

Analysis of non-small cell lung cancer histopathology images using deep learning



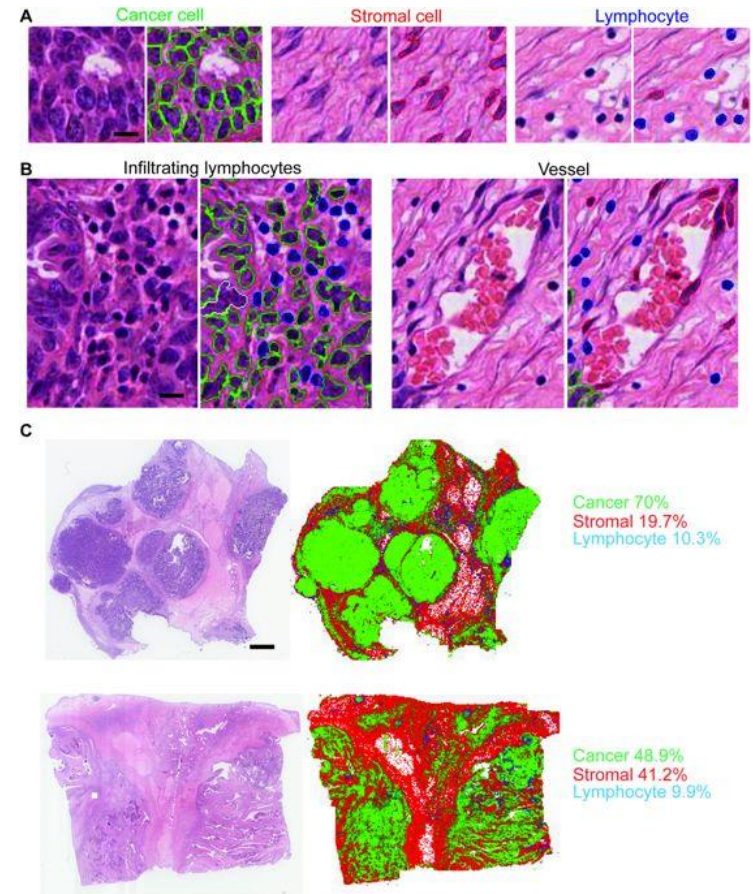
Saman Farahmand
10/12/2018

Project Goal

Developing a deep learning model to predict patients diagnosed with tumors from microscopic images of tissue biopsies and genomic biomarkers (TCGA)

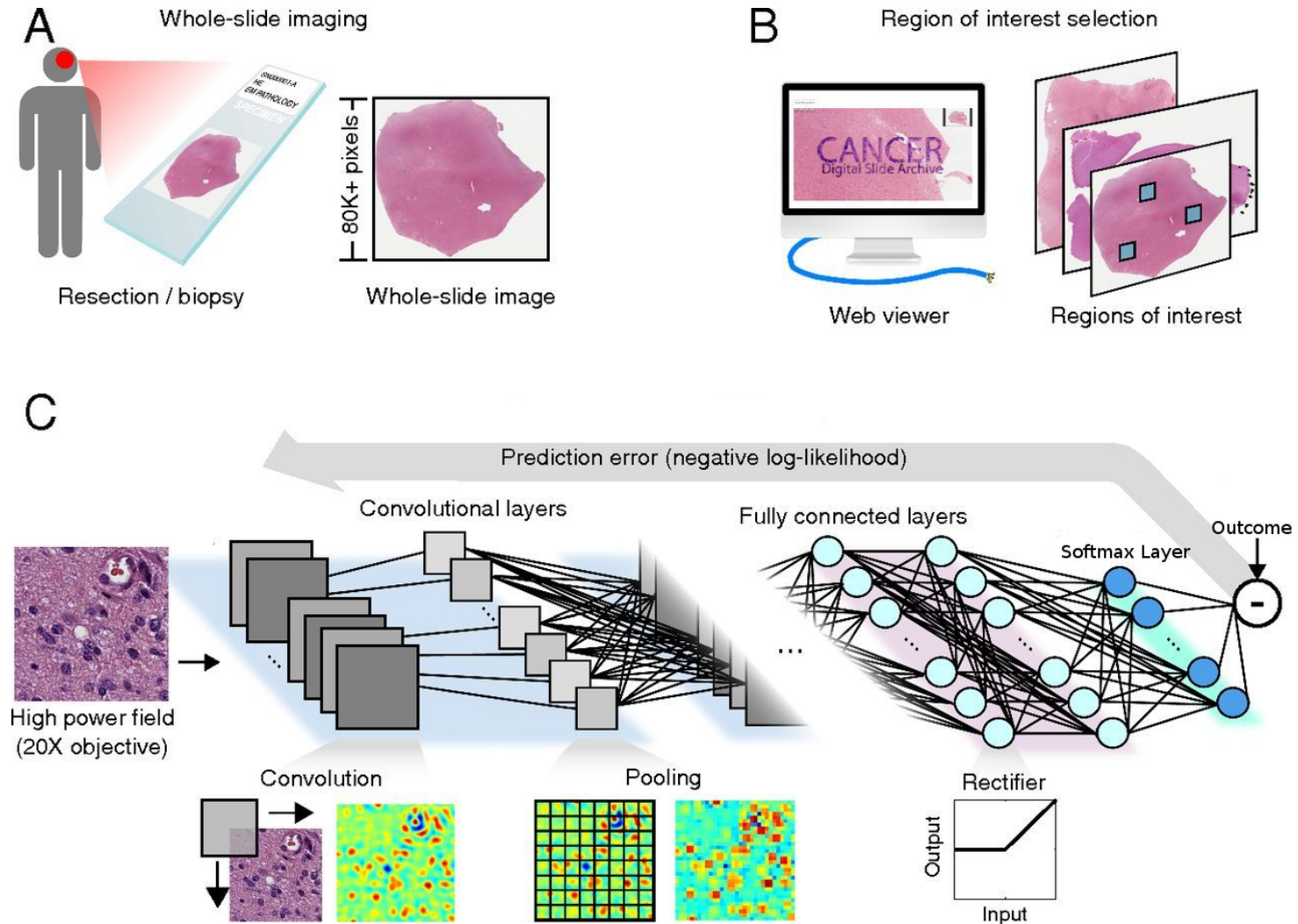
Introduction

- Hematoxylin and Eosin stained biopsy slides
- Widely used stain in medical diagnosis and is often the gold standard by pathologist
- Large whole-slide images are generated by digitizing H&E-stained glass slides



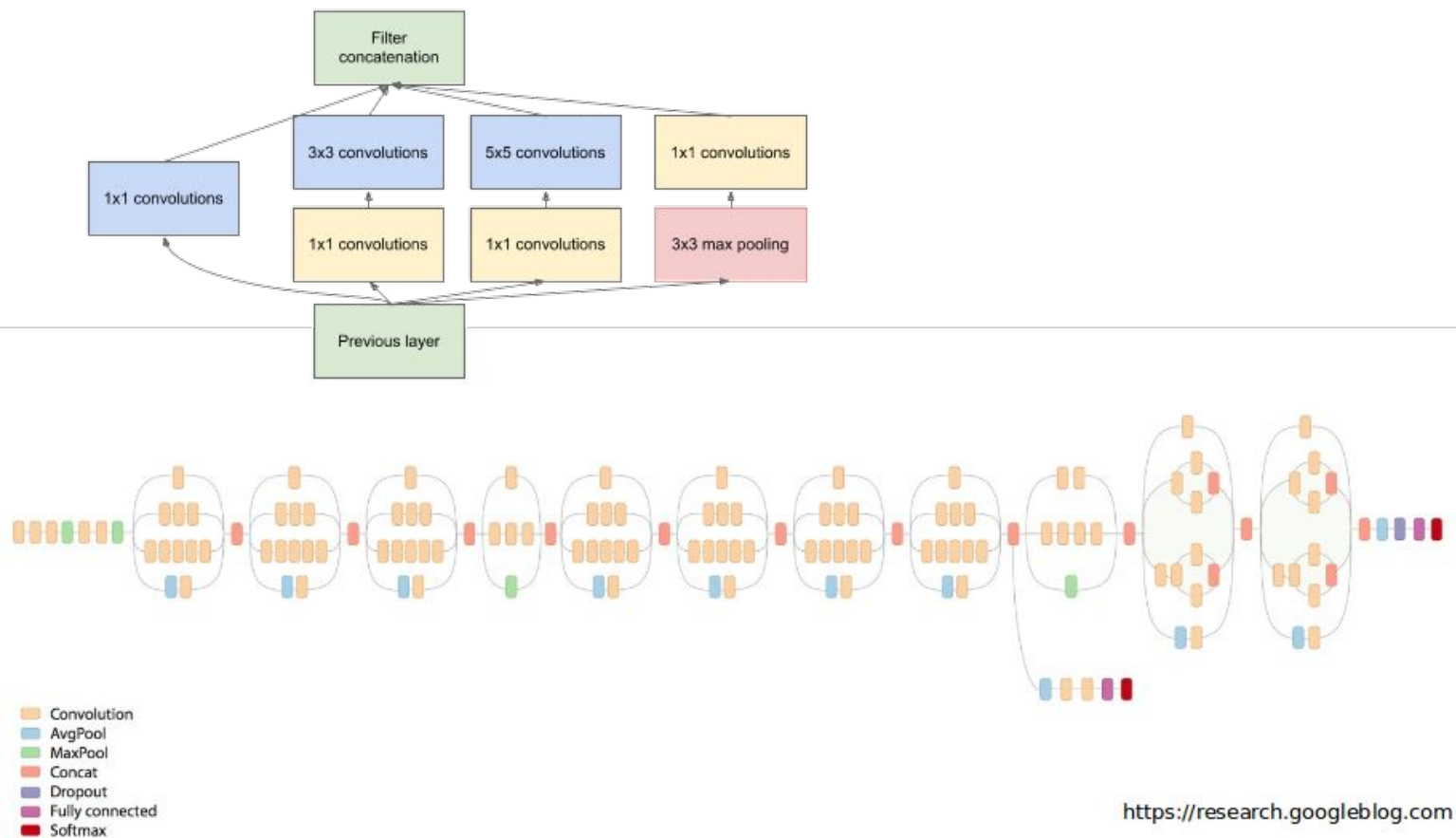
Introduction

- Slides are used to train a ConvNet to predict patient outcome



Inception ConvNet Architecture

- Best performance in the ImageNet
- Inception modules
- Mixture of convolutions of the current layer
- Let the model pick what's best
- Allowing the model to recover both local feature via smaller convolutions and high abstracted features with larger convolutions

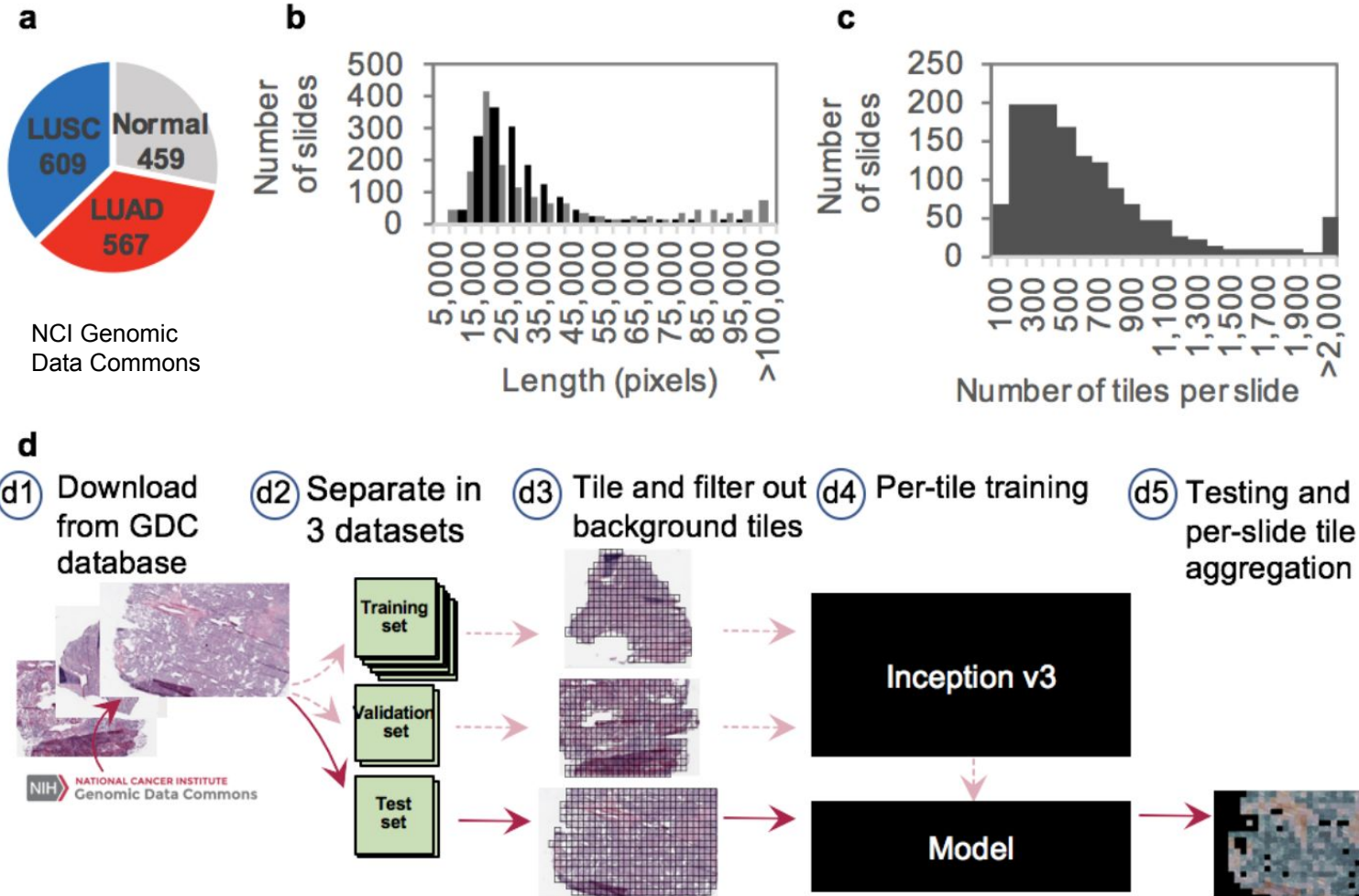


Szegedy, Christian, et al., Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

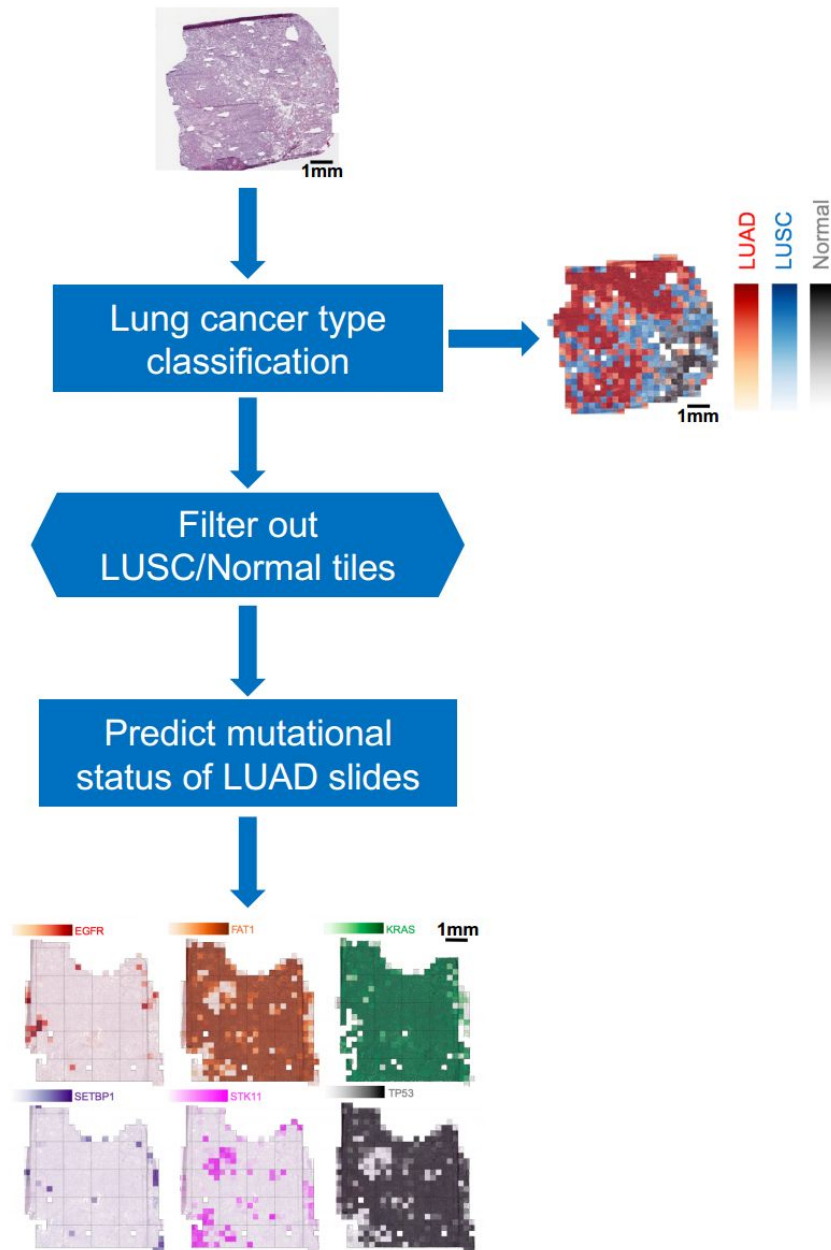
Training & Classification

- Training Strategies:
 - Transfer learning on inception v3
 - Use a pre-trained model architecture and weights to save time
 - Over 14 million images contain over 20 thousand categories
 - Fine-tune the parameters of the last layer of the network on the data via back propagation
 - Training the entire inception v3 network
- Different Classifiers:
 - Normal vs Tumor
 - LUAD vs LUSC
 - Normal vs LUAD vs LUSC
 - Mutations: STK11, EGFR, SETBP1, TP53, FAT1, KRAS, KEAP1, LRP1B, FAT4, NF1
 - Trying to determine the association between histology and genomic heterogeneity

Data and Strategy



Data and Strategy



Data and Strategy

	Training	Validation	Testing
Normal	132,185 / 332	28,403 / 53	28,741 / 74
Primary tumor	556,449 / 825	121,094 / 181	121,059 / 170

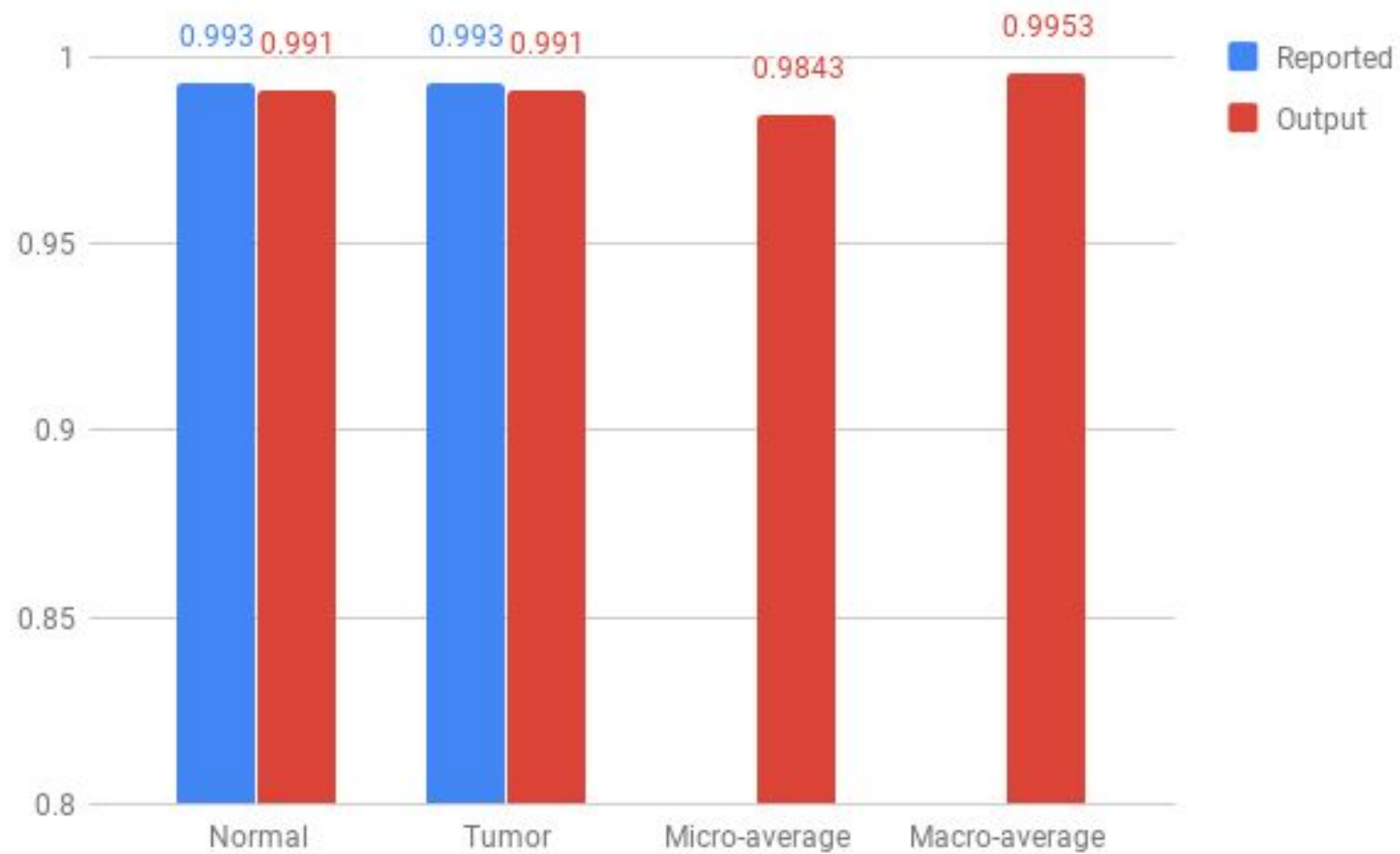
	Training	Validation	Testing
LUAD	255,975 / 403	55,721 / 85	55,210 / 79
LUSC	300,474 / 422	65,373 / 96	65,849 / 91

Gene										
mutated	TP53	LRP1B	KRAS	KEAP1	FAT4	STK11	EGFR	FAT1	NF1	SETBP1
%Patients	50	34	28	18	16	15	12	11	11	11

Pre-processing

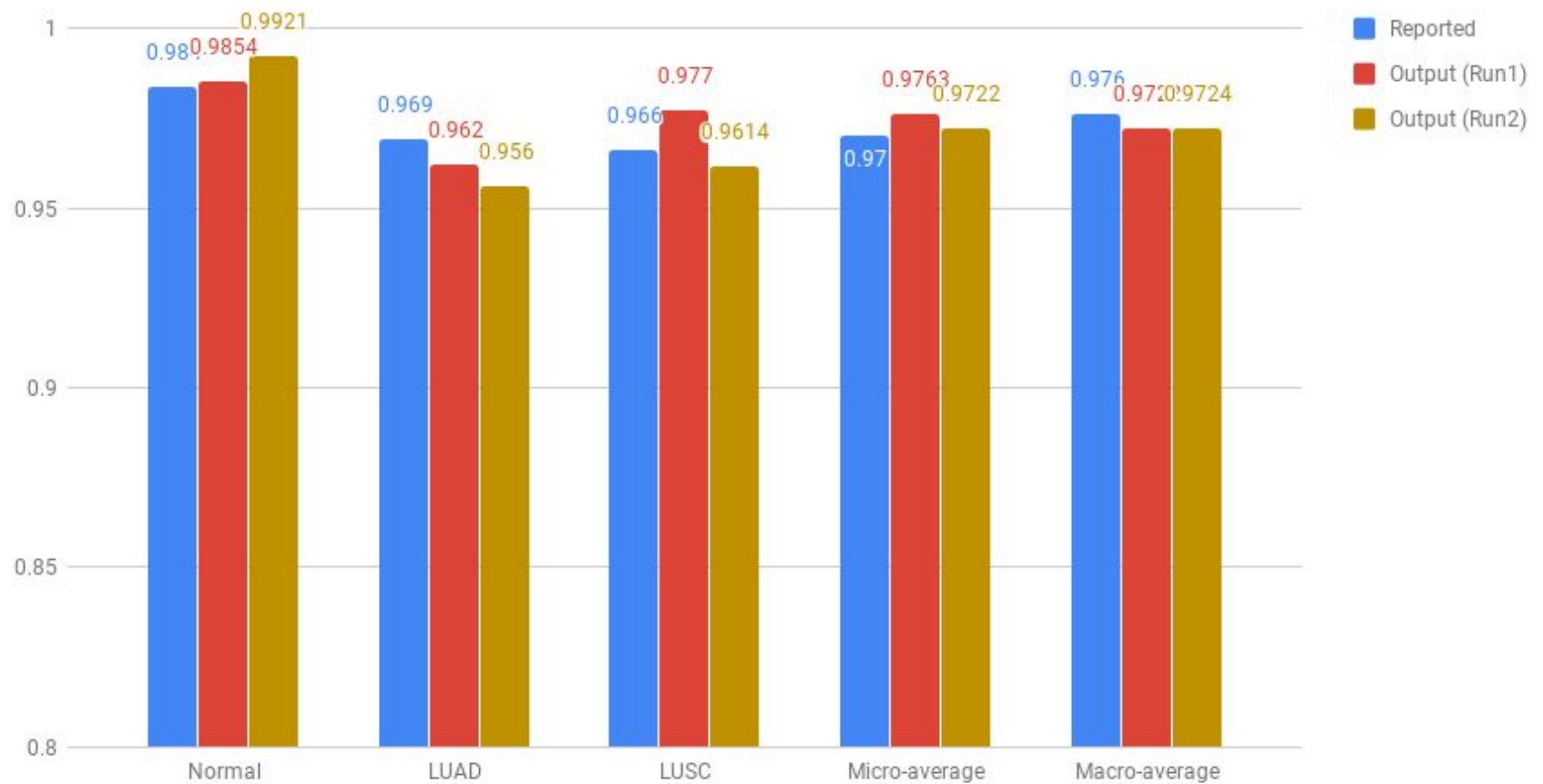


Result



Result

AUC 3 Classes. Normal vs LUAD vs LUSC (Average Probabilities)



Result

AUC Mutations - Average Probabilities



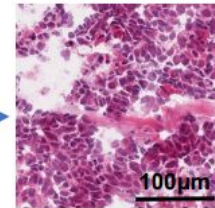
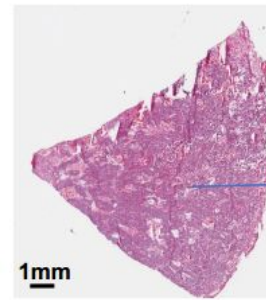
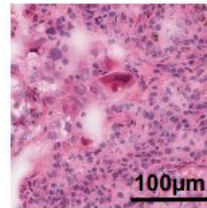
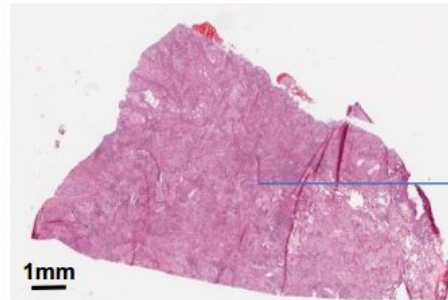
Result

AUC Mutations - Percentage Classified

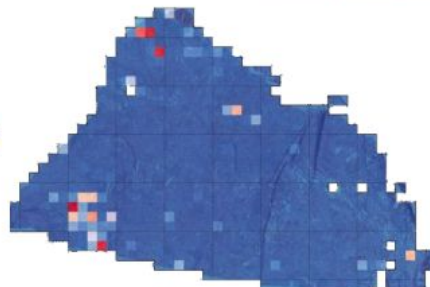


Result

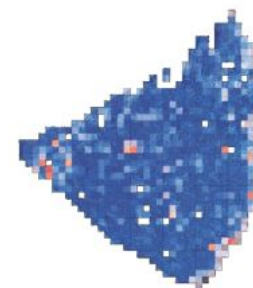
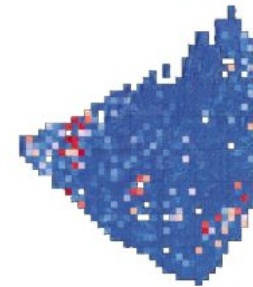
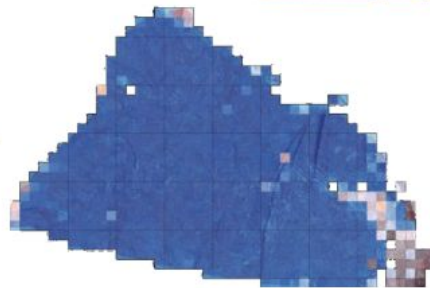
**TCGA
image**



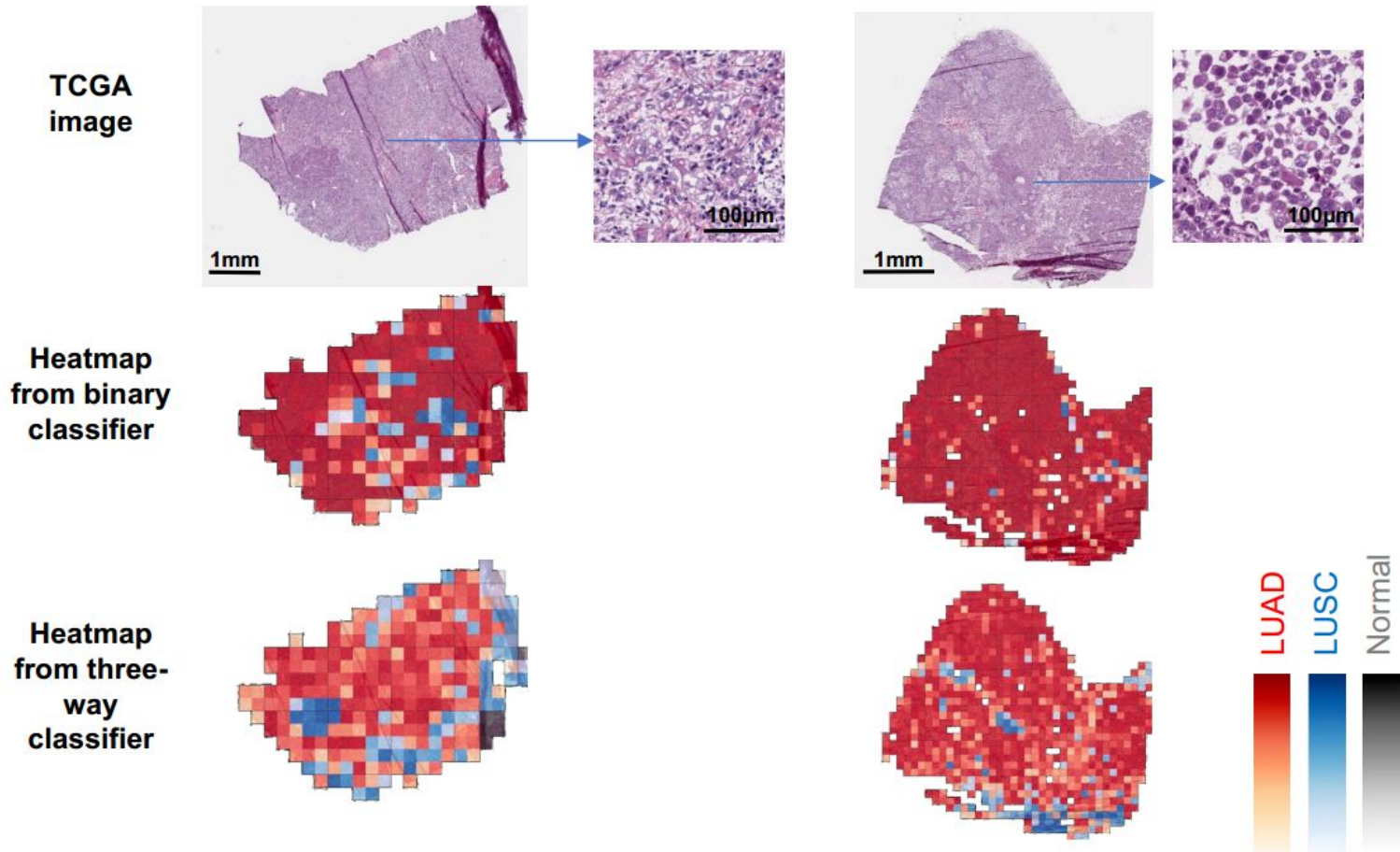
**Heatmap
from binary
classifier**



**Heatmap
from three-
way
classifier**



Result



Result

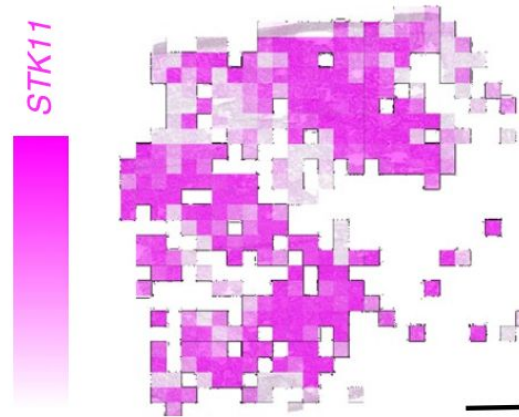
b

LUAD



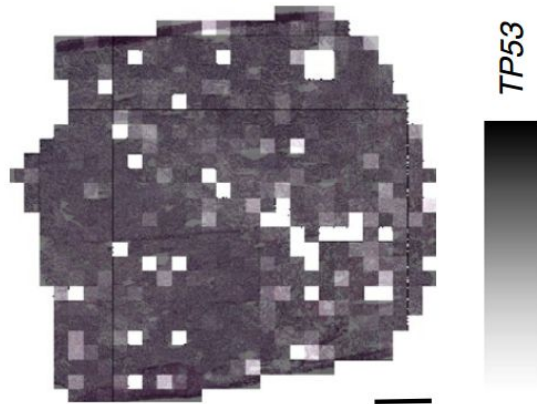
c

LUAD



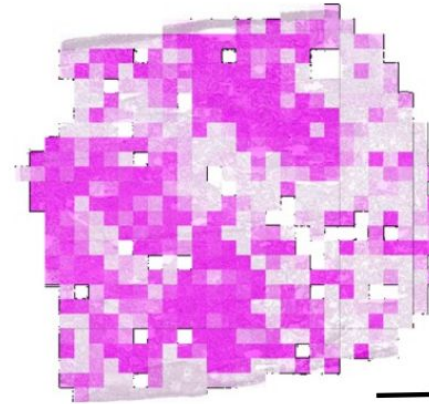
d

All



e

All



Future Works

- Try to improve the mutation-based classifier by addressing the unbalanced data
 - Replacing cross-entropy loss function with a weighted cross-entropy loss function
- Try to use this pipeline on breast cancer tumors to find spatial features of slides
 - Considering SNVs

Thank you