Fully-automated Peri-tubular Capillary Counting for Renal Allograft Rejection by Deep Neural Net Method: Extra-validation in Giga-pixel Immunostaining Pathologic Images

Young-Gon Kim   Gyuheon Choi   Heoungjeong Go   Hyunna Lee   Namkug Kim[*]
A-Reum Lee   BeomHee Park   Yong Won Cho   Sungwon Ham   Joon Beom Seo
Asan Medical Center & University of Ulsan College of Medicine
Seoul, Republic of Korea

Abstract

Pathologic scoring on renal allograft specimen mainly depends on pathologists' visual scoring, which would be time consuming and sensitive to inter or intra-observer variations. In this study, our aims are two folds; one is to propose a fully-automated system to find feasible regions of interest (ROIs) and count their number of C4d positive and negative in PTC on each feasible ROI by convolutional neural net (CNN) method in the giga-pixel immunostaining pathologic slide images. The other is to validate whether AI-assisted labeled data by a detection network is feasible. Our results showed that the performance in terms of area under curve (AUC) for classification model to find feasible ROIs is 0.9601 and the sensitivities to detect positive and negative in PTC is 0.9413 and 0.8523 at mean of false positive (FP)s 1.5 and 4 per ROI, respectively. In addition, we proved that AI-assisted labeled data is feasible by showing that the sensitivities were increased to 0.9522 and 0.8864 at the same mean FPs, respectively.

1   Introduction

In pathology, detection of lymph node metastasis of breast cancer with deep learning has been mainly studied [1], [2]. Because the demand for kidney transplantation is increasing worldwide, decision on kidney transplant rejection mainly depends on factors counting all C4d positives (stained) and negatives (unstained) in peri-tubular capillaries (PTCs). In the whole slice image, however, it is practically impossible for pathologists to quantify all PTC samples since image size is too big (100K x 200K) and its task is laborious, time consuming, and susceptible to intra and inter-observers' variation in clinical workflow. Therefore, we proposed a fully-automated system to select ROIs and detect PTCs. In addition, we validated whether AI-assisted labeled data is feasible to get more labeled data efficiently.

2   Materials and Methods

This study is approved by institutional review board of Asan Medical Center (AMC), Seoul, Korea. A total of 380 needle biopsy cases of renal allograft were collected for the study.

---

[*]Corresponding author: namkugkim@gmail.com

2.1 Materials

All slides were sampled from Department of Pathology at AMC from 2009 to 2016. Whole-slide images for 380 C4d Immunohistochemistry slides were obtained using Pannoramic 250 Flash digital slide scanner (3DHISTECH, Budapest, Hungary) with a 20x objective lens (specimen-level pixel size, 0.221um x 0.221um), and were anonymized before analysis and labeling tasks. All slides were randomly split into subset 1 (200 slides) and subset 2 (180 slides) datasets to validate feasibility of AI-assisted labeled data. In ROI classification, candidate ROIs of which size is 1024 x 1024 were extracted from subset 1 at randomly selected regions. All candidates are labeled as feasible ROIs 2,134 and non-feasible ROIs 769 by two pathologists in consensus. For PTC detection, C4d positive 549 and negative 1,274 in PTCs from the subset 1 were manually drawn by two pathologists in consensus. To validate feasibility of AI-assisted labeled data, candidate C4d positive 1,274 and negative 2,239 in PTCs were detected in feasible ROIs of subset 2 with the detection network trained by only subset 1. To get mass of labeled data, pathologists corrected them efficiently.

2.2 Methods

First, the background area is effectively removed by Otsu's thresholding, and then all windows having a certain size are scanned for all regions of the cell regions to classify feasible and non-feasible ROIs. All windows selected in previous step are classified as feasible or non-feasible ROI by a detection model. Then, C4d positive and negative PTC in the feasible ROIs are detected by classification model trained by another type of labeled data. In training the model for detection, a simple way to make detection model improved is proposed by adding a certain size of margin around the mask data while training the model.

2.2.1 Detection with Various Size of Margin

Faster RCNN algorithm was used to train a deep learning based detection model with pre-trained network (ResNet-50). Enlarged mask region with various sizes of margin from 0 to 70 with 10 intervals leads to better performance for PTC detection, which were compared in terms of free-response ROC curve (FROC) analysis.

2.2.2 Feasibility of AI-assisted Labeled Data

To get AI-assisted labeled dataset, detection of C4d positive and negative in PTC in feasible ROIs were performed for the subset 2 by the first and second models trained by only subset 1. Then, those datasets were corrected by pathologists efficiently. Comparison among accuracies of models trained by labeled data by only pathologists, AI-assisted labeled data, and combination of all data were conducted in terms of FROC analysis.

3 Results

3.1 Detection with Various Size of Margin

An example of labeled C4d positive in PTC with various sizes of margin was shown in Fig. 1. Fig. 2 shows the FROC analyses for both classes. The models with size of margin 40 for both classes showed the highest performance.
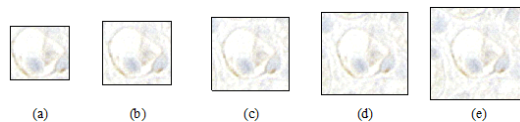

Figure 1: An example of labeled positive PTC with various margin. (a) margin 0, (b)-(e) margin 10 to 40 at step size 10.
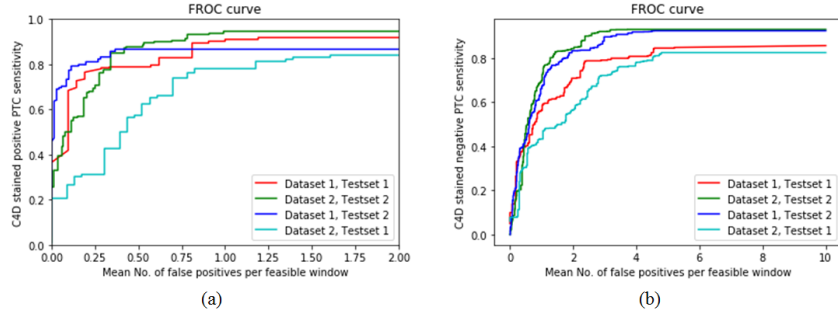
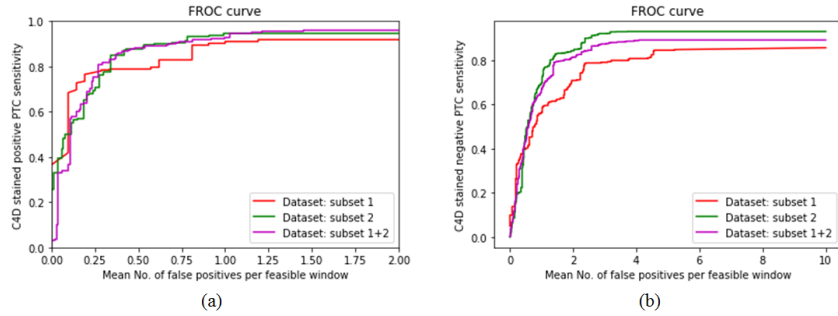Figure 2: FROC comparisons at different size of margin. (a) results for positive, (b) negative in PTC.

Figure 3: FROC comparisons for validation of combination of hard labeled and re-labeled data. (a) results for positive, (b) negative in PTC.

3.2 Feasibility of AI-assisted Labeled Data

To get the AI assisted dataset, the detection and classification model were used where the performance of the classification model in terms of area under curve was 0.9522 and that of the detection model in terms of recall and precision was 0.8823, 0.9312 (positive in PTC) and 0.8023, 0.7101 (negative in PTC), respectively. Overall FROC performances are shown in Fig. 3. Detection performance by combination of both labeled datasets for positive in PTC is higher (sensitivity 0.9522) than others (0.9413, 0.9272) and for negative in PTC is also higher (0.8864) than others (0.8523, 0.8193).

4 Conclusion

We proposed a detection method with CNN to automatically select feasible ROIs and count positive and negative in PTC. Experimental results of detection with size of margin 40 showed the best accuracy. In addition, AI-assisted labeled data generated from the model could be useful to increase number of training dataset.

Acknowledgments

References

[1] Teresa Araújo, Guilherme Aresta, Eduardo Castro, José Rouco, Paulo Aguiar, Catarina Eloy, António Polónia, and Aurélio Campilho. Classification of breast cancer histology images using convolutional neural networks. *PloS one*, 12(6):e0177544, 2017.

[2] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes van Diest, Bram van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM van der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210, 2017.