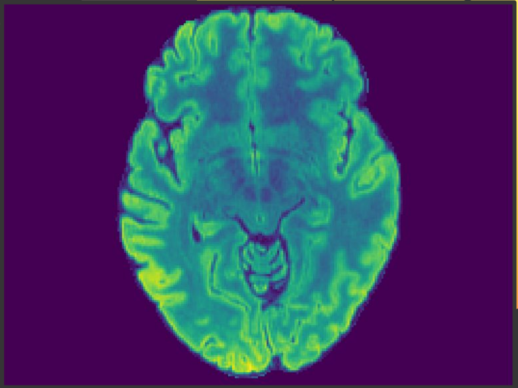
Data Science Bootcamp
Fall 2025

Tumor or no?

GBM or no?
IDH mutation?



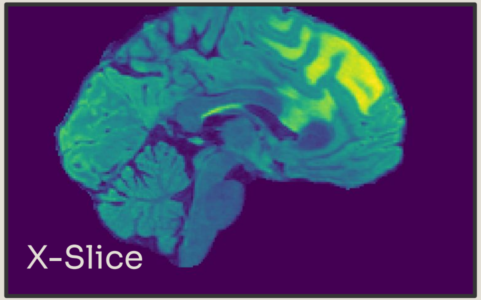
Brain Tumor Diagnosis

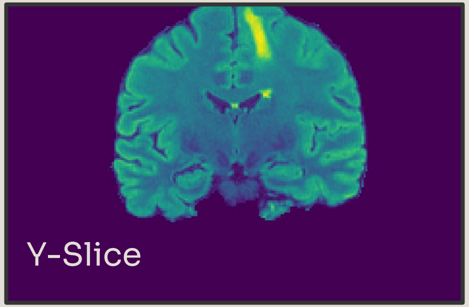
Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

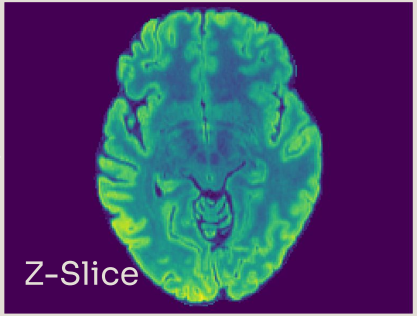
Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

Introduction: Gliomas

- Most common primary brain tumors in adults
- Early and accurate identification is vital for surgery, prognosis and treatment
- Traditional grading usually requires biopsy and molecular testing
- MRI offers rich structural and physiological information, but interpretation remains challenging and operator-dependent

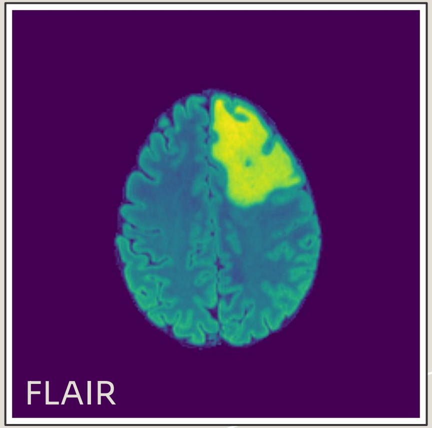


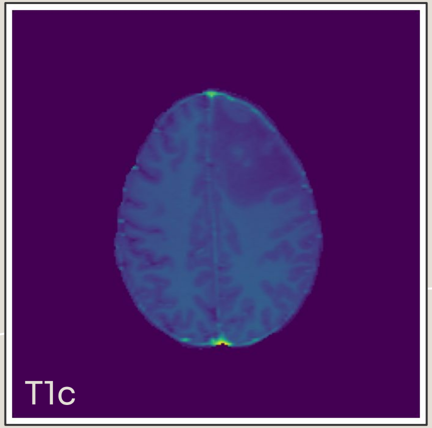


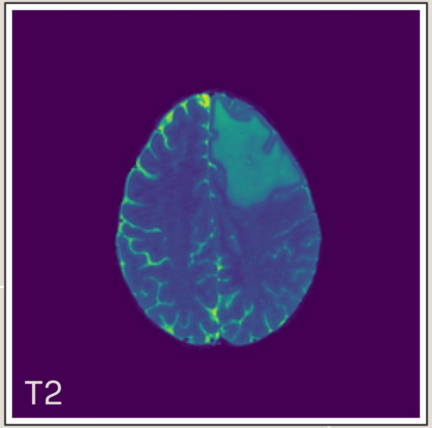


Cancer Imaging Archives Image Dataset

- ~ 500 patients
- All patients have various tumor types of severity/grades 2 - 4
- Each patient has multiple image sequences, such as T1c, T2, and FLAIR.
- Tumor segmentation image used as a mask for both tumor type and labels for binary classification neural networks.
- The dataset also included in depth metadata regarding patient information such as sex, age, and tumor severity.



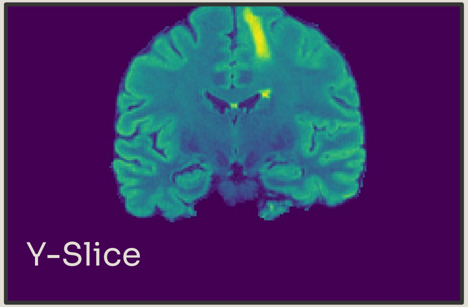


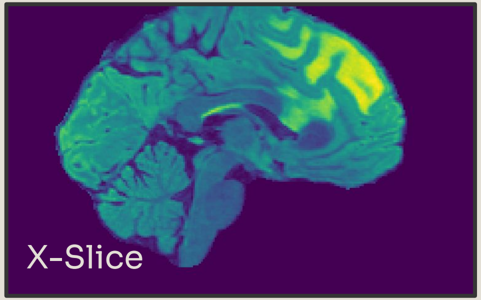


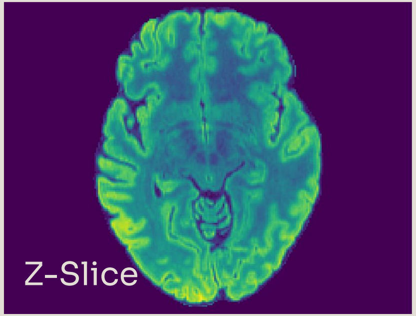
Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

RAS Orientation

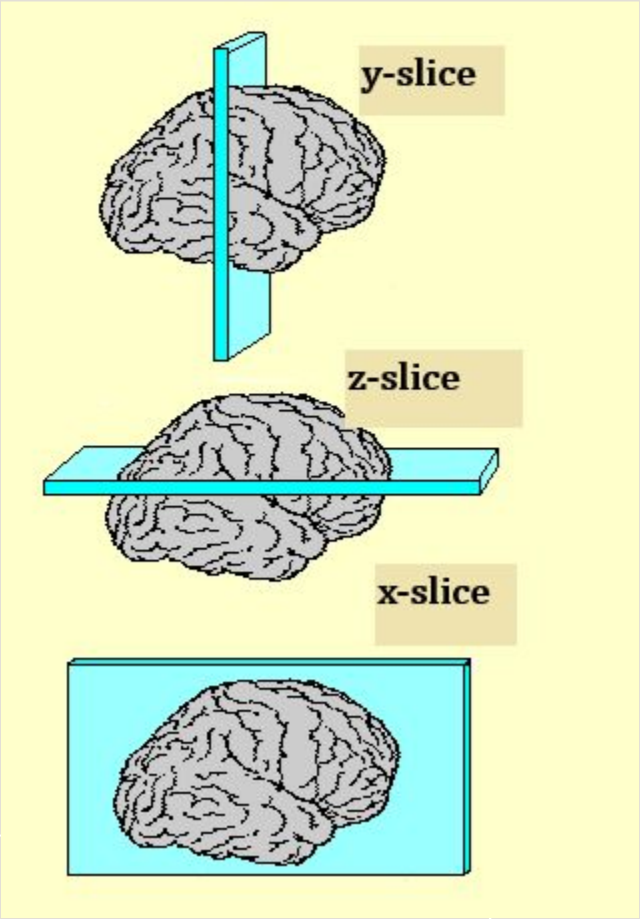
- All MRI images are 3D and we adopt the RAS orientation in this project.







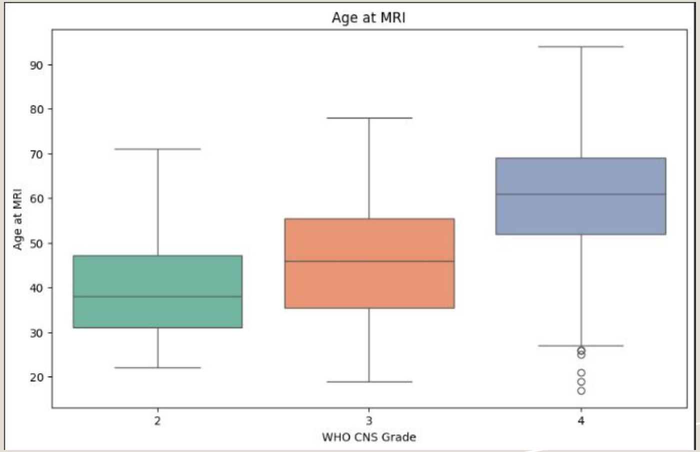
*sample mri slices read from the dataset



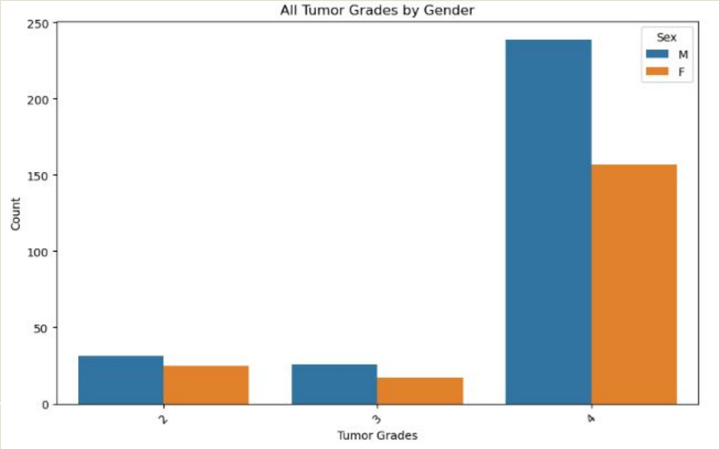
Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

Cancer Imaging Archives Metadata

- Metadata features: Sex, Age at MRI, tumor grade, IDH mutation, 1p/19q codeletion status.
- Among the 495 unique patients, 80% are in grade 4 and roughly 10% each in grade 2 and grade 3. To make sure we have a more balanced dataset*, we apply data augmentation on mri slices in grade 2/3.



Age distribution per grade



Gender in each grade

Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

Overview of the Problems and Our Approaches

- Diagnosis of presence of tumors can be time consuming + human variability
- Constructing a labeled dataset for tumor detection requires manual segmentation, which limits scalability.

➡ **3D Tumor vs No Tumor model**: predicting presence of tumors + accurate 3D masks of their extent within the brain.

➡ **2D Tumor vs No Tumor model**: gives probability that a mri image contains a tumor-like mass; useful when 3D data not available.

Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

Overview of the Problems and Our Approaches

- Early detection of Glioblastomas (most aggressive type) is better for treatment planning, clinical trial eligibility, and patient counseling.

→ **Glioblastoma vs Other Gliomas Classifier**

- IDH mutation is very important in predicting tumor behavior but molecular testing can be invasive and costly.

→ **IDH Mutation Classifier**

- 2D models using z-slices with maximum tumor size
- Metadata features included (e.g. 1p/19q codeletion)
- Could be useful in making tumor grade prediction

Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

Three-Component Pipeline

MRI Data Preprocessing

Objective:
Prepare MRI volumes for downstream deep-learning analysis.

Purpose:
Standardizes raw MRI data to ensure reliable learning and prevents domain variability from affecting inference.

Tumor Detection

Objective: Detect the presence of a tumor on MRI scans.

Relevance: General screening & segmentation

Methods: Two independent CNN-based classifiers (2D & 3D CNN)

Glioma Grading

Objective: Distinguish Grade 4 vs. Grade 2–3 gliomas.

Relevance: Precision diagnosis among glioma patients

Methods: Two independent CNN-based classifiers

- Glioblastoma vs. Other gliomas
- IDH mutant vs. IDH-wildtype

Binary Classification - 2 Models

Goal: Produce a 3D probability map and mask showing where tumor is present in the brain.

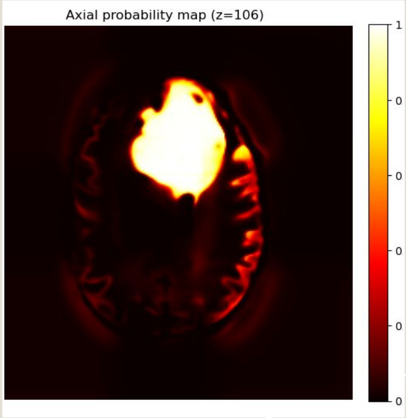
Input: Four preprocessed MRI volumes (FLAIR, T1, T1c, T2).

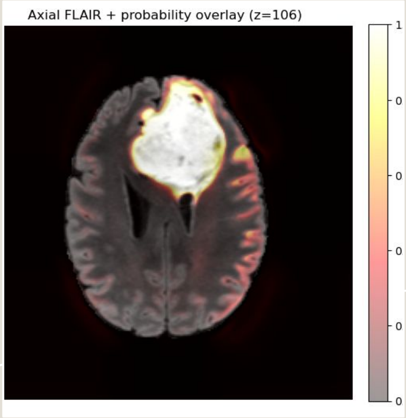
Output: One 3D map where each voxel has a value in [0,1], giving the probability that that location is tumorous.

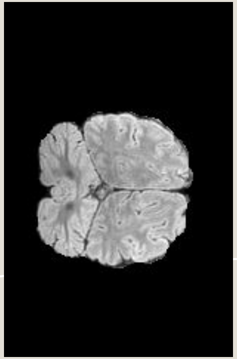
2D Classifier:

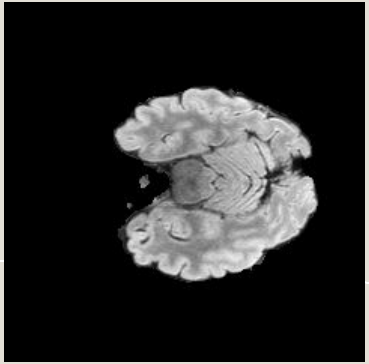
Purpose: Score 2D slices of Brain MRIs for tumor/no-tumor classes.
Can be applied to historic images where the 3D data is not available.

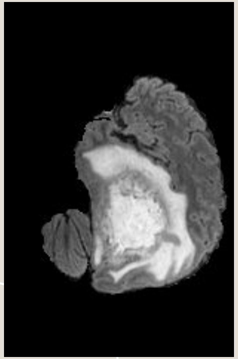
2D ResNet18:
Input: 8-bit Grayscale images (FLAIR)
Output: Classification score [0,1].











Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

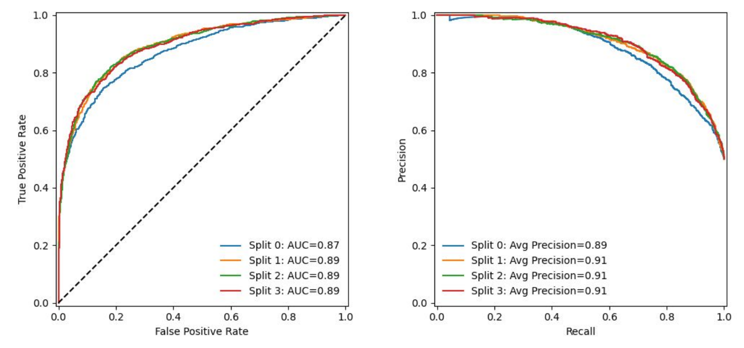
2D Fine-tuned Residual Network - Tumor vs No Tumor

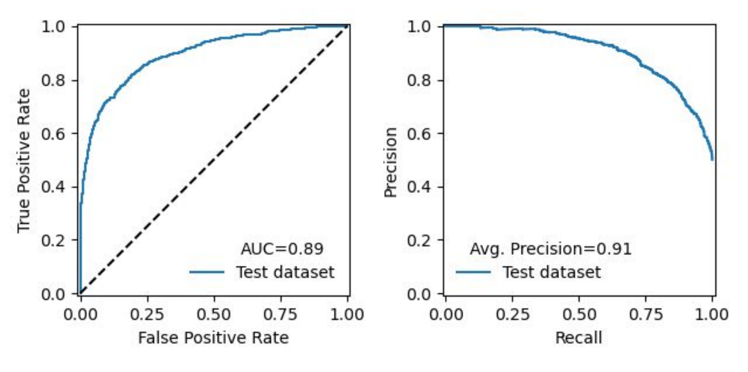
Uses pre-trained **resnet18** from torchvision.models

- Modify dense layer to a single binary output with dropout regularization.
- Freeze residual block parameters during training.
- 4-fold cross validation on training set (80%).
    - AUC ~0. 89
    - Avg. Precision ~0.90

Fold	0	1	2	3
Validation Accuracy	79%	76%	**82%**	81%

Performance on validation sets across folds





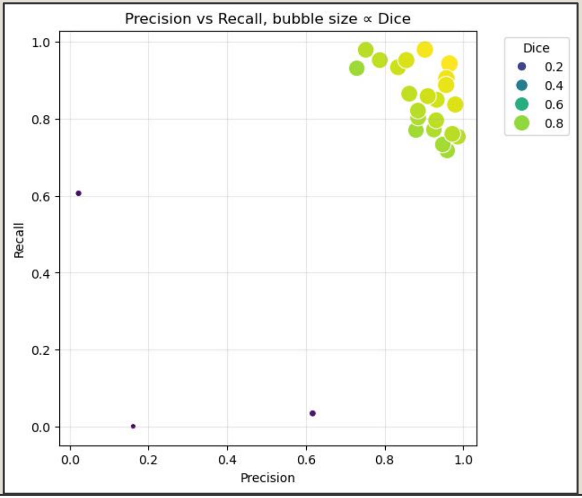
Performance of model from split 2 on the test set

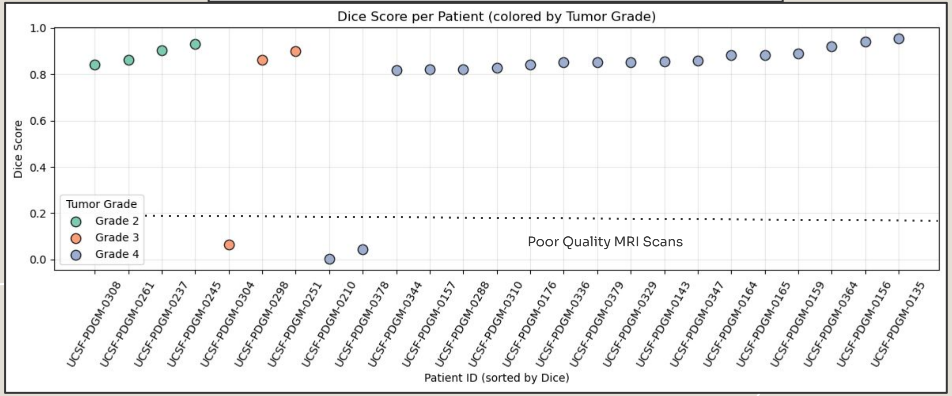
Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

3D CNN - Tumor vs No Tumor

Metric	Performance
Mean Dice	~0.83
Precision	~86%
Recall	~82%
Patients w/ Dice > 80%	20 / 25
Low-Quality Scans	3/25
Overall Score	Stable across grades 2–4

$$Dice = \frac{2TP}{FP + 2TP + FN}$$





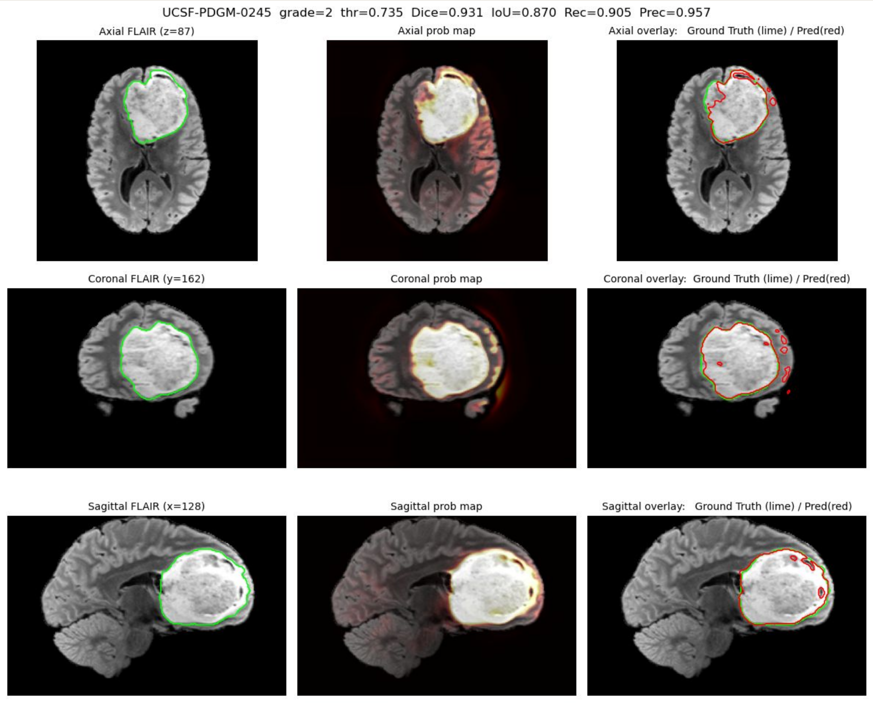
Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

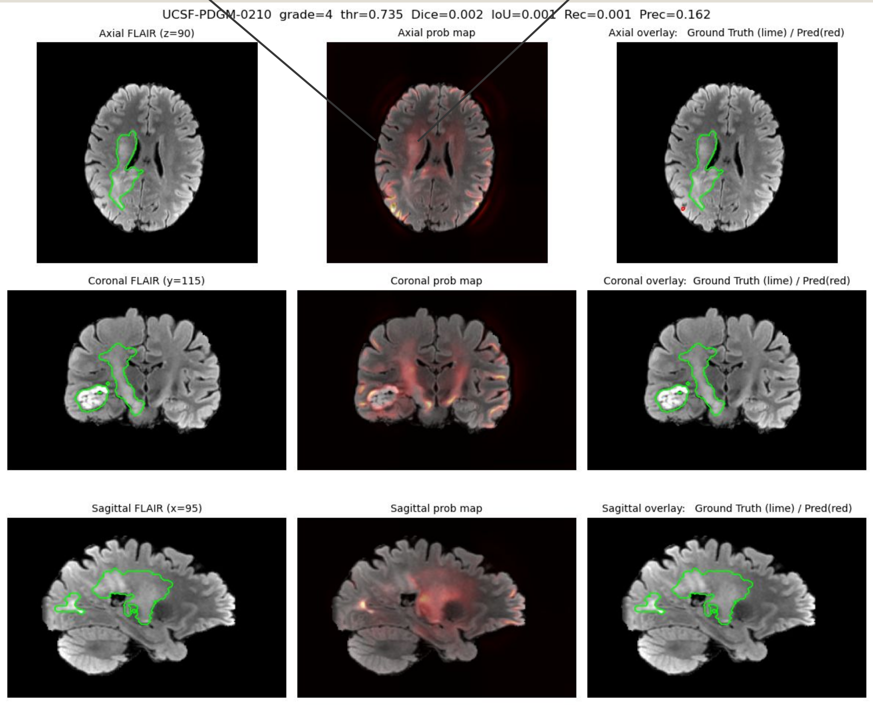
Tumor Hardly Visible in FLAIR

Good Quality Image

Non-uniform intensity

Poor Quality Image





Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

Glioma Grading– **Glioblastoma vs others, IDH mutation identifier**

Both models take in 2d z-slices with maximum tumor area and output probabilities of whether it's Glioblastoma (resp IDH mutated). The first model incorporates metadata features such as age, gender, IDH mutation and 1p/19 codeletion status. The IDH model includes only age and gender.

Glioblastoma vs. Other Gliomas (GBM vs OG)

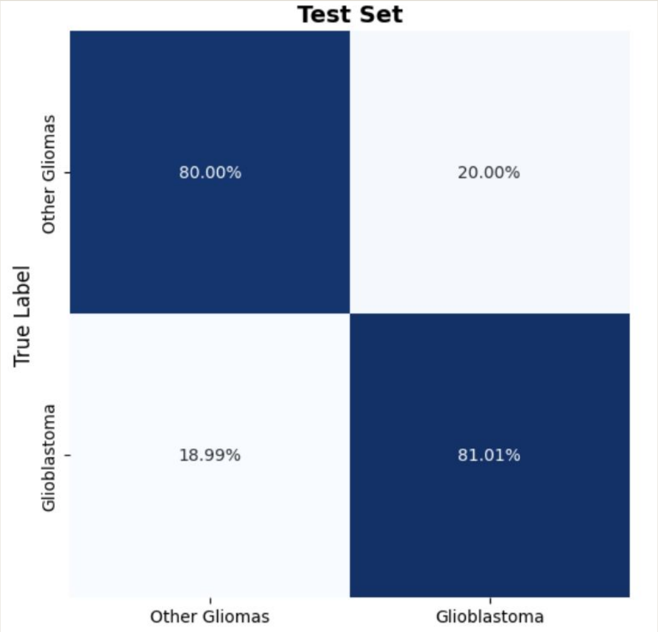
Metric	Performance
Precision	~86%
Recall	~81%
Validation Accuracy	~86%
Test Balanced Accuracy	~81%
Test AUROC*	0.79
F1 Score	~87%

IDH mutation classifier

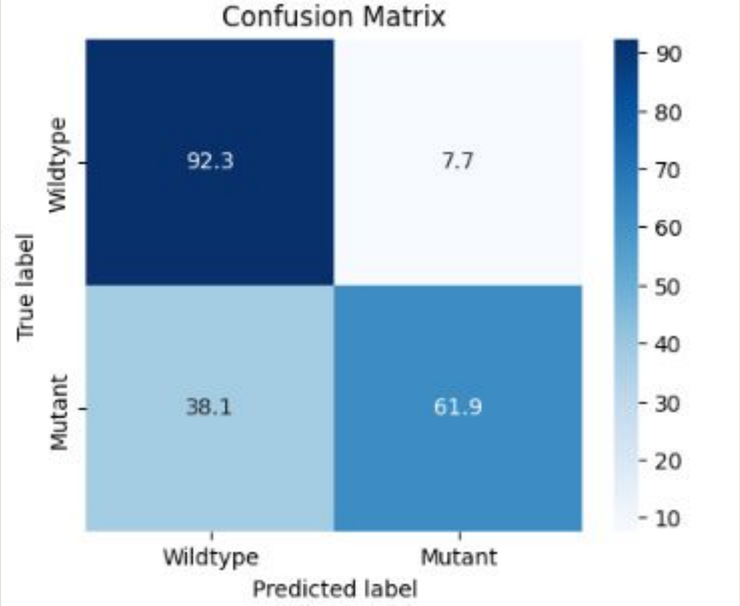
Metric	Performance
Precision	~79%
Recall	~85%
Validation Accuracy	~88%
Test Balanced Accuracy	~77%
Test AUROC*	0.81
F1 Score	~78%

Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

Main Results- **Glioblastoma vs others, IDH mutation identifier**



Confusion Matrix for **GBM vs OG**



Confusion Matrix for **IDH model**

Chris Ewasiuk, Jacob Johnson, Wenwen Li, Suo-Jun Tan

Future Directions

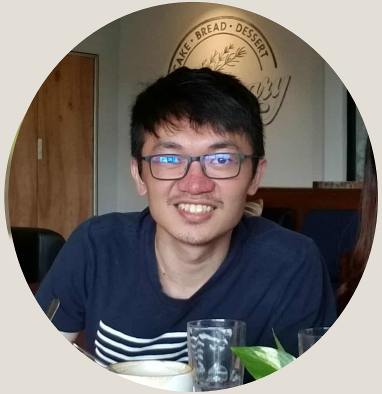
- Improve featured-based classifiers by gathering more quality data (especially for grade 2/3 tumors). Including other molecular markers could be helpful in distinction between grade 2 and 3.
- Incorporate the different models into a single model capable of performing both tumor vs no-tumor classification and tumor type identification on images determined to have a tumor.
- Expand from binary to full WHO grade classification (II / III / IV):
    - Incorporate additional MRI modalities
    - Explore 3D CNNs + volumetric tumor masks.

Meet the team

Chris Ewasiuk
UC Santa Cruz

Jacob Johnson
UC Santa Cruz

Wenwen Li
Florida State University



Suo-Jun Tan
University of Utah