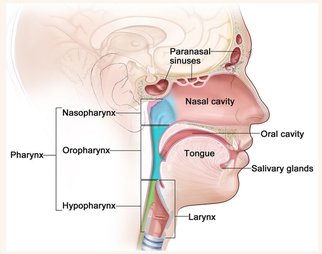
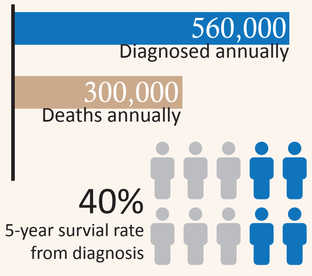
Patch-based Molecular Subtype Classification of Head and Neck Cancer with Neural Networks

Wanyi Qian, Guoli Yin, Frances Liu
Stanford University, Spring 2017, CS 231N Convolutional Neural Networks for Visual Recognition

Introduction



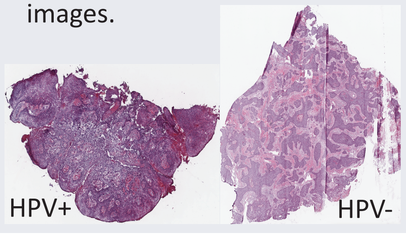
Head and neck cancer (HANC) ranks the seventh most common type of cancer in the US. Effective treatment of HANC often relies on correctly identifying the cancer subtypes, but this is very difficult to achieve given the inherent heterogeneity of the head and neck squamous cell carcinoma (HNSCC). Recently, tumors that test positive for human papillomavirus (HPV) seems to be more reponsive to chemotherapy and radiation than the HPV-negative cases. Previously, genotype variations have been found among different subtypes, but no researches are on molecular level classification.

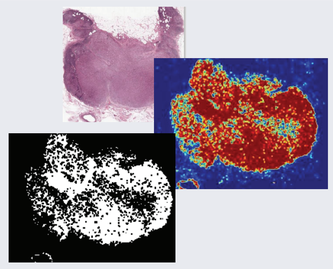


We apply deep learning with Convolutional Neural Networks (CNN) to whole slide images (WSI) of tumor sections. This has been done on breast and lung cancer, but not HANC. Our goal for this project is to explore molecular level subtypes and capability of deep learning for classification.

Problem Statement

- HPV is an important risk factor for HANC. We use CNN to examine if there are any molecular-level differences between HPV-positive and negative subtypes of HANC using hematoxylin and eotin stained pathology images.

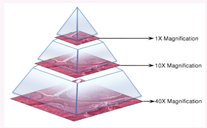


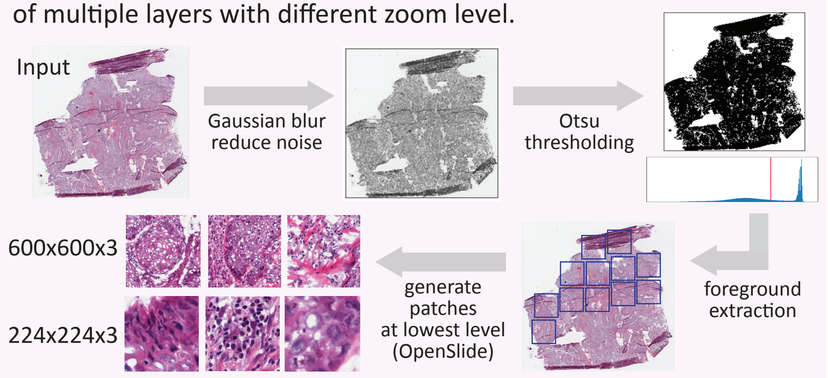


- We evaluate our model based on the accuracy, and attempt to compare the performance of different models with different patch sizes based on their accuracy

Dataset

Our data comes from The Cancer Genome Atlas (TCGA) curated by the National Institute of Health (NIH). The complete dataset for HANC contains 582 patient cases. For computational efficiency, we select 86 HPV-negative and 79 HPV-positive samples for training and testing. Pathology images are loaded in svs format, which comprises of multiple layers with different zoom level.



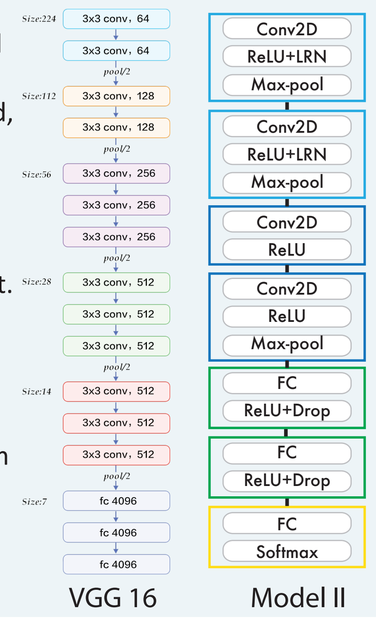


Methods

Pre-processing the data includes foreground extraction in the WSI and generating patches from discriminative regions. Two patch sizes are used, 224x224 for VGG, and 600x600 for model II. The patches inherit the WSI's class label.

The first model used is VGG16 with weights pretrained on ImageNet. We freeze the first 10 Conv2d/Max-Pooling llayers and fine-tune the last block of layers with our training and validation data.

The second model is derived from Hou et al. 2016. The architecture is simpler and designed primarily for processing lung cancer pathology images in binary classification.



Findings

Model	Optimizer	Patch size	Learning rate	Accuracy
Model II	RMSProp	600x600x3	1e-5	0.6269
Model II	SGD	600x600x3	1e-5	0.5706
Model II	Adam	600x600x3	1e-5	0.58
VGG		224x224x3	1e-3	0.58
Best		400x400x3		0.77

Obeservations so far:

- Random sampling of patches doesn't help increase accuracy.
- We have tried patch size of 128, 224, and 600. Larger patch sizes tend to have more predictive power.
- Assuming all patches are discriminative according to their image label, this introduces noisy label problem.

Steps going forward:

- Run VGG model with updated patches.
- Explore expectation maximization and construct discriminative masks for better patch extraction.
- Find ways for meaningful contribution given the limitations of the dataset.

Conclusion

Results show that the deep learning models shows prediction power and its accuracy is at comparable levels with similar studies (0.6-0.7) (Hou et al. 2016). The combination of larger patch size and customized model architecture seems to render more accurate predictions.

The main problem with the patch-based pathology image classification is the lack of ground truth labels for the individual patches. We are not able to obtain masks delineating (strictly) regions with predictive power within the WSIs. For this reason, many patches are falsely assigned labels solely based on its parent slide, thereby suppressing the CNN's ability to learn.

We hope to continue exploring expectation maximization and noisy label learning methods in the following week.