Hybrid Brain Tumor Analysis in MRI: Xception Classification + Class-Conditional Classical Segmentation

CS 555 - Introduction to Visual Information Processing

Final Term Project

Maxime Argenson

Saturday December 13, 2025

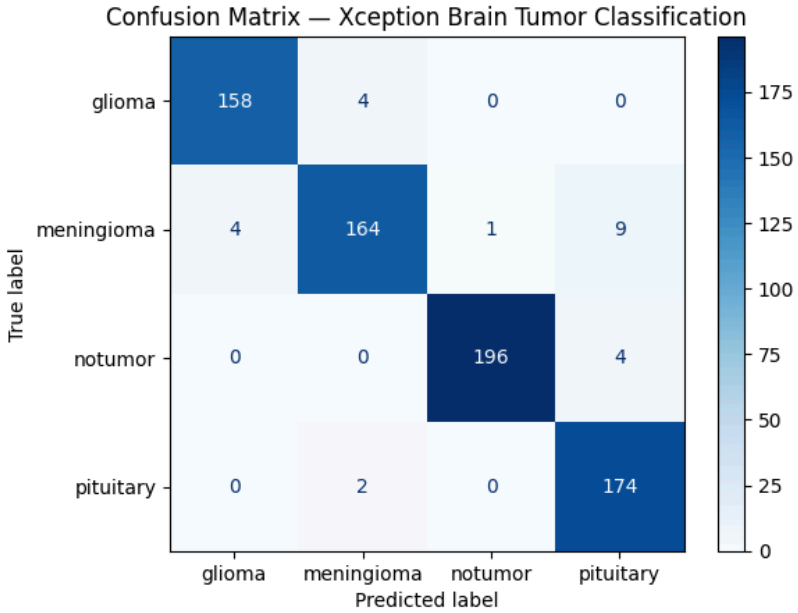
Problem Statement

Magnetic Resonance Imaging (MRI) is widely used to detect and characterize brain tumors, but interpreting scans is time-consuming and often complicated by noise, scanner variability, and subtle differences between tumor types. The objective of this project was to build an automated pipeline for brain tumor detection and subtype classification, and to explore tumor localization/segmentation as a secondary step. The work uses the Brain Tumor MRI Dataset from Kaggle (7,153 images total: 1,621 glioma, 1,775 meningioma, 1,757 pituitary, 2,000 no tumor) and targets four-class prediction: glioma, meningioma, pituitary, and no tumor.

The project initially proposed a direct comparison between classical image processing segmentation and U-Net-based deep segmentation. However, the dataset provides classification labels only (no pixel-level masks), making supervised segmentation evaluation unrealistic: without truth outlines, a segmentation model couldn't be trained using standard metrics. In parallel, early experiments with a "one-size-fits-all" classical segmentation approach were unreliable because tumor appearance varies substantially across classes and even across scans within the same class. As a result, the project was re-scoped into a hybrid approach: (1) perform multi-class classification using deep learning, then (2) apply a class-conditional classical segmentation/localization plan only when a tumor class is predicted, using different assumptions per tumor type.

For classification, transfer learning was used to improve performance versus a CNN trained from scratch. An Xception backbone pretrained on ImageNet was fine-tuned on the MRI dataset with an 80/10/10 train, test, validation split. The final model achieved 0.9665 test accuracy with 0.9665 precision and 0.9665 recall (test loss 0.1072). The confusion matrix indicates strong class separation overall, with the most notable errors occurring between visually

similar tumor types, especially meningioma misclassified as pituitary (9 cases), while no-tumor predictions were highly reliable (196/200 correct in the test set). In contrast, segmentation/localization results were assessed qualitatively through visual overlays, because the dataset lacks ground-truth tumor boundaries.



This project is important because reliable automated classification can support clinical operations, reduce radiologist workload, and provide consistent second opinions, especially in settings where specialist availability is limited.

Literature Review

Automated analysis of brain MRI has been widely studied. Early approaches relied primarily on classical image processing techniques, such as intensity thresholding, morphological operations, and region-based segmentation (Gonzalez & Woods, 2018). While these methods are computationally efficient and interpretable, their performance is often limited by MRI-specific challenges, including noise and non-uniformity.

More recently, convolutional neural networks (CNNs) have become the dominant technique for medical image classification. Transfer learning has been widely adopted for its strong performance, where models pretrained on large-scale natural image datasets are fine-tuned for medical tasks. Architectures such as Xception, which use depthwise separable convolutions, have demonstrated strong performance while maintaining computational efficiency (Xception, 2017).

Method

Dataset Organization and Labeling

The classification model was trained on the Brain Tumor MRI Dataset from Kaggle, organized into four folders corresponding to the target classes: glioma, meningioma, pituitary, and no tumor. A Pandas DataFrame was constructed to store (1) each image file path and (2) its class label, allowing the dataset to be loaded using Keras' "flow_from_dataframe" function. The final dataset contained 7,153 total images: glioma (1,621), meningioma (1,775), pituitary (1,757), and no tumor (2,000). Class counts were visualized to confirm the dataset was moderately balanced with a slight majority of no-tumor examples.

To ensure unbiased evaluation, the dataset was split into 3 subsets: 80% training, 10% test, 10% validation. Splits were performed using stratified sampling so that each subset preserved the original class proportions. Stratification is important for multi-class classification because it prevents a subset from being accidentally under-represented in a particular class, which could distort both training dynamics and evaluation metrics.

All images were resized to 299×299, matching the default input resolution expected by Xception. MRIs are inherently grayscale but the network expects 3-channel input, therefore images are loaded in RGB mode. Keras ImageDataGenerator was used to create minibatches for

training and evaluation. For the training generator, the following changes were applied to the images: small rotations, small zoom, brightness scaling, small translations. This is to help the model generalize better to unseen images, therefore no such changes were made to the test and validation sets. Also all images were normalized using the official Xception preprocessing function

Deep Learning - Model Architecture

My classifier used Xception as a pretrained feature extractor to produce a compact, high-level feature vector from each image. A custom classification head was then added:

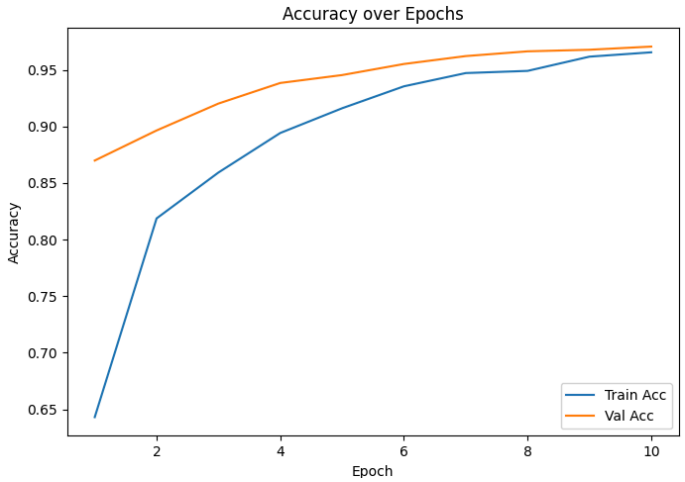
Layer (type)	Output Shape	Param #
xception (Functional)	(None, 2048)	20,861,480
flatten (Flatten)	(None, 2048)	0
dropout (Dropout)	(None, 2048)	0
dense (Dense)	(None, 128)	262,272
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 4)	516

Total params: 21,124,268 (80.58 MB)

Trainable params: 262,788 (1.00 MB)

Non-trainable params: 20,861,480 (79.58 MB)

A final softmax layer was used to output a probability distribution for the 4 classes. The model was trained using categorical cross-entropy loss which I thought would be best for 4-class one hot labels). I used an Adamax optimizer with a learning rate 1e-3 during head training and 1e-4 during fine-tuning. After training, the best model was evaluated on the held-out test set exactly once, to estimate generalization. Reported metrics include test loss, accuracy, precision, and recall.



Classical Segmentation

Classical image processing techniques were used to attempt to localize tumor regions within MRI scans after classification. An initial attempt to design a single algorithm to successfully segment tumor types was ineffective. The different classes of tumors were vastly different in appearance, including differences in size, brightness, shape, location within the brain, and contrast enhancement. As a result, a “one-size-fits-all” segmentation approach was extremely sensitive to noise and frequently failed to isolate the true tumor region. To solve this problem, segmentation was re-framed as a class-conditional task. Rather than attempting to segment tumors blindly, the system first predicts the tumor subtype using the trained classifier. Then, a designated algorithm is applied specifically to the classification.

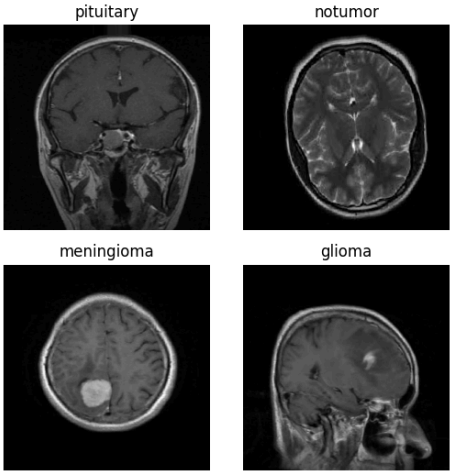
All three segmentation algorithms (glioma, meningioma, pituitary) operated on grayscale MRI images normalized to the range [0,1]. Contrast Limited Adaptive Histogram Equalization (CLAHE) was used to enhance local contrast and prevent excessive noise. Noise reduction was used to suppress high-frequency noise while preserving edges, specifically Gaussian Blurring, median and bilateral filtering. Additionally, opening and closing operations are applied to remove small spots and fill gaps in candidate regions.

For the pituitary tumor segmentation, I noted that they were typically small, bright, and centrally located near the base of the brain. This relatively consistent anatomical position allowed for a region-of-interest (ROI)–based approach. First, a few anatomically motivated ROIs were defined in normalized image coordinates, covering the central and lower brain regions. For each ROI, CLAHE-enhanced intensities were processed using a morphological filter to emphasize small, bright structures. Two thresholds were applied, a strict "seed" threshold to identify high-confidence bright pixels, and a relaxed "grow" threshold to expand the candidate region. Then connected components that intersect the seed region were retained. After that each candidate component was scored using intensity contrast, circularity, solidness, and distance from the ROI center.

When beginning meningioma segmentation it was clear that meningiomas often appeared near the brain boundary and were commonly confused with skull or dura-related artifacts. Because of this, a head/brain mask was extracted using Otsu thresholding and morphological cleanup to suppress background regions. A distance-to-boundary transform was computed from the brain mask. Two candidate regions were generated: bright regions within the brain interior and bright regions near the inner boundary of the brain mask. Then a morphological cleanup was applied to merge and refine candidate regions. Finally, connected components were scored using relating area, contrast, circularity, solidity, average distance to brain boundary, and degree of boundary contact. The component with the highest composite score is selected as the meningioma mask.

Gliomas typically have darker cores and in some cases, a light ring around the edges. First, a brain mask was applied to remove background regions. Then a low-percentile intensity threshold was applied to identify dark-core candidates. Then morphological opening and closing

remove small speckles and clarify core regions. Finally each connected component was evaluated using: relative area, contrast between inside and outside the component, evidence of a surrounding white edge, and distance from the image center. The best-scoring component was selected.



Results

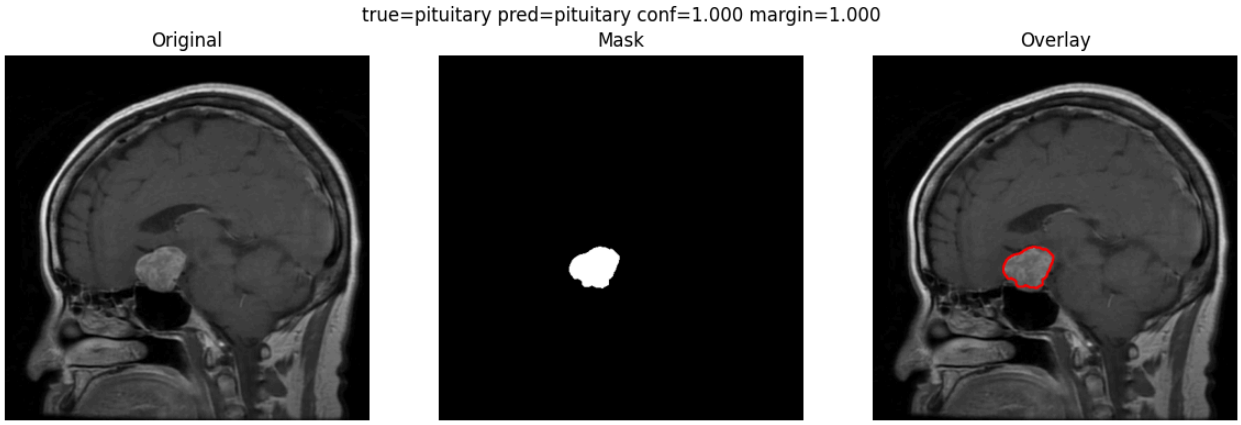
Deep Learning Classifier

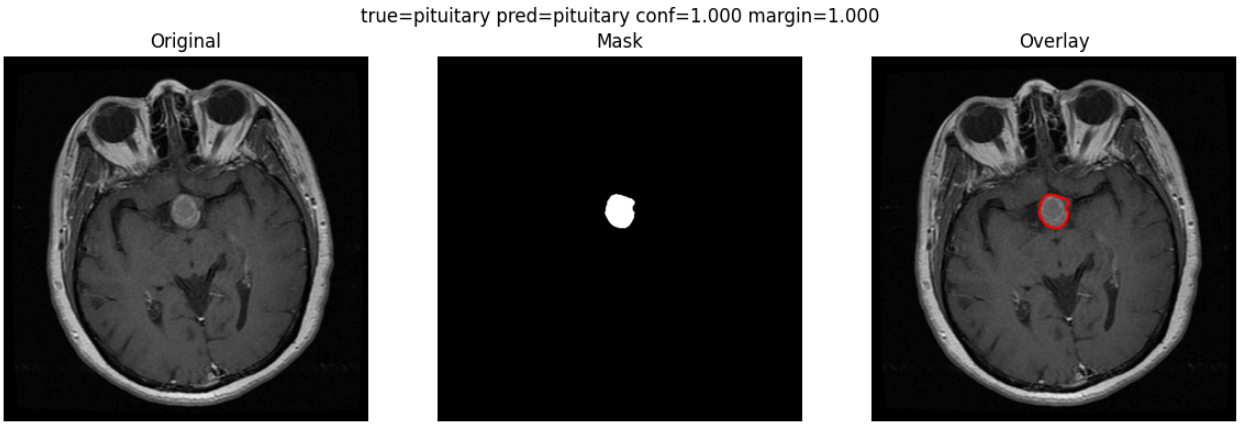
The fine tuned Xception-based classifier demonstrated strong generalization performance on the held-out test set. Final evaluation metrics are summarized below:

```
Test loss: 0.1072
Test accuracy: 0.9665
Test precision: 0.9665
Test recall: 0.9665
```

An overall accuracy of 96.65% indicates that the transfer learning approach is highly effective for multi-class brain tumor classification on this dataset. Precision and recall closely match accuracy, suggesting that performance is well balanced across classes and not dominated by a single majority class. Also, the confusion matrix shows that most errors occur between visually similar tumor subtypes, rather than between tumor and no-tumor classes.

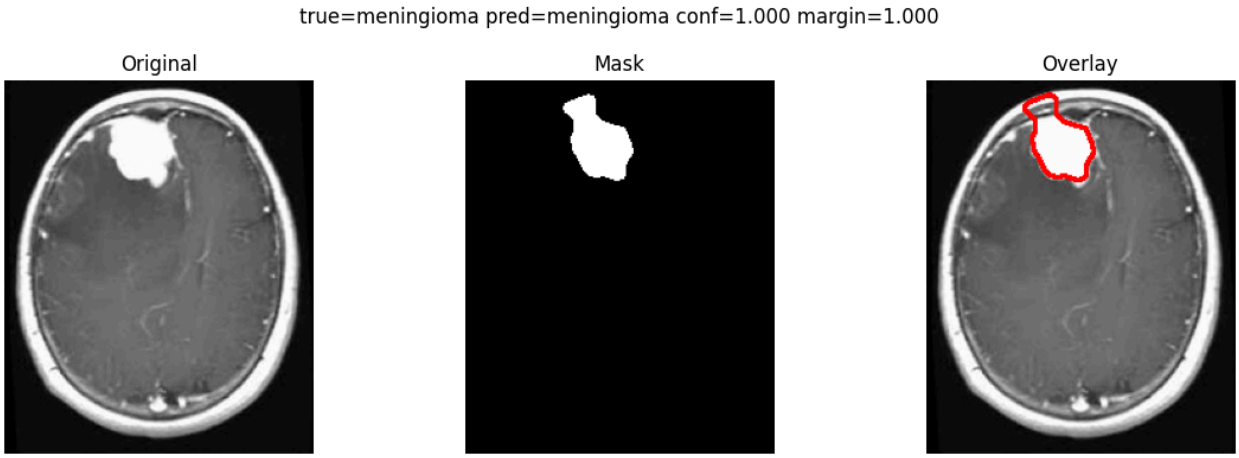
The classifier serves as the decision gate for the remainder of the system. Only images predicted as glioma, meningioma, or pituitary are passed to the corresponding classical image segmentation part. Unlike the classification stage, there's no metric to measure segmentation performance because the dataset does not provide the actual location of the tumors. Instead, segmentation performance was assessed through visual inspection of segmentation overlays.

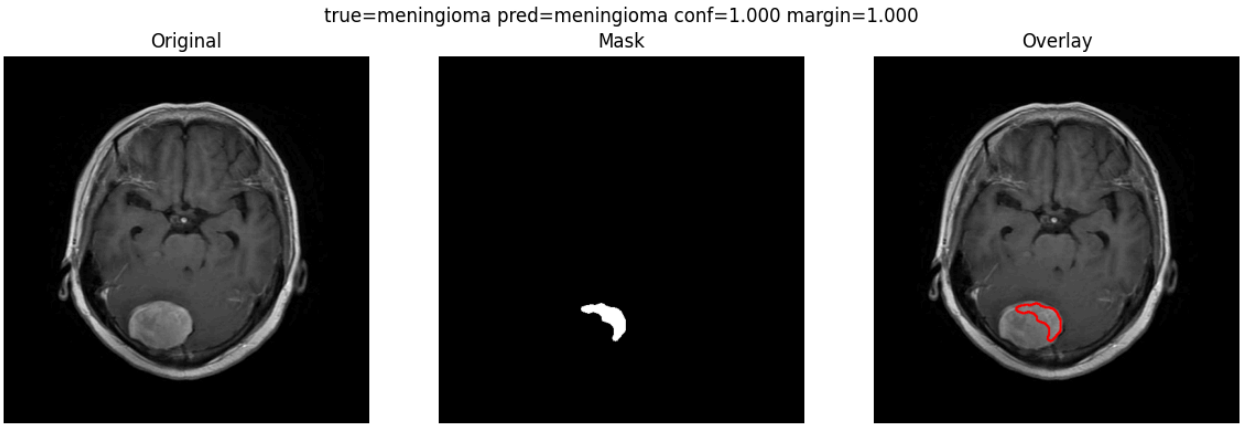




Pituitary tumor segmentation produced the most consistent and visually correct results among the three tumor types. In many cases, the algorithm successfully isolated a compact, bright region located near the center and lower portion of the brain, corresponding to the expected anatomical location of the pituitary gland. The ROI-based strategy effectively suppressed irrelevant bright structures elsewhere in the brain. This was extremely useful due to

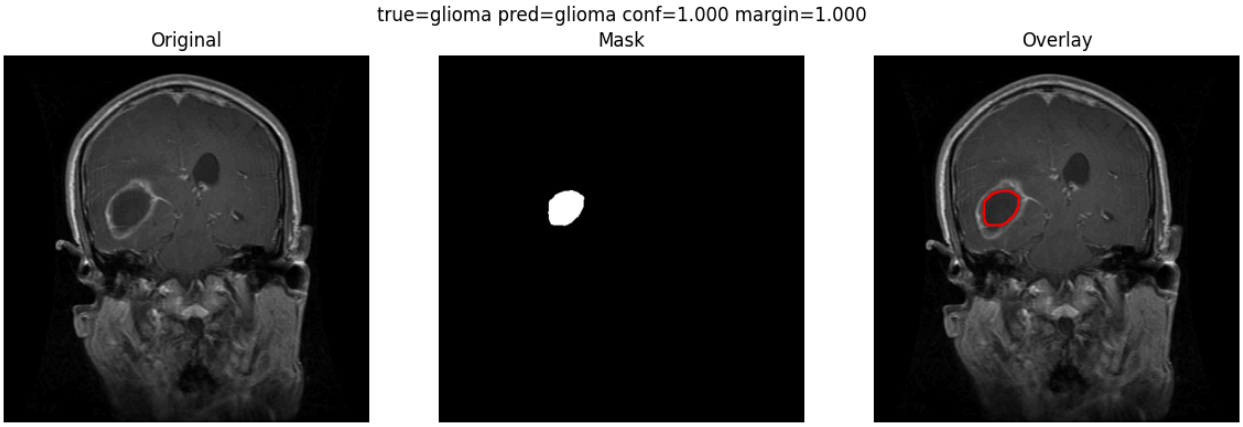
the high level of noise on the axis where the pituitary tumor resides (eyes, nostril, tissue). False positives were relatively rare when the tumor exhibited enough contrast. However, very small tumors with weak contrast were sometimes missed entirely. Overall, pituitary tumors benefited most from the incorporation of anatomical location, making them well suited for classical segmentation.

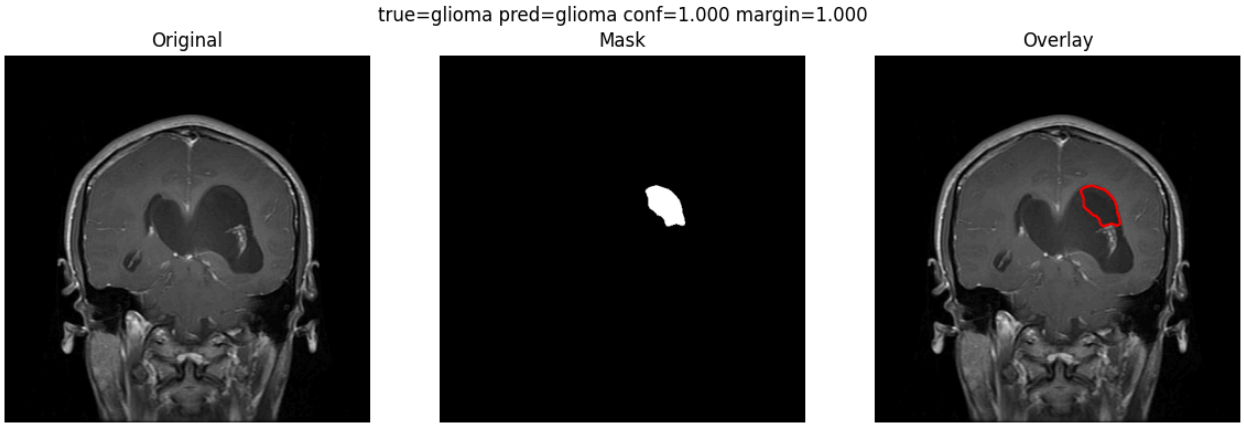




Meningioma segmentation showed mixed performance. Many meningiomas were correctly localized as bright, relatively compact regions near the brain boundary. Boundary-aware scoring helped distinguish true meningiomas from interior brain tissue while solidity and circularity constraints reduced false detections. However, while the tumor itself was often found, frequently borders were wrongly detected. The borders would either stretch past the

tumor region, or only encapsulate a fraction of it. Additionally, in low-contrast scans, tumors blended into surrounding tissue and were not clearly separable. These results show the difficulty of identifying meningiomas using only intensity based tactics.





Glioma segmentation proved to be the most challenging among the three tumor types as they often present inconsistent contrast patterns across scans. In the majority of cases, at least a part of the tumor was detected, especially when the tumor was significantly darker. However, noisy regions such as parts of the face were occasionally misidentified as dark tumor cores. Tumors with diffuse or infiltrative appearance were poorly captured. Non–ring-enhancing gliomas lacked enough boundary contrast for reliable localization. These challenges demonstrate the limitations of classical segmentation when applied to complex tumor morphology.

Conclusion & Discussion

Though this project was proposed as a comparison between classical image processing–based segmentation and deep learning–based segmentation using U-Net, during development, it became clear that the available dataset, which contains classification labels but no pixel-level tumor annotations, does not support a realistic evaluation of supervised segmentation models. As a result, the project was restructured into a pipeline that combines deep-learning-based classification with class-conditional classical segmentation for tumor localization. This revised objective was successfully achieved. A transfer-learning-based classifier using an Xception backbone reached 96.65% test accuracy, with equally high precision and recall. The classifier reliably distinguishes between glioma, meningioma, pituitary, and no tumor cases. Building on this foundation, classical segmentation techniques were applied selectively and tailored to each tumor class, producing visually plausible tumor localization overlays in many cases. The segmentation results, while qualitative, provide important insights. Tumor classes with location and appearance (pituitary tumors) were best with classical segmentation. In contrast, tumors with high morphological variability (gliomas) proved difficult to localize using intensity- and morphology-based heuristics alone. For future work, techniques such as Grad-CAM–based pseudo-labeling or classifier-guided attention maps could bridge the gap between classification and segmentation when labels are limited. In conclusion, this project demonstrates that when segmentation labels are unavailable, high-accuracy classification combined with class-specific classical localization offers a practical and interpretable alternative to pure segmentation models.

References

- **Kaggle Brain Tumor Dataset**
- **Xception Model**
- **Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing 3e**