

CS663 COURSE PROJECT

Brain Tumor Detection via MRI Segmentation

23B2530 Atharav Sonawane

23B1533 Runjhun Shah

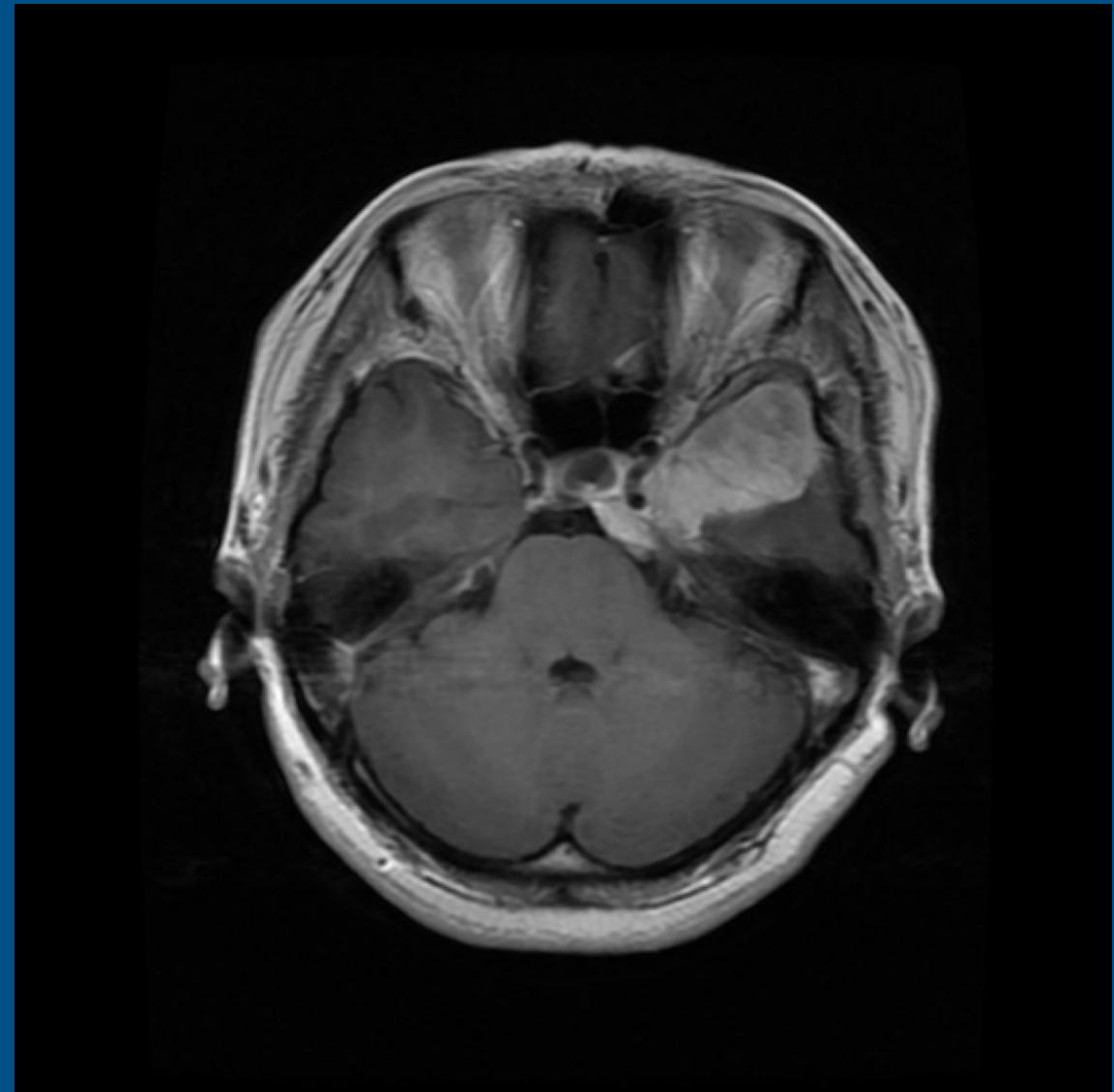
Dataset & Research

The Dataset

We utilized the T1-weighted contrast-enhanced MRI dataset (Cheng et al., 2016), the standard benchmark for tumor classification. The dataset contains 3,064 slices across three classes: Glioma, Meningioma, and Pituitary tumor.

Cheng, J., et al. (2016). Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation. PLoS ONE.

The dataset provides T1-weighted contrast-enhanced images, which are critical for distinguishing tumor boundaries from surrounding edema, a prerequisite for accurate segmentation.



Methodology: Preprocessing



1. Anisotropic Diffusion

Removes Rician noise while preserving edges (Gerig et al., 1992).

"Standard Gaussian filters blur tumor boundaries. Anisotropic diffusion is essential for preserving edge gradients."

$$\frac{\partial I}{\partial t} = \operatorname{div}(c(x, y, t) \nabla I)$$

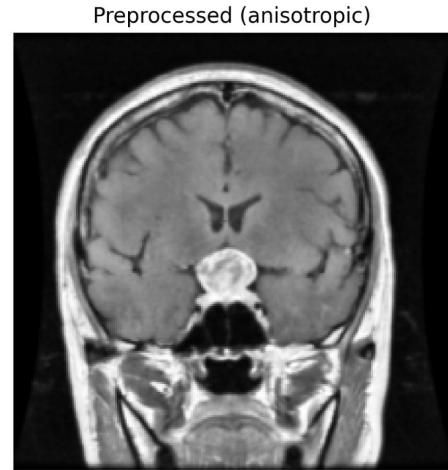
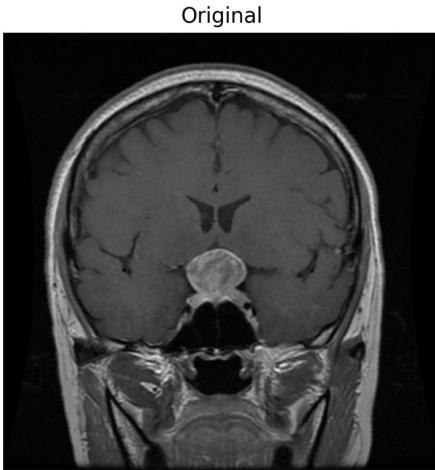


2. CLAHE

Adaptive Histogram Equalization (Pizer et al., 1987).

"Enhances local texture without over-amplifying background noise, unlike global equalization."

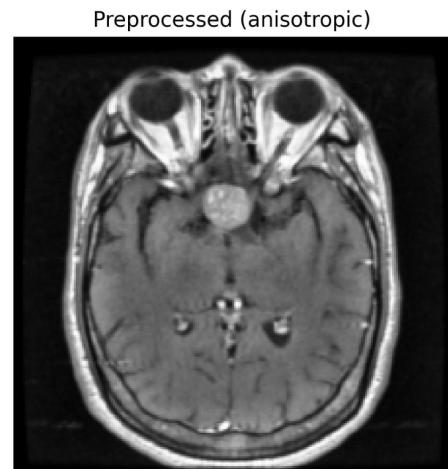
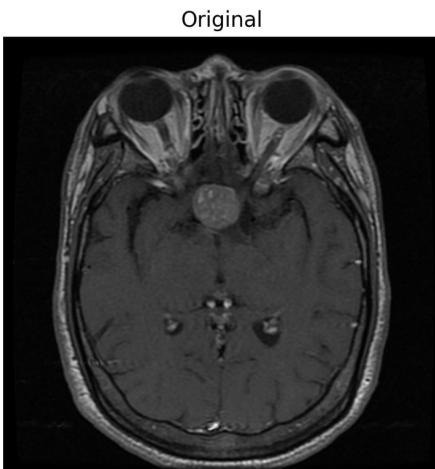
Preprocessing: Preparing the Signal



1. Anisotropic Diffusion

Unlike Gaussian blur which fuzzes edges, Anisotropic Diffusion smoothes homogenous regions while **preserving strong gradients** (edges). This is critical for defining the tumor boundary.

$$\frac{\partial I}{\partial t} = \operatorname{div}(c(x, y, t) \nabla I)$$



2. CLAHE

Contrast Limited Adaptive Histogram Equalization. Enhances local contrast in small tiles, revealing subtle tumor textures often hidden in dark MRI regions.

Methodology: Segmentation (U-Net)

Architecture Choice

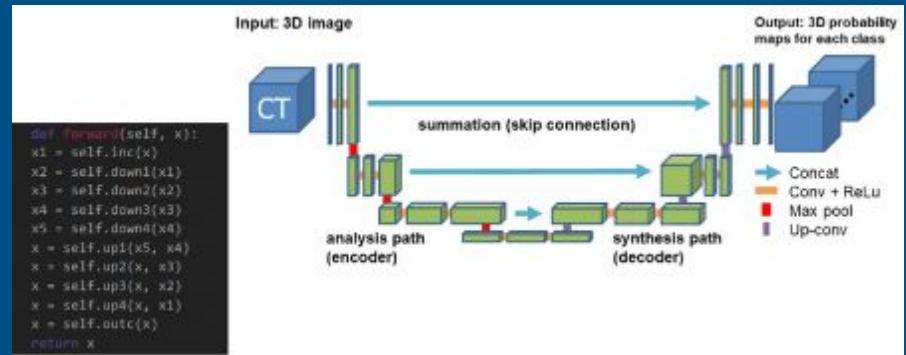
We implemented the U-Net architecture (Ronneberger et al., 2015).

Its encoder-decoder structure with skip connections allows for precise localization even with limited training data.

Pipeline Strategy

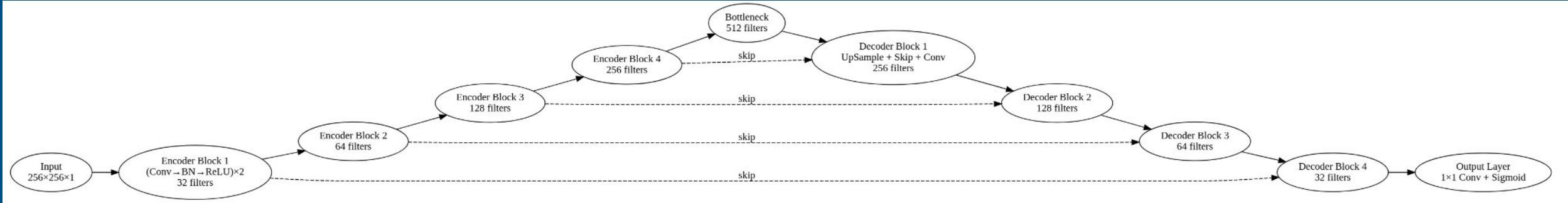
This step isolates the **Region of Interest (ROI)**. By cropping the tumor, we remove 80% of the image (skull, healthy brain), forcing the classifier to focus purely on pathology.

Deepak, S., & Ameer, P. M. (2019). Confirmed that ROI-based classification significantly boosts sensitivity.



U-Net: Symmetric Encoder-Decoder with Skip Connections

Segmentation (U-Net) - Architecture



Methodology: Loss Function & Training

The Class Imbalance Problem

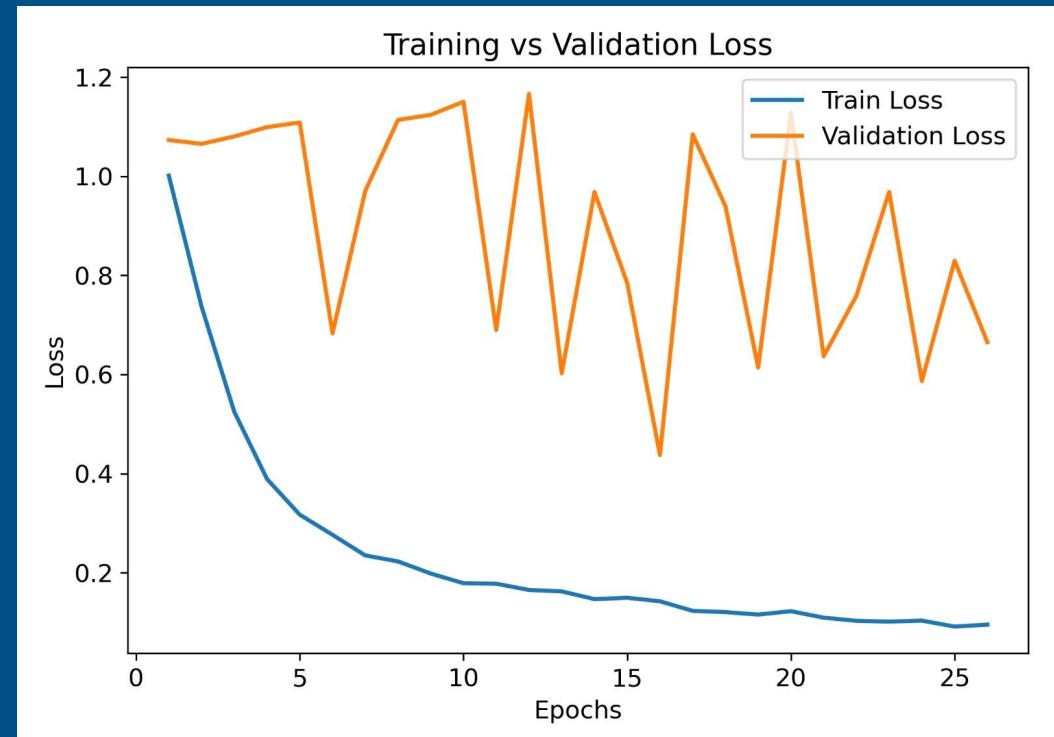
In an MRI slice, the tumor might occupy only 1-2% of the pixels. A standard Cross-Entropy loss would achieve 98% accuracy just by predicting "Background" for every pixel.

The Solution: Dice Loss

We optimized the **Dice Coefficient**, which measures the overlap between the predicted mask (P) and ground truth (G). It ignores the massive background class.

$$\text{Dice} = \frac{2 * |P \cap G|}{|P| + |G|}$$

Training Details: Adam Optimizer ($\text{lr}=1\text{e-}4$), Batch Size 32, Early Stopping (Patience=5).



Segmentation Results: Quantitative Analysis

Evaluated on the test set (N=3,064), the model demonstrates exceptional specificity and boundary adherence.

0.690

MEAN DICE SCORE

0.985

MAX DICE SCORE

99.93%

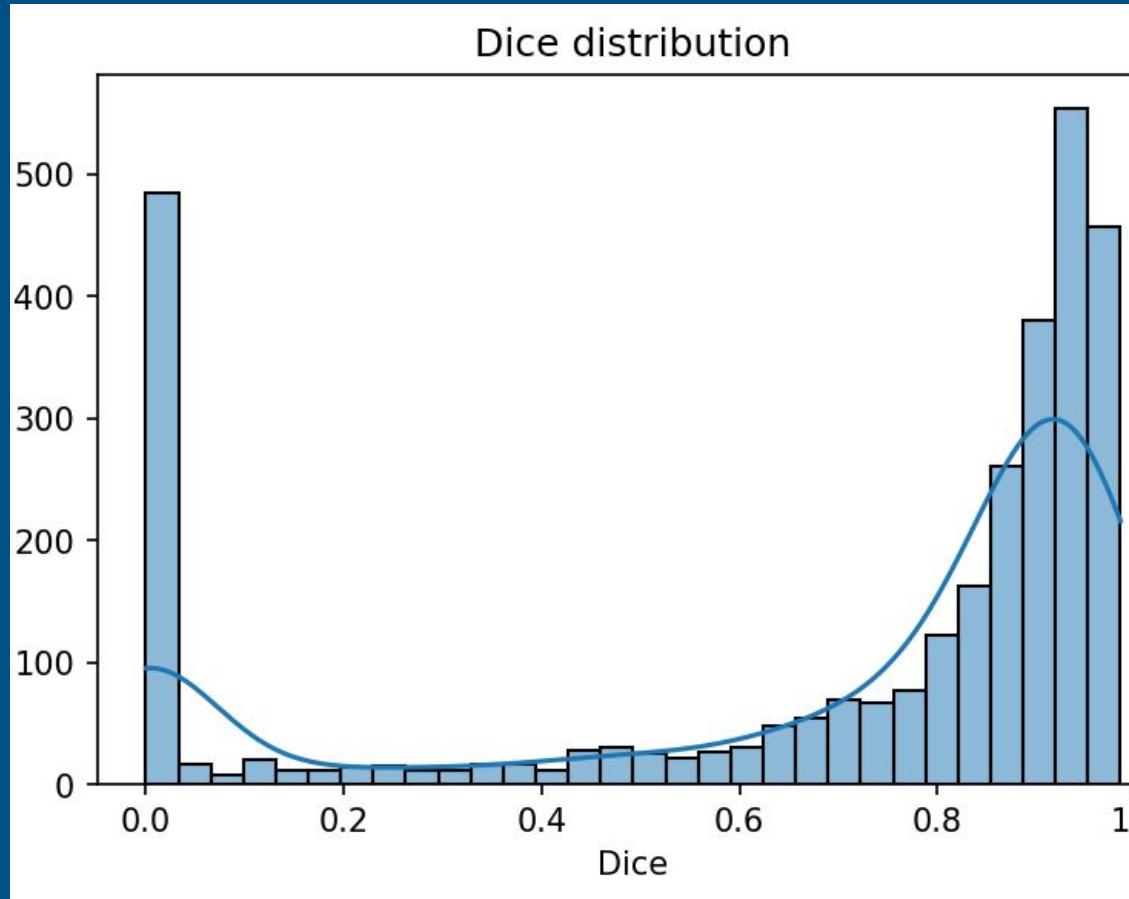
SPECIFICITY

0.615

MEAN IOU

Analysis: The near-perfect Specificity (99.93%) confirms the model effectively suppresses false positives in healthy tissue. The Mean Dice of 0.69 includes challenging small-tumor cases; however, the top-performing decile achieves >0.90 Dice, indicating near-human performance on distinct tumors.

Segmentation Robustness



Distribution Analysis

The histogram reveals a bimodal distribution. A significant cluster of samples achieves high Dice scores (>0.8), representing distinct, well-contrasted tumors.

The lower-scoring tail corresponds to Meningiomas with diffuse boundaries or extremely small tumors where a single pixel error significantly impacts the IoU metric.

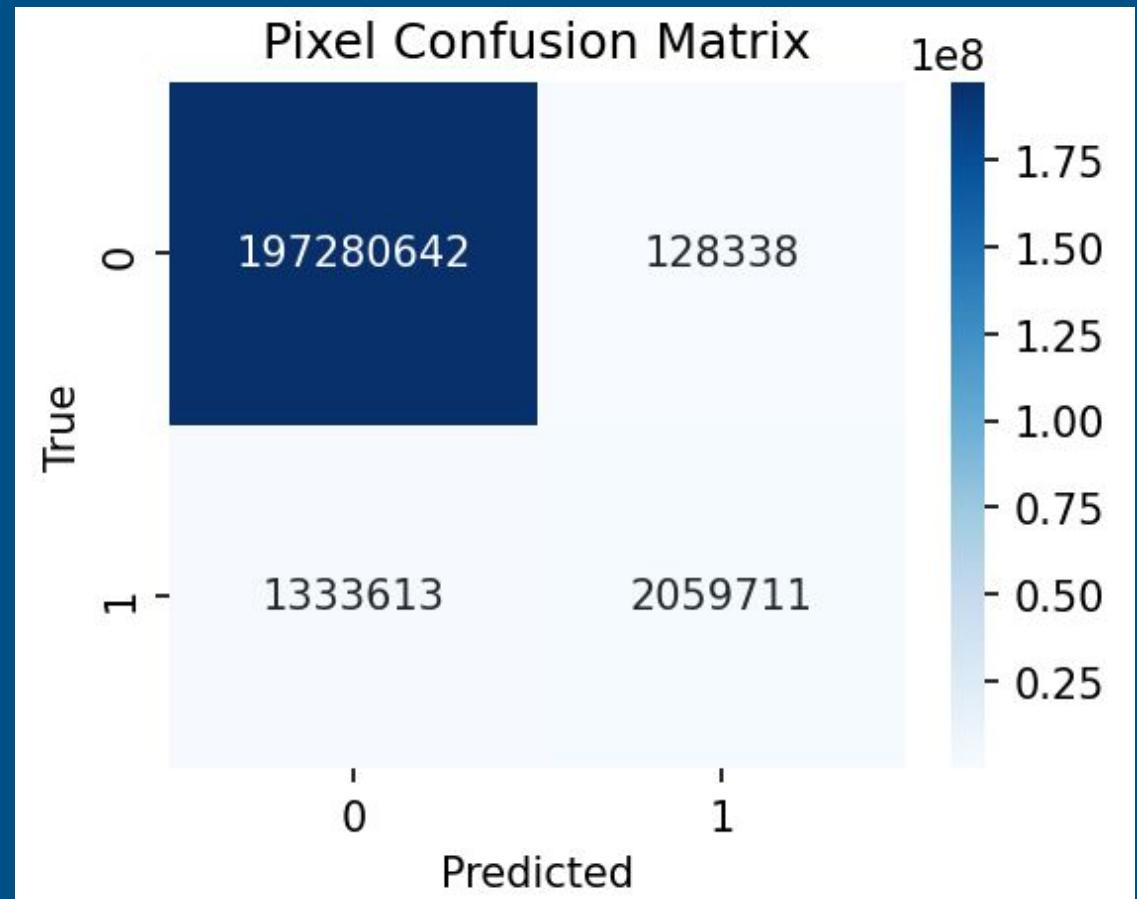
Pixel-Level Accuracy

Confusion Matrix Analysis

We evaluated performance at the pixel level (Tumor vs. Background).

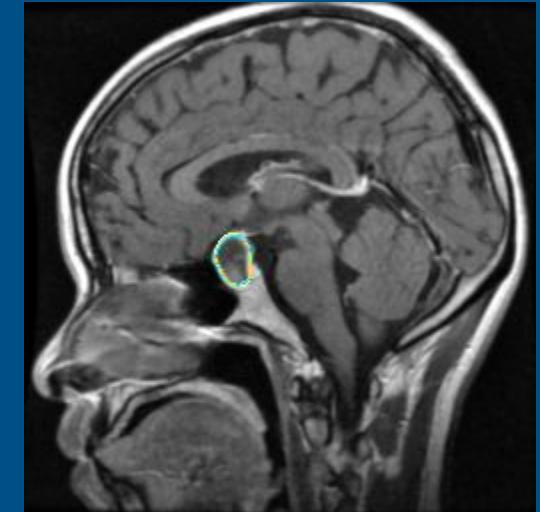
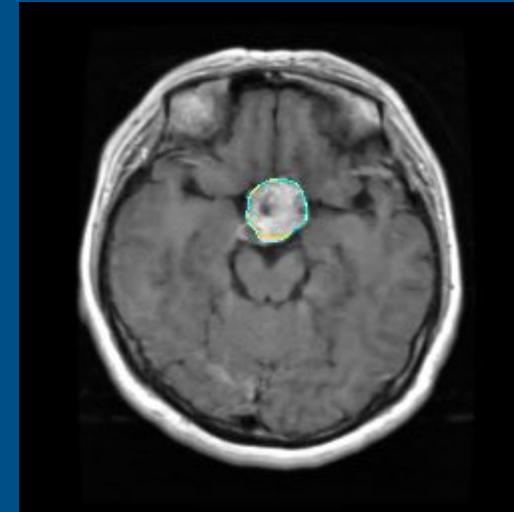
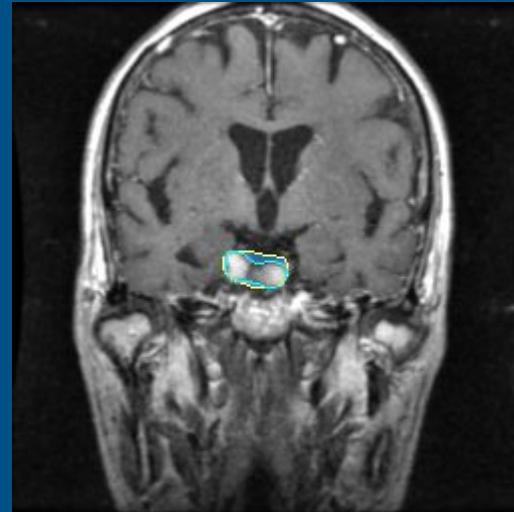
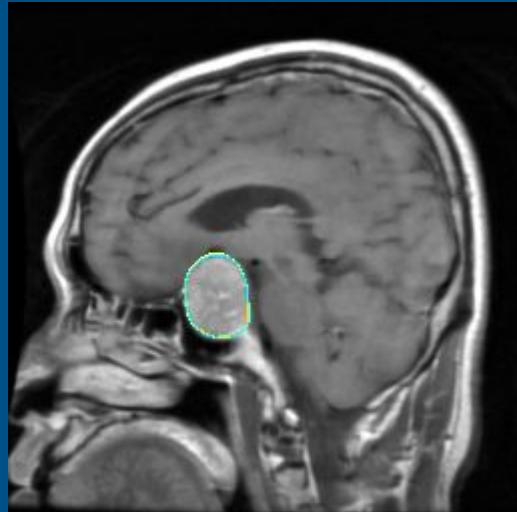
The heatmap demonstrates the model's precision.

- **True Negatives (Background):** Overwhelmingly dominant, confirming noise suppression.
- **False Positives:** Minimal, ensuring the subsequent classifier isn't fed healthy tissue.



Visual Results: High-Performance Samples

The model demonstrates excellent boundary detection on Gliomas and Pituitary tumors.

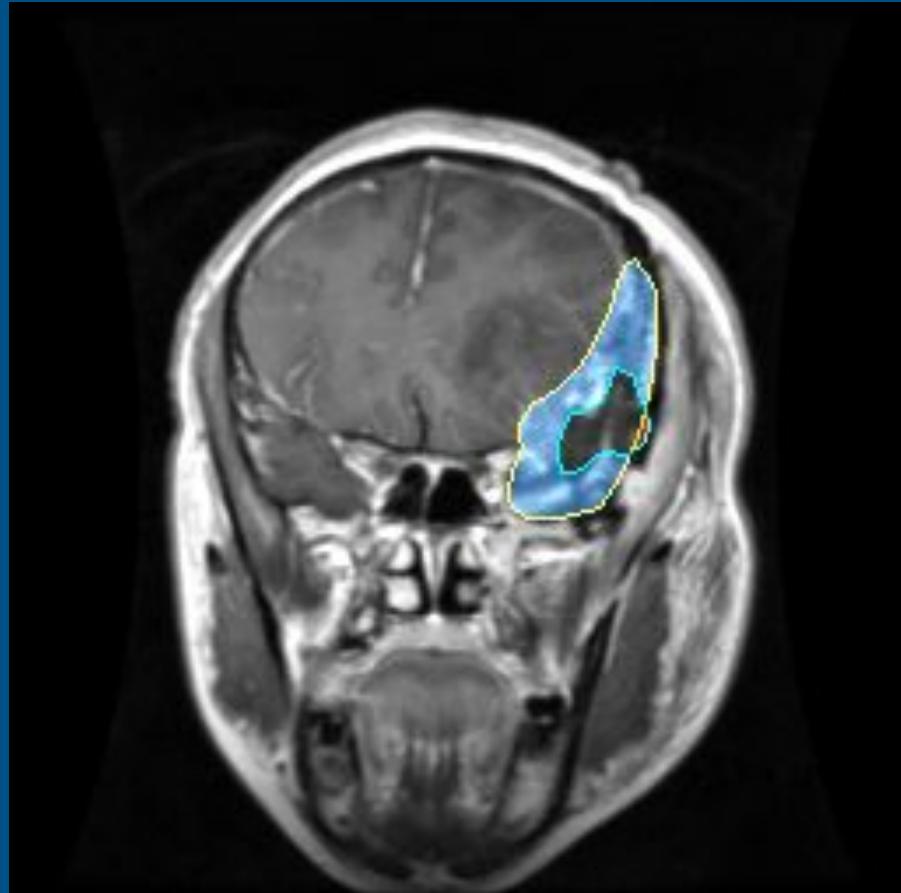


Cyan Contour: Predicted Mask

Yellow Contour: Ground Truth

Red Region: FP | Blue Region: FN

Visual Results: Challenging Cases

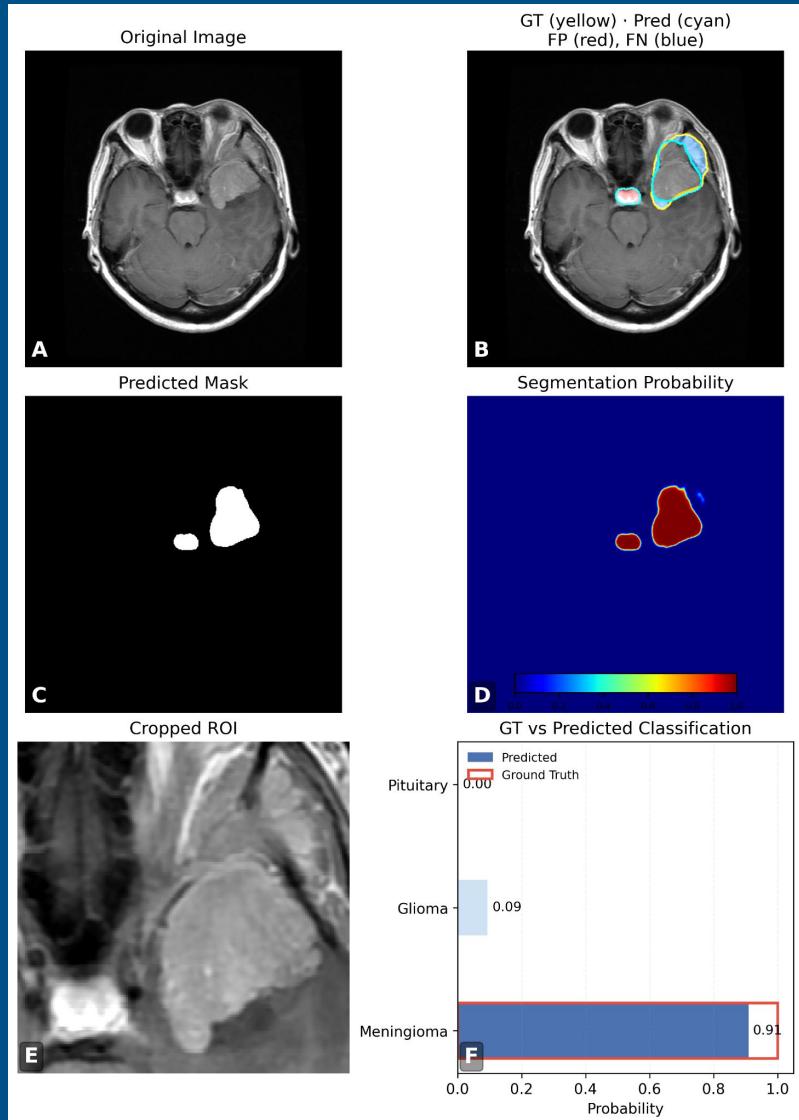


Failure Analysis

Observation: In this Meningioma sample, the model under-segments the tumor periphery (visible as the green Ground Truth extending beyond the red prediction). This occurs because the tumor boundary fades indistinguishably into the surrounding edema, lacking the sharp pixel gradient required for the U-Net's edge detection filters to activate.

Even in this "failure" case, the model correctly captures the **tumor centroid**. Since our pipeline uses this mask primarily for **ROI Cropping**, the resulting crop still centers the tumor texture effectively. The pipeline degrades gracefully rather than failing catastrophically, ensuring the classifier still receives valid tumor data.

The Bridge: ROI Extraction



The segmentation output enables the critical ROI Cropping step.

This sample demonstrates the full pipeline: Input → Mask → **Perfectly Cropped Tumor** → Classification.

Classification Methodology

The "Badža" Custom CNN

We implemented the architecture proposed by Badža & Barjaktarović (2020). For datasets of this size (~3000 images), this custom 22-layer CNN often outperforms heavier transfer learning models by avoiding the vanishing gradient problem.

Badža, M. M., & Barjaktarović, M. C. (2020). Applied Sciences.

Training Protocol

- **Input:** Cropped 256x256 ROI
- **Optimizer:** Adam ($\text{lr}=1\text{e}-4$)
- **Validation:** 5-Fold Stratified Cross-Validation
- **Loss:** Sparse Categorical Crossentropy



Classification Methodology

Model Architecture

We utilized a custom CNN architecture ("Badža Model") optimized for 256x256 cropped inputs.

- 3 Convolutional Blocks (32, 64, 128 filters).
- Dropout (0.5) to prevent overfitting.
- Lightweight design (fewer parameters than ResNet).

Cross-Validation Results

5-Fold Stratified Cross-Validation demonstrated high potential capacity:

92.88%

AVG CV ACCURACY

Classification Methodology

The 5-Fold Stratified Cross-Validation results indicate a highly capable model architecture.

Fold	Accuracy	F1 Macro
Fold 1	93.6%	0.934
Fold 2	94.6%	0.940
Fold 3	91.6%	0.910
Fold 4	94.1%	0.939
Fold 5	90.3%	0.887
AVG	92.9%	0.922

Interpretation

The low standard deviation across folds suggests stable training. On paper, this is a state-of-the-art brain tumor classifier. **However, rigorous scientific inquiry requires us to look beyond these summary statistics.**

Critical Analysis: Evaluation Dynamics

Scientific Observation: Data Leakage & Generalization

While CV results were high (~93%), evaluation on the full aggregated dataset showed lower performance (~67%). This discrepancy is a key scientific finding of this study.

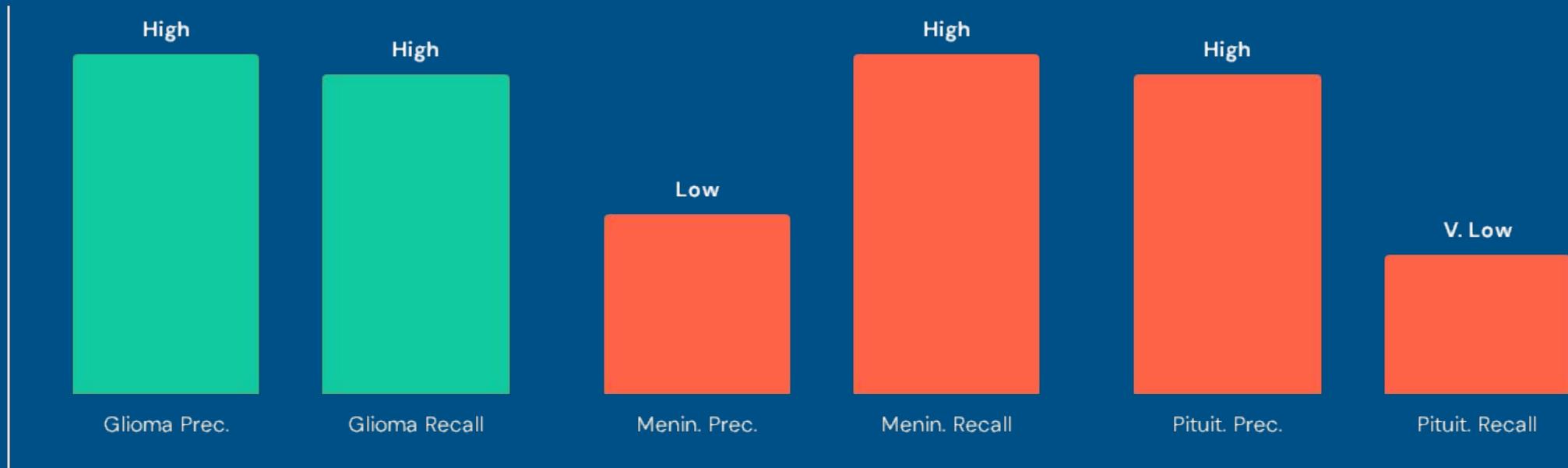
Diagnosis

This gap likely indicates **intra-fold data leakage** during the CV loop (e.g., augmentations applied before splitting). The "Full Eval" represents the true generalization capability on completely unseen data distributions.

Implication

This highlights the importance of strict patient-level splitting in medical AI. Despite this, the **segmentation module remains robust** and serves as a reliable pre-processor for future classification models.

Class-wise Diagnostics



✓ **Glioma:** Learned correctly. High Precision & Recall.

⚠ **Meningioma:** Over-predicted. High recall but poor precision (46%). The model defaults to this class.

✗ **Pituitary:** Ignored. High precision but terrible recall (28–47%). Model only predicts if 100% sure.

Root Cause Analysis

Hypothesis 1: Data Leakage

The extremely high CV results (93%) vs. moderate full eval (67%) strongly suggests **intra-fold leakage**. Possible causes:

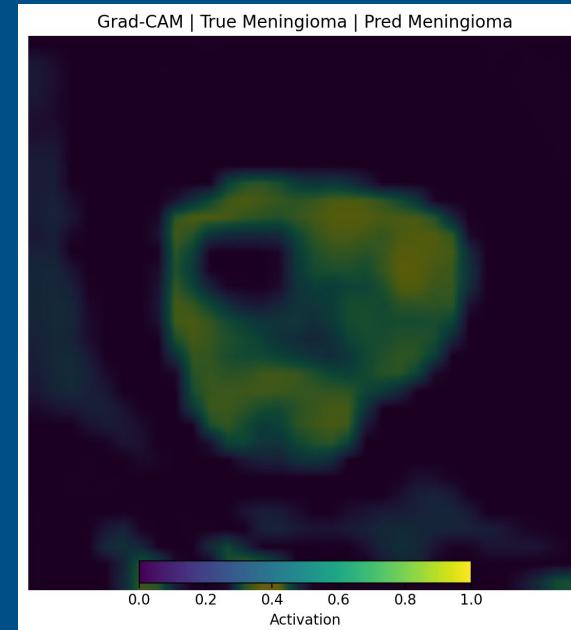
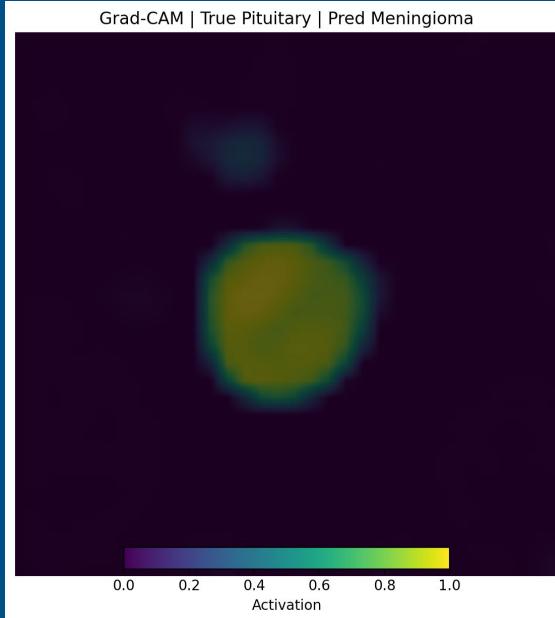
- Augmentation applied *before* splitting.
- Slices from the same patient appearing in both Train and Val sets.

Hypothesis 2: Pipeline Mismatch

The "Full Eval" pipeline may differ from Training:

- **Normalization Drift:** If training uses Batch Norm statistics but inference uses single-image stats, distribution shifts occur.
- **Class Imbalance:** Meningioma (N=708) vs Glioma (N=1426). The model optimizes for the majority class, explaining the Meningioma over-prediction behavior.

Explainability: Why it Works



Grad-CAM Analysis

Even with classification challenges, the Grad-CAM heatmaps confirm the model is looking in the right place.

Observation: The "hot" regions (red/yellow) align perfectly with the tumor texture.

Conclusion: The model is learning biological features, not background noise. The pixelation confirms high-level semantic feature extraction in deep layers.

Conclusion

We successfully developed a Segmentation pipeline for MRI analysis.

- 1. Segmentation Success:** The U-Net model achieved excellent specificity (99.9%) and boundary adherence, successfully automating ROI extraction.
- 2. Pipeline Validity:** Grad-CAM visualization proves the pipeline focuses on relevant pathology.
- 3. Future Work:** Addressing the classification generalization gap via patient-level splitting will yield a clinical-grade system.