

PREDICTING GLIOMA GRADING

- IDENTIFYING GLIOMA GRADING THROUGH BEST COMBINATIONS OF GENES, MUTATIONS, AND FEATURES

Nandini Muresh
Slides: 1, 5, 11, 12,
13

Anveetha Suresh
Slides: 7, 8, 9, 10,
11

Aashna Kothari
Slides: 1, 2, 3, 4, 6

Francesca Palladino
Slides: 1, 2, 3, 4, 5

CGS3342.001

AREA OF RESEARCH

GLIOMA GRADING

LGG

Lower-Grade Gliomas

- Tumors on glial cells
- Mostly malignant/cancerous
- Slow growing
- Grades I - III
- Life expectancy: 5 to 13 years
- Affect younger people

GBM

Glioblastoma Multiforme

- Tumors on glial cells
- Always malignant/cancerous
- Fast Growing
- Grade IV
- Life expectancy: 14 to 16 months
- Affect older people

Both

- Mutations may occur randomly through lifetime or could be genetic
- Exact causes are unknown
- Because of this methods to detect predictors can be extremely useful

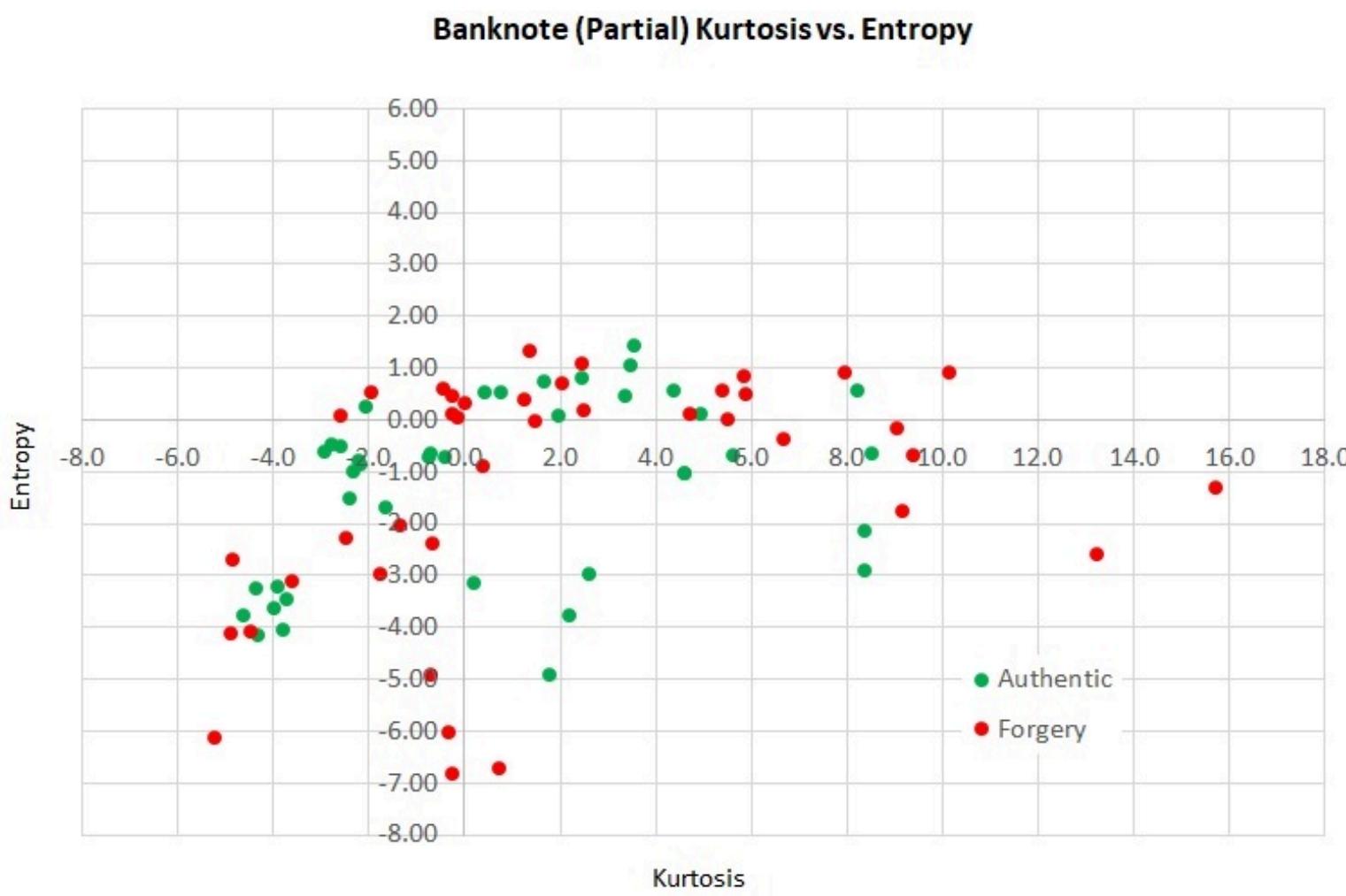
IMPORTANCE

- Funded by The Cancer Genome Atlas (TCGA) Project, supported by the **National Cancer Institute (NCI)**
- Of interest to **oncologists**, doctors who specialize in cancer treatment, and other medical professionals
 - Increased accuracy of glioma grading
 - Creation of new glioma therapies
- Of interest to **neuroscientists**
 - Glioma cells interaction with neurons
 - Glioma cells began to “hijack” activity in the brain to drive tumor growth
 - What factors influence whether a person has LGG or GBM



LITERATURE REVIEW #1

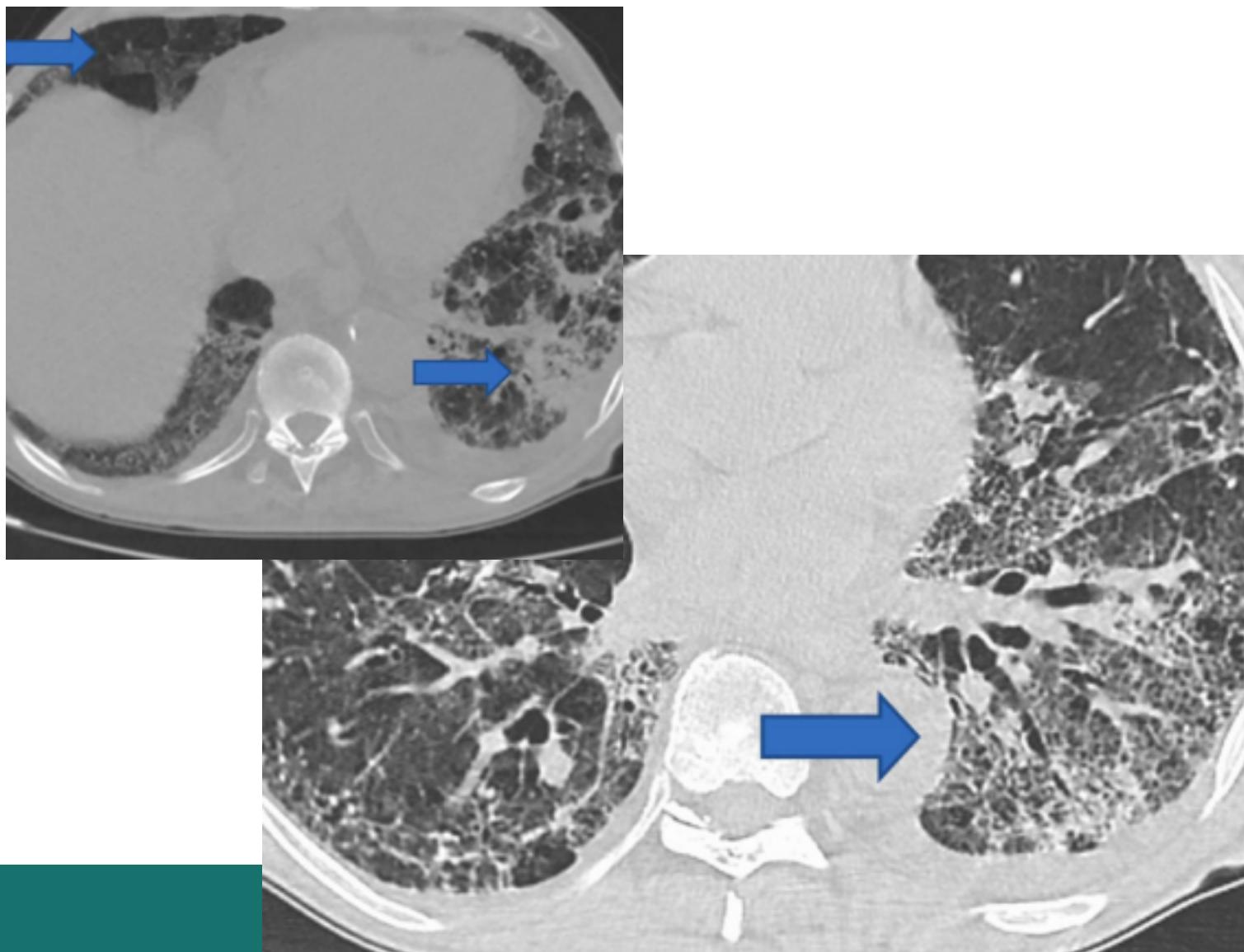
JAMES MCCAFFREY (2018): "HOW TO DO NEURAL BINARY CLASSIFICATION USING KERAS"



- Scanned images to see if they are authentic (0) or forged (1)
- 4 input nodes, 2 hidden layers with 8 neurons each, 1 output node with a sigmoid activation
- 99.27% accuracy
 - 500 epochs for training fixed all code
 - 20 epochs for initial training
 - 60 epochs and 80 to train further
 - 90 epochs to track accuracy, precision, and recall

McCaffrey, James. "How to Do Neural Binary Classification Using Keras." *Visual Studio Magazine*, 30 Aug. 2018, visualstudiomagazine.com/articles/2018/08/30/neural-binary-classification-keras.aspx.

LITERATURE REVIEW #2



BARSTUGAN, OZKAYA, OZTURK (2020): “CORONAVIRUS (COVID-19) CLASSIFICATION USING CT IMAGES BY MACHINE LEARNING METHODS”

- Study focuses on early detection of COVID-19
- Used 4 different datasets from 150 CT (Computed Tomography) images
- Along with 2 fold cross validation, used Grey Level Co-occurrence Matrix, Local Directional Pattern, & Grey Level Run Length Matrix
- Best classification accuracy was 99.68% using 10-fold cross validation

Barstugan, Mucahid, et al. “Coronavirus (COVID-19) Classification Using CT Images by Machine Learning Methods.”

arXiv.Org, 20 Mar. 2020, arxiv.org/abs/2003.09424v1.

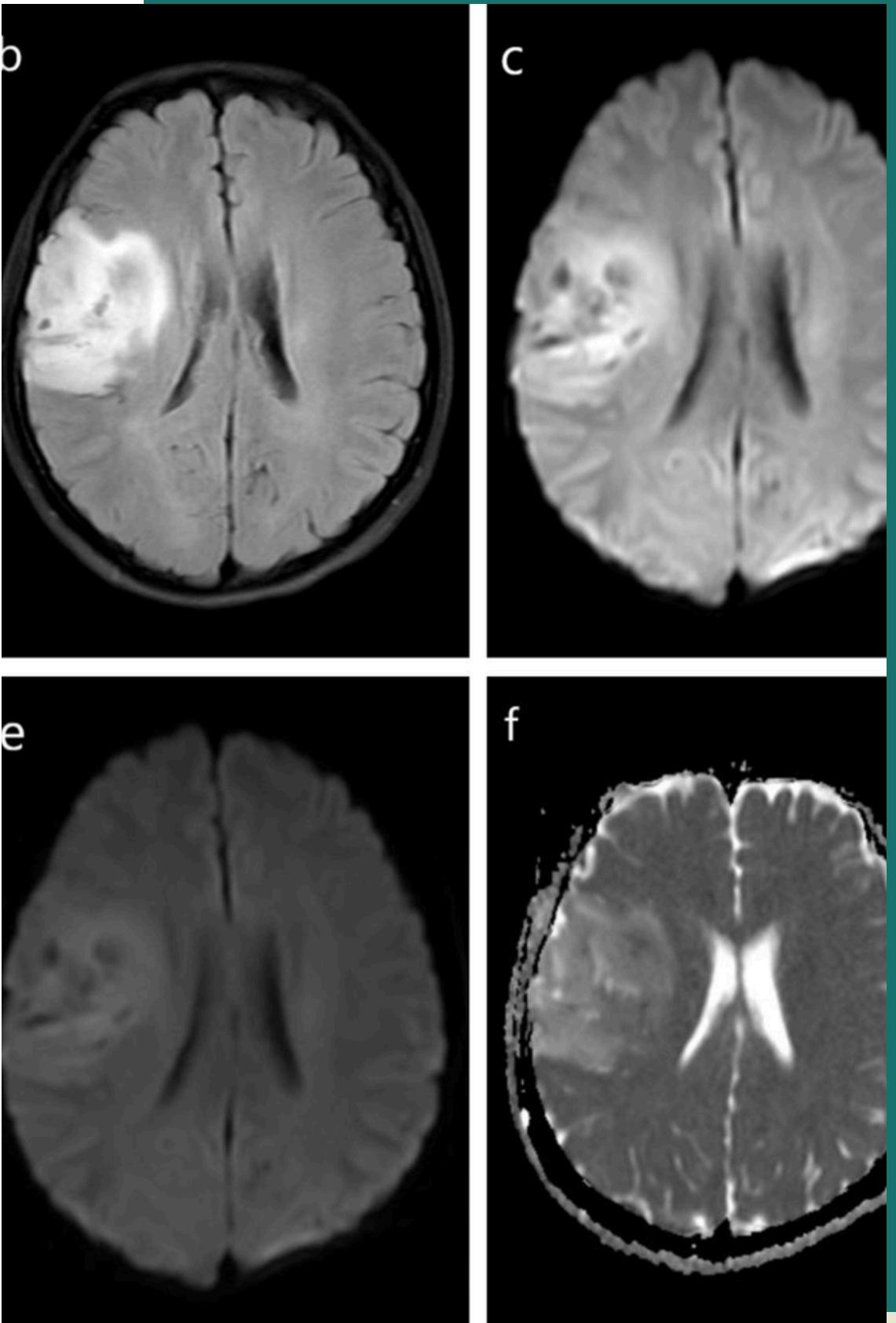
OVERVIEW + JUSTIFICATION

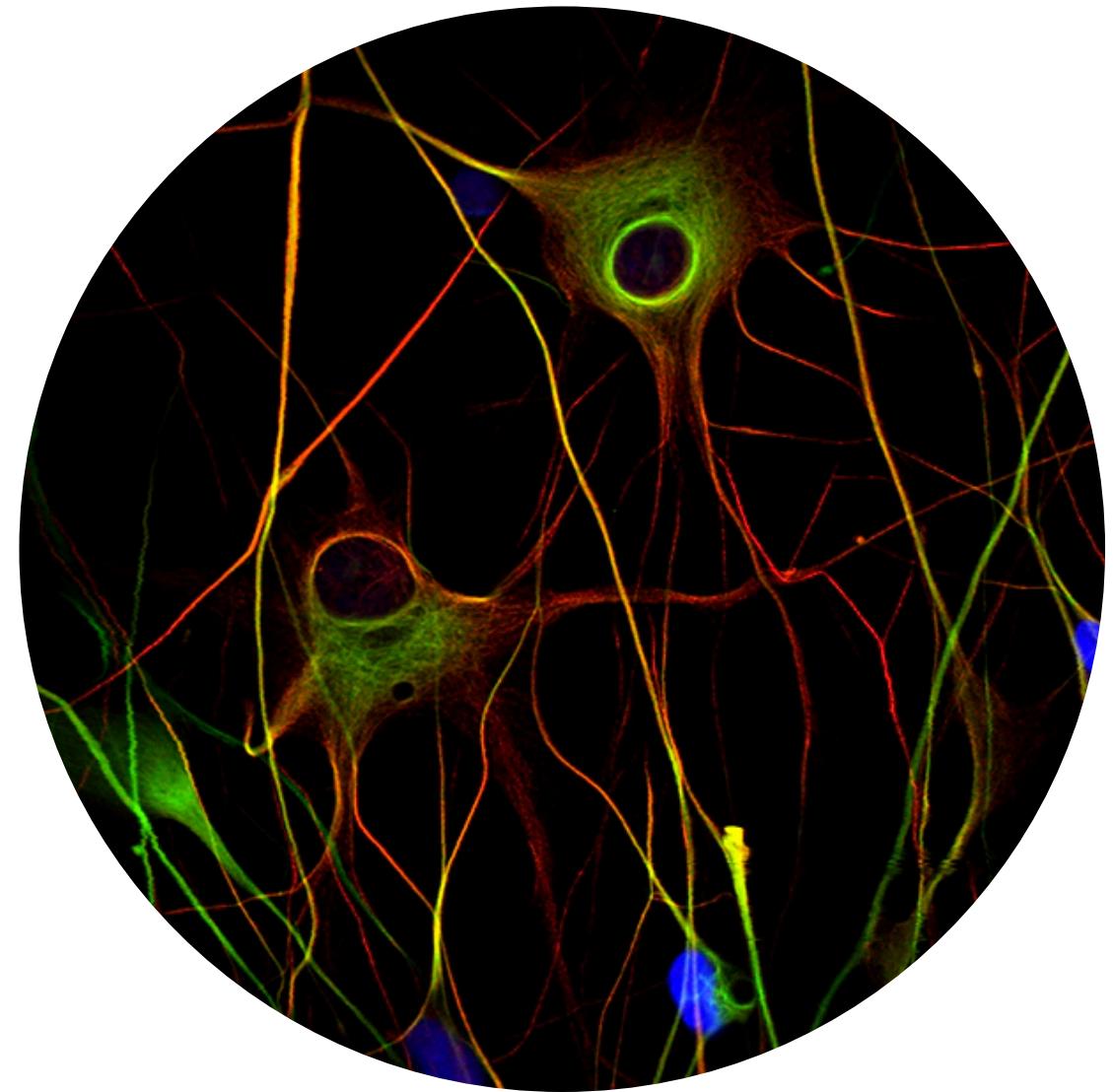
Overview:

- Methodology: 2-fold cross-validation
- Dataset is split into 2 parts:
 - One part is used for training
 - One part is used for testing
 - Process is repeated with roles switched
- Process ensures that each observation is used in training and testing

Justification:

- Scored 10.0 on usability
- Wanted to work with binary data
- We were curious to see if mutations and different sets of genes could help predict glioma grading





EXAMPLE STIMULI

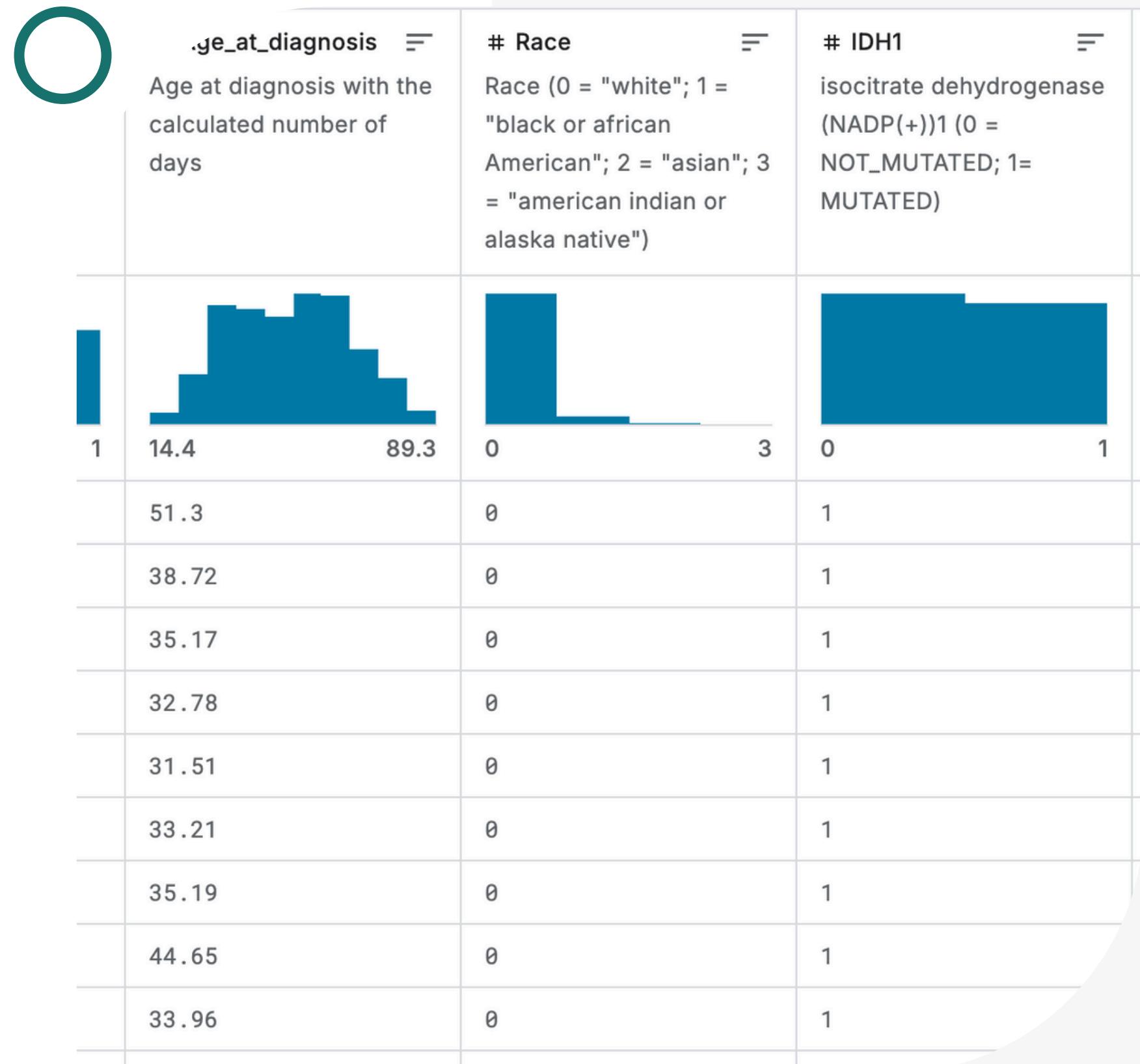
Predictors (Stimuli): Genomic and clinical features including genetic mutation markers such as IDH1, TP53, ATRX, as well as demographic variables like age at diagnosis, gender, and race. These features serve as the input stimuli for the model during training and evaluation.

Target Variable: Tumor Grade, encoded as a binary outcome indicating low-grade (0) or high-grade glioma (1)

Grade	Gender	Age_at_diagnosis	Race	IDH1	TP53	ATRX	PTEN	EGFR	CIC	...	FUBP1	RB1	NOTCH1	BCOR	CSMD3	SMARCA4	GRIN2A	IDH2	FAT4	PDGFRA
0	0	0	51.30	0	1	0	0	0	0	...	1	0	0	0	0	0	0	0	0	
1	0	0	38.72	0	1	0	0	0	0	1	...	0	0	0	0	0	0	0	0	
2	0	0	35.17	0	1	1	1	0	0	0	...	0	0	0	0	0	0	0	0	
3	0	1	32.78	0	1	1	1	0	0	0	...	0	0	0	0	0	0	0	1	
4	0	0	31.51	0	1	1	1	0	0	0	...	0	0	0	0	0	0	0	0	

DATA SET

- TCGA Brain Glioma Grading Dataset
- Found on Kaggle
- Data derived from Cancer Genome Atlas and funded by National Cancer Institute
- MRI pixel data
- 23 variables
 - 20 genes and markers (mutated and non-mutated
 - IDH1, TP53, ATRX, etc.
 - Demographic information
 - Gender
 - Age
 - Race
- Glioma grading as target result



ML PROCEDURES

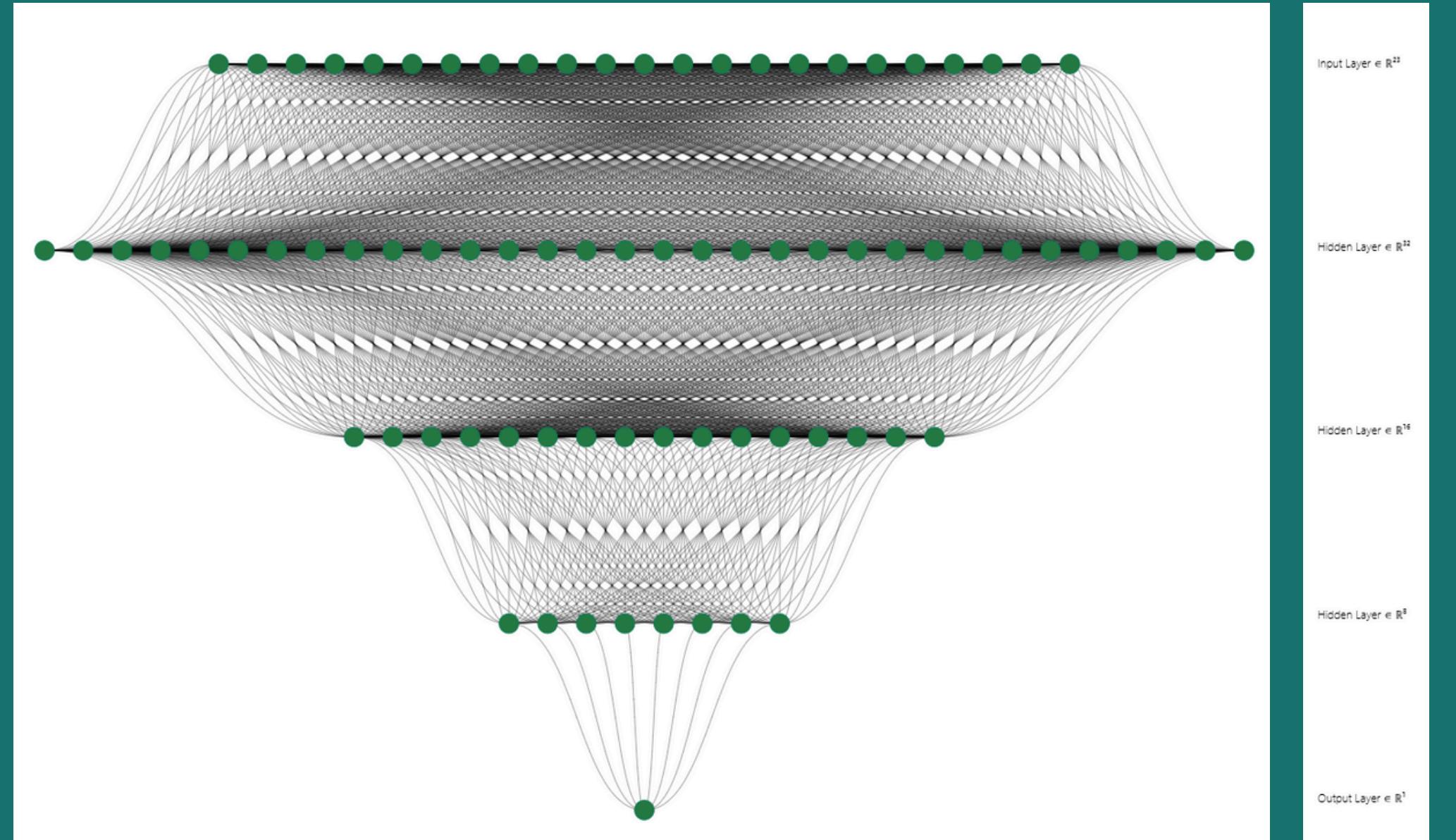
SOLUTIONS OF THE PROBLEMS

Classification Problem

- 23 features
- categorical features are one-hot encoded

Epoch, Hidden Units, and Hidden Layers

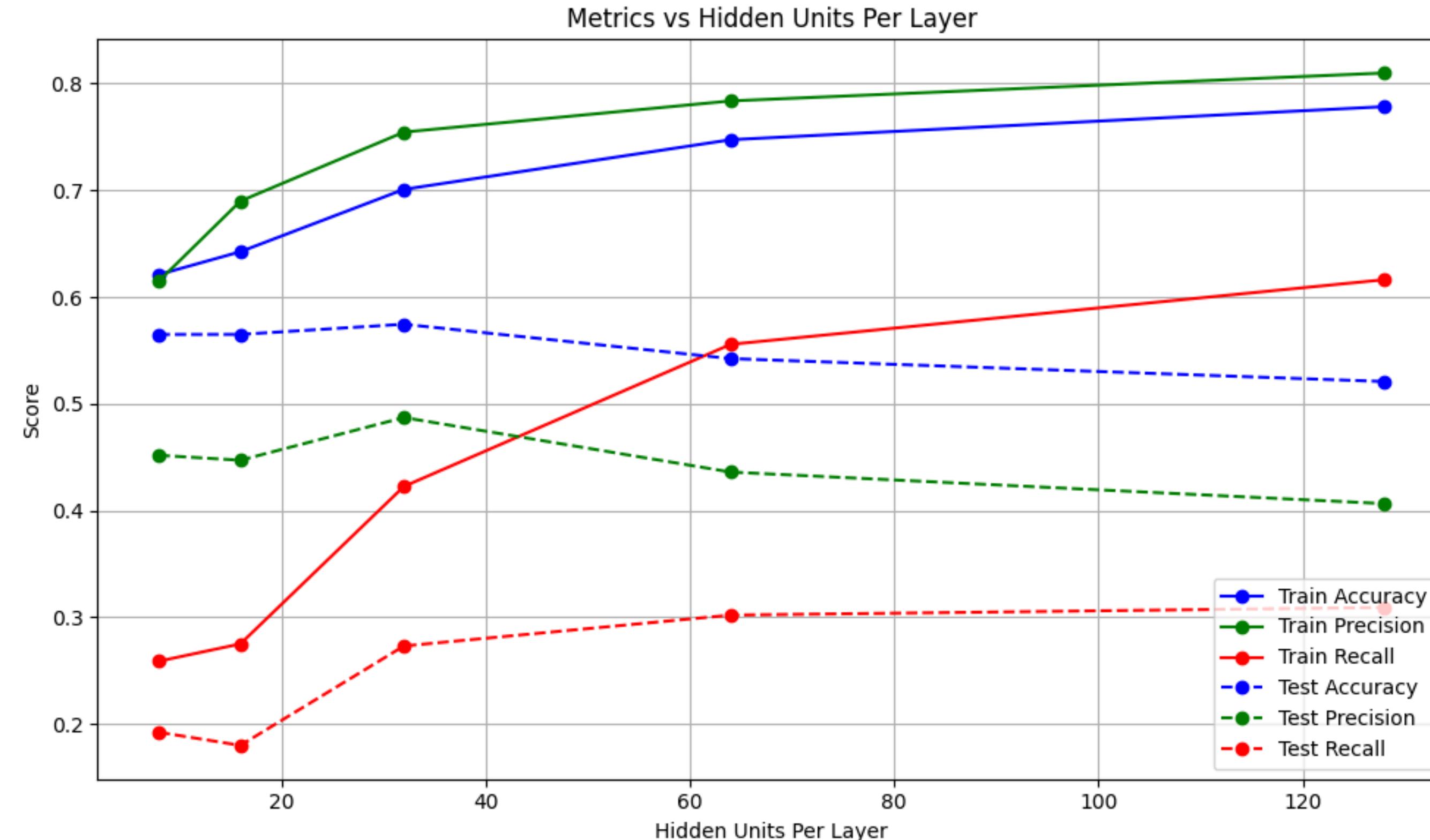
- By modifying epoch, hidden unit and hidden layers, we can adjust how well the model is trained.



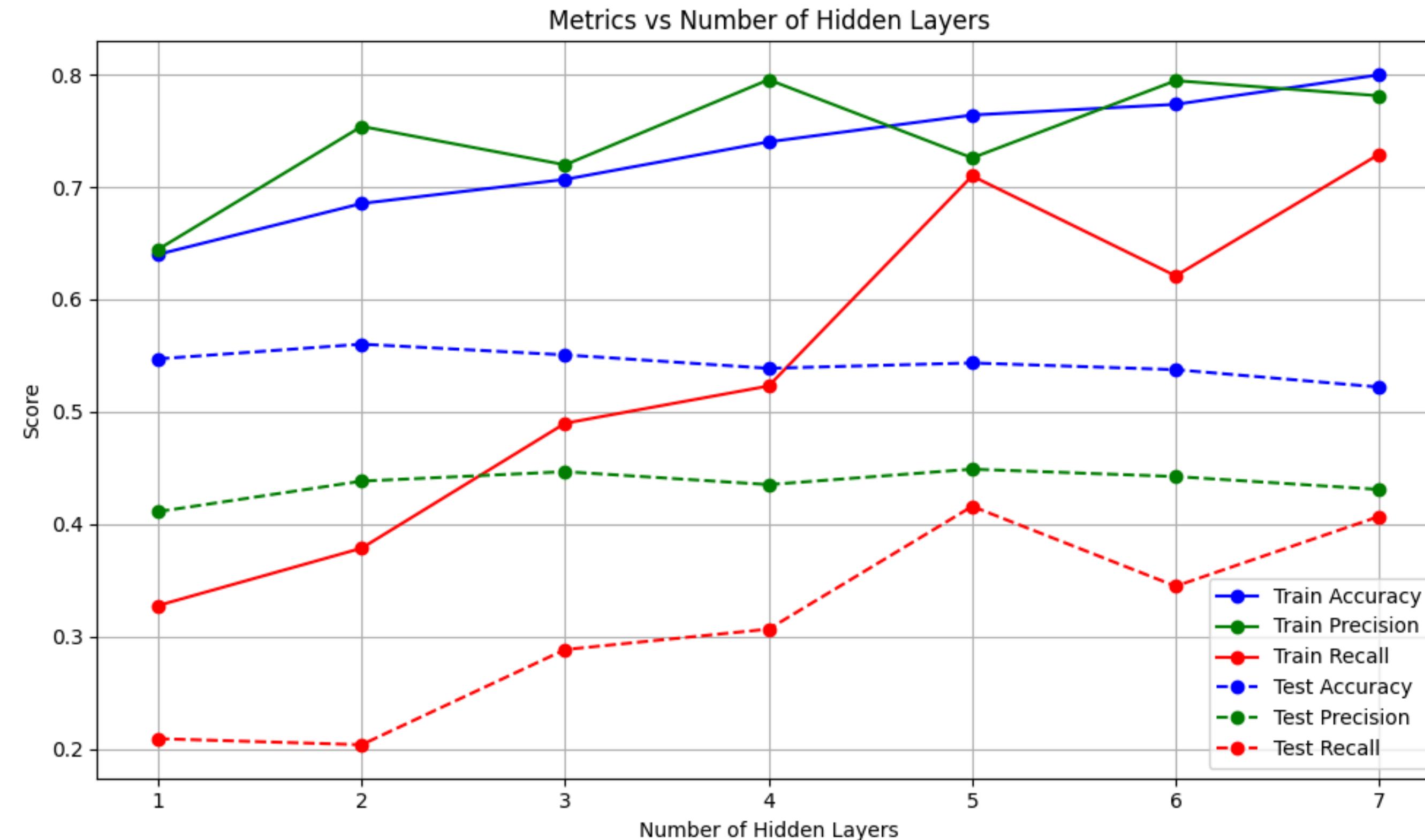
RESULTS: METRICS FOR EPOCHS



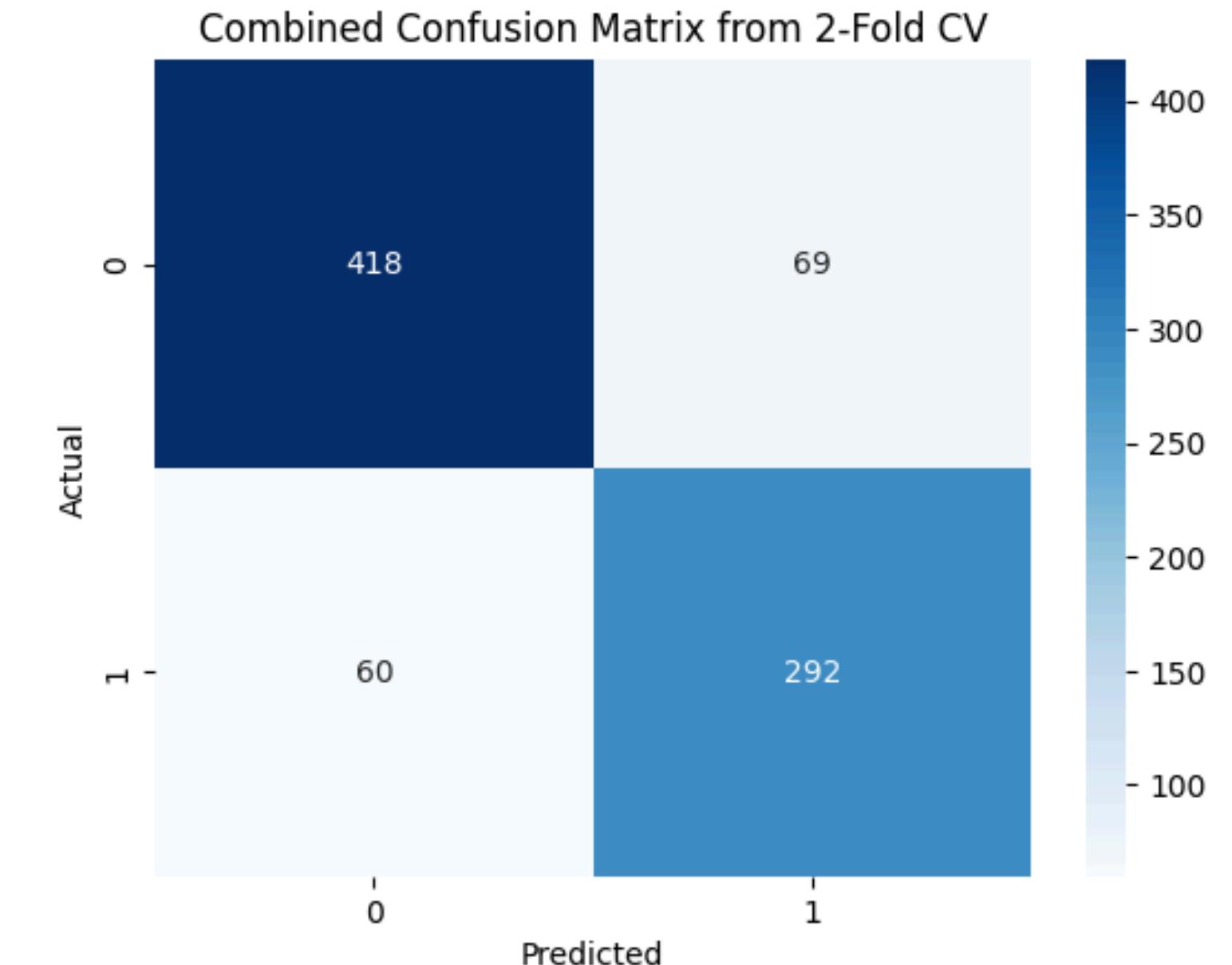
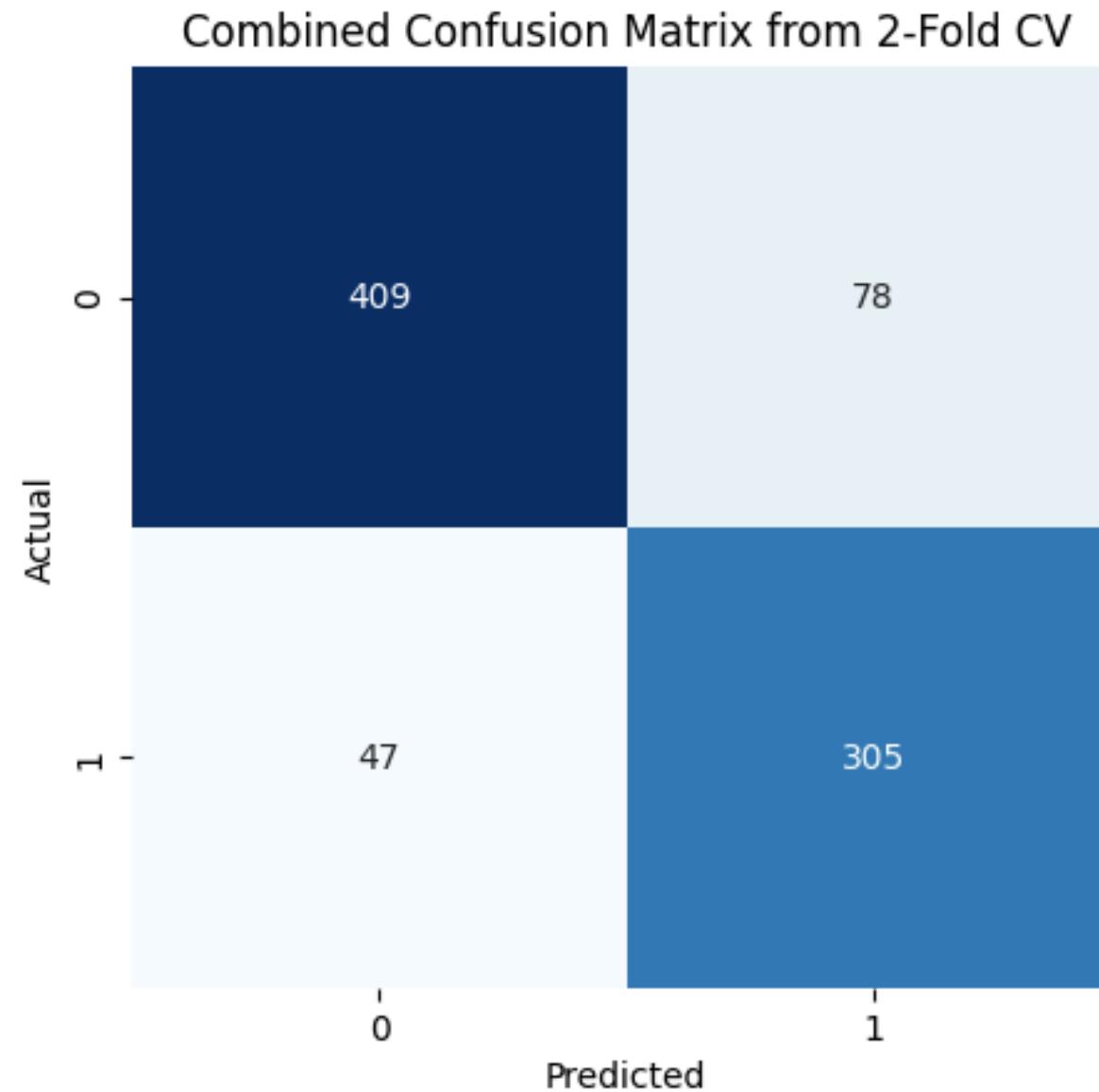
RESULTS: HIDDEN UNITS



RESULTS: HIDDEN LAYERS



FINDINGS



Average Metrics Across Folds			
3 hidden layers	Precision	0.799716	
32 units maximum	Recall	0.894970	
20 epochs	F1-Score	0.844517	
	Accuracy	0.861734	

Average Metrics Across Folds			
5 hidden layers	Precision	0.809315	
64 units maximum	Recall	0.829846	
20 epochs	F1-Score	0.818992	
	Accuracy	0.846250	

LITERATURE RECALL

McCaffery (2018)

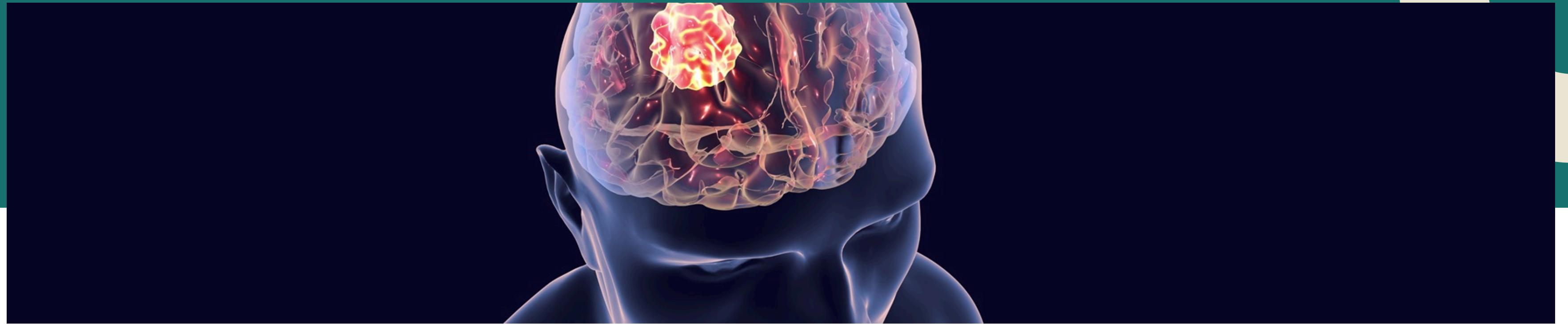
- Tabular data (Banknote stats) with hyperparameters
 - 500 epochs
 - 2 hidden layers
 - 8 units each
- Evaluated accuracy only
- Feed forward neural network
- Binary
- 99.27% accuray

Barstugan (2020)

- Image data (CT scans)
- 10 fold CV evaluation
- ML netwoks
- feature extraction
- Binary
- 99.68% accuracy

Glioma Grading

- Tabular data (Glioma data)
- 2 fold CV evaluation
- Feed forward neural network
- Binary
- 86.17% accuracy



FINAL MESSAGE

IDENTIFYING GLIOMA GRADING THROUGH BEST COMBINATIONS OF GENES, MUTATIONS, AND FEATURES

Our project has an 86% accuracy rate of identifying and classifying Glioma grading. Through manipulating features which are genetic and demographical. This proves how machine learning can be of aid in a medical sense with more fine tuning. As we are able to increase the model's performance, we can catch and treat glioma earlier.

Works Cited

- Barstugan, Mucahid, et al. "Coronavirus (COVID-19) Classification Using CT Images by Machine Learning Methods." *arXiv.Org*, 20 Mar. 2020, arxiv.org/abs/2003.11421.v1.
- CCMJ. "What Is Glioblastoma (GBM)?" *Cleveland Clinic*, 23 Apr. 2025, my.clevelandclinic.org/health/diseases/17032-glioblastoma.
- Deacu, Mariana, et al. "Prognostic Factors of Low-Grade Gliomas in Adults." *Current Oncology (Toronto, Ont.)*, U.S. National Library of Medicine, 30 Sept. 2022, pmc.ncbi.nlm.nih.gov/articles/PMC9600247/.
- McCaffrey, James. "How to Do Neural Binary Classification Using Keras." *Visual Studio Magazine*, 30 Aug. 2018, visualstudiomagazine.com/articles/2018/08/30/neural-binary-classification-keras.aspx.
- McCaffrey, James. "Neuroscientists Win The Brain Prize 2025 for Pioneering Research into Brain Cancer." *News*, 6 Mar. 2025, www.news-medical.net/news/20250305/Neuroscientists-win-The-Brain-Prize-2025-for-pioneering-research-into-brain-cancer.aspx.
- "Neuroscientists Win The Brain Prize 2025 for Pioneering Research into Brain Cancer." *News*, 6 Mar. 2025, www.news-medical.net/news/20250305/Neuroscientists-win-The-Brain-Prize-2025-for-pioneering-research-into-brain-cancer.aspx.
- Northwestern Medicine. "Overview: Understanding Low Grade Glioma." *Northwestern Medicine*, www.nm.org/conditions-and-care-areas/neurosciences/comprehensive-care-for-low-grade-gliomas/low-grade-glioma#:~:text=Low%20grade%20gliomas%20are%20brain%20tumors%20that%20develop,from%20one%20of%20several%20types%20of%20glial%20cells. Accessed 1 May 2025.
- Zj. "Clinical_gliomagrading." *Kaggle*, 16 Dec. 2023, www.kaggle.com/datasets/tanshihjen/clinical-gliomagrading?resource=download.



THANK YOU

● FOR YOUR ATTENTION