AUTOMATIC CLASSIFICATION ON PATIENT-LEVEL BREAST CANCER METASTASES

Sanghun Lee, Sangjun Oh, Kyuhyoung Choi, Sun Woo Kim

Deep Bio Inc.

ABSTRACT

Automatic diagnosis of breast cancer is a challenge that promises more accessible healthcare. In this paper, we describe the process of predicting slide-level cancer metastasis with machine learning techniques. First, a whole slide image is split into smaller patches which are classified for cancer by a model based on DenseNet, a Deep Neural Network with established performance. Next, the patch-level results are aggregated into a confidence map, which then goes through DBSCAN, a clustering algorithm, to reveal morphological features of cancerous regions. Finally, the minimal number of slides with the highest representative power is selected through independent repetitions of train-validation cycles with XGBoost. The resulting slide-level results from the trained XGBoost determine the pN stages of individual patients.

***Index Terms*—** Breast cancer, pN stage, deep neural network, DenseNet, DBSCAN, XGBoost

1. INTRODUCTION

Diagnosis of pathological whole slide images is not a trivial task; not only is the process time-consuming, but the results are also highly dependent on the pathologist's skill level, and even the most experienced pathologists may have discrepancies on viewing the same slide. Such circumstances may lead to unexpected human errors, which an automated diagnosis assistance system may be able to help prevent. This is one of the ends that the Camelyon17 challenge [1] aims to provide for by predicting pN-stages of 100 patients. The challenge dataset consists of 5 hematoxylin and eosin stained slide images of different lymph nodes from each subject. Five patient-level classes of pN stage, namely pN0, pN0(i+), pN1mi, pN1, and pN2, are automatically determined by 5 slide-level metastases, which are negative, micro-metastases, macro-metastases, isolated tumor cells (ITCs). Therefore, it is critical to classify each slide correctly.

Recently, Convolutional Neural Networks (CNNs), one of the Deep Neural Network models, have been showing revolutionary results on image recognition and classification tasks.[2] In ImageNet Large Scale Visual Recognition Competition (ILSVRC), advanced CNN models have shown state-of-the-art performance on classification of 1000 classes such as cats, dogs, etc. Since CNNs are not limited to a specific image domain, it can also be applied to the field of digital pathology. As an example, in a pathological study of breast cancer, CNN was used to achieve state-of-the-art results.[3]

In this paper, our automated diagnosis process consists of 3 steps. The first step is to classify patches whether these contain cancer or not. This is necessary due to the size of the whole slide images, considering that average size of whole slide image is about 200,000 by 100,000 pixels. Here, the patch classifier is trained on patches of 304 by 304 pixels, randomly generated from the annotated cancerous and non-cancerous regions. The second step is the hard example mining process. Upon observing the initial performance of generated heatmaps, additional normal patches were extracted from the training slides into the training dataset. The final step is to extract morphological features from each heatmap, and train XGBoost [4] , a boosting algorithm, to classify slide-level metastases. The patient-level pN stages are determined based on 5 slide metastases.

2. METHODS

The overall procedure, as illustrated in Figure 2, is as follows:

- Preprocess datasets, including patch extraction.
- Train patch-level classifier with a deep learning model based on DenseNet.
- Optimize the training data distribution with hard example mining.
- Extract morphological features from individual heatmaps clustered with DBSCAN algorithm. [5]
- Sample representative training set and train slide-level classifier with XGBoost.

2.1. Preparation of patch-level data set

The data set for the training patch classifier comes from both Camelyon16 and Camelyon17 dataset (abbreviated to the '16 and the '17 dataset, respectively). Assuming that the '16

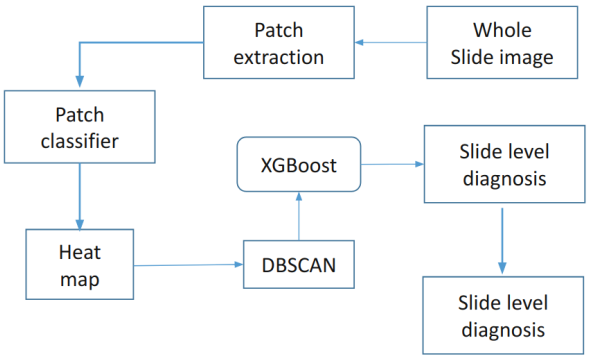


Fig. 1. whole procedure

dataset contains two different medical centers, total seven different stain styles are included. Cancerous and non-cancerous patches in '17 dataset are extracted from the slides containing annotations. Patches with cancer are generated inside annotation areas, while normal patches are extracted outside annotation areas, randomly without intersection. To examine various statistics, we balance the number of patches almost equally between cancer and normal patches. The resulting total number of patches in training, validation, test are 490,261, 245,127, and 245,127 respectively.

2.2. Train patch-level classifier

The patch classifier, as illustrated in Figure 3, is a model based on DenseNet [6], using 4 densely connected blocks. To address the problems of max-pooling as discussed in [7], the fully connected layers at the end are replaced by a fully convolutional layer. The initial learning rate is 0.1 and is reduced by one tenth per 10 epochs, and the optimizer is SGD with the decay of 1e-4.

2.3. Hard example mining with heatmap

By using the patch classifier, each whole slide image is transformed to a heatmap which considers a 304-by-304 pixel patch as a single pixel, as shown in Figure 3. Since normal patches may not be sufficient to represent the entire distribution of the normal types, additional normal patches are chosen from the heatmap regions that disagree the most with the reference annotation. Finally, the same patch classifier is trained again with the dataset with the additionally extracted normal types.

2.4. Extracting morphological features

To classify slide-level metastases, morphological features from heatmap are extracted by DBSCAN algorithm as shown in Figure 4. Per each of the three largest clusters within a

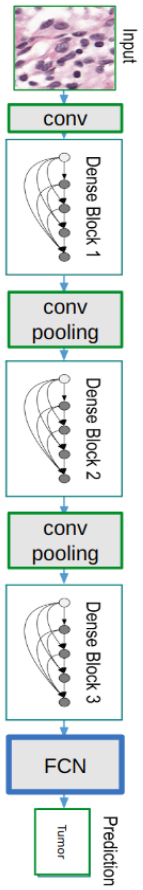


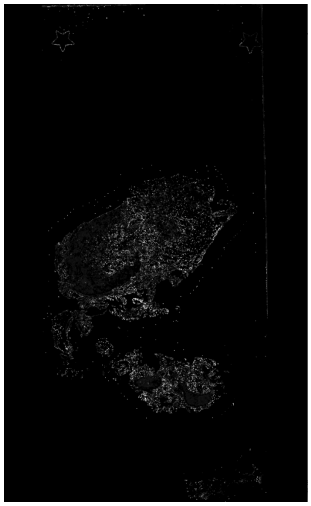
Fig. 2. Patch classifier

slide, features such as the major axis, mean probability, maximum probability, and minimum probability are extracted.

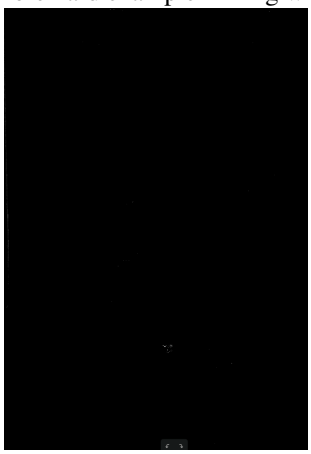
2.5. Training and predicting slide-level classifier with XGBoost

The extracted slide features are used to train XGBoost as the slide-level classifier. To capture sufficient cases of metastases while avoiding overfitting, we limit the size the training set to the 185 slides from 37 patients, using the remaining 315 slides from 63 patients as the validation set. XGBoost is trained on the randomly sampled training set of the given size, and then evaluated on the validation set. The sampling-training cycle is continued repeatedly and independently to find the training set of the highest representative power, which we define as having the best validation accuracy on both micro and macro metastases.

The remaining task of predicting patient-level pN-stage is automatically determined by slide-level metastases predictions.



(a) Heatmap before hard example mining without threshold



(b) Heatmap after hard example mining above threshold

Fig. 3. Regions in white show high predicted probability of cancer. Hard example mining shows significant decrease of false alarms.

3. RESULTS

Patch-level classifier shows 0.99 ROC and 0.99 PR-AUC in both validation and test patches. The optimal threshold in the validation patch set is 0.65, as chosen for the highest F1 score. With this optimal threshold, accuracy, recall, specificity, and precision yields 0.99, 0.98, 0.99 and 0.99, respectively, in the validation set.

False alarms in the heatmap level decrease significantly upon hard example mining, which compare the initially generated heatmaps against the ground-truth annotations.

Slide-level accuracy is 0.904 and 0.912 in the validation slides and the entire 500 slides. The kappa score for the validation 315 slides is 0.94. The validation slides for XGBoost differs from the 500 test slides used for the submission.

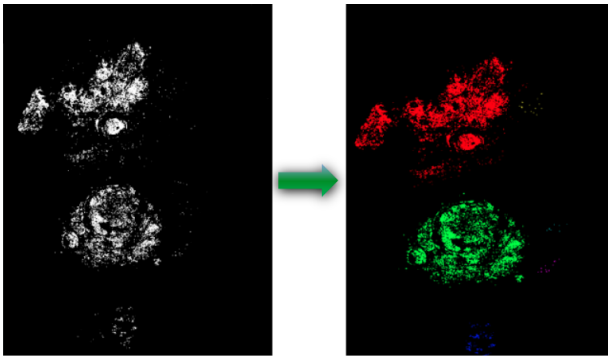


Fig. 4. The original heatmap(left) and the clusters after applying DBSCAN(right)

4. DISCUSSION

The prevalent approaches to predicting pN-stage focus on patch-level classifications. This is because the whole slide image is too big to be trained under the memory restrictions of the hardware, while the number of whole slide images is too small to train. A crucial limitation in this approach is that, although patch classifier reaches 0.99 accuracy, false alarms still occur in a certain ratio, making it difficult to catch ITC cases sensitively in slide-level. Hence we focus on the accuracy in negative and micro, macro metastases.

In our attempt previous to the current approach, overfitting proved to be the most serious risk, as shown by the difference between the validation and the test score. To avoid such issue, we have decreased the number of layers in the patch classifier down to one third, and selected a training set of the minimal size. The larger size of the validation set is expected to more adequately represent the actual distribution of the test set.

The annotation process is still time-consuming and expensive. In the field of pathology, it is costly to get a sufficient number of annotated data as the golden standard. Therefore, considering the limited number of annotated slides, unsupervised or at least semi-supervised learning is necessary to overcome the problem of annotation cost.

5. CONCLUSION

The main change in our algorithm lies in the structure of the patch-classifier. We still pertain to patch-level to slide-level approach, and therefore the same limitations still remain.

In future work, we will try unsupervised and semi-supervised learning to overcome the limited number of annotated slides.

6. REFERENCES

[1] Oscar Geessink, Péter Bándi, Geert Litjens, and Jeroen van der Laak, "Camelyon17: Grand challenge on cancer metastasis detection and classification in lymph nodes," 2017.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., pp. 1097–1105. Curran Associates, Inc., 2012.

[3] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep Learning for Identifying Metastatic Breast Cancer," *ArXiv e-prints*, June 2016.

[4] Tianqi Chen and Carlos Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2016, KDD '16, pp. 785–794, ACM.

[5] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu, "A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. 1996, KDD'96, pp. 226–231, AAAI Press.

[6] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[7] S. Sabour, N. Frosst, and G. E Hinton, "Dynamic Routing Between Capsules," *ArXiv e-prints*, Oct. 2017.