bottleneck" building block for ResNet50.

original mapping is rewritten as $F(x) + x$. The formula of $F(x) + x$ is implemented by feed-forward neural networks with "shortcut connections" which perform the *identity* mapping (Fig. 4(A)). For deeper nets, a *bottleneck* design which uses a stack of 3 layers instead of 2 for each residual function is proposed (Fig. 4(B)). The ResNet50 consists of 16 "bottleneck" building blocks and takes as input a {3, 224, 224} RGB image.

The training of ResNet50 from scratch requires a large number of training images to avoid over-fitting. However, because of the paucity of histology images in our dataset, we adopt a transfer learning strategy [32], [33] and use ResNet50 pre-trained on the ImageNet dataset [34]. We remove the top layer of the network and add a softmax classifier with 4 neurons, then, we resize the patches of 512×512 pixels and 128×128 pixels to 224×224 pixels for fine-tuning two modified networks as original feature extractors and the trained networks are denoted as ResNet50-512 and ResNet50-128 respectively. 2048-dimensional features of patches can be obtained from the *GlobalAveragePooling* layer of ResNet50.

D. SCREENING PATCHES

The strategy for sampling patches from histology images is described in Section3.2. The purpose of this section is to introduce the method of screening discriminative 128×128 pixels patches based on certain machine learning algorithms and ResNet50-128.

1) CLUSTERING

The patches of 128×128 pixels may not contain sufficient diagnostic information. For example, there will be some patches with large areas of fat cells and stroma, and normal patches extracted from malignant histology images. Our purpose is to select discriminative patches that have the same label as the source images and have enough breast cells to diagnose whether the cells have become cancerous.

In order to screen discriminative patches in batches, we refer to the idea proposed by Zhu *et al.* [35] to aggregate the patches into different clusters based on their phenotypes. In order to reduce the computation, we rescale the patches of 128×128 pixels to smaller size thumbnail images of 32×32 pixels and concatenate row pixels to get 1024-dimensional features respectively. Then, we use the principal component analysis (PCA) to preserve 200 components of patches to present their phenotypes prior to K-means algorithm. After clustering, we get K distinguishing phenotype clusters.

2) SELECTING CLUSTERS

The candidate clusters obtained by K-means include patches with distinguishing phenotypes respectively. In order to get clusters with more discriminative patches, we employ the ResNet50-128 fine-tuned by all 128×128 pixels patches for screening clusters. Because most of the small patches extracted from histology images with same label have similar rich phenotypes, the ResNet50-128 is more sensitive to them. Therefore, we predict all patches in each cluster through ResNet50-128 and select the top-k clusters with average classification accuracy, that is, clusters including more discriminative patches. Finally, the patches in the selected clusters are used to retrain the ResNet50-128, and the network with new weights is denoted as ResNet50-cluster which is trained by patches contained highly discriminative information, and can extract more representative features of patches.

E. IMAGE-WISE CLASSIFICATION

For the 4-class classification of breast cancer histology images, the sampling strategy of two kinds of patches, the screening method of 128×128 pixels patches and feature extractors based on ResNet50 have been introduced above. Then, we rescale the extracted patches of 512×512 pixels and selected patches of 128×128 pixels corresponding to each image in the training set, and feed them into the fine-tuned ResNet50-512 and ResNet50-cluster respectively to obtain the 2048-dimensional features group, which can represent the cells and tissue structures information of the image. In order to obtain the final feature of an image, we employ the P-norm pooling fusion method [36] and the formulation is as follows:

$$f_p(\mathbf{v}) = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{v}_i^p \right)^{\frac{1}{p}} \tag{1}$$

Here, N represents the number of patches, v_i denotes the 2048-dimensional feature of the *i-th* patch and $P = 3$ is used in our paper. At last, the image-wise features of histology images in the training set are used to train the SVM classifier for 4-class breast cancer histology classification.

IV. EXPERIMENTS

In this section, a serial contrast experiments and performance comparison with other breast histology images classification methods are described.

A. PERFORMANCE EVALUATION

In order to evaluate and compare the performance of our approach of classifying breast histology images into four categories of normal tissue, benign lesion, *in situ* carcinoma and

invasive carcinoma quantitatively, we use accuracy, macro-F [37] related to recall and precision, and confusion matrix as evaluation metrics. The calculation formulas are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

$$Macro\text{-}F = \frac{1}{n}\sum_{i=1}^{n} F_i \tag{5}$$

Here, TP (true positive) is the number of positives cases that are classifies as positive. Analogously, TN, FN and FP represent the numbers of true negatives, false negatives and false positives respectively. The recall represents the percentage of positive samples that are correctly classified, which is more clinically relevant. Macro-F, also known as macro-averaging, is used to evaluate the performance of multi-classification globally and is computed by first computing the F-scores for the n categories then averaging these per-category scores to compute the global means. The confusion matrix is a specific contingency table that allows visualization of the performance

B. SETUP AND RESULTS

We use normalized breast histology images described in Section 2 to conduct experiments. The ResNet50 pre-trained by ImageNet is provided by *Keras* deep learning framework and is fine-tuned by Adam optimization method. The batch size and the learning rate are set as 16 and 0.0001 respectively. The machine learning algorithms used for clustering and the SVM classifier are implemented with *opencv* and *Sklearn* library respectively. Experiments are conducted on a computer equipped with a NVIDIA GeForce GTX TITAN X.

According to the sampling strategy in Section 3.2, the numbers of patches of 512 × 512 pixels and 128 × 128 pixels extracted from the breast histology images in the training set are shown in Table 2.

TABLE 2. The numbers of two kinds of patches extracted from the images in the training set.

Class	Training set	512×512 pixels	128×128 pixels
Normal	55	1925	10560
Benign	69	2415	13248
In situ	63	2205	12096
Invasive	62	2170	11904
Overall	249	8715	47808

The smaller patches extracted from four categories of histology images are aggregated into 7 clusters separately. Then, we select the top-4 clusters with average classification accuracy using the ResNet50-128. The results are given in Table 3 and the number of selected patches of 128 × 128 pixels is 25201.

TABLE 3. The results of clustering and screening (the selected cluster is represented by the bold in the table).

Clusters	Normal	Benign	*In situ*	*Invasive*
0	3051	**2996**	**806**	**2015**
1	770	2236	**3042**	1837
2	**722**	**1701**	2648	**2650**
3	2542	**887**	1224	**1987**
4	**1114**	799	2330	1069
5	**1271**	3588	**1449**	513
6	**1090**	**1041**	**597**	**1833**
Selected	4197	6625	5894	8485

1) IMAGE-WISE CLASSIFICATION

We use the normalized breast histology images in the test set to verify the approach proposed in this paper. The procedure of the experiment is as follows:

a) The sampling strategy introduced in Section 3.2 is used to extract contiguous non-overlapping patches of 128 × 128 pixels and patches of 512 × 512 pixels with 50% overlap from the test images.

b) The ResNet50-cluster fine-tuned by patches of 128 × 128 pixels in the selected clusters is sensitive to more discriminative patches, therefore, we use the network to predict the smaller patches and select patches with classification probability higher than a set threshold.

c) We rescale the extracted patches of 512 × 512 pixels and selected patches of 128 × 128 pixels corresponding to each test image to 224 × 224 pixels, and feed them into the fine-tuned ResNet50-512 and ResNet50-cluster respectively to obtain the 2048-dimensional features group.

d) We employ the 3-norm pooling method to compute the final feature of each image and make final classification by using SVM.

TABLE 4. The numbers of two kinds of patches extracted from the images in the test set and the number of screened patches of 128 × 128 pixels.

Class	test set	512×512 pixels	128×128 pixels	Selected
Normal	9	315	1728	1058
Benign	9	315	1728	656
In situ	9	315	1728	769
Invasive	9	315	1728	1582
Overall	36	1260	6912	4065

The numbers of patches extracted from the histology images in the test set are shown in Table 4. The patches of 128 × 128 pixels are predicted using the ResNet50-cluster, and the patches with classification probability higher than 90% are retained. The number of screened patches is also shown in Table 4. The test images are classified into four classes by using the trained SVM and the confusion matrix of the result are given in Table 5. Four test images are classified into wrong categories, three of which belong to the extended

TABLE 5. The confusion matrix of our model.

Actual label \ Predicted label	Normal	Benign	*In Situ*	*Invasive*
Normal	7	2	0	0
Benign	0	9	0	0
In Situ	0	1	8	0
Invasive	1	0	0	8

test set, and the remaining one labeled as normal is classified as benign. The image-wise accuracy of the initial test set and overall test set is 95% and 88.89% respectively.

According to the confusion matrix, precision, recall and F-score of each class can be obtained respectively, as shown in Table 6. The value of macro-F calculated according to formula (5) is 89.14%.

TABLE 6. The performance of our model.

Results	Normal	Benign	*In situ*	*Invasive*
precision	0.875	0.75	1.0	1.0
recall	0.78	1.0	0.89	0.89
F-score	0.825	0.857	0.942	0.942

2) IMAGE-WISE CLASSIFICATION BASED ON TWO SIZES OF PATCHES WITHOUT SCREENING

In order to verify the impact of the screening process of 128 × 128 pixels patches on the multi-classification performance of breast cancer histology images, we use the ResNet50-512 and ResNet50-128 as the feature extractors without the process of fine-tuning ResNet50-128 with selected discriminative patches of 128 × 128 pixels, and then, train the SVM with all sampled patches. In the test phase, we use all the sampled patches to extract features through the feature extractors, and then use the P-norm pooling method to obtain the image-level feature of each image for classification.

TABLE 7. The confusion matrix of multi-classification based on two sizes of patches without screening.

Actual label \ Predicted label	Normal	Benign	*In Situ*	*Invasive*
Normal	7	1	1	0
Benign	0	8	1	0
In Situ	0	1	8	0
Invasive	1	0	0	8

The confusion matrix of the method are given in Table 7. The image-wise accuracy of the overall test set is 86%.

TABLE 8. The performance of multi-classification based on two sizes of patches without screening.

Results	Normal	Benign	*In situ*	*Invasive*
precision	0.875	0.8	0.8	1.0
recall	0.78	0.89	0.89	0.89
F-score	0.825	0.843	0.843	0.942

The precision, recall and F-score of each class are shown in Table 8. After calculation, the macro-F is 86.33%.

The performance of the method without the process of screening discriminative patches of 128 × 128 pixels is reduced, because some patches with large areas of fat cells and stroma, sparse breast cells, and normal patches extracted from malignant histology images bring certain commonality between normal and malignant images.

3) IMAGE-WISE CLASSIFICATION BASED ON LARGER PATCHES

In our scheme, patches of 128 × 128 pixels and 512 × 512 pixels are extracted from breast histology images to include both cell-level information and tissue structures features. In order to verify the effect of the method on multi-classification performance, we use fixed-size patches of 512 × 512 pixels to train feature extractors and classifier.

TABLE 9. The confusion matrix of multi-classification based on patches of 512 × 512 pixels.

Actual label \ Predicted label	Normal	Benign	*In Situ*	*Invasive*
Normal	7	1	1	0
Benign	0	7	2	0
In Situ	1	1	7	0
Invasive	1	0	0	8

TABLE 10. The performance of multi-classification based on patches of 512 × 512 pixels.

Results	Normal	Benign	*In situ*	*Invasive*
precision	0.78	0.78	0.7	1.0
recall	0.78	0.78	0.78	0.89
F-score	0.780	0.780	0.738	0.942

The confusion matrixes of the experiment are shown in Table 9. The image-wise accuracy of the overall test set is 80.56%. The precision, recall and F-score of each class are shown in Table 10. According to the formula (5), the calculated result of macro-F is 81.0%.

As can be seen from the results, the performance of the method has dropped significantly. Here, the F-score of

TABLE 11. Comparative results of the recall and accuracy.

Test	Method	Recall				Acc
		Normal	Benign	*In Situ*	*Invasive*	
Initial	Proposed	0.8	1.0	1.0	1.0	0.95
	Araújo	0.8	0.6	1.0	1.0	0.85
Extended	Proposed	0.75	1.0	0.75	0.75	0.81
	Araújo	0.75	0.75	0.5	0.75	0.69
Overall	Proposed	0.78	1.0	0.89	0.89	0.89
	Araújo	0.78	0.67	0.78	0.89	0.78

invasive carcinoma class is unchanged, because patches of 512 × 512 pixels contained continuous tissue structures information are used to differentiate *in situ* and *invasive* carcinomas mainly. The absent of smaller patches contained cell-level features leads to the confusion of classifying normal, benign and *in situ* carcinoma histology images.

4) DISCUSSION

In our work, we extract smaller patches of 128 × 128 pixels and larger patches of 512 × 512 pixels from the breast histology images to contain cell-level and tissue-level features, then, we screen discriminative 128 × 128 pixels patches based on clustering algorithm and CNN. Through comparative experiments, it is proved that the two methods proposed in this paper can effectively improve the performance of multi-classification of breast histology images.

We compare the results of our approach with the benchmark method proposed in [20] (CNN+SVM) and the comparative result is shown in Table 11. Araújo *et al.* used the same dataset as us and extracted patches of 512 × 512 pixels. They employed a CNN of their own designed and achieved a best accuracy of 77.8% of multi-classification with augmented dataset. It can be seen that our approach has a substantial improvement in accuracy and recall compared with the benchmark scheme, especially in the classification of benign and *in situ* carcinoma images.

In addition, Rakhlin *et al.* [38] employed several traditional CNNs as feature extractors and gradient boosted trees classifier. Golatkar *et al.* [39] extracted patches that are rich in nuclei and used fine-tuned Inception-v3. They achieved 87.2% and 85% accuracy respectively for 4-class classification used 400 H&E stained breast histology images in an extended dataset released for Breast Cancer Histology Challenge (BACH). As seen, our approach are competitive compared to other state-of-the-art methods.

V. CONCLUSION

In this paper, we propose an effective method to classify the H&E stained breast histology images into four classes: normal tissue, benign lesion, *in-situ* carcinoma and *invasive* carcinoma. Due to the atypia of cancerous cells and the difference in tissue morphology and structures between *in situ* carcinoma and *invasive* carcinoma, we extract two kinds of patches of 512 × 512 pixels and 128 × 128 pixels from the histology images to contain different levels features. We design a process to screen more discriminative patches of 128 × 128 pixels automatically based on several machine learning algorithms and CNN. ResNet50 is used as feature extractor to extract features from patches, P-norm pooling is used to get final features of images and SVM is employed for final image-wise classification. Our scheme achieves 95% accuracy on the initial test set compared to 85% accuracy of the benchmark work [20]. In addition, the validity of our methods is demonstrated through a series of experiments.

REFERENCES

[1] N. Christian. (2018). *What You Need to Know About Breast Cancer*. [Online]. Available: https://www.medicalnewstoday.com/articles/37136.php

[2] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.

[3] L. He, L. R. Long, S. Antani, and G. Thoma, "Computer assisted diagnosis in histolopathology," *Sequence Genome Anal., Methods Appl.*, vol. 3, pp. 271–287, 2010.

[4] L. He, L. R. Long, S. Antani, and G. R. Thoma, "Histology image analysis for carcinoma detection and grading," *Comput. Methods Programs Biomed.*, vol. 107, no. 3, pp. 538–556, 2012.

[5] J. G. Elmore *et a.*, "Diagnostic concordance among pathologists interpreting breast biopsy specimens," *JAMA*, vol. 313, no. 11, p. 1122, 2015. doi: 10.1001/jama.2015.1405PMID:25781441.

[6] M. Veta, J. P. W. Pluim, M. A. Viergever, and P. J. van Diest, "Breast cancer histopathology image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1400–1411, May 2014.

[7] 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, 2013.

[8] 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.

[9] P. Filipczuk, T. Fevens, A. Krzyzak, and R. Monczak, "Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies," *IEEE Trans. Med. Imag.*, vol. 32, no. 12, pp. 2169–2178, Dec. 2013.

[10] P. Wang, X. Hu, Q. Liu, X. Zhu, and Y. Li, "Automatic cell nuclei segmentation and classification of breast cancer histopathology images," *Signal Process.*, vol. 122, pp. 1–13, May 2016.

[11] S. Naik, S. Doyle, A. Madabhushi, M. Feldman, J. Tomaszewski, and S. Agner, "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *Proc. 5th IEEE Int. Symp.*, May 2008, pp. 284–287.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[13] J. Gu *et al.* (2015). "Recent advances in convolutional neural networks." [Online]. Available: https://arxiv.org/abs/1512.07108

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.

[15] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

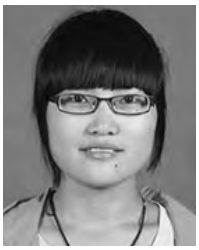
[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, Dec. 2012, pp. 1097–1105.

[17] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[18] F. A. Spanhol, L. S. Oliveira, L. Heutte, and C. Petitjean, "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016.

[19] F. A. Spanhol, L. S. Oliveira, L. Heutte, and C. Petitjean, "Breast cancer histopathological image classification using convolutional neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2560–2567.

[20] T. Araújo *et al.*, "Classification of breast cancer histology images using convolutional neural networks," *PLoS ONE*, vol. 12, no. 6, p. e0177544, 2017.
[21] A. Pêgo and P. Aguiar. (2015). *Bioimaging*. [Online]. Available: http://www.bioimaging2015.ineb.up.pt/dataset.html
[22] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck. (2016). "Deep learning for identifying metastatic breast cancer." [Online]. Available: https://arxiv.org/abs/1606.05718
[23] B. E. Bejnordi *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, Dec. 2017.
[24] B. E. Bejnordi *et al.*, "Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images," *J. Med. Imag.*, vol. 4, no. 4, p. 044504, 2017.
[25] J. Makki, "Diversity of breast carcinoma: Histological subtypes and clinical relevance," *Clin. Med. Insights, Pathol.*, vol. 8, pp. 23–31, Dec. 2015.
[26] A. Vahadane *et al.*, "Structure-preserved color normalization for histological images," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 1012–1015.
[27] M. Macenko *et al.*, "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro (ISBI)*, Jun./Jul. 2009, pp. 1107–1110.
[28] E. Reinhard, M. Adhikhmin, P. Shirley, and B. Gooch, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, Sep./Oct. 2001.
[29] A. K. Mary. (2018). *Breast Cancer*. [Online]. Available: https://www.merckmanuals.com/home/women-s-health-issues/breast-disorders/breast-cancer
[30] K. He, X. Zhang, J. Sun, and S. Ren, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[31] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5353–5360.
[32] J. Yosinski, J. Clune, H. Lipson, and Y. Bengio, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.
[33] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.
[34] J. Deng, W. Dong, L.-J. Li, K. Li, L. Fei-Fei, and R. Socher, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[35] X. Zhu, J. Yao, J. Huang, and F. Zhu, "WSISA: Making survival prediction from whole slide histopathological images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 7234–7242.
[36] Y. L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 111–118.
[37] Y. Yang, "An evaluation of statistical approaches to text categorization," *Inf. Retr.*, vol. 1, nos. 1–2, pp. 69–90, 1999.
[38] A. Rakhlin, A. Shvets, A. A. Kalinin, and V. Iglovikov, "Deep convolutional neural networks for breast cancer histology image analysis," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, 2018, pp. 737–744.
[39] A. Golatkar, D. Anand, and A. Sethi, "Classification of breast cancer histology using deep learning," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, 2018, pp. 837–844.



YUQIAN LI was born in Henan, China, in 1993. She received the B.S. degree from the School of Computer Science and Technology, Anhui University, Hefei, in 2016. She is currently pursuing the M.S. degree with the University of Science and Technology of China. Her research interests include deep learning, image processing, and histology image classification and segmentation.

JUNMIN WU received the graduate degree in computer science and the Ph.D. degree in computer science and engineering from the University of Science and Technology of China, in 1994 and 2005, respectively, where he is currently an Associate Professor with the Department of Computer Science and Technology. His research interests include computer architecture, virtualization technologies, cluster computing, multi-core computing, and simulating. His works have been published in several international conferences and journals.

QISONG WU received the B.S. degree in medicine from the Changwei Medical College, in 1983. He has been involved in pathological diagnosis and teaching for 30 years and has rich experience in clinical pathology diagnosis. He is currently the Director and Chief Physician of the Department of Pathology, Suzhou Kowloon Hospital Shanghai Jiao Tong University School of Medicine. He has written more than 30 medical papers. His research interests include cytological diagnosis, pathology, human anatomy, and immunohistochemical diagnosis.