Neurocomputing 366 (2019) 46–53

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

EMS-Net: Ensemble of Multiscale Convolutional Neural Networks for Classification of Breast Cancer Histology Images

Zhanbo Yang[a,b], Lingyan Ran[b], Shizhou Zhang[b], Yong Xia[a,b,*], Yanning Zhang[b]

[a] Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 518057, China
[b] National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China

ARTICLE INFO

Article history:
Received 17 January 2019
Revised 25 July 2019
Accepted 25 July 2019
Available online 30 July 2019

Communicated by Pingkun Yan

MSC:
00-01
99-00

Keywords:
Breast cancer
Microscopy image
Convolutional neural networks
Multiscale
Ensemble model

ABSTRACT

Histology images analysis resulted from needle biopsy serves as the gold standard for breast cancer diagnosis. Deep learning-based classification of breast tissues in histology images, however, is less accurate, due to the lack of adequate training data and ignoring the multiscale structural and textural information. In this paper, we propose the Ensemble of MultiScale convolutional neural Networks (EMS-Net) to classify hematoxylin–eosin stained breast histopathological microscopy images into four categories, including normal tissue, benign lesion, in situ carcinoma, invasive carcinoma. We first convert each image to multiple scales, and then use the training patches cropped and augmented at each scale to fine-tune the pre-trained DenseNet-161, ResNet-152, and ResNet-101, respectively. We find that a combination of three fine-tuned models is more accurate than other combinations, and use them to form an ensemble model. We evaluated our algorithm against three recent methods on the BACH challenge dataset. It shows that the proposed EMS-Net algorithm achieved an accuracy of $91.75 \pm 2.32\%$ in the five-fold cross validation using 400 training images, which is higher than the accuracy of other three algorithms, and also achieved an accuracy of 90.00% in the online verification using 100 testing images.

1. Introduction

Breast cancer is the leading type of cancer in women, resulting in 1.68 million new cases and 522,000 deaths in 2012 worldwide, which account for 25.16% of all cancer cases and 14.71% of cancer-related death [1]. Accurate diagnosis of breast cancer is crucial for appropriate treatment and preventing further deterioration. Several diagnostic tests have been used, including physical exam, mammography, magnetic resonance imaging (MRI), ultrasound, and biopsy. Among them, histology image analysis resulted from needle biopsy serves as the gold standard. During the diagnostic procedure, cellular regions of hematoxylin–eosin (H&E) stained histopathological microscopy images are evaluated by specialists to determine the type of breast tissues, including normal tissue, benign lesion, in situ carcinoma, invasive carcinoma. However, due to the complexity and enormous size of histopathological microscopy images, detecting carcinoma by biopsy requires a high degree of skill and concentration, and is time-consuming, labor-intensive, and prone to operator bias. The grand challenge on BreAst Cancer Histology images (BACH) [2] organized in conjunction with the 15th international conference on image analysis and recognition (ICIAR 2018) recognized the complexity of this problem and the increasing demand for the development of automated detection and diagnosis tools.

Automated classification of H&E stained breast histopathological microscopy images is challenging in two aspects. First, there are huge intra-class variances and inter-class similarities in microscopy images, e.g., the hard mimics from benign lesion which carry similar morphological appearance with carcinoma. Second, histopathological microscopy images usually have an extremely high resolution, and hence contain rich structural and detailed information that is hard to be characterized effectively. To address both challenges, various handcrafted visual features have been designed to characterize breast tissues [3–5]. Based on these features, many machine learning techniques, such as the support vector machine (SVM) [6] and compressed sensing (CS) [7], have been used for automated tissues classification.

Recently, deep convolutional neural networks (DCNNs) have been widely acknowledged as one of the most powerful tools for image classification, since they have distinct advantages over traditional solutions in providing an end-to-end feature

* Corresponding author at: National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China.
E-mail addresses: yxia@nwpu.edu.cn, yxia@ieee.org (Y. Xia).

https://doi.org/10.1016/j.neucom.2019.07.080
0925-2312/

extraction-and-classification framework to free users from the troublesome handcrafted feature extraction [8,9]. The success of DCNNs has prompted many investigators to employ this deep learning technique for histopathological microscopy image classification [10–12]. However, despite their prevalence, a single DCNN model has limited ability to extract discriminative features and result in sub-optimal classification performance. Therefore, ensembles of DCNN architectures have been designed to learn the representation of histopathological microscopy images from different perspectives for more accurate classification.

In this paper, we propose the Ensemble of MultiScale convolutional neural Networks (EMS-Net) for the automated classification of H&E stained breast histopathological microscopy images. We first convert each histopathological microscopy image to multiple scales, and then use the training patches cropped and augmented at each scale to fine-tune three pre-trained DCNNs, i.e. DenseNet-161 [13], ResNet-152 [14], and ResNet-101 [14], respectively. We expect that the DCNNs fine-tuned at multiple scales can extract diverse and complementary visual features to represent those images from different perspectives. Then an optimal subset of fine-tuned DCNNs was selected and combined to form the EMS-Net. We evaluated the proposed EMS-Net algorithm on the BACH Challenge dataset and achieved an accuracy of $91.75 \pm 2.32\%$ in the five-fold cross validation using 400 training images and an accuracy of 90.00% in the online verification using 100 testing images.

There are three main contributions of this paper. First, the EMS-Net is proposed to extract diverse and complementary visual features in H&E stained breast histopathological microscopy images by ensemble of multiple DCNNs at multiscale, which is important for analysis of histopathological images. Second, the proposed method reduces time complexity effectively by selecting the optimal ensemble of fine-tuned deep models to form the EMS-Net, which is meaningful for clinical application. Third, the proposed method achieves an accuracy of 90.00% in the online verification of the BACH Challenge classification task.

2. Related work

In this section, we review related microscopy images classification algorithms from three aspects.

2.1. Handcrafted features-based algorithms

The early study on histopathological microscopy image analysis is mostly based on handcrafted features [3–6]. Khan et al. [3] proposed the geodesic geometric mean of regional covariance descriptors as an image-level feature for nuclear atypia grading in breast histopathological microscopy images. Vink et al. [6] proposed a modified AdaBoost algorithm to create two detectors, which focus on different characteristics in the appearance of nuclei for nucleus detection. Although these methods are easy to implement and train, they often suffer from limited performance due to the heterogeneity of the biological architectures and lack of heuristics to guide the feature extraction.

2.2. Deep learning-based algorithms

The emergence and development of deep learning introduced many breakthroughs in many computer vision applications, including the classification of natural and medical images [15,16]. Many well-performed DCNN architectures, such as AlexNet [17], VGG [18], GoogLeNet [19], ResNet [14], and DenseNet [13], have been designed for the ILSVRC ImageNet classification task [20]. Among them, ResNet [14] and DenseNet [13] are the most prevalent network backbone structure with the superior performance to other machine learning approaches. ResNet uses skip-connections to bypass the non-linear transformations and adds an identity projection to each block. Thus, it solves the problem of vanishing gradient and enables users to design extremely deep network architectures. Based on the idea of skip-connections, DenseNet is efficient in computation and parameters, and achieves state-of-the-art performance by directly connecting each feature layer to all subsequent ones.

Nowadays, DCNN models have been widely applied to the analysis of histopathological microscopy images. Mehta et al. [12] added a parallel branch to U-Net [21], and thus constructed the so-called Y-Net for simultaneous segmentation and classification of microscopy images. Campanella et al. [22] proposed the deep multiple instance learning (MIL) framework for classification and localization of microscopy images. Nazeri et al. [10] proposed a two-stage DCNN architecture, which includes a “patch-wise” network and a “image-wise” network, to classify breast microscopy images into four categories. To complement the merits of deep learning and traditional methods, Araújo et al. [11] proposed a patches-based “DCNN + SVM” model to address the same classification problem. In this model, a DCNN is designed for feature extraction and SVM is used as classifier. Despite the improved performance, these DCNN-based algorithms have not achieved the same performance on routine histopathological microscopy image classification as they have done in the ImageNet challenge. The suboptimal performance is attributed mainly to the fact that the training dataset available for these tasks is too small and a deep model is prone to overfit those training data.

2.3. Deep ensemble learning algorithms

To further improve the performance of DCNN-based algorithms, those models, which have suboptimal performance due to overfitting, were viewed as weak learner and then combined together to form an ensemble model, which is expected to be a strong learner. Makarchuk and co-workers [23] extracted image features from microscopy patches using ResNet34 [14] and DenseNet201 [13], and then fed the obtained features to a XGBoost classifier [24]. Rakhlin et al. [25] classified microscopy images in a similar way, but using several DCNN architectures and the gradient boosted trees. Chennamsetty et al. [26] formed an ensemble of three DCNNs, each of which was pre-trained on different preprocessing regimes, and thus achieved the 1st place on the BACH challenge at the first stage. Although produced further performance gain, these deep ensemble learning algorithms analyzes microscopy images at a single scale, without considering the partial but important structural and textural information at multiple scales. Since specialists often use 10X, 20X, and 40X lenses alternately when inspecting microscopy images, the multiscale information in those images must play a critical role for classification.

3. BACH dataset

The ICIAR 2018 grand challenge on BACH dataset [2] was used for this study. This dataset is composed of 400 H&E stained breast histopathological microscopy images with a size of 2048×1536 for training a classification model and 100 similar microscopy images with the same size for testing. Training images have four class labels including normal, benign, in situ carcinoma and invasive carcinoma (see Fig. 1). Each of four categories contain 100 training images. The 100 test images were officially presented for online verification, and their labels are not available.

4. Method

The proposed EMS-Net algorithm consists of three steps: (1) multiscale image patch extraction, (2) training multiple DCNN

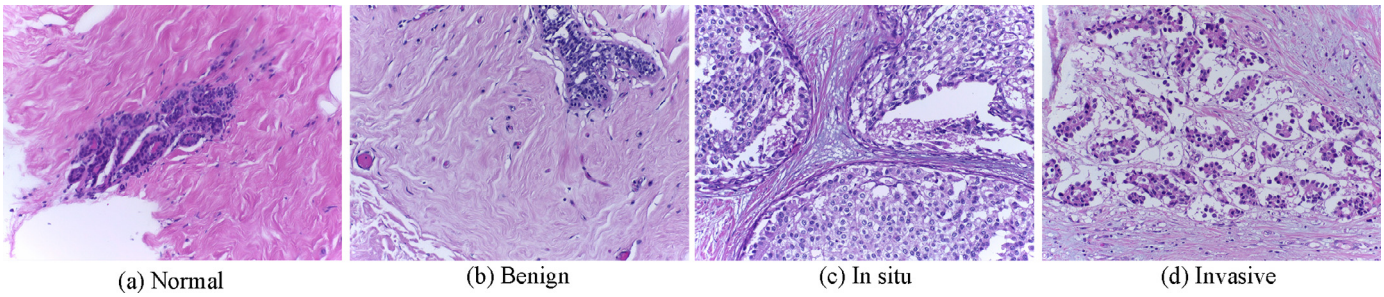

Fig. 1. Four H&E stained breast histopathological microscopy images from different categories.

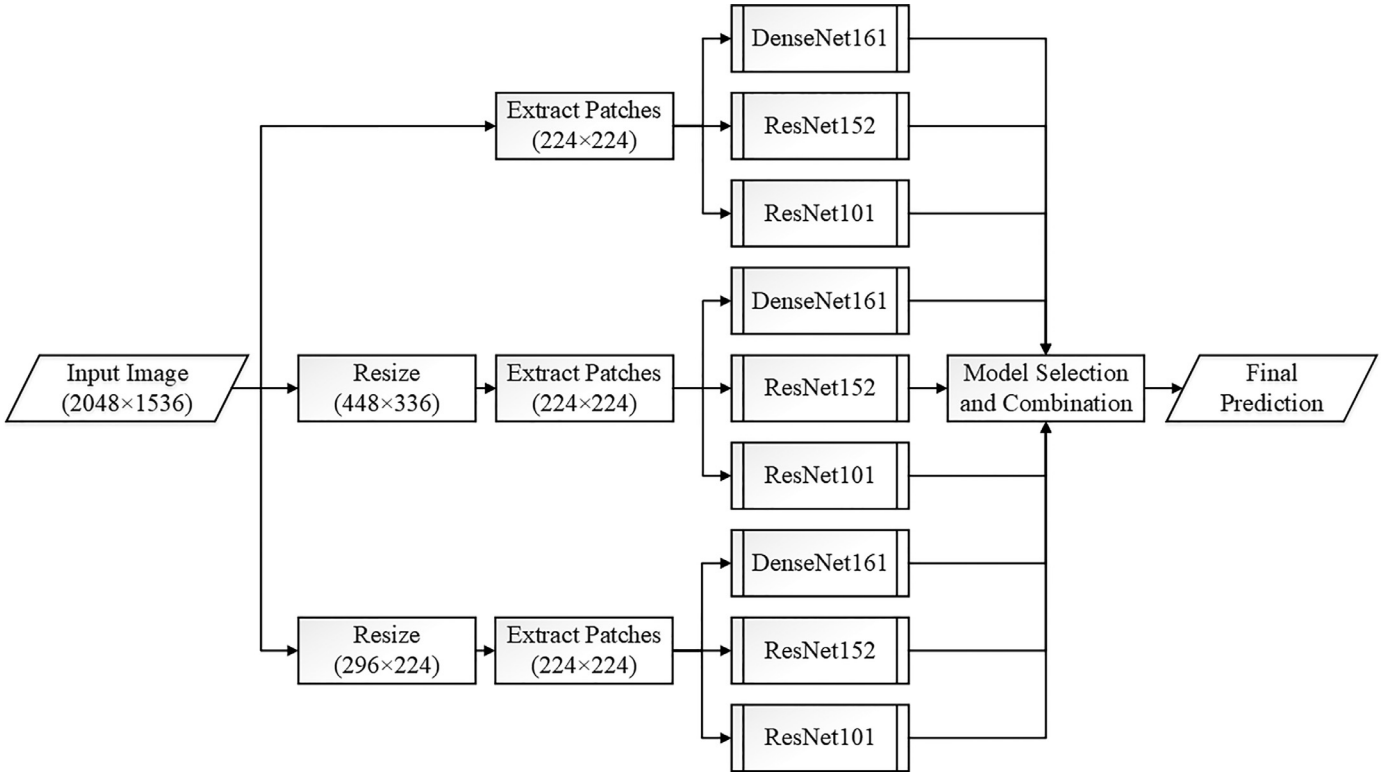

Fig. 2. Diagram of the proposed EMS-Net algorithm.

models, and (3) model selection and combination. The diagram that summarizes this algorithm is shown in Fig. 2. We now delve into each of these three steps.

4.1. Multiscale image patch extraction

Each image has a size of 2048×1536 (Scale I). To use the multiscale information in it, we resize the image to 448×336 (Scale II) and 296×224 (Scale III), respectively. To extract image patches at each scale, we slide a $w \times h$ window with a stride of s on the original and resized images. Since the pre-trained DCNN models used for this study take 224×224 images as input, we set the patch size to 224×224. Then, the number of patches we extracted from an image of size $H \times W$ is

$$N = \left(\left\lfloor \frac{W-w}{s} \right\rfloor + 1\right) \times \left(\left\lfloor \frac{H-h}{s} \right\rfloor + 1\right) \quad (1)$$

where $\lfloor\rfloor$ denotes rounded down. We set the stride to 224 at Scale I, 28 for training data extraction and 56 for testing data extraction at Scale II, and 9 for training data extraction and 18 for testing data extraction at Scale III. Note that the images at Scale II and Scale III were divided into partly overlapped patches to generate more training data. Next, the intensities of each microscopy patch are standardized to zero mean and unit variance.

To alleviate the overfitting of DCNN models, we employ two data augmentation methods to expand the size and increase the diversity of the training dataset. First, each patch is transformed into eight patches by combining the $k \cdot \pi/2$ rotations, where $k = \{0, 1, 2, 3\}$, and with/without vertical reflections. Second, random color perturbations [27] have been applied each patch. Consequently, we have 259,200 augmented microscopy patches at Scale I, 216,000 patches at Scale II, and 43,200 patches at Scale III to train each DCNN model and enable them to learn rotation invariant, mirroring invariant, and color invariant image representations.

4.2. Training multiple DCNNs

Due to their outstanding performance in image classification, we choose ResNet-101 [14], ResNet-152 [14] and DenseNet-161 [13] as the backbone networks for this study.

ResNet-101 consists of 101 learnable layers, including a convolutional layer with the kernel size of 7×7, a 3×3 max pooling layer, four residual modules, which have 3, 4, 23, and 3 triple-layer residual blocks, respectively, and an average pooling layer followed

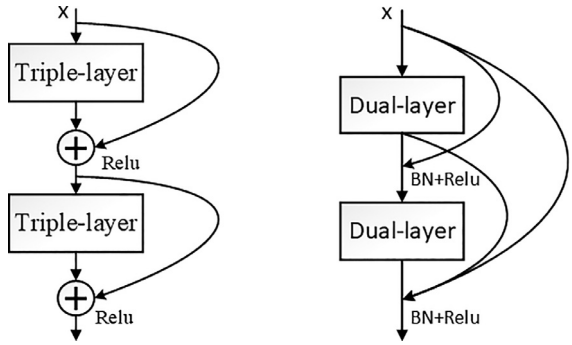

Fig. 3. Architectures of backbone networks used for this study. Left: residual blocks, right: dense blocks.

by a fully-connected layer. As show in the left part of Fig. 3, a triple-layer residual block is different from conventional deep convolutional structure in the existence of a skip-connection between the input and output, which serves as an identity projection. Using skip-connections can alleviate significantly the vanishing gradient issue, and thus enable the network architecture to become extremely deep. ResNet-152 has a similar architecture, but there are 3, 8, 36, and 3 triple-layer residual blocks in four residual modules, respectively. Similar to ResNet models, DenseNet-161 consists of 161 learnable layers, including a convolutional layer with the kernel size of 7×7, a 3×3 max pooling layer, four dense modules, three transition layers, and a global average pooling layer followed by a fully-connected layer. Those four dense modules contain 6, 12, 36, and 24 dual-layer dense blocks, respectively. Each of the first three dense modules is followed by a transition layer, which is a 1×1 convolutional layer followed by 2×2 average pooling. As show in the right part of Fig. 3, skip-connections are used to connect any layer to all subsequent layers in a dense block. Different from ResNet, DenseNet concatenates feature maps produced in previous layers instead of adding them, and hence is able to achieve improved performance in many image classification tasks with fewer parameters.

To adapt each pre-trained DCNN model to our four-category image classification problem, we keep only four neurons in the last fully-connected layer and remove the other output neurons and the weights attached to them. Thus when pre-trained DCNN models are fine-tuned, their inputs are microscopy image patches of size, which are the augmented copies of the patches extracted from resized microscopy images, and their outputs are four-dimensional vectors, representing the probabilities of inputs belonging to each of normal, benign, in situ carcinoma and invasive carcinoma cases.

Since the number of training microscopy patches is small, each DCNN model has been pre-trained and converged on the ImageNet challenge dataset. The pre-trained models we used from PyTorch [28]. To fine-tune each DCNN model, we adopted the Adam optimizer [29] to minimize the cross entropy loss, and set the batch size to 8, learning rate to 0.0001 with a decay of 10% every 10 training epochs, and maximum epoch number to 30.

4.3. Model selection and combination

We use the augmented microscopy patches at each scale to fine-tune each pre-trained DCNN model, and thus obtained nine fine-tuned models, which are denoted by A to I. As shown in Table 1, the letters outside the brackets are the indexes of nine fine-tuned DCNN models. To reduce the complexity of our algorithm, our aim is to select a subset of fine-tuned models and combined them together to form an ensemble model.

Table 1
Indexes of nine fine-tuned DCNN models and their classification accuracy (%) on the validation dataset.

Scales	Architectures		
	DenseNet-161	ResNet-152	ResNet-101
Scale I	A(72.50)	B(73.75)	C(72.50)
Scale II	D(96.25)	E(93.75)	F(90.00)
Scale III	G(92.50)	H(80.00)	I(81.50)

Table 2
Optimal subsets of k fine-tuned DCNN models and their computation time (h) and ensemble classification accuracy (%) on the validation dataset.

k	Optimal subset	Computation time (h)	Accuracy (%)
1	{*D*}	25.5	96.25
2	{*D, E*}	53.5	97.50
3	{***D, E, G***}	**58.5**	**98.75**
4	{*D, E, F, G*}	76.5	98.75
5	{*A, D, E, F, G*}	106.5	98.75
6	{*D, E, F, G, H, I*}	86	97.50
7	{*A, D, E, F, G, H, I*}	116	97.50
8	{*A, B, D, E, F, G, H, I*}	148	95.00
9	{*A, B, C, D, E, F, G, H, I*}	169	92.50

Since there are only nine fine-tuned models, we have totally $2^9 - 1 = 511$ possible subsets to form ensemble models, which we can evaluate exhaustively. To this end, we randomly selected 40 images from the training dataset to form validation set. We extract and augment microscopy patches from 40 validation images using the aforementioned methods and input them each fine-tuned model to get the prediction. As shown in Table 1, the values inside the brackets are the classification accuracy of the corresponding models, which is the percentage of correctly classified images by the models. For each possible model subset, first the image category probabilities from each model are determined by the ratio of number of patches belongs to this category to the number of patches extracted from this image, which is named majority voting. Then, these category probabilities are averaged and the category with the largest value is assigned as the label of the image. The performance of each subset is defined as the accuracy on validation images, which is the percentage of correctly classified images by the subset.

Table 2 gives the optimal model subsets that contain k ($k = 1, 2, \ldots, 9$) fine-tuned models and the corresponding computation time and classification accuracy. It shows that the ensemble models that contain 3 (D, E, and G), 4 (D, E, F, and G) or 5 (A, D, E, F, and G) fine-tuned models achieved the highest classification accuracy of 98.75%. Considering the lower computational cost, the D–E–G combination of models achieved best comprehensive performance. The advantage comes from three aspects. First, considering the outstanding performance of DenseNet-161 and ResNet-152 in image classification, and the appropriate resized scales of input image, model D, E and G can extract more discriminative features and improve the model performance for classification of histopathological microscopy images. As we can see in Table 1, those achieved better classification accuracy than other models individually. Second, model D is DenseNet-161 architecture and model E is ResNet-152 architecture, which both fine-tuned at Scale II (448×336), and model G is DenseNet-161 architecture fine-tuned at Scale III (296×224), that include different network architectures and input image with multiple scales. Therefore the D–E–G combination of models can extract diverse and complementary visual features to represent those images from different perspectives, and improves the classification performance of combined model. Third, except for the ResNet-101 has worse representative capability than DenseNet-161 or ResNet-152, models fine-tuned at Scale I can only

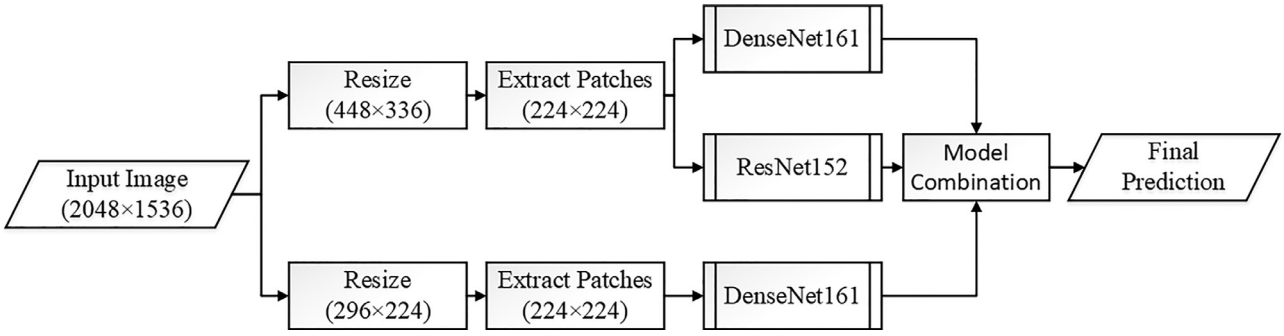

Fig. 4. Diagram of our ensemble model.

Table 3
Performance (mean ± standard deviation) of proposed EMS-Net algorithm on BACH training images with five-fold cross validation.

Category	Accuracy (%)	Precision (%)	Recall (%)	AUC (%)
Normal	95.98 ± 2.89	89.27 ± 9.36	97.00 ± 2.25	98.78 ± 0.88
Benign	94.24 ± 2.82	97.65 ± 4.71	79.00 ± 9.75	96.40 ± 5.08
In situ	96.01 ± 2.66	91.36 ± 8.81	94.00 ± 4.90	98.38 ± 1.50
Invasive	97.26 ± 0.90	92.57 ± 3.22	97.00 ± 4.00	98.50 ± 1.94
All	**91.75 ± 2.32**	–	–	–

Table 4
Accuracy (%) of proposed EMS-Net algorithm and three recent methods on BACH training images.

Method	Accuracy (%)
DCNN+SVM, 2017 [11]	77.80
Pre-trained VGG-16, 2018 [2]	83.00
Ensemble of three DCNNs, 2018 [2]	87.00
EMS-Net (proposed)	**91.75**

Table 5
Combined leader-board at first and second stages of BACH Challenge (ranked according to accuracy on Part A).

Position	Participants	Accuracy (%)
1	bamboo	93.00
2	Heechan Yang	92.00
3	NPU_SAIIP(we)	**90.00**
4	nishuone	89.00
5	zhuangjin2018	89.00
6	jerome.rony.1	89.00
7	amartel	89.00

receive local information but not integral information due to the large size of image at Scale I (2048×1536) and small input patch with size of 224×224, which may influence the classification performance of models. Therefore, the D–E–G combination of network achieved the best performance.

For the sake of model simplicity and time efficiency, we choose the subset of the fine-tuned model D, E, and G to form our ensemble model, which has been shown in Fig. 4.

5. Experiments and results

Since the testing images in the BACH dataset were not equipped with class labels, we evaluated the proposed EMS-Net algorithm using 400 training images with the five-fold cross validation. Specifically, we divided the BACH training images into five subsets randomly and equally, each containing 20 images from each of four categories. In each experiment, we used one subset of images for testing and the other images for training. In the training stage, microscopy patches were extracted from 320 training images at three scales and augmented to fine-tune the model D, E, and G, respectively. In the testing stage, microscopy patches were extracted from each testing image and then fed to the ensemble model given in Fig. 4. The predictions made by model D, E, and G on all those patches were combined to determine the classification of the test image. The performance of the proposed method in differentiating each category of images was measured by accuracy, precision, recall, and the area under the receiver operating characteristic curve (AUC), and the overall accuracy was measured by the percentage of correctly classified images.

Table 3 gives the accuracy, precision, recall, and AUC of the proposed algorithm in differentiating each category of images and the overall classification accuracy. It shows that it is easy for the proposed algorithm to separate invasive carcinoma tissues from others, but difficult to separate benign tissues from others. Nevertheless, we achieved an average AUC of more than 96.40% in all categories and an overall accuracy of 91.75 ± 2.32%, which demonstrate the effectiveness of the proposed algorithm in the classification of breast histopathological microscopy images.

Next, we compare the proposed EMS-Net algorithm to three recent methods: (1) using a custom DCNN for feature extraction and a SVM for classification [11], (2) using a pre-trained VGG-16 together with a variety of data augmentation methods [2], and (3) using ensemble of three pre-trained DCNNs, which was fine-tuned on image level using microscopy images with a much reduced size [2]. Table 4 shows the overall accuracy of those three methods, which have been adopted in the literature, and the accuracy of our algorithm. It reveals that proposed algorithm is substantially more accurate than those three methods on this image classification task.

Moreover, besides 400 labeled training images, the organizers of BACH challenge also provide 100 microscopy images without labels for online testing. We submit the classification results, which we obtained on those testing images, to the official website, and the organizers calculate the accuracy of our algorithm. We combine the leader-board at first and second stages and display it in Table 5. It shows that our algorithm achieved an accuracy of 90.00% in the online validation on testing images. Note that the BACH challenge includes two parts, that aim at the classification of H&E stained histopathological microscopy images and segmentation of breast cancer whole slide tissue images (WSI) respectively. Since the EMS-Net algorithm is proposed to solve the BACH challenge classification task, we ranked only according to the accuracy on Part A. By this way, the proposed EMS-Net algorithm won the 3rd place on the leader-board[1] at the time we submitted the prediction to the official website.

To further demonstrate the validity of the proposed EMS-Net, we also design the experiment on another histopathological

[1] Available at: https://iciar2018-challenge.grand-challenge.org/evaluation/results/.

Table 6
Performance of the proposed EMS-Net algorithm on BreakHis dataset.

Category	Accuracy (%)	Precision (%)	Recall (%)	AUC (%)
Benign	99.75	100.00	99.20	99.99
Malignant	99.75	99.64	100.00	99.99
All	**99.75**	–	–	–

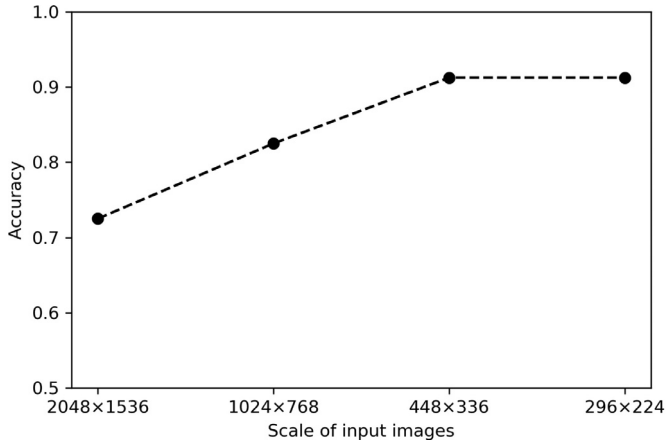

Fig. 5. Accuracy of DenseNet-161 versus scale (size) of input images.

dataset named Breast Cancer Histopathological Database (BreakHis), which is composed of 2480 benign and 5429 malignant microscopic images of breast tumor tissue with 700×460 pixels and 3 channels collected from 82 patients using different magnifying factors (40X, 100X, 200X, and 400X). To match the resolution of images in BACH dataset, we use microscopic images with 40X magnifying factor for experimentation, which contains 625 benign and 1370 malignant samples. We randomly select 20% of microscopic images for testing and other images for training. Images with 40X magnifying factor in BreakHis dataset correspond to images with Scale II in BACH dataset, and we down sample them to 510×335 corresponding to images with Scale III in BACH dataset. Due to the BreakHis dataset is a two-class dataset (benign or malignant), we keep only two neurons in the last fully-connected layer in each model and fine-tune them. As shown in Table 6, we achieved an AUC of 99.99% in all categories and an overall accuracy of 99.75%, which further demonstrates the effectiveness of the proposed algorithm in the classification of breast histopathological microscopy images.

6. Discussion

6.1. Choice of multiple scales

A major contribution of this work is the joint use of important structural and textural information at multiple scales. To this end, we mimicked the visual inspection conducted by specialists, in which 10X, 20X, and 40X lenses are used alternately, and size each microscopy image from 2048×1536 (Scale I) to 448×336 (Scale II) and 296×224 (Scale III), respectively. To validate the choice of those scales is acceptable, we selected DenseNet-161 as a case study and evaluated its performance on testing microscopy images, which have been resized to 1024×768, 448×336, and 296×224, respectively. We adopted the aforementioned settings of patch extraction, data augmentation, and model fine-tuning. The accuracy of DenseNet-161 versus the size of input images was plotted in Fig. 5. It shows that DenseNet-161 achieved the highest classification accuracy when the input images were resized to 448×336, and 296×224 (i.e. at Scales II and III). This result is consistent with the conclusion drawn from our model selection process, where the DCNN models in the global optimal subset were trained at Scales II and III.

Table 7
Accuracy (%) of nine fine-tuned DCNNs, the ensemble of three architectures trained at each scale, and the ensemble of each architecture trained at three scales at training stage.

Scales	Architectures			
	DenseNet-161	ResNet-152	ResNet-101	Ensemble models
Scale I	72.50(8)	73.75(7)	72.50(9)	71.25
Scale II	96.25(1)	93.75(3)	90.00(4)	96.25
Scale III	92.50(2)	80.00(5)	81.50(6)	90.00
Ensemble models	92.50	92.50	88.75	92.50

Table 8
Accuracy (%) of nine fine-tuned DCNNs, the ensemble of three architectures trained at each scale, and the ensemble of each architecture trained at three scales at testing stage.

Scales	Architectures			
	DenseNet-161	ResNet-152	ResNet-101	Ensemble models
Scale I	71.25(8)	71.25(9)	72.50(7)	71.25
Scale II	91.25(1)	90.00(3)	90.00(4)	91.25
Scale III	91.25(2)	87.50(5)	82.50(6)	88.75
Ensemble models	91.25	88.75	88.75	90.00

Table 9
Accuracy (%) of the ensemble of three models trained at each scale, and the fine-tuned integral models at each scale at testing stage.

Scales	Architectures	
	Ensemble of fine-tuned models	Fine-tuned integral models
Scale I	71.25	73.75
Scale II	91.25	91.25
Scale III	88.75	90.00

6.2. Performance of nine fine-tuned DCNNs

The proposed EMS-Net algorithm uses an ensemble of an optimal subset of fine-tuned DCNN models, which has been selected via exhaustive searching all possible combinations. Table 7 and Table 8 give the accuracy of each model, the ensemble of three architectures trained at each scale, and the ensemble of each architecture trained at three scales at training stage and at testing images, respectively. At the middle part of Tables 7 and 8, the values outside the brackets are the classification accuracy, and the numbers inside the brackets are the ranks of the corresponding model among nine models. It shows that the results of ensemble of different models are consistent at training stage and at testing images. About network architectures, DenseNet-161 performs best on this task, whereas ResNet-101 performs worse than other two models. Meanwhile, using the images at Scale II results in higher classification accuracy, and using the images at Scale I results in worse accuracy. This conclusion is in line with the results plotted in Fig. 5.

In addition, for comparing fine tuning integral model with fine tuning each model individually, we combine three architectures (i.e., DenseNet-161, ResNet-152 and ResNet-101) to form integral model and fine tune it at each scale. We combine the three architectures by concatenating the outputs of last fully-connected layers to form a 12-D feature vector and adding a fully-connected layer with 4-D output. As shown in Table 9, when the combined models are fine-tuned integrally using the augmented patches at single scale, they perform slightly better than the combinations of three individually trained models, but still perform worse than the combination of the D–E–G with accuracy of 91.75%. That is due to the fine-tuned integral models can only extract features at single

scale, however, the combination of the D–E–G can extract diverse features at multiple scales.

6.3. Time complexity

The proposed EMS-Net algorithm involves patch augmentation, fine-tuning nine pre-trained DCNN models, and model selection. Therefore, it is much time-consuming in the training stage, especially in fine-tuning nine models. In the test stage, however, the time is mainly cost by the forward computation of three DCNN models, which is relatively fast.

The experiment was performed on an Ubuntu 14.04 64bit system (Intel Core i7-4790 CPU 3.2 GHz, NVidia GTX 1080Ti GPU and 32 GB memory). It took about 169 h in the training stage and less than 5 s to apply the proposed algorithm to classify each microscopy image. Although training the model is time-consuming, it can be done offline. The fast online testing suggests that our approach could be used in a routine clinical workflow.

7. Conclusion

In this paper, we propose the EMS-Net algorithm to classify H&E stained breast histopathological microscopy images into four categories including normal tissue, benign lesion, in situ carcinoma, and invasive carcinoma. This algorithm has been evaluated on the BACH Challenge dataset, and achieved an accuracy of 91.75% in the five-fold cross validation on training images and an accuracy of 90.00% in the online validation on testing images. Our future work will focus on introducing the attention mechanism to identify and use the most discriminative regions in each microscopy image.

Declaration of Competing Interest

The authors declare that there is no conflict of interests regarding the publication of this article.

Acknowledgment

This work was supported in part by the Science and Technology Innovation Committee of Shenzhen Municipality, China, under Grant JCYJ20180306171334997, in part by the National Natural Science Foundation of China under Grant 61771397, in part by the Fundamental Research Funds for the Central Universities under Grant 3102019G2019KY0001, in part by the Seed Foundation of Innovation and Creation for Graduate Students in Northwestern Polytechnical University under Grants ZZ2019029, and in part by the Project for Graduate Innovation team of Northwestern Polytechnical University. We appreciate the efforts devoted by the organizers of the ICIAR 2018 Grand Challenge on BreAst Cancer Histology to collect and share the data for comparing automatic detection and diagnosis algorithms for histopathological microscopy image.

References

[1] B. Stewart, C.P. Wild, et al., World Cancer Report 2014, The International Agency for Research on Cancer, 2014.
[2] G. Aresta, T. Araújo, S. Kwok, S. Saketh Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan, et al., BACH: Grand Challenge on Breast Cancer Histology Images (2018), arXiv:1808.04277.
[3] A.M. Khan, K. Sirinukunwattana, N. Rajpoot, Geodesic geometric mean of regional covariance descriptors as an image-level descriptor for nuclear atypia grading in breast histology images, in: Proceeding of the International Workshop on Machine Learning in Medical Imaging, Springer, 2014, pp. 101–108.
[4] X. Qi, F. Xing, D.J. Foran, L. Yang, Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set, IEEE Trans. Biomed. Eng. 59 (3) (2012) 754–765.
[5] J. Barker, A. Hoogi, A. Depeursinge, D.L. Rubin, Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles, Med. Image Anal. 30 (2016) 60–71.
[6] J.P. Vink, M. Van Leeuwen, C. Van Deurzen, G. De Haan, Efficient nucleus detector in histopathology images, J. Microsc. 249 (2) (2013) 124–135.
[7] H. Kong, Z. Lai, X. Wang, F. Liu, Breast cancer discriminant feature analysis for diagnosis via jointly sparse learning, Neurocomputing 177 (2016) 198–205.
[8] H. Jia, Y. Xia, Y. Song, D. Zhang, H. Huang, Y. Zhang, W. Cai, 3D APA-Net: 3D adversarial pyramid anisotropic convolutional network for prostate segmentation in MR images, IEEE Trans. Med. Imaging (2019) Published online.
[9] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, W. Cai, Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT, IEEE Trans. Med. Imaging 38 (4) (2018) 991–1004.
[10] K. Nazeri, A. Aminpour, M. Ebrahimi, Two-stage convolutional neural network for breast cancer histology image classification, in: Proceeding of the International Conference Image Analysis and Recognition, Springer, 2018, pp. 717–726.
[11] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, A. Campilho, Classification of breast cancer histology images using convolutional neural networks, PLoS One 12 (6) (2017) 1–14.
[12] S. Mehta, E. Mercan, J. Bartlett, D. Weave, J.G. Elmore, L. Shapiro, Y-Net: joint segmentation and classification for diagnosis of breast biopsy images (2018), arXiv:1806.01313.
[13] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 2261–2269.
[14] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[15] Y. Xie, J. Zhang, Y. Xia, Semi-supervised adversarial model for benign-malignant lung nodule classification on chest CT, Med. Image Anal. 57 (2019) 237–248.
[16] J. Zhang, Y. Xie, Y. Xia, C. Shen, Attention residual learning for skin lesion classification, IEEE Trans. Med. Imaging (2019) Published online.
[17] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proceeding of the Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[18] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition (2014), arXiv:1409.1556.
[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.
[20] J. Deng, W. Dong, R. Socher, L. Li, K. Li, F.-F. Li, ImageNet: a large-scale hierarchical image database, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2009, pp. 248–255.
[21] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, et al., U-Net: deep learning for cell counting, detection, and morphometry, Nat. Methods 16 (1) (2018) 67–70.
[22] G. Campanella, V.W.K. Silva, T.J. Fuchs, Terabyte-scale Deep Multiple Instance Learning for Classification and Localization in Pathology (2018), arXiv:1805.06983.
[23] A. Pimkin, G. Makarchuk, V. Kondratenko, M. Pisov, E. Krivov, M. Belyaev, Ensembling neural networks for digital pathology images classification and segmentation, in: Proceeding of the International Conference Image Analysis and Recognition, Springer, 2018, pp. 877–886.
[24] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the Twenty-Second ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016, pp. 785–794.
[25] A. Rakhlin, A. Shvets, V. Iglovikov, A.A. Kalinin, Deep convolutional neural networks for breast cancer histology image analysis, in: Proceeding of the International Conference Image Analysis and Recognition, Springer, 2018, pp. 737–744.
[26] S.S. Chennamsetty, M. Safwan, V. Alex, Classification of breast cancer histology image using ensemble of pre-trained neural networks, in: Proceeding of the International Conference Image Analysis and Recognition, Springer, 2018, pp. 804–811.
[27] Y. Liu, K. Gadepalli, M. Norouzi, G.E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P.Q. Nelson, G.S. Corrado, et al., Detecting Cancer Metastases on Gigapixel Pathology Images (2017), arXiv:1703.02442.
[28] PyTorch. (https://pytorch.org/).
[29] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization (2014), arXiv:1412.6980.

Zhanbo Yang received the B.S. degree in information and computing science from Northwestern Polytechnical University (NPU), Xi'an, China, in 2017. He is currently working toward the M.E. degree in computer science and technology from the School of Computer Science and Engineering, NPU. His research interests include medical data analysis and deep learning.

Lingyan Ran received his Bachelor degree and Ph.D. degree from Northwestern Polytechnical University (NWPU), Xi'an China, in 2011 and 2018. Earlier, he was a visiting scholar in Stevens Institute of Technology from 2013 to 2015. His research interests include image classification and semantic segmentation.

Shizhou Zhang received a B.E. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 2010 and 2017, respectively. Currently, he is with Northwestern Polytechnical University as an assistant professor. His research interests include content-based image analysis, pattern recognition and machine learning, specifically in the areas of deep learning, image classification and image parsing.

Yong Xia received the B.E., M.E., and Ph.D. degrees in computer science and technology from Northwestern Polytechnical University (NPU), Xi'an, China, in 2001, 2004, and 2007, respectively. From 2007 to 2013, he was a Postdoctoral Researcher at the Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney. Since 2014, he has been a Professor at the School of Computer Science and Engineering, NPU. His current research interests include medical image analysis, computer-aided diagnosis, pattern recognition, and machine learning.

Yanning Zhang received the Ph.D. degree in computer science and technology from Northwestern Polytechnical University (NPU), Xi'an, China, in 1996. She is currently a professor at the School of Computer Science and Engineering, where she is also the Director of the Shaanxi Key Lab of Speech & Image Information Processing (SAIIP). She is also the Assistant President, NPU. Her research interests include pattern recognition, image analysis, multimedia and deep learning.