Towards Semantic Encoding in Digital Pathology using Weakly Supervised Learning Strategies

Eitan Halper-Stromberg[1], Alex Baras[1]

[1]Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD

Background

Model architectures (ResNet, DenseNet, etc) pre-trained on the ImageNet dataset (transfer learning) has become a conventional approach in digital pathology. This convention has resulted in good performance over a variety of studies, though the latent representation of a given image carries no semantic meaning and is difficult to manage considering tiling approaches for larger images can result in thousands of tiles. In this study we generated more compact encodings of input images that carried with it semantic meaning. Subsequently, we tested whether the proposed semantic encoding strategies generalized to classification machine learning tasks in this imaging context.

Methods

We used whole slide imaging (WSI) of conventional H&E slides from The Cancer Genome Atlas (TCGA) cohort. The 11,766 WSI were tiled at 12 different magnification levels and encoded using DenseNet201 pretrained on on ImageNet. This resulted in roughly 3 million tiles encoded into the 1920-dimensional feature vector produced by DenseNet201. These tiles were subsequently labeled by the diagnosis of the WSI from which they were extracted; here we used both the original TCGA codes (33 classes) and a more granular mapping of the diagnostic entities within the TCGA to NCI-Thesaurus ontology (53 classes). This is weak supervision since, to a variable degree, not every tile extracted from a WSI is representative of the diagnosis of the whole WSI. We employed a simple fully connected network (FCN), with drop-out, trained with DenseNet201 encodings as inputs to predict the above-described weakly supervised tile label for both TCGA and NCIT. The output of these FCN represent a semantic encoding of the content within a given tile. To test encodings, manually curated datasets from the TCGA tiles representative of the WSI label were generated for the following classification tasks: (i) lung squamous cell carcinoma vs lung adenocarcinoma (ii) adrenal cortical carcinoma vs renal clear cell carcinoma (iii) seminoma versus non-seminoma of the testes. An orthogonal approach (random-forest classifier) to the FCN was used as the classifier in these tasks and evaluated in 6-fold cross validation. In each task above, an out-group of random tiles was included in the classification task.

TCGA Label	Tiles Used
Liver hepatocellular carcinoma	311,779
Lung adenocarcinoma	285,455
Sarcoma	247,685
Head and Neck squamous cell carcinoma	226,137
Stomach adenocarcinoma	225,334
Brain Lower Grade Glioma	203,306
Testicular Germ Cell Tumors	184,853
Lung squamous cell carcinoma	174,271
Skin Cutaneous Melanoma	139,984
Thyroid carcinoma	122,302
Esophageal carcinoma	113,396
Uterine Corpus Endometrial Carcinoma	111,385
Breast invasive carcinoma	96,316
Colon adenocarcinoma	88,709
Kidney renal clear cell carcinoma	70,781
Prostate adenocarcinoma	67,583
Bladder Urothelial Carcinoma	63,569
Kidney renal papillary cell carcinoma	57,961
Glioblastoma multiforme	57,177
Adrenocortical carcinoma	38,115
Thymoma	37,328
Pheochromocytoma and Paraganglioma	37,092
Cervical squamous cell carcinoma and endocervical adenocarcinoma	32,528
Diffuse Large B-cell Lymphoma	29,538
Pancreatic adenocarcinoma	27,199
Uterine Carcinosarcoma	16,920
Rectum adenocarcinoma	13,710
Kidney Chromophobe	11,856
Mesothelioma	9,664
Ovarian serous cystadenocarcinoma	638
Cholangiocarcinoma	9,301
Uveal Melanoma	5,744
Total	3,117,616

Table 1: Number of Tiles Generated for each TCGA Label (NCI-T Numbers and labels not displayed)

Results

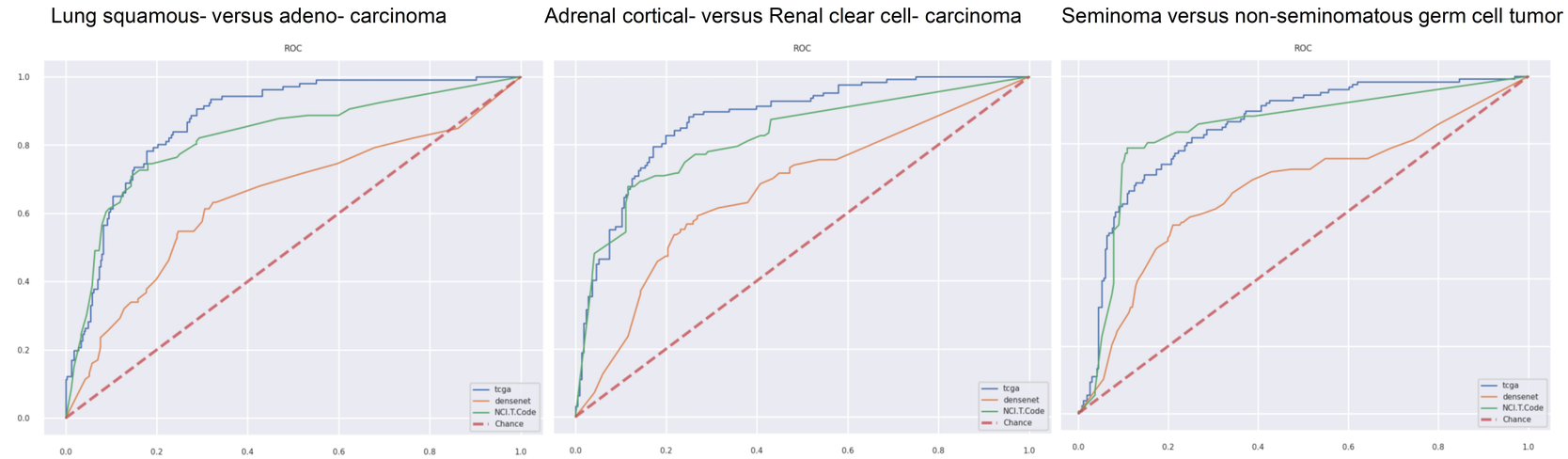
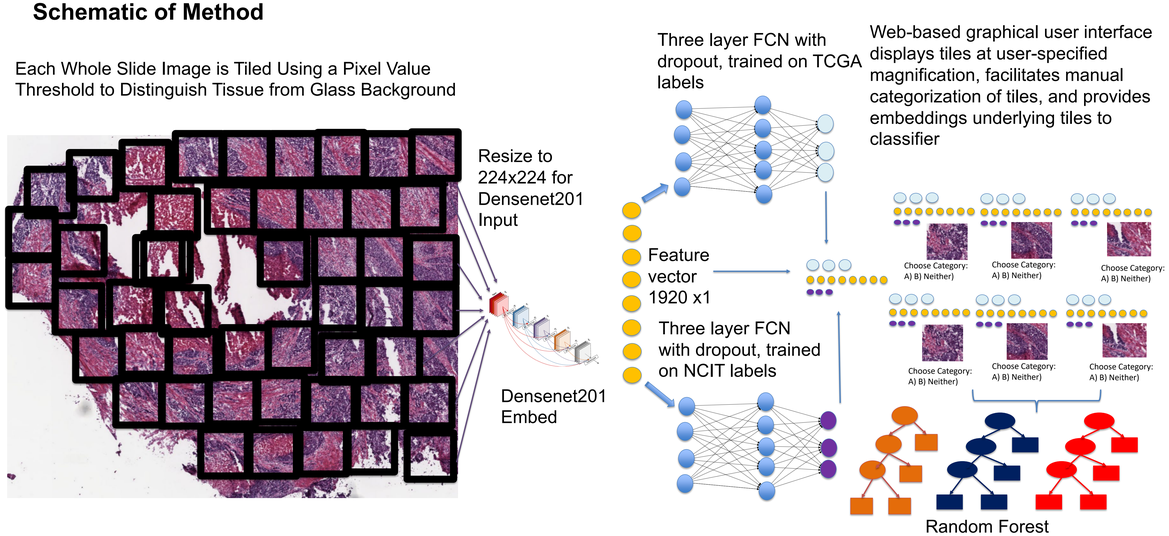


Figure 1: Receiver Operating Characteristic Curves For Three Classification Tasks Contrasting Semantic Encodings from TCGA, NCIT, and Densenet201

In each of the classification tasks examined, the semantic encodings from the TCGA and NCIT labels exhibited increased performance relative to Densenet201 pretrained on ImageNet. For lung squamous cell carcinoma versus lung adenocarcinoma, the classifiers derived from DenseNet201, TCGA, and NCIT encodings had AUCs of 0.66, 0.87, and 0.81, respectively. For adrenal cortical carcinoma versus renal clear cell carcinoma the classifiers based on DenseNet201, TCGA, and NCIT encodings had AUCs of 0.68, 0.88, and 0.85, respectively. For seminoma versus non-seminoma the classifiers based on DenseNet201, TCGA, and NCIT encodings had AUCs of 0.75, .90, 0.91, respectively. (see ROC plots)

Schematic of Method



Conclusion

Classifiers using TCGA and NCIT derived semantic encodings outperformed classifiers using Densenet201 derived encodings in sets of manually curated tiles from a subset of the tiles used to generate encodings. These findings suggest that hybrid strategies of combining transfer learning with domain specific learning may generate more optimal encoding strategies in digital pathology.

Contact

ehalper2@jhmi.edu
https://www.linkedin.com/in/eitan-halper-stromberg-md-phd-4a314bb/
baras@jhmi.edu
https://www.linkedin.com/in/alexander-baras-md-phd-03986586/