Deep Learning in Automated Breast Cancer Diagnosis by Learning the Breast Histology from Microscopy Images

Qiangqiang Gu[1]; Naresh Prodduturi[1]; Steven N Hart[1]; Chady Meroueh[2]; Thomas J Flotte[2]

[1]Mayo Clinic, Department of Quantitative Health Sciences
[2]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN

Abstract

Background: Breast cancer is one of the most common cancers in women. With early diagnosis, some breast cancers are highly curable. However, the concordance rate of breast cancer diagnosis from histology slides by pathologists is unacceptably low. Classifying normal versus tumor breast tissues from breast histology microscopy images is an ideal case to use for deep learning and could help to more reproducibly diagnose breast cancer.

Methods: Using 42 combinations of deep learning models, image data preprocessing techniques, and hyperparameter configurations, we tested the accuracy of tumor versus normal classification using the **B**re**A**st **C**ancer **H**istology (BACH) dataset.

This approach included two steps. We first tested the patch-level validation accuracy of tumor versus normal classification for 16 combinations of nonspecialized deep learning models, image data preprocessing techniques, and hyperparameter configurations, and chose the model with the highest patch-level validation accuracy. Then we computed the slide-level validation accuracy of the selected models and compared them with 26 hyperparameter sets of a pathology-specific attention based multiple-instance learning model.

Results: Two generic models (One-Shot Learning and the DenseNet201 with highly tuned parameters) achieved 94% slide-level validation accuracy compared to only 88% for the pathology-specific model.

Conclusions: The combination of image data preprocessing and hyperparameter configurations have a direct impact on the performance of deep neural networks for image classification. To identify a well-performing deep neural network to classify tumor versus normal breast histology, researchers

should not only focus on developing new models specifically for digital pathology, since hyperparameter tuning for existing deep neural networks in the computer vision field could also achieve a high prediction accuracy.