

833

A SCALABLE DEEP LEARNING FRAMEWORK FOR RAPID AUTOMATED ANNOTATION OF HISTOLOGIC AND MORPHOLOGIC FEATURES FROM LARGE UNLABELED PAN-CANCER H&E DATASETS

David Soong*, David Soong, David Soong, Anantharaman Muthuswamy, Clifton Drew, Nora Pencheva, Maria Jure-Kunkel, Kate Sasser, Hisham Hamadeh, Suzana Couto, Brandon Higgs. *Genmab, Plainsboro, NJ, USA*

Background Recent advances in machine learning and digital pathology have enabled a variety of applications including predicting tumor grade and genetic subtypes, quantifying the tumor microenvironment (TME), and identifying prognostic morphological features from H&E whole slide images (WSI). These supervised deep learning models require large quantities of images manually annotated with cellular- and tissue-level details by pathologists, which limits scale and generalizability across cancer types and imaging platforms. Here we propose a semi-supervised deep learning framework that automatically annotates biologically relevant image content from hundreds of solid tumor WSI with minimal pathologist intervention, thus improving quality and speed of analytical workflows aimed at deriving clinically relevant features.

Methods The dataset consisted of >200 H&E images across >10 solid tumor types (e.g. breast, lung, colorectal, cervical, and urothelial cancers) from advanced disease patients. WSI were first partitioned into small tiles of 128 μ m for feature extraction using a 50-layer convolutional neural network pre-trained on the ImageNet database. Dimensionality reduction and unsupervised clustering were applied to the resultant embeddings and image clusters were identified with enriched histological and morphological characteristics. A random subset of representative tiles (<0.5% of whole slide tissue areas) from these distinct image clusters was manually reviewed by pathologists and assigned to eight histological and morphological categories: tumor, stroma/connective tissue, necrotic cells, lymphocytes, red blood cells, white blood cells, normal tissue and glass/background. This dataset allowed the development of a multi-label deep neural network to segment morphologically distinct regions and detect/quantify histopathological features in WSI.

Results As representative image tiles within each image cluster were morphologically similar, expert pathologists were able to assign annotations to multiple images in parallel, effectively at 150 images/hour. Five-fold cross-validation showed average prediction accuracy of 0.93 [0.8–1.0] and area under the curve of 0.90 [0.8–1.0] over the eight image categories. As an extension of this classifier framework, all whole slide H&E images were segmented and composite lymphocyte, stromal, and necrotic content per patient tumor was derived and correlated with estimates by pathologists ($p < 0.05$).

Conclusions A novel and scalable deep learning framework for annotating and learning H&E features from a large unlabeled WSI dataset across tumor types was developed. This automated approach accurately identified distinct histomorphological features, with significantly reduced labeling time and effort required for pathologists. Further, this classifier framework was extended to annotate regions enriched in lymphocytes, stromal, and necrotic cells – important TME contexture with clinical relevance for patient prognosis and treatment decisions.

<http://dx.doi.org/10.1136/jitc-2021-SITC2021.833>