

Mycobacterial Detection and Localization on WSI Using Machine Learning

Chady Meroueh, MD¹; Jun Jiang, PhD²; Thomas Flotte, MD¹

¹ Department of Laboratory Medicine & Pathology, Mayo Clinic, Rochester, Minnesota, USA

² Department of AI and Informatics, Mayo Clinic, Rochester, Minnesota, USA

ABSTRACT:

Introduction:

Identification of mycobacteria in microscopic slides remains labor-intensive. Recently, object detection models (such as mitosis) have been introduced into clinical practice. However, detection of mycobacteria could be problematic due to the smaller size and lack of defined structures. We aim to evaluate the feasibility of using machine learning for mycobacterial detection within histological slides.

Method:

We collected 138 AFB-Stained slides from Mayo Clinic (2016-2018, 26 positive and 112 negative slides, 40x, Philips scanner). Five slides were annotated by 2 annotators with a Cohen's kappa of 86.9% (consensus by a senior pathologist), with the remainder annotated by a single pathologist. The initial dataset was divided using a ratio of 70 /15/15 (D0, Train: {627 patches, 819 instances}, Validation: {27 patches, 19 instances}, held-out Test: {36 patches, 26 instances}, Patch = 640x640 pixels). An out-of-the box YOLOv5 algorithm was trained in a transfer learning manner with weights initiated from a publicly available model (YOLOv5x, ImageNet). Hyperparameter tuning with mosaic data augmentation were used to train multiple models.

Result:

Using test-time image augmentation, an ensemble of 4 models (M1) showed a Precision 0.855 (Recall = 0.621, F1 = 0.72). This ensemble was then used to detect Mycobacteria on previously non-annotated patches (D1, Train: {531 patches, 1118 instances} + 504 negative patches). A single new model M2 (dataset = D0+D1) showed improvement in Recall (0.724), and F1 score (0.77).

Furthermore, 13 negative test whole slides were inferenced through our models (M1 and M2). A slide is considered positive if a single Mycobacteria is detected on any of the patches. M2 (Precision = 0.43, Recall = 0.75, time = 77 minutes, 1 T4 GPU) showed an improvement in all metrics compared to M1 (Precision = 0.28, Recall = 0.5, time = 250 minutes).

Conclusion:

In our preliminary testing, our trained model showed promising results for mycobacterial detection from our small dataset. This baseline model will prove extremely helpful in automatic annotation of numerous slides from multiple institutions and scanned using multiple scanners, in order to iteratively enrich our dataset for generalizability and real-world application.