

A DEEP LEARNING METHOD FOR DETECTING AND CLASSIFYING BREAST CANCER METASTASES IN LYMPH NODES ON HISTOPATHOLOGICAL IMAGES

Kaiwen Xiao¹, Zichen Wang¹, Tong Xu¹, Tao Wan^{1,2}

¹ DeepCare Inc., Beijing, China

² School of Biological Science and Medical Engineering, Beihang University, Beijing, China
 {txu,twan}@deepcare.com, taowan@buaa.edu.cn

ABSTRACT

We present an automated computerized approach for the identification and classification of metastatic breast cancer in digital whole slide images of sentinel lymph node biopsies through a combination of deep learning-based method and image processing techniques. In this work, we applied a simple thresholding based segmentation method to automatically detect the tissue regions mostly likely containing cancer metastasis and exclude image background, such as white space, in order to reduce computational cost. A ResNeXt based deep learning network was adopted in our work due to the fact of high efficiency and stable network structure. This 101-layer deep convolutional neural network was trained using a large amount of image patches from positive and negative regions of the histopathological images to make patch-level predictions to distinguish tumor-patches from non-tumor patches. The patch-level predictions were used to generate tumor probability maps, in which the morphological operations were performed to identify the most likely tumor regions with a maximum connectivity. An ensemble learning method was utilized to differentiate 5 different pN-stage of lymph node metastasis for every patient in the test dataset. The performance of classification task were assessed by using five class quadratic weighted kappa metric adopted by the Camelyon Grand Challenge 2017 (Camelyon17). This work suggested that our method could be useful in developing a computational diagnostic tool for discriminating breast cancer metastases with different pathologic stages from digital breast histopathological images, which might enable objective alternative for cancer diagnosis.

Index Terms— breast cancer metastases, lymph nodes, histopathological image; deep learning, convolutional neural networks

1. INTRODUCTION

Breast cancer is the most frequently diagnosed cancer and is the leading cause of cancer death among women [1]. A breast biopsy is a diagnostic procedure that can definitely determine

if the suspicious area is malignant (cancerous) or benign (non-cancerous). During biopsy, samples of tissue are removed from breast to produce stained histology slides, which are observed under microscope and graded by pathologists. The TNM system [2] is an internationally accepted means to classify the extent of cancer spread in patients with a solid tumor, which is one of the most important tools for clinicians to select a suitable treatment option and to obtain an indication of prognosis. In breast cancer, the TNM staging takes into account the size of the tumor (T-stage), whether the cancer has spread to the regional lymph nodes (N-stage), and whether the tumor has metastasized to other parts of the body (M-stage). The lymph nodes in the axilla are the first place breast cancer is likely to spread. Metastatic involvement of lymph nodes is one of the most important prognostic factors in breast cancer. Lymph nodes are surgically removed and examined microscopically when cancer has spread to the lymph nodes. However, the diagnosis of breast cancer metastases in lymph nodes is time-consuming and suffers subjective variances. Most importantly, small metastases are very difficult to detect and sometimes they are missed. Thus there is clearly a need for development of a automated breast cancer detection method that might avoid inter- and intra-reader variability thus improving the consistency of the decision-making process.

With the recent advent of whole slide digital scanners and advances in computational power, it is now possible to use digitized histopathological images and computer-aided image analysis to facilitate breast diagnosis and prognosis [3]. Analogous to the role of computer-aided diagnosis (CAD) algorithms in medical imaging and diagnostic radiology to complement the opinion of a radiologist, CAD algorithms have begun to be developed in biological images for disease detection, diagnosis, and prognosis prediction to complement the opinion of the pathologist [4, 5, 6, 7]. For instance, a number of techniques have been developed on automated nuclei segmentation in routinely stained breast histopathological images using methodologies, such as adaptive thresholding [8], region growing [9], watershed [10], graph cuts [11], active contour [12], statistical model [13], and deep learning [14]. Further, various intensity, morphology, and textural features

have been implemented for malignancy detection in breast cancer histopathology [9, 15, 16]. Such CAD systems on the breast histopathology could be used for increased diagnosis accuracy and reduced subjective variability of breast cancer grading as well the workload of the pathologists [17].

The primary goal of this paper is to develop an automated computerized method to detect and classify breast cancer in digitized histopathology using deep learning techniques. A pre-processing step using a simple thresholding segmentation method was adopted to eliminate the non-cancer background, such as white space appearing around tissue region, in order to focus on processing tissue regions and further increase computational efficiency. This step was efficiently performed through a high throughput parallelization strategy to speed up processing of large numbers of histopathological images [18]. The ResNeXt (ResNet v4) architecture was the chosen network for initial detection of breast cancer metastases in lymph nodes by producing tumor probability map on the digital pathological images. Finally, a ensemble learning method was utilized to distinguish every patient into 5 pN-stages, including pN0(no micro-metastases or macro-metastases or isolated tumor cells (ITCs) found), pN0(i+)(Only ITCs found), pN1mi(micro-metastases found, but no macro-metastases found, pN1(metastases found in 1-3 lymph nodes), and pN2(metastases found in 4-9 lymph nodes).

The rest of this paper is organized as follows. In Section 2, the description of dataset and evaluation metrics are give. Section 3 describes the details of the method. The experimental results are presented in Section 4. Finally, the concluding remarks are drawn in Section 5.

2. DATASET AND EVALUATION METRICS

Camelyon17 focus on the detection and classification of breast cancer metastases in lymph nodes, which is same as Camelyon16. The difference between Camelyon16 and Camelyon17 is that we need to focus on the results of each different tissue slice from a patient. Since the histological assessment of lymph node metastases is an essential part of TNM classification, Cameyon17 focus on the pathologic N-stage (pN-stage). The data in the Camelyon17 contains whole-slide images (WSI) of hematoxylin and eosin (*H&E*) stained lymph node sections. The ground truth is provided as lesion-level with detailed annotations of metastases in WSI and a patient-level with a pN-stage label per patient 5 medical centers. For training, 100 patients are provided and another 100 patients for testing. To compose a pN-stage, the number of positive lymph nodes (i.e. nodes with a metastasis) are counted. There are four categories of lymph node, including macro-metastases, micro-metastases, isolated tumor cells, and normal. The task is to automatically determine per patient with 5 pN-stages (pN0, pN0(i+), pN1mi, pN1, and pN2). For the evaluation of the results, we used five class quadratic

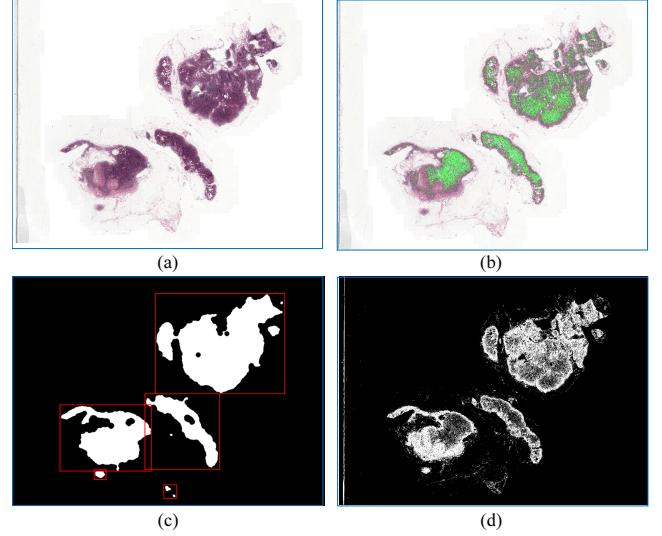


Fig. 1. Image preprocessing step for tissue segmentation. (a) Original Image (b) Cell-intensive areas (c) Binarized image (d) Regions of interest.

weighted kappa where the classes are the pN-stages.

3. METHODS

The task of Camelyon17 is to classify if the histopathological images contains tumor and to determine a pN-stage for every patient in test dataset. Our framework for detecting cancer metastases in sentinel lymph nodes can be modularized into three components, and they image preprocessing, tile-based classification using deep learning method, and classification with different pN-stage.

3.1. Image Preprocessing

It seems that we cannot put the whole images into memory because the smallest WSI has 1000M in the train datasets and the region of interest accounts for only 80% of a WSI. So we did foreground extraction through a certain algorithm. First, we get the entire image in low magnification and use the Otsu method to binarize the image. After that, the picture of the noise, black holes and artificial labels(noise) were eliminated according to the characteristics of these pictures via corrosion and expansion algorithm, which are used to get reasonable region to split images into patches. The flowchart of image pre-processing step is illustrated in Figure 1.

To reduce program computational running time, we offer two strategies: (1) Divide the all images task into N parts and transport them to different CPU processes; (2) Obtain the coordinate of each WSI and cut the target tiles by using parallel computing, which could save 70% of the time. Since the Camelyon challenge dataset was generated from five separate institutions, we wanted to eliminate any stain variability that

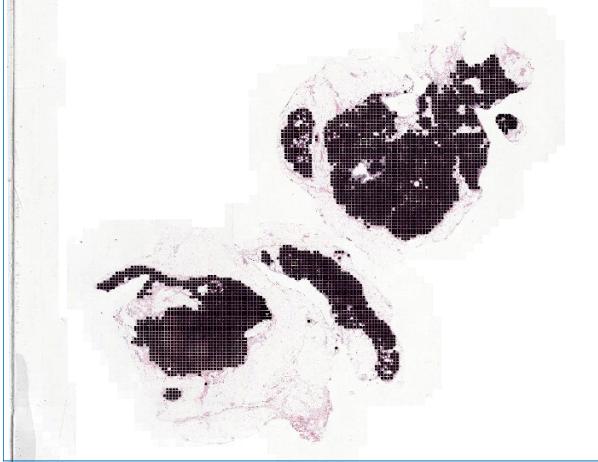


Fig. 2. Image patches generated from pre-processed histopathological image.

could negatively affect training and testing. By performing stain normalization, we can perform a non-linear color feature mapping that maps all of the respective training images by to a target stain. Without elimination of the stain variability, the heterogeneity of the stains will introduce bias to our model, inducing a bimodal stain vector color distribution.

Unlike others, we do not use more advanced and accurate network models, but rather work on data augmentation. We set the step size to be smaller than the picture size, which means that the data set will overlap. We used 512×512 size to cut the data set, through data augmentation to obtain the 3,000,000 patches in training set, as shown in Figure 2. At the same time we focus on strengthening the image of macrophages very similar to the cancer texture in negative samples, thus avoiding excessive noise in the heatmap. Meanwhile, we could delete the too white patches by setting the mean and variance threshold of the RGB, most of which are included fat cells.

3.2. Cancer Metastasis Detection Framework

After creating the training set tiles, we trained a convolutional neural network to discriminate between tissue and tumor tiles. Namely, we chose ResNeXt (ResNet v4) architecture with inputs sized $224 \times 224 \times 3$ (the default) to assess the value of initializing from existing models pre-trained on another domain. For each input patch, we used random crop and rotation (from MXNet) to generalize neural network. We label a patch as tumor if cancer area accounted for 40% of the entire picture. We explored the influence of the number of parameters by reducing the number of filters per layer while keeping the number of layers constant. For the choice of picture size, we did a series of experiments. If you choose too large size, will make the final resolution of heatmap significantly reduced, resulting in late classification model is difficult to achieve better results; if selected small size, will make the heatmap of the

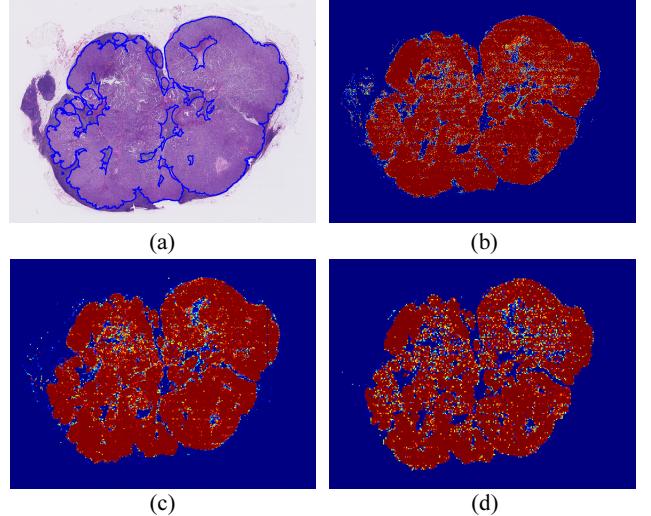


Fig. 3. Image preprocessing steps for tissue segmentation. (a) Ground truth; (b)Size of 256×256 (Noisy); (c) Size of 300×300 (middle-Noisy); (d) Size of 512×512 (clear).

noise point of a wide range. This is due to the presence or absence of adipose tissue and normal cell tissue in the cancer area where the actual image is calibrated. If the cut size is too small, the normal tissue image will enter the positive sample datasets, causing the model to become chaotic. Details can be found in Figure 3.

3.3. Slide Classification

The goal of classification is to discriminate 5 different pN-stages of lymph node metastasis for each patient study. Multi-class classification is a challenging problem, thus we performed two slide classification tasks, one to determine three categories into macro, micro and ITC/negative, and the other to distinguish ITC from normal lymph nodes.

3.3.1. Three-class classification

We apply ensemble algorithms to train a model to divide input heatmaps into three categories, as shown in Figure 4. 4100-dimensional features are extracted from each heatmap result as the training data, 4096-dimensional features are extracted by convolutional neural network and the others are local features of tumor region. we extract four features from each region, which are diameter of region, area of region, the cell number in the region and the length of the major axis of the ellipse that has the same normalized second central moments as the region. Then the largest (the value of diameter is the priority criteria, the value of area comes second) region's local features are selected as the features of the whole slide. We trained a convolutional neural network to classify three categories of lymph node and chose VGG16 architecture with inputs sized $224 \times 224 \times 3$ (the default) to assess the value of

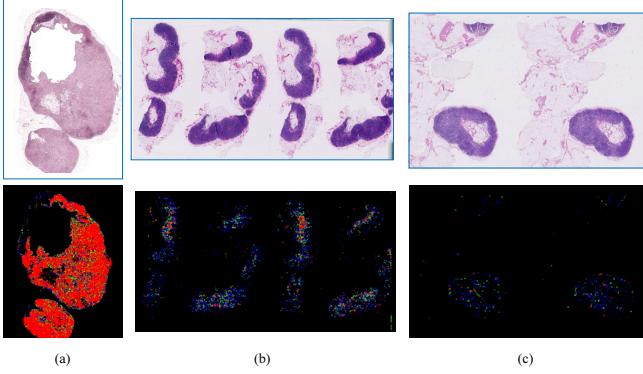


Fig. 4. Illustration for 3-class classification. (a) macro-metastases; (b) micro-metastases; (c) isolated tumor cells. The first line is original image, and the second line is the resulting image.

initializing from existing models pre-trained on another domain. And then we extract 4096-dimensional features from 7th fully connected layer of VGG16 of each image.

3.3.2. Two-class classification

To distinguish the ITC from normal lymph nodes, we trained another model based on the statistical features of prediction data and 4096-dimensional feature extracted by CNN. We did not compute local features for tumor regions in this task due to no difference between ITC and normal lymph nodes on heatmaps. We extract 7 statistical features of prediction data (the probability of each patch), they are: 1) Number of nonzero elements, 2) Maximum probability, 3) Sum of probabilities, 4) Average value of all probabilities, 5) Variance of data, 6) Standard deviation of data, 7) Average absolute deviation of data. We choose the same CNN architecture as three-classification and the implement details are also the same. Then we extract 4096-dimensional features from 7th fully connected layer of VGG16 of each image. Finally, we trained a classification model based on random forest algorithm.

4. EXPERIMENTAL RESULTS

4.1. Experimental Design

In the tumor detection stage, we trained our networks with stochastic gradient descent in MXNet , with running on four NVIDIA Pascal GPUs and i7-6850K. We used RMSProp with momentum of 0.99 and decay of 0.99. The initial learning rate was 0.05, with a decay of 0.5 every 2 million examples. For refining a model pretrained on ImageNet, we used an initial learning rate of 0.002. For the classification, we ran our model on one NVIDIA Pascal GPUs and i7-6850K. The RMSProp with decay of 0.0005 was used. The initial learning rate was 0.05. For refining, a model pretrained on ImageNet

Table 1. Two-class classification performance via precision, recall, F-score.

Class	Precision	Recall	F-score
ITC	0.97	0.91	0.94
Normal	0.91	0.97	0.94
Average	0.94	0.94	0.94

was utilized with an initial learning rate of 0.00005. There are three main parameters involved in the random forest based classifier. The number of base estimators in the ensemble was 25000, the number of features to draw from training data to train each base estimator was 64, and the minimum number of samples required to split an internal node was 5.

4.2. Results

We performed three-class and two-class classification tasks to discriminate breast cancer metastases in Lymph Nodes with different pN stages. The performance of classification task were assessed by using five class quadratic weighted kappa metric adopted by the Camelyon Grand Challenge 2017. In the three-class classification task, the best average kappa score was 99.54% using 5-fold cross validation scheme. In the two-class classification task, we also employed the 5-fold cross validation scheme and achieved the average kappa score of 89.57%. Table 1 lists the classification results measured by precision, recall, F-score.

5. CONCLUSION

In this paper, we developed a computerized image analysis method to automatically determine patients diagnosed with metastatic breast cancer with different pathological lymph node classes (pN-stage). We derived a new double-network model to make predictions for the lesion-based detection task and the slide-based classification task. For each slide, we computed two types of features, including handcrafted (local statistical) features and CNN-based extracted features. The integration of these two features could provide comprehensive representations for the identification and classification of metastatic breast cancer on histopathological images. The quantitative results using quadratic weighted kappa metric demonstrated that our method achieved good classification performance in distinguishing breast cancer metastases in histopathological images, which could be useful in establishing a computational diagnostic tool for diagnosing breast cancers.

6. REFERENCES

- [1] J.Ma and A.Jemal, “Breast cancer statistics, breast cancer metastasis and drug resistance,” Springer New York, pp. 1–18, 2013.

- [2] J. R. Benson, "The tn m staging system and breast cancer journal =," .
- [3] M. Gurcan, L. 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.
- [4] A. Madabhushi, S. Agner, A. Basavanhally, S. Doyle, and G. Lee, "Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data," *Comput. Med. Imag. Graph.*, vol. 35, no. 7-8, pp. 506–514, 2011.
- [5] 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.
- [6] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review-current status and future potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97–114, 2014.
- [7] X. Zhang, H. Dou, T. Ju, J. Xu, and S. Zhang, "Fusing heterogeneous feature from stacked sparse autoencoder for histopathological image analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 99, pp. doi: 10.1109/JBHI.2015.2461671, 2015.
- [8] S. Petushi, F. Garcia, M. Haber, C. Katsinis, and A. Tozeren, "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer," *BMC Med. Imag.*, vol. 6, no. 14, pp. doi:10.1186/1471-2342-6-14, 2006.
- [9] A. Basavanhally, S. Ganesan, S. Agner, J. Monaco, M. Feldman, J. Tomaszewski, G. Bhanot, and A. Madabhushi, "Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 3, pp. 642–653, 2010.
- [10] M. Veta, P. van Diest, R. Kornegoor, A. Huisman, M. Viergever, and J. Pluim, "Automatic nuclei segmentation in H&E stained breast cancer histopathology images," *PLOS ONE*, vol. 8, no. 7, pp. doi: 10.1371/journal.pone.0070221, 2013.
- [11] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 841–852, 2010.
- [12] J. Xu, A. Janowczyk, S. Chandran, and A. Madabhushi, "A high-throughput active contour scheme for segmentation of histopathological imagery," *Med. Image Anal.*, vol. 15, no. 6, pp. 851–862, 2011.
- [13] X. Zhang, F. Xing, H. Su, L. Yang, and S. Zhang, "High-throughput histopathological image analysis via robust cell segmentation and hashing," *Med. Image Anal.*, vol. 26, no. 1, pp. 306–315, 2015.
- [14] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imaging.*, vol. 35, no. 1, pp. 119–130, 2016.
- [15] M. Dundar, S. Badve, G. Bilgin, V. Raykar, R. Jain, O. Sertel, and M. Gurcan, "Computerized classification of intraductal breast lesions using histopathological images," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 7, pp. 1977–1984, 2011.
- [16] H. Irshad, "Automated mitosis detection in histopathology using morphological and multi-channel statistics features," *J. Pathol. Inform.*, vol. 4, pp. doi:10.4103/2153-3539.112695, 2013.
- [17] M. Veta, J. pluim, P. van Diest, and M. Viergever, "Breast cancer histopathology image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1400–1411, 2014.
- [18] T. Kurc, X. Qi, D. Wang, F. Wang, G. Teodoro, L. Cooper, M. Nalisnik, L. Yang, J. Saltz, and D. Foran, "Scalable analysis of big pathology image data cohorts using efficient methods and high-performance computing strategies," *BMC Bioinformatics*, vol. 16, no. 399, pp. doi: 10.1186/s12859-015-0831-6, 2015.