Efficient cancer classification from tissue slide images using CNNs

DS-GA 1006: 11-30 Project Update

Joyce Wu (jmw784)
Eduardo Fierro (eff254)
Raul Delgado Sanchez (rds491)

November 30, 2017

Update

Since the last update, we have now successfully beaten our advisors' macro AUC of 0.97, and achieved an macro AUC of 0.976 on Lung Cancer with a smaller model that is much faster to train.

Furthermore, we have successfully finished tiling and sorting the breast and kidney cancer images. With the same hyperparameters from our best model with lung cancer, we have trained a new model to achieve a macro AUC of 0.985 with kidney cancer. We are in the process of training a model for breast cancer classification with the same settings.

We have also successfully implemented Google's Inception V3. With transfer learning, using weights pre-trained on ImageNet, we have achieved around 0.90 AUC. We are also planning to run a trial where we train all of the parameters (rather than just the last linear layer). However, because our other model is performing better at this point, we are not as concerned about tuning Inception V3 to the same performance as our advisors.

At this point, we are primarily working on evaluation, visualization, and documentation for the poster and final report. An example visualization can be seen in Figure 1 on the next page. The argmax prediction is depicted by the color, and the probability score for that prediction is represented by the opacity (a little difficult to see). We are also looking into how to visualize the features of the model.

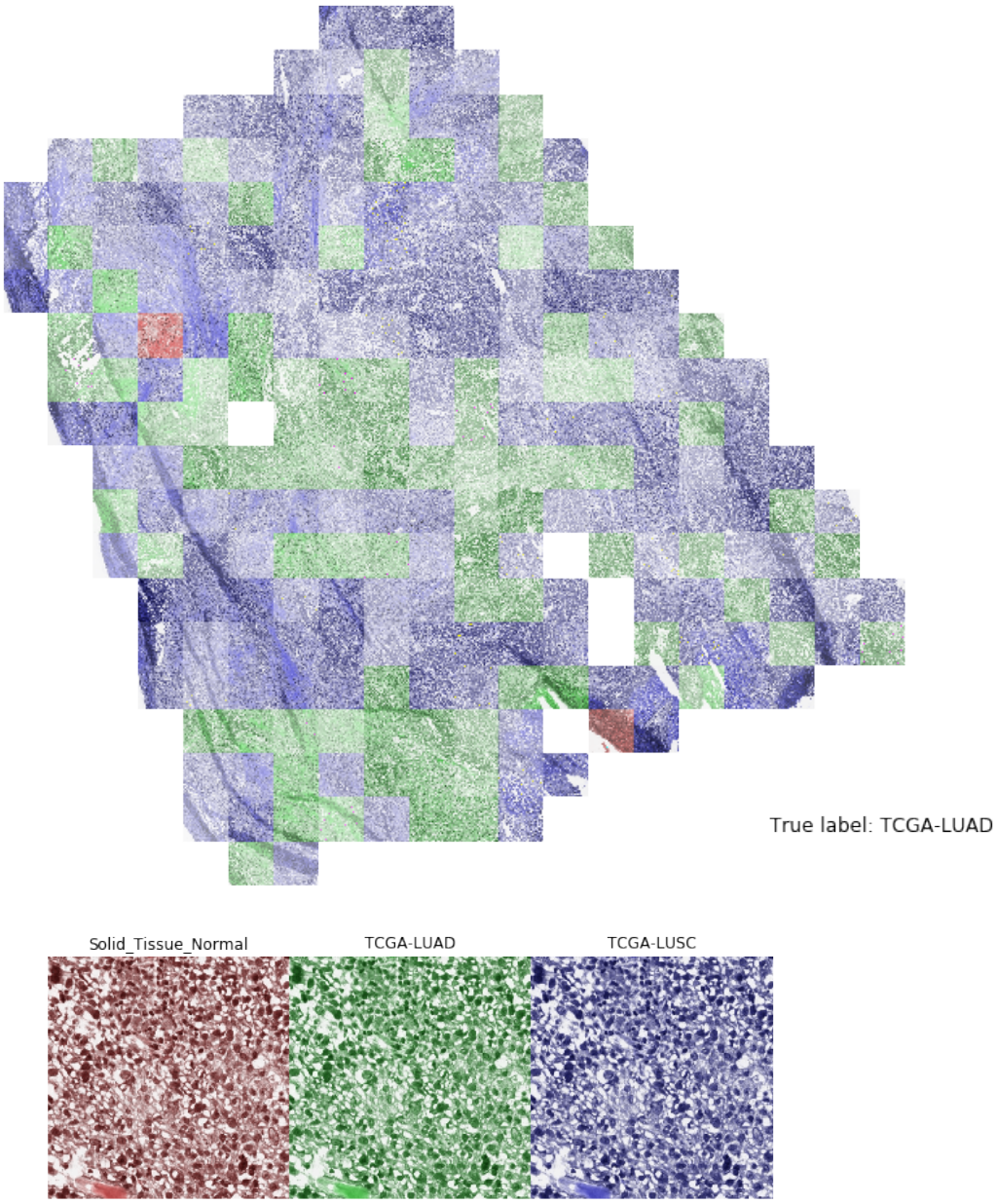

Figure 1: Slide prediction from individual tiles