Efficient cancer classification using Convolutional Neural Networks

Joyce Wu, Eduardo Fierro, Raul Delgado Sanchez

NYU Center for Data Science Capstone Project

Objectives

- Replicate the work previously done by our advisors on lung cancer to strengthen the validity of their work
- Build a simpler neural network architecture that has comparable or better performance with deeper models such as Google's Inception V3
- Apply this neural network architecture to other cancers, such as kidney cancer and breast cancer

Introduction

Advances in deep learning have facilitated great improvements for the task of image processing and classification. In particular, convolutional neural networks (CNNs) first proposed by Yann Lecun [1] have quickly risen to the state-of-the-art on almost all image based tasks.

We sought to build a CNN architecture for the task of diagnosing cancer subtypes based on tissue slide images. While this task is relatively easy for pathologists, as cancerous and non-cancerous tissue are quite visually different, building and training such a model is a starting point for applying the method to more difficult tasks - such as mutation detection.

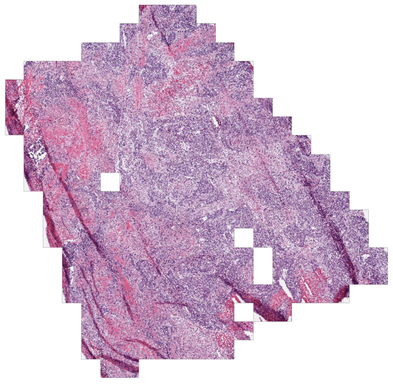


Figure 1: Example tissue slide image of lung cancer

As the data size for medical images is extremely large (20k resolution), we employed a method to grid the larger slide images into smaller tiles for training. This assumes that the label for the entire slide is consistent for each tile within the slide. Final prediction is done by aggregating scores of each tile for a particular slide. This method was proposed in [2] for the same task on lung cancer. Using several classic machine learning models such as Random Forest, SVM, and Naive Bayes, they achieved AUC $\sim$ 0.87 for distinguishing between cancer and non-cancerous tissue, and $\sim$ 0.78 for distinguishing between the cancerous subtypes.

Hyperparameters and methods

1. Tiling and sorting
2. Data augmentation

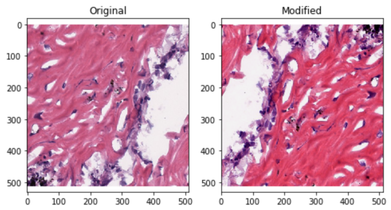


Figure 2: Example tile with augmentation

3. Weight initialization
4. Nonlinearity
5. Optimizer
6. Tile aggregation for prediction
7. Evaluation

Results

Table 1: Macro AUC performance for various cancers

Dataset	Lung	Kidney	Breast
Validation	0.977	0.985	0.994
Test	0.947	0.995	0.996

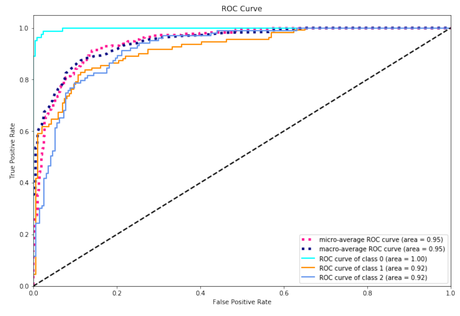
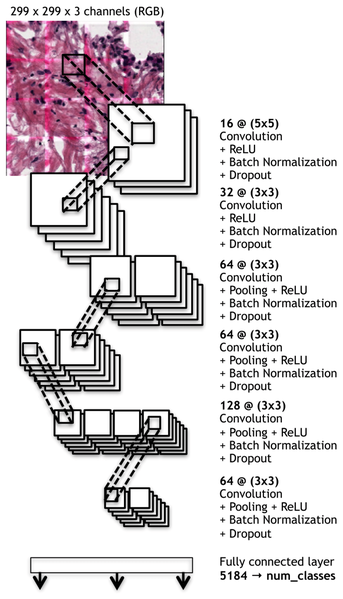


Figure 3: Lung cancer - ROC curve on test set

The model was quite good at distinguishing cancerous tissue from non-cancerous tissue, and performed slightly worse distinguishing between the cancerous subtypes.

Final architecture



Visualization

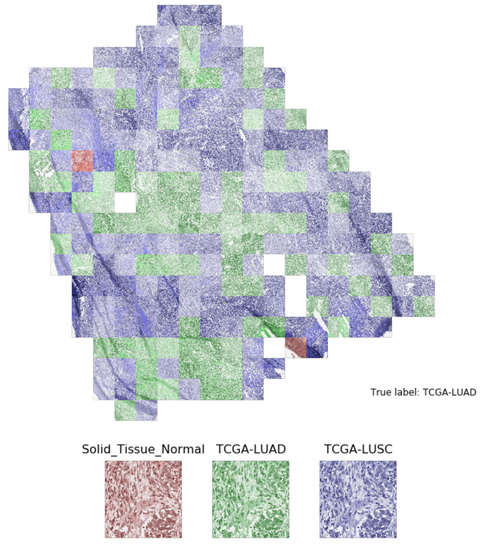


Figure 4: Aggregated prediction using tiles of a slide

In Figure 4, the confidence score of the predicted class for each tile is proportional to the opacity.

Conclusions

One can see a significant drop in performance between validation and test for lung cancer when compared to kidney cancer (which actually did better on test). This may be because we had $\sim$ 60% more tile data for kidney cancer in the training set (due to difference in available data and the number of tiles that were dropped from background percentage).

Data augmentation helped with performance, although it greatly slowed down training as the random augmentation was calculated each time a batch was loaded.

Overall, performance was quite good with AUC $\sim$ 0.998 for distinguishing between cancerous and non-cancerous tissue and AUC $\sim$ 0.92 - 0.99 for distinguishing between cancerous subtypes. In general, this is a significant improvement over the machine learning methods in [2]. Furthermore, the same framework was successfully applied to several different cancers!

Future work

- We spent a lot of time setting up the framework for tiling, sorting, training, augmentation, and aggregation with many different options, but could have done more tuning and run more experiments to achieve even better results.
- Application of this framework to even more types cancers and mutations of cancer.

References

[1] Y. LeCun and Y. Bengio.
Word-level training of a handwritten word recognizer based on convolutional neural networks.
IEEE, 1994.

[2] K. Yu et al.
Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features.
Nature Communications, 2016.

Acknowledgements

Thank you to our advisors Aristotelis Tsirigos, Nicolas Coudray, Narges Razavian from the NYU School of Medicine, Applied Bioinformatic Laboratories for their guidance and support with our project.

NYU Center for Data Science