

Radiogenomic Analysis of Glioblastoma with Deep Learning Techniques

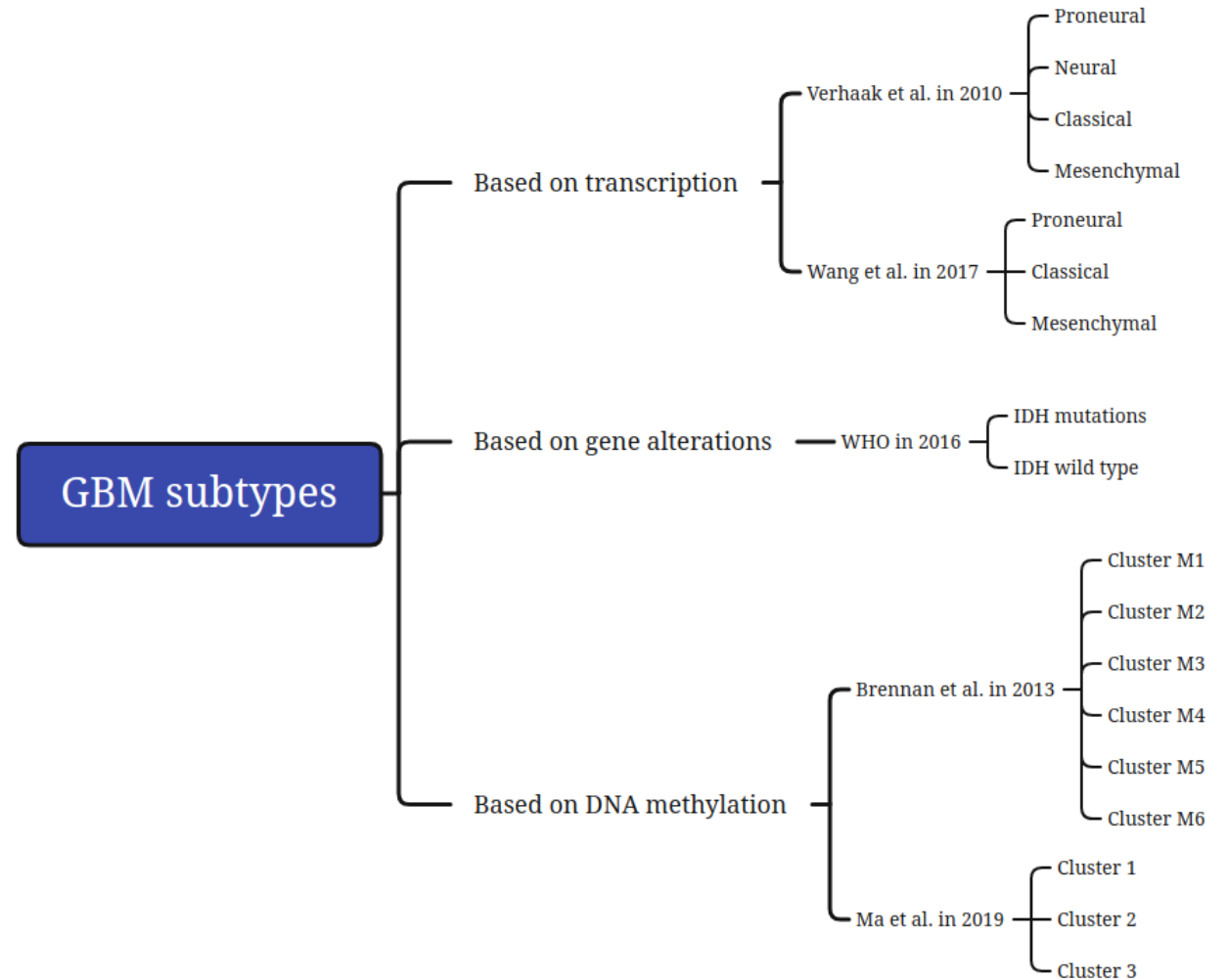
01/17/2021

Yichong Wang

Brief Introduction

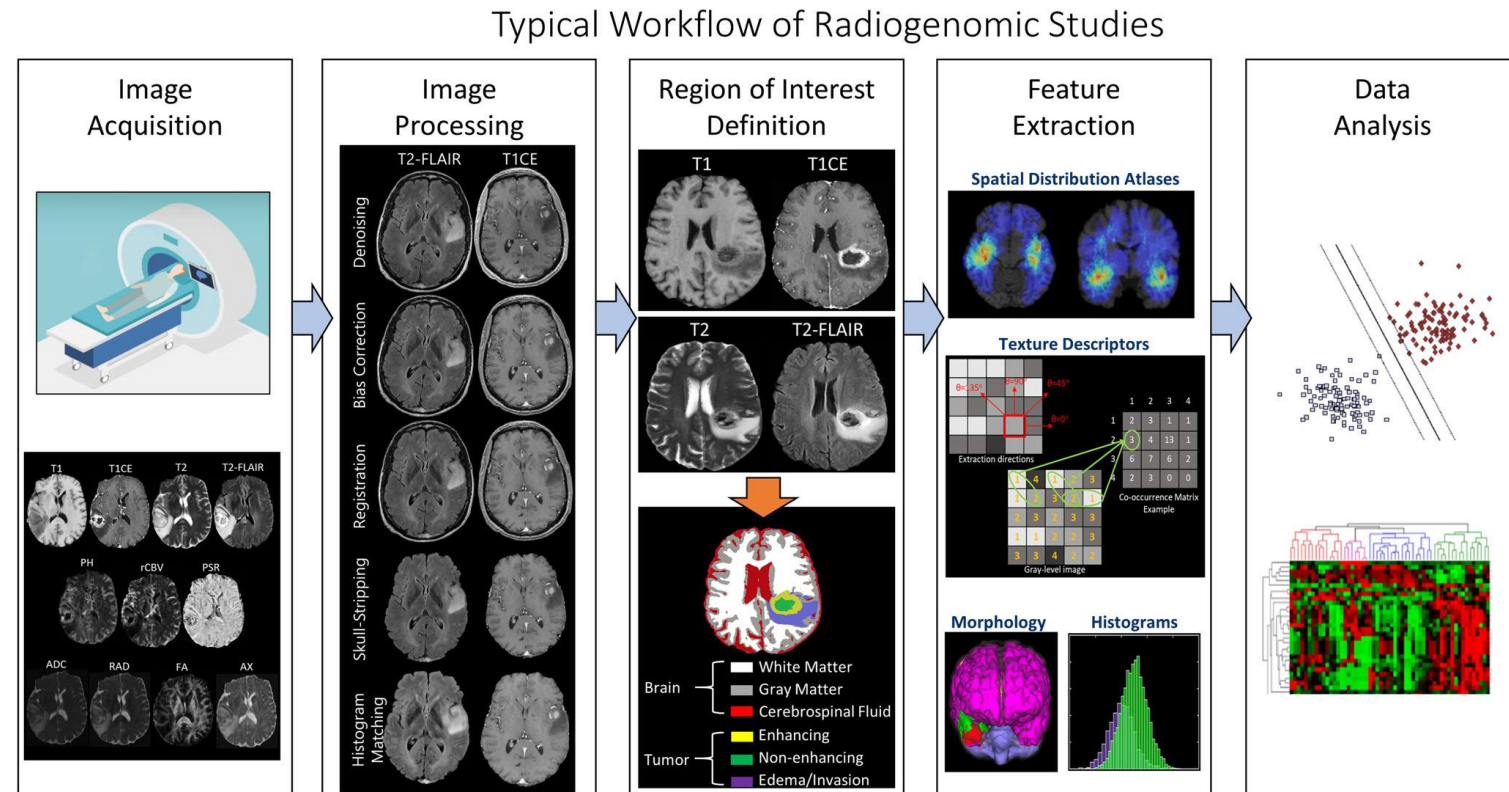
Glioblastoma

- Glioblastoma (GBM) is a highly aggressive adult brain tumor and the deadliest brain tumor that originates in glial cells.
- A high degree of heterogeneity in both space and time
 - Do not respond well to prescribed therapy because therapy targets the entire tumor rather than specific genetic sub-regions.
- Three main criteria to divide GBM subtypes
 - The most common subtype we use in lab is Varhaak et al. research in 2010



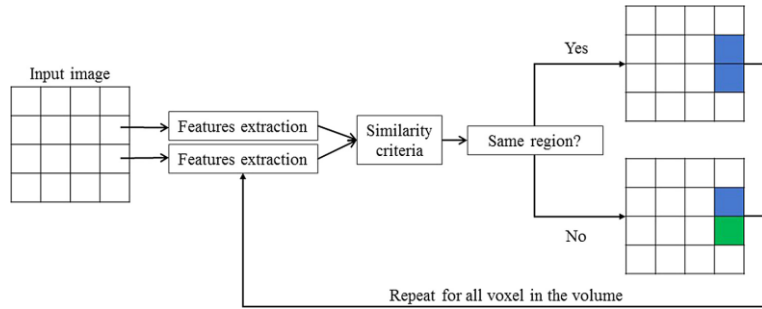
Radiogenomics

1. Image acquisition;
2. Image processing, including noise/artifact reduction, intensity and/or orientation standardization, coregistration of the multiparametric MRI scans
3. **ROI definition** using manual annotation or automatic segmentation
4. **Feature extraction** based on human-engineered (conventional radiomics) or deep-learning approaches
5. **Data analysis**, involving machine/deep-learning methods for feature selection, classification, and cross-validation

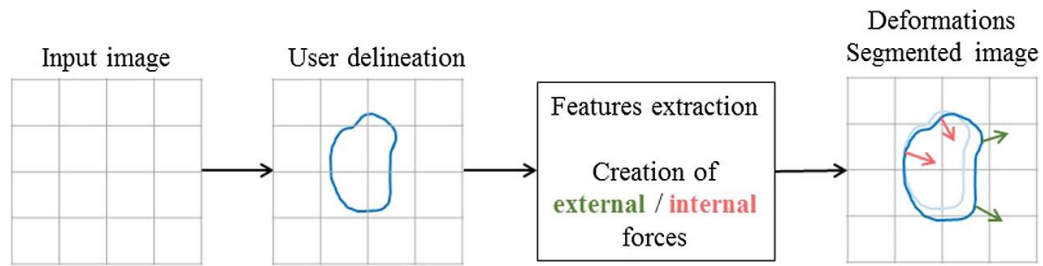


Progress made to-date

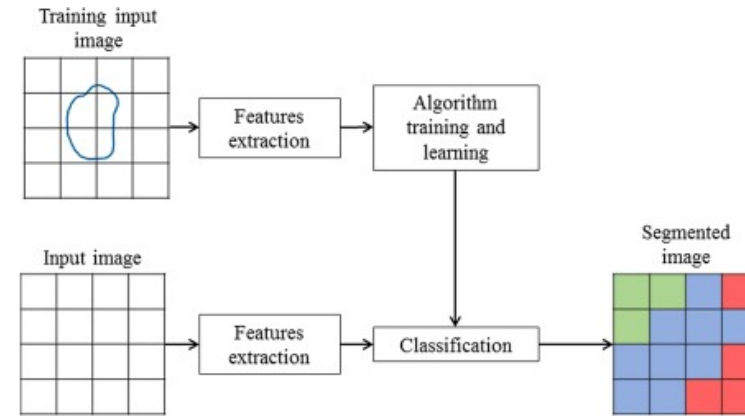
Segmentation Result



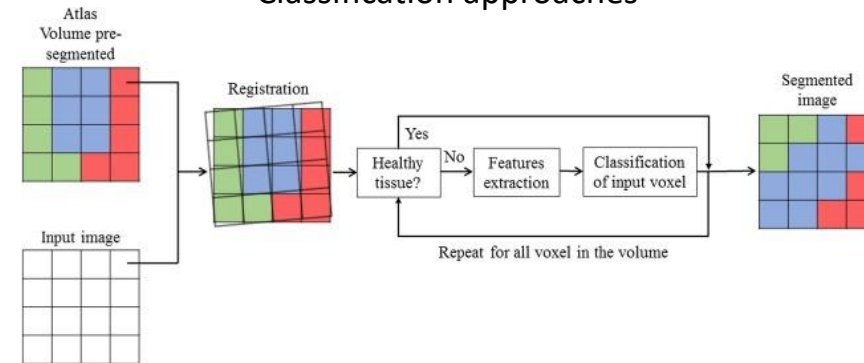
Region-based Approaches



Edge-Based Approaches



Classification approaches

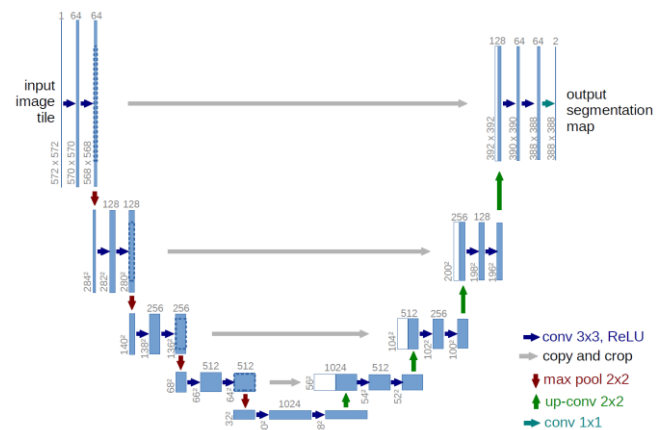


Atlas-based approaches

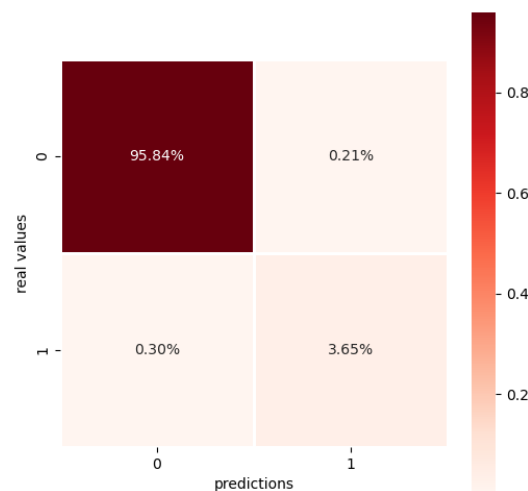
Segmentation Result



Left: MRI Image, Middle: Ground Truth, Right: Semgmentation Results



U-net



Confusion Matrix of all image segmentation results

Dataset: 39 Patients
MRI Series: T1,T2,FLAIR
Plane: **AX**, COR, SAG
Totally 18528 Images
Train : Test = 4 : 1
Train times: 225248

Sensitivity and Specificity:
Actual Values

	Actual Values	
	Positive	Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

Sensitivity>95%

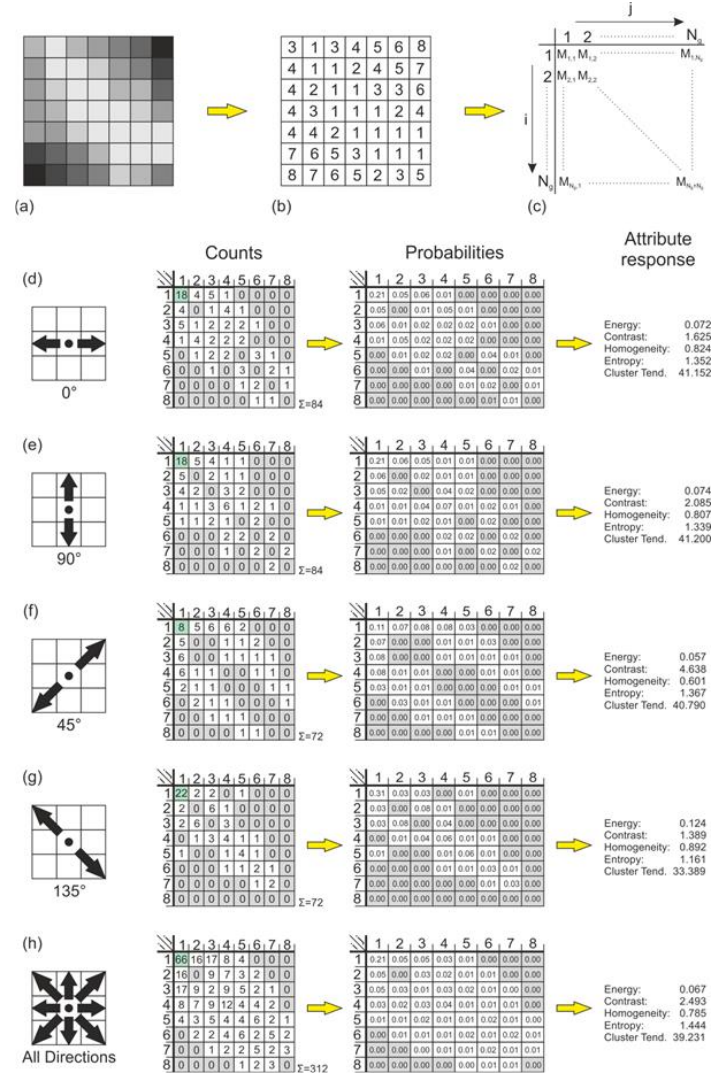
Figure 1

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

Feature Extraction

- In the feature Extraction procedure, we apply **conventional** CV method, using specific equation to calculate features to put image information from 2D to 1D
- At first, I use my own code to calculate features, but the speed is slow. Right now, we use pyradiomics module to extract features from ROI
- **Pyradiomics** have 120 features, most of them are texture features.
- The most classical texture feature design is using **GLCM**



Type of features

- Shaped Based Features

— Area

- Preimeter

- **Circularity**

- Irregularity

- Shape Index

- Mean

- Variance

- Standard Division

- Median

- Skewness

- Kurtosis

- Range

- Pixel Orientation

Contrast

Correlation

Entropy

Energy or Uniformity

- Cluster shade

Sum of square: Variance

Inverse difference Moment

Sum Average: Mean

Sum Variance

Sum Entropy

Difference Variance

Difference Entropy

- Inertia

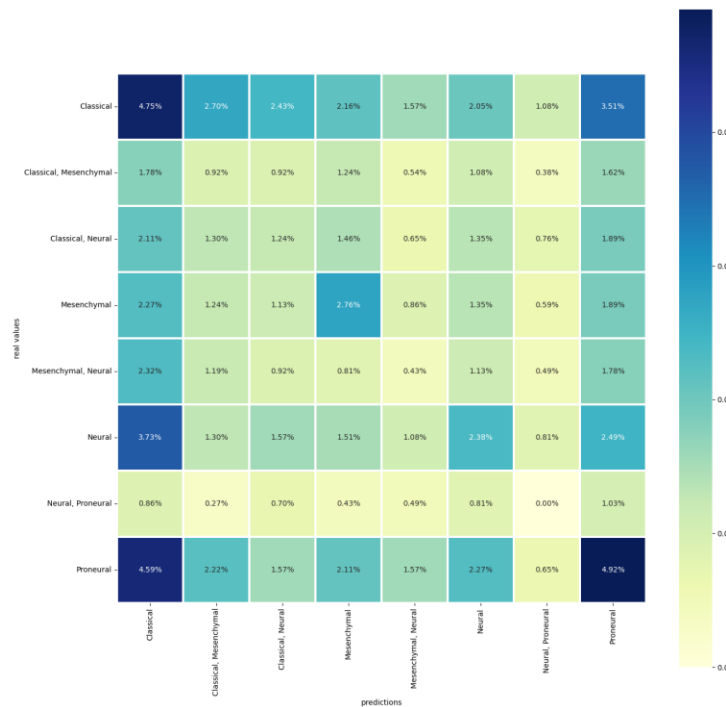
- Cluster prominence

Classify GBM Subtypes

Decision Tree:



Original Data



PCA Data (99% information)

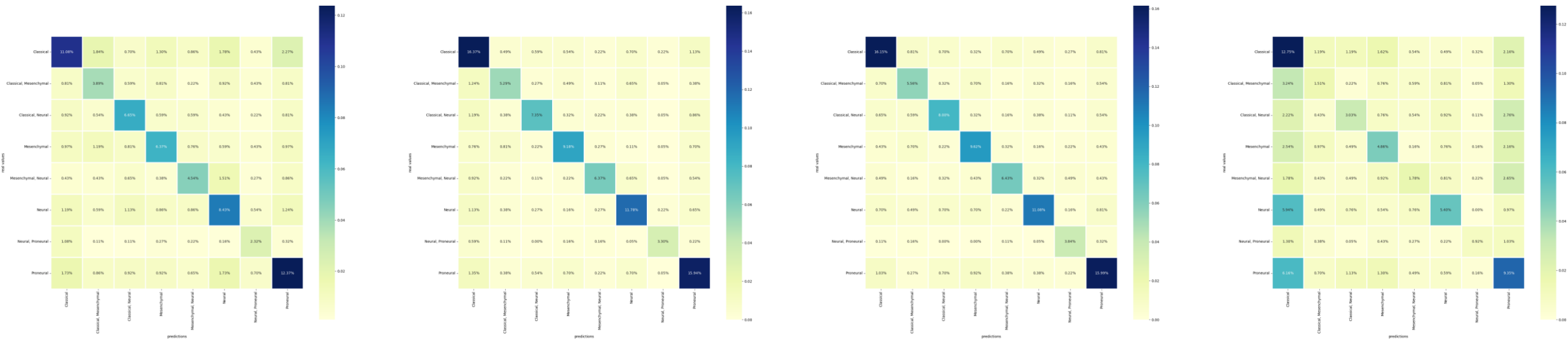


PCA Normalized Data

Need Normalize before PCA

Classify GBM Subtypes

Overview:

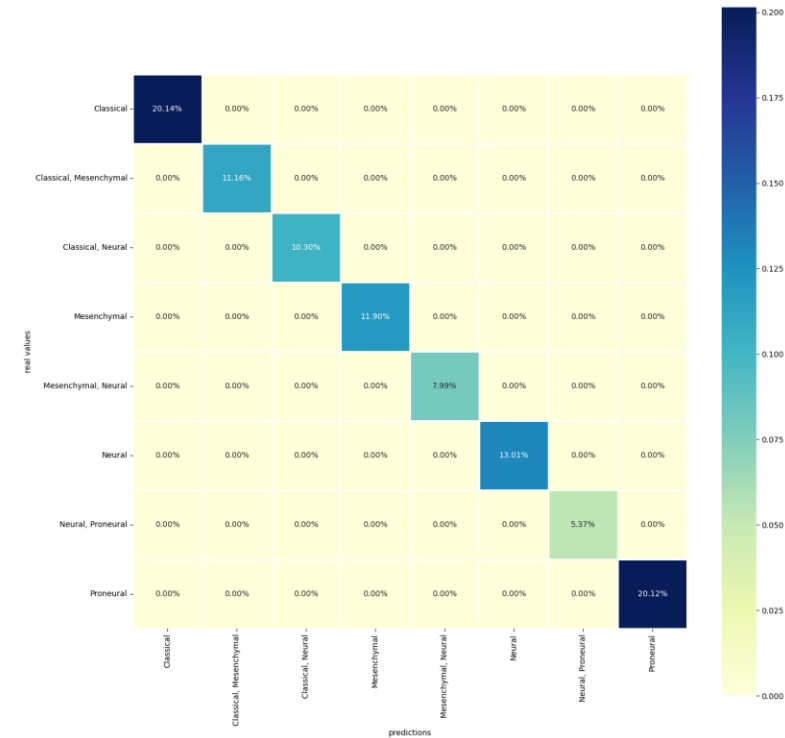
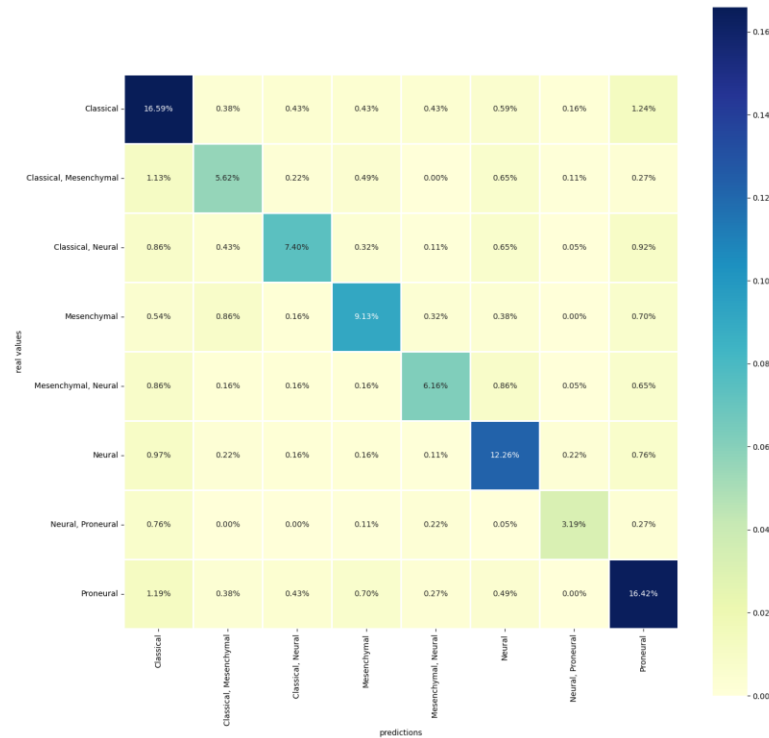


Decision Tree: 56% Random Forest: 77% MLP: 80% Linear SVM: 40%

MLP have the best preference in classifying this dataset
This dataset is not linear relationship

Classify GBM Subtypes Problems

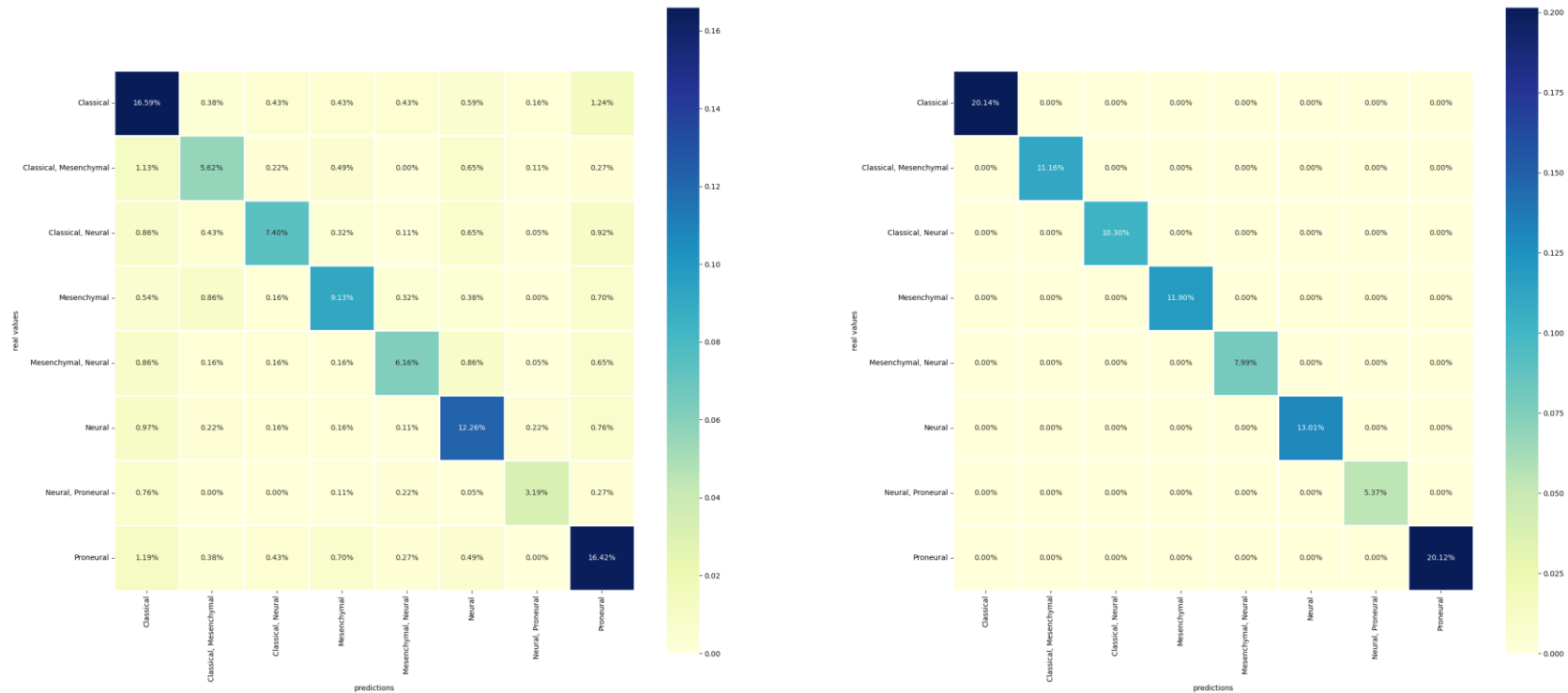
Overfit:



Even add random forest tree numbers into 500, the overfit question still exist (Train accuracy = 100%)
This problem exist in every algorithms I have ever tried

Classify GBM Subtypes Problems

Imbalance:

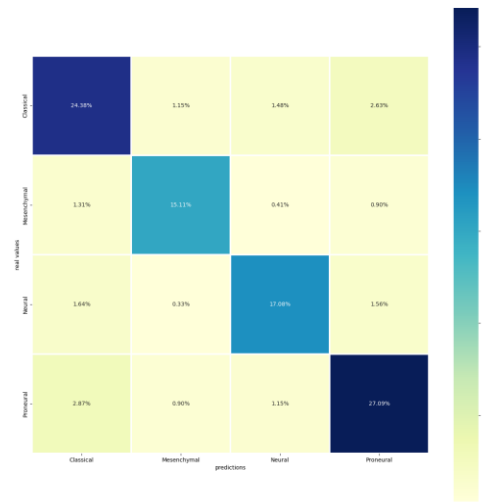


Data Imbalance, the Classical and Proneural is the most subtypes of GBM, leading to data imbalance, which may lead classical and proneural type prediction sensitivity is higher, but others will be lower

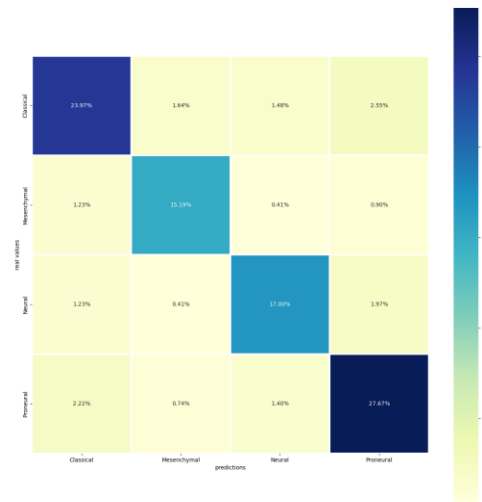
Classify GBM Subtypes Problems



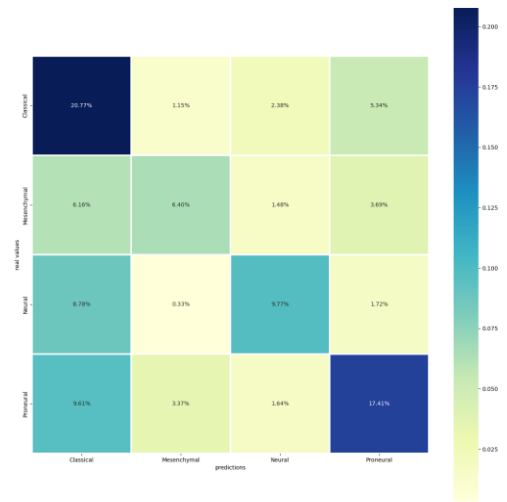
Decision Tree: 68%



Random Forest: 84%



MLP: 88%



Linear SVM: 54%

Because one GBM cancer in one patient have different subtypes in different subregions, using two subtypes combined into a tuple as a label is very not scientific

Future Plan

Improvements and Other direction

- Using Transformer Network replace U-net (CNN) and Pyradiomics to do feature extraction to do classification
- Using gene data to map radiomic features with gene data.
- Predict key gene mutation or not.
- Learn Qt to make UI

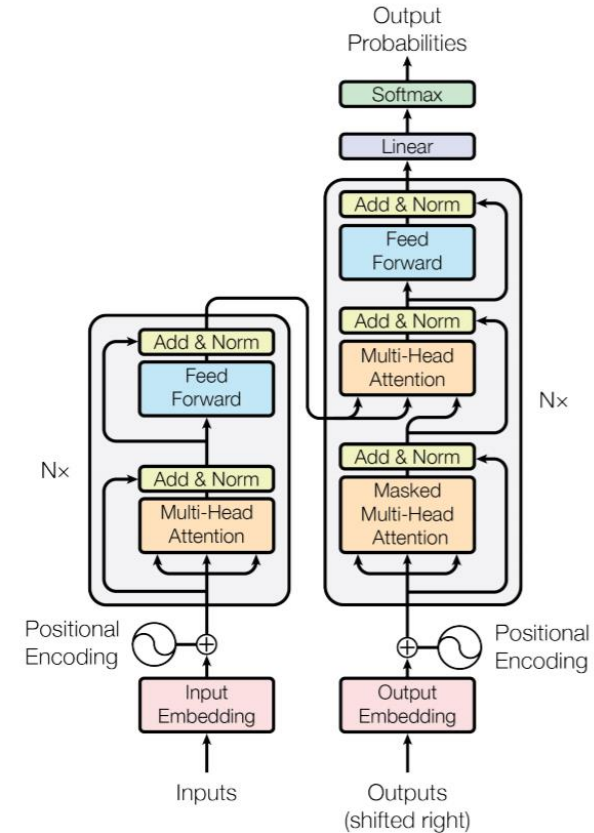
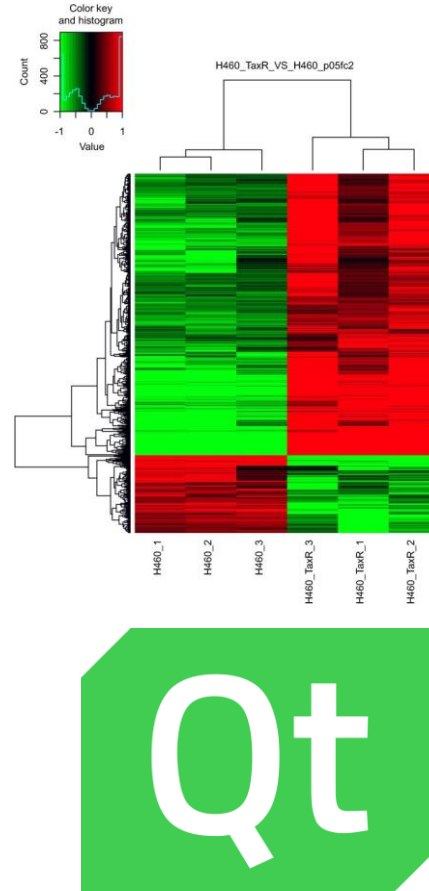


Figure 1: The Transformer - model architecture.