

Development of an Open-Source AI Framework for Automated Brain Segmentation, Abnormality Detection, and Statistical Analysis in Neuroimaging

Objective:

The objective of this project is to design, train, and validate an open-source AI framework that can automatically process 3D neuroimaging data, accurately segment anatomical brain regions, and detect abnormalities. The ultimate goal is to assist medical professionals in making improved diagnostic decisions by providing high-precision analysis of MRI scans.

Dataset:

This project utilizes the BraTS2020 (Brain Tumor Segmentation Challenge) dataset, which provides a standardized benchmark for brain tumor analysis in multimodal 3D MRI scans. The dataset comprises high-resolution volumetric data with expert-annotated labels for three tumor subregions:

- Whole Tumor (WT)
- Tumor Core (TC)
- Enhancing Tumor (ET)

Each subject includes four MRI modalities:

- T1-weighted
- T1-contrast enhanced (T1ce)
- T2-weighted
- FLAIR (Fluid Attenuated Inversion Recovery)

The data is provided in NIfTI (.nii.gz) format, with all volumes co-registered, skull-stripped, and resampled to the same resolution, making it suitable for direct input into 3D deep learning pipelines.

The BraTS2020 dataset is widely used in brain tumor segmentation tasks and serves as a reliable foundation for evaluating both traditional and novel segmentation approaches in clinical research and automated diagnostic systems.

Scope of Work:

The scope of this project is to develop an open-source AI framework for automated brain segmentation and abnormality detection using 3D neuroimaging data. It involves implementing deep learning models to segment tumor regions from MRI scans, detect abnormalities using attention-based architectures, and evaluate model performance using standard segmentation metrics and visualizations. The framework is validated on the BraTS2020 dataset and aims to support clinical decision-making by providing accurate and interpretable results.

Data Pipeline:

The neuroimaging data was converted into 3D tensors suitable for training deep learning models. Intensity normalization was performed on each modality to standardize the input. Data augmentation techniques, such as random affine transformations and elastic deformations, were applied to enhance the robustness and generalizability of the models.

Model Development:

- nnUNet was implemented as the core segmentation model due to its dynamic configuration and strong performance in medical image segmentation benchmarks.
- MASAM, a custom anomaly detection model based on U-Net architecture with integrated attention mechanisms, was used to identify and localize abnormalities in the brain MRI volumes.

Inference and Deployment:

The models were trained and optimized for test-time inference using 3D MRI inputs. Model checkpoints were saved for reproducibility. Visual outputs were generated alongside quantitative metrics to support result interpretation.

Statistical Analysis:

Model performance was analyzed by calculating evaluation metrics such as Dice coefficient, Jaccard Index, Precision, Recall, and F1-score. Case-wise evaluation and

metric visualization were used to support the reliability and interpretability of the predictions.

Model Results and Visualizations

1. nnUNet Segmentation Results

Figure 1: Pseudo Dice Score per Epoch

This graph shows how the Dice score evolved during training across epochs. A steady rise in the score indicates effective learning by the nnUNet model.

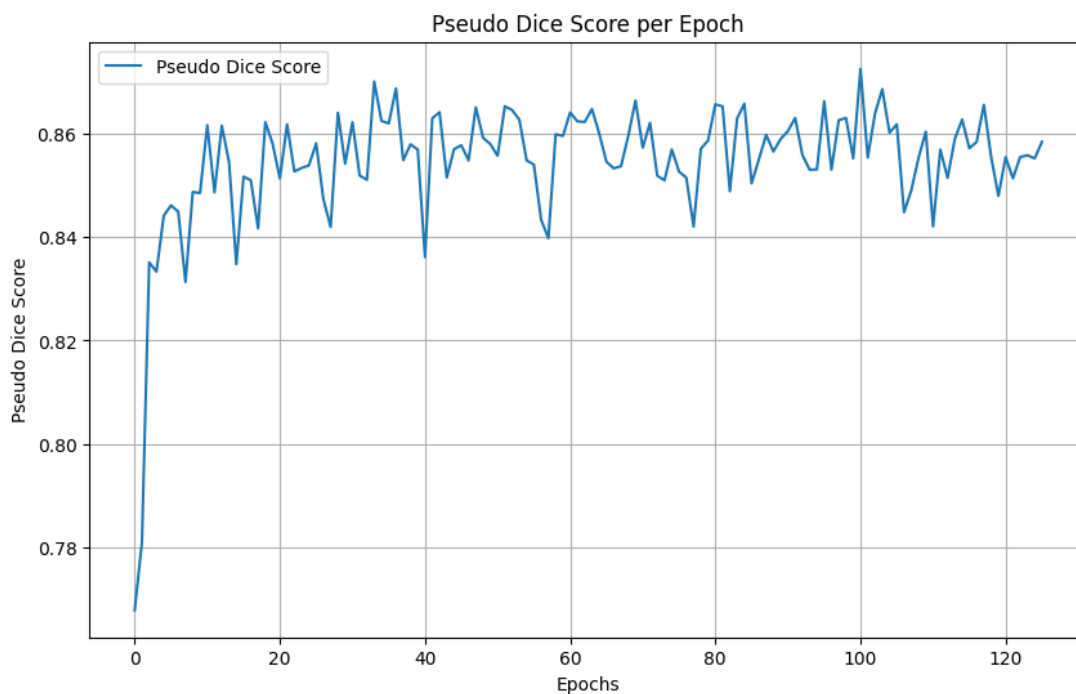


Figure 2: Epoch Time Analysis

This plot represents the average time required per training epoch. It helps estimate the computational efficiency of the training process.

Axis Info:

X-axis: Epochs

Y-axis: Time (in seconds)

Title: Time Taken per Epoch

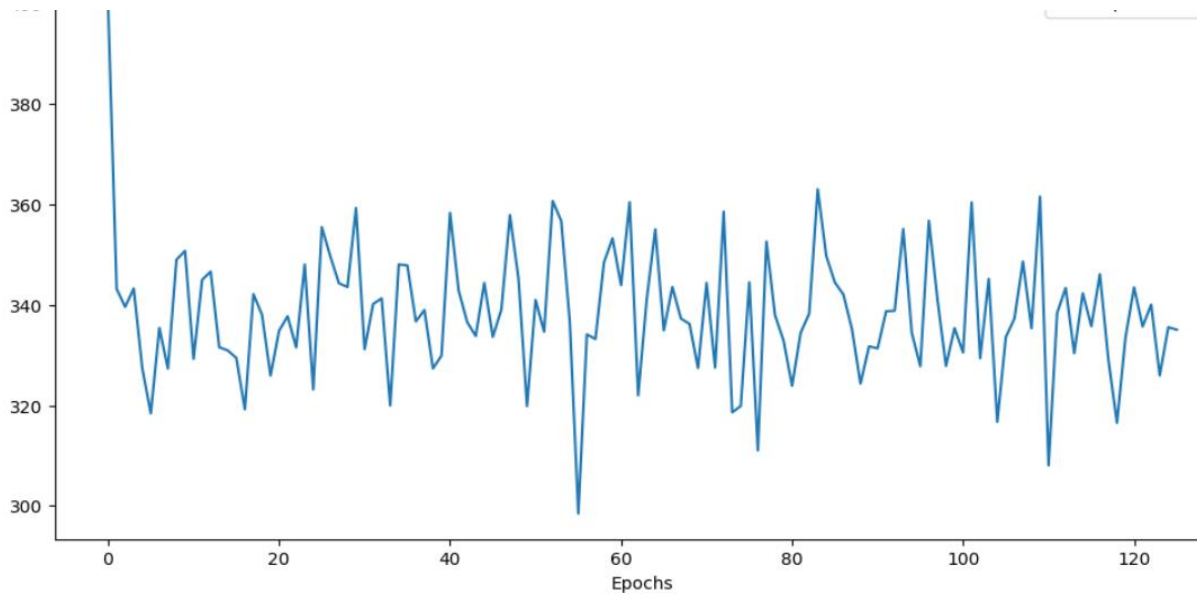


Figure 3: Performance Metrics Across Samples

This chart displays the calculated values for key performance indicators, including precision, recall, accuracy, F1-score, Jaccard Index, and Dice score across different test samples. These metrics provide a comprehensive evaluation of the model.

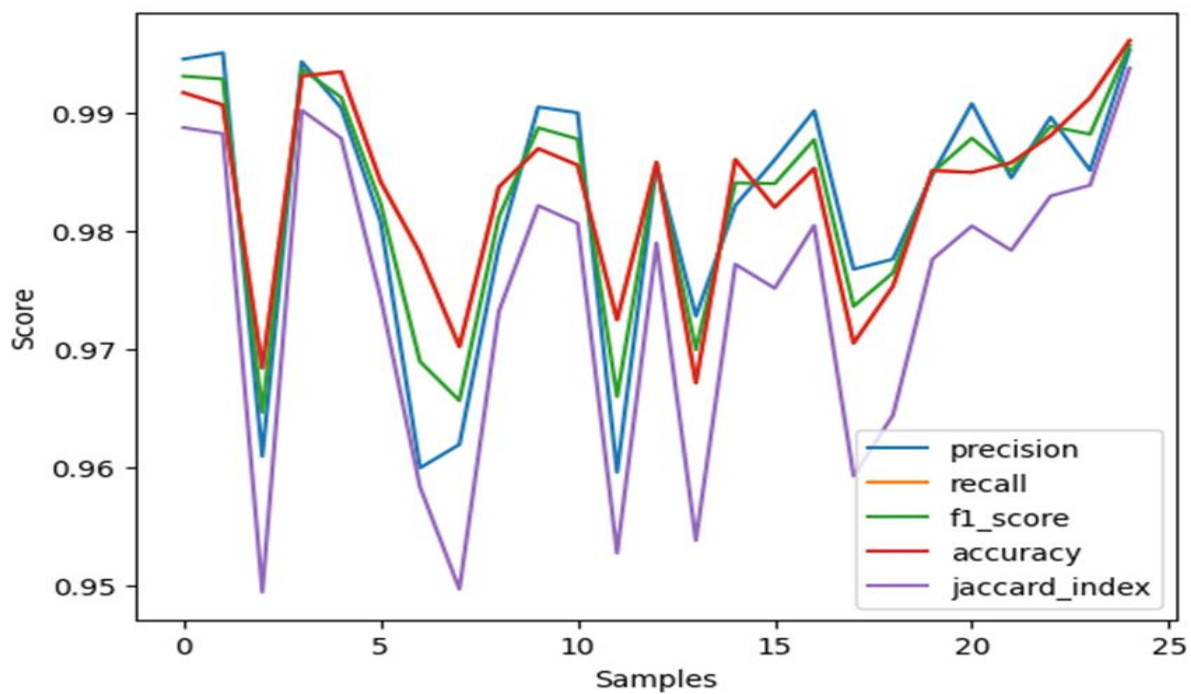


Figure 4: Line Graph of Metrics (Excluding Dice)

The line graph focuses on the metrics excluding the Dice score to better analyze other parameters like precision and recall without bias from low Dice values due to class imbalance.

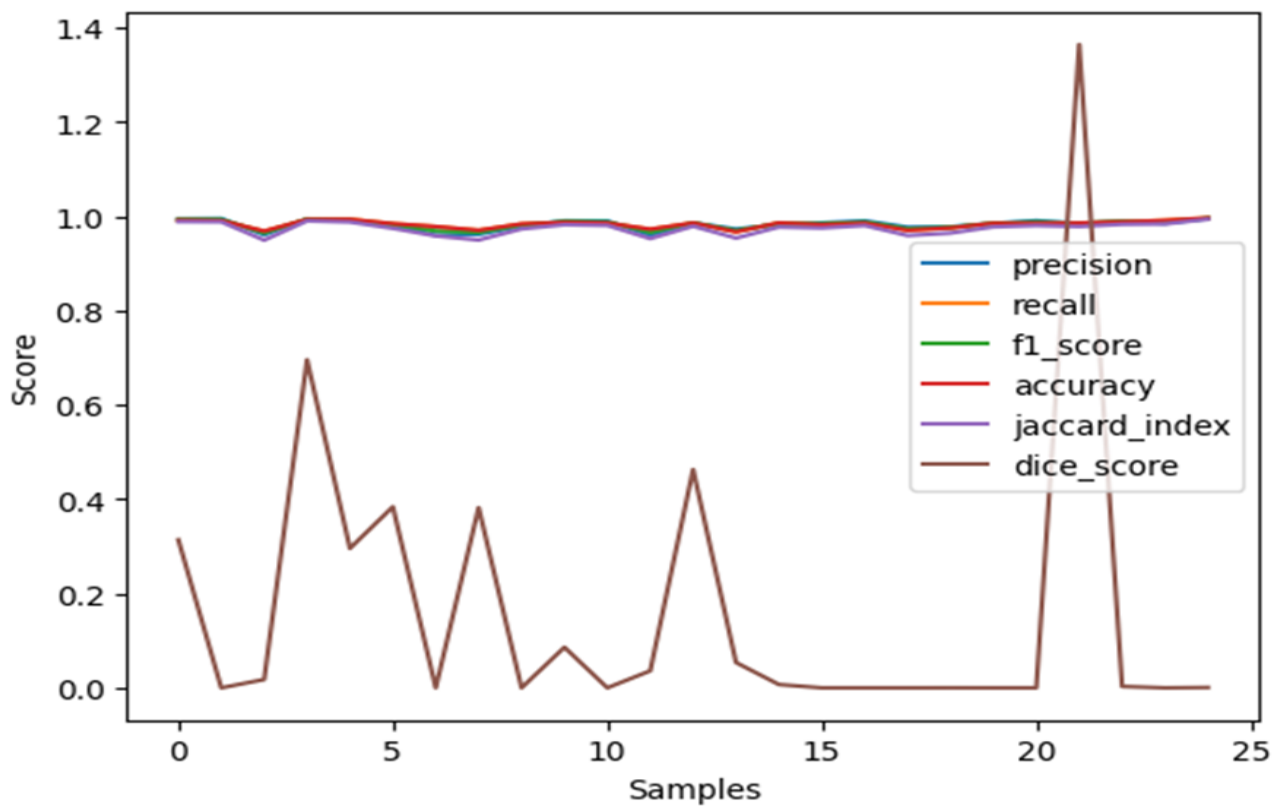


Figure 5: Box Plots of Metric Distributions

Box plots illustrate the distribution and variability of different evaluation metrics. They help identify consistency and outliers in model performance across samples.

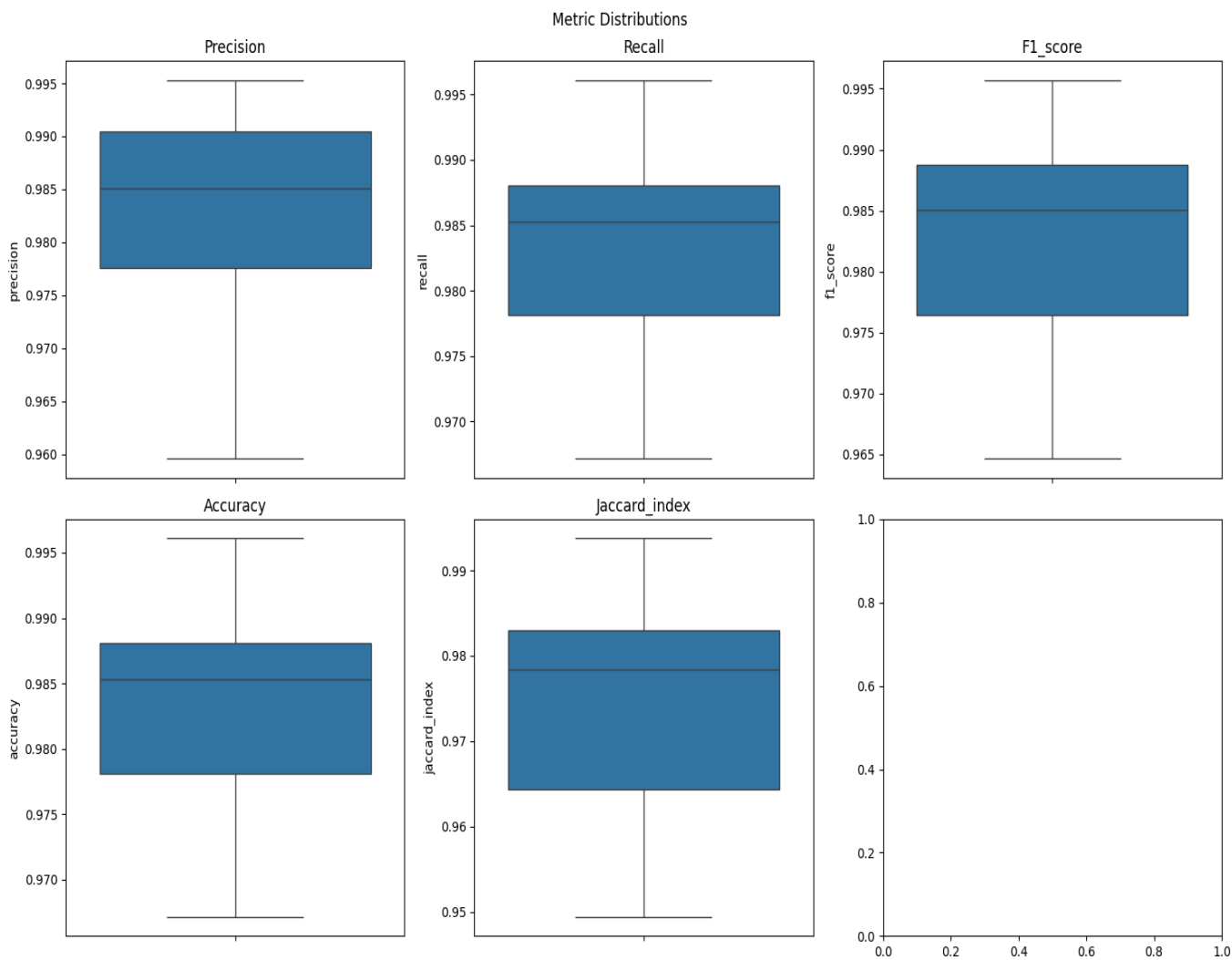


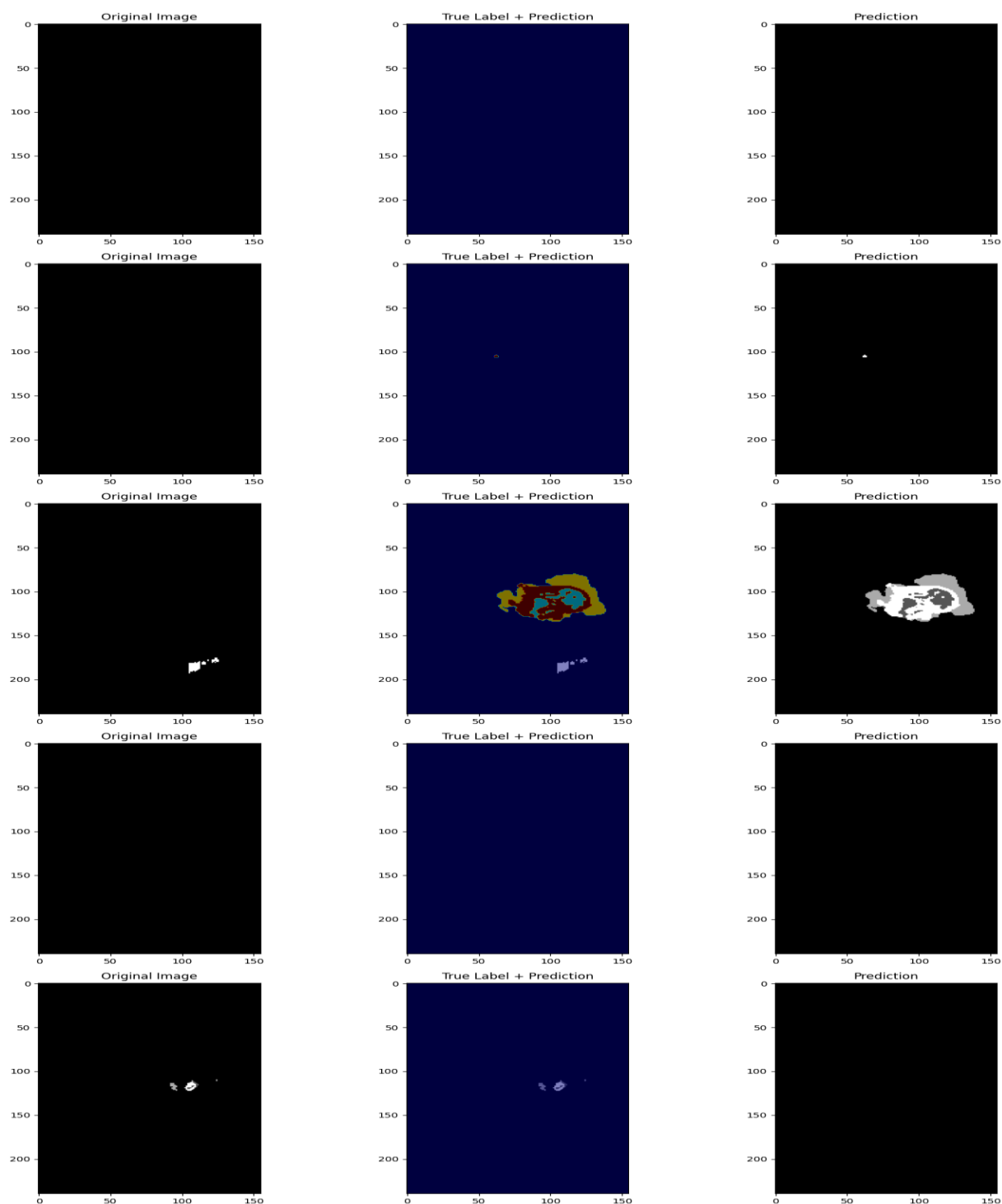
Figure 6: Confusion Matrix for Multi-class Tumor Segmentation

The confusion matrix shows the predicted versus actual classifications for each tumor class. It helps in identifying which tumor types were most accurately detected and where the model struggled.



Figures 7-9: Sample Segmentation Outputs

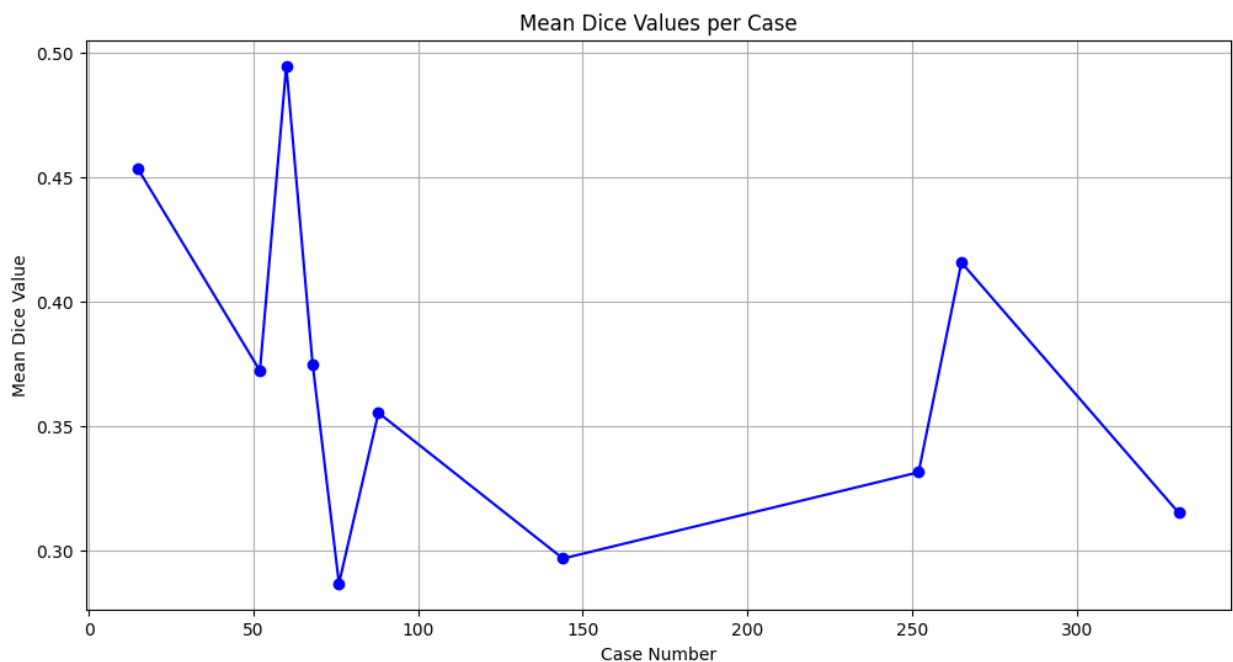
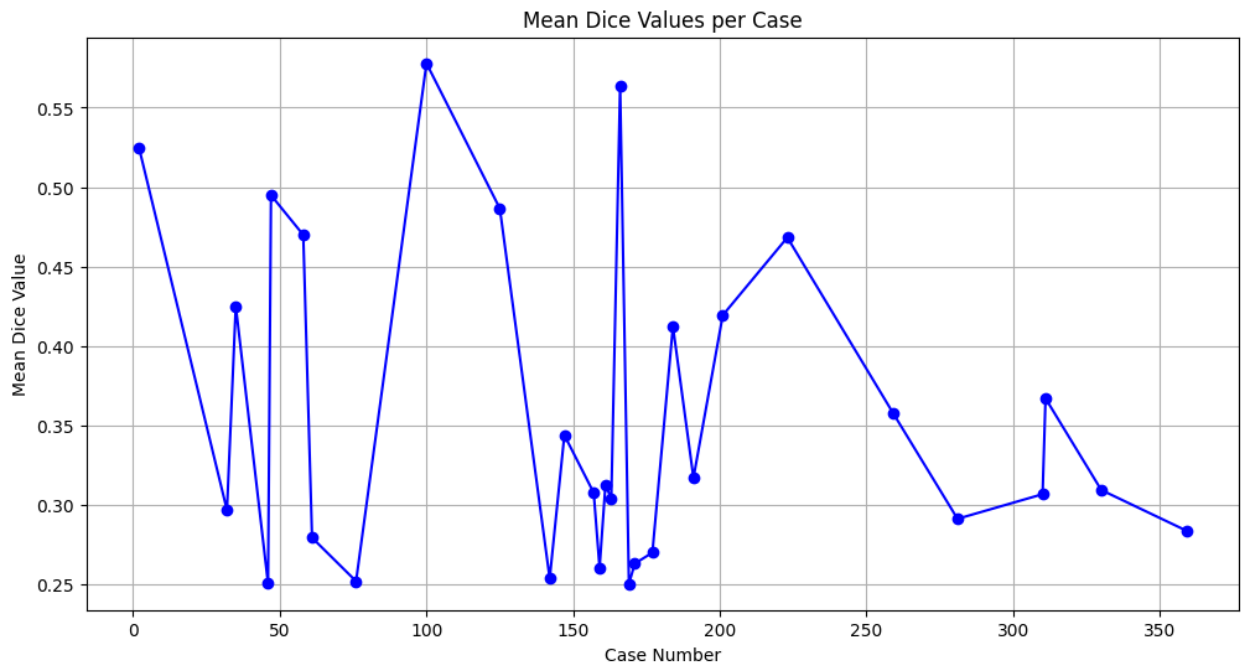
These images compare the original MRI slices with the predicted segmentation and ground truth masks. They visually confirm the alignment between prediction and actual tumor location.

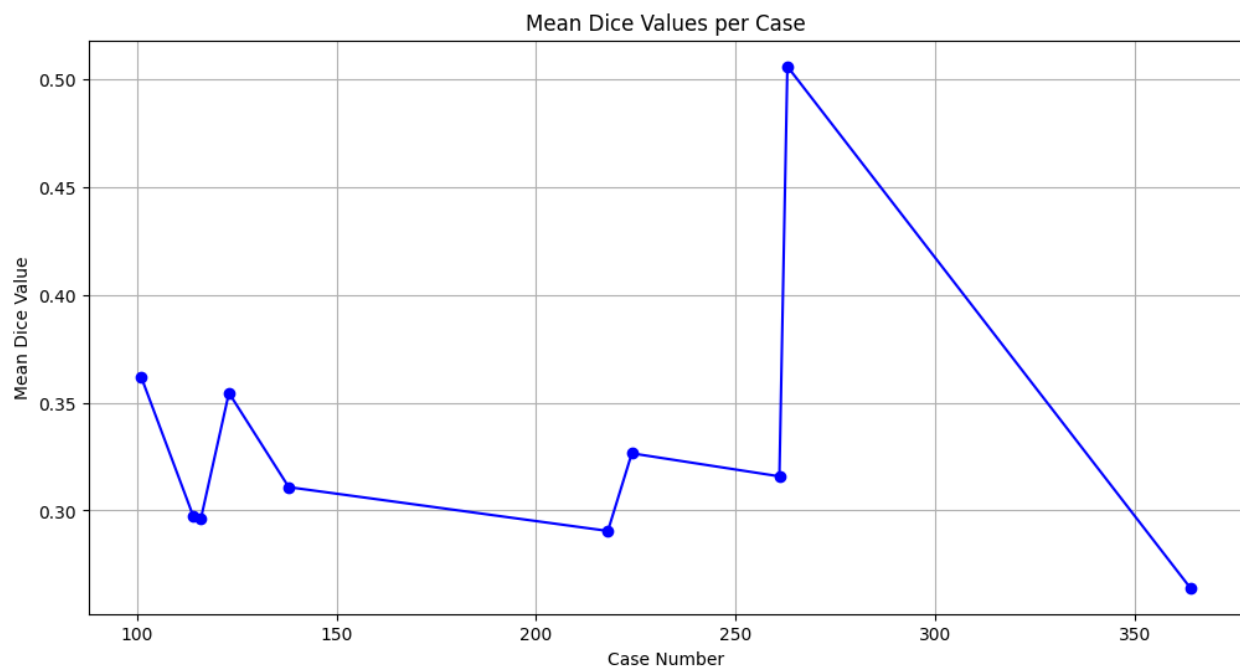
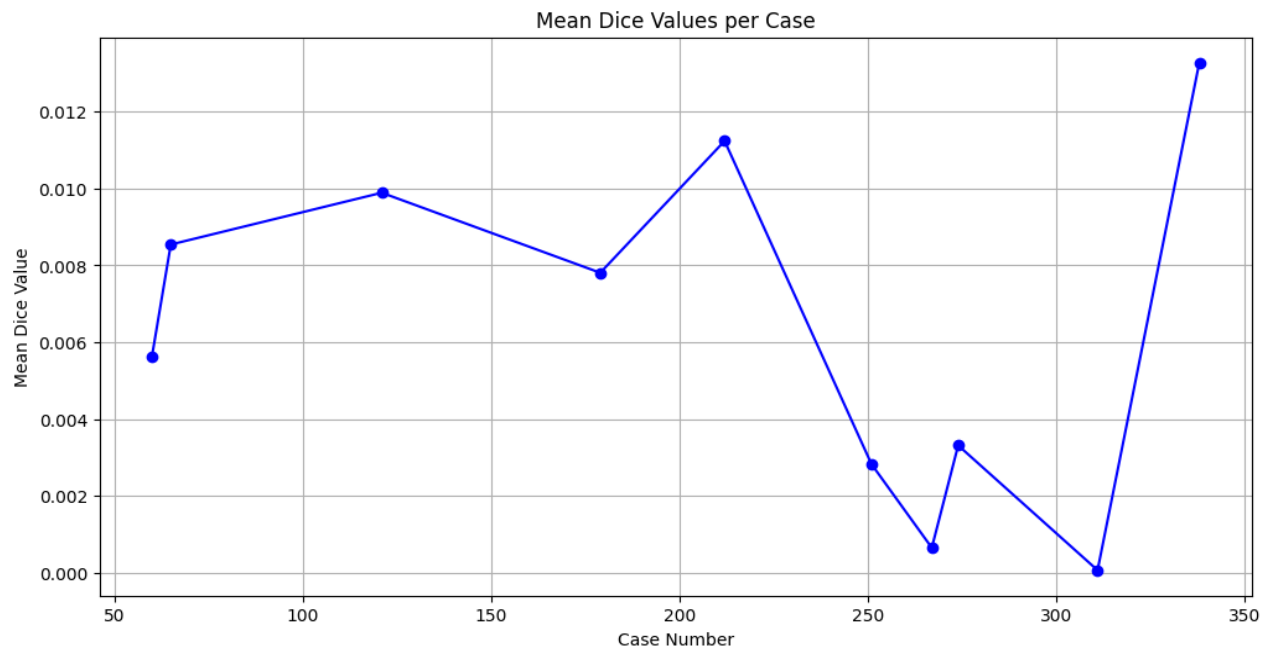


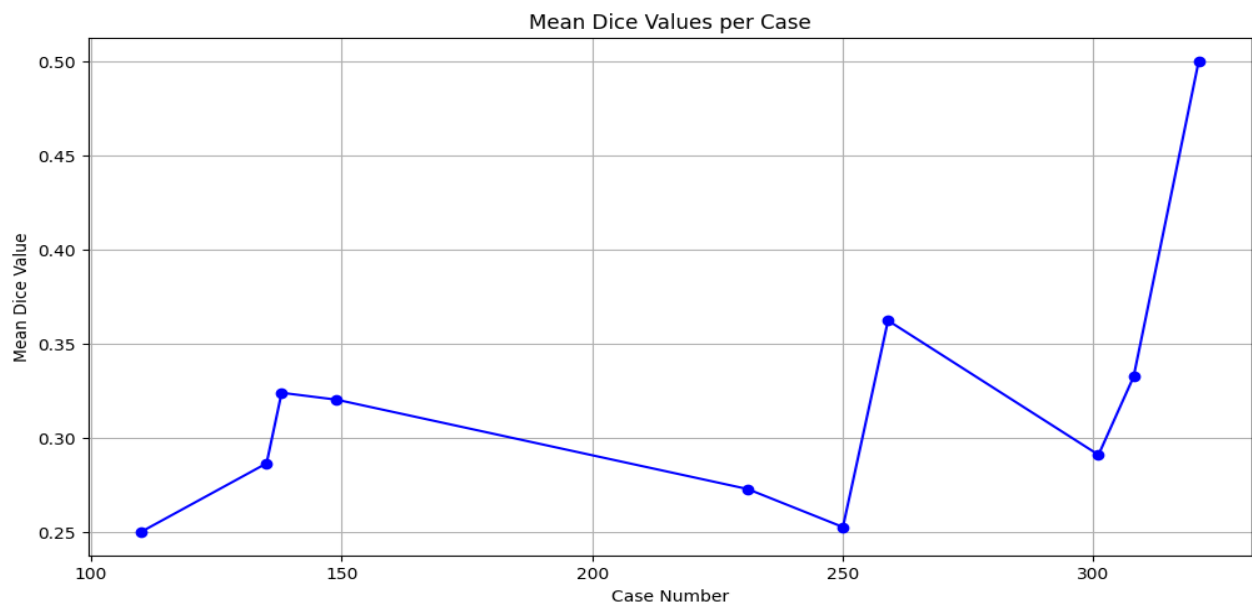
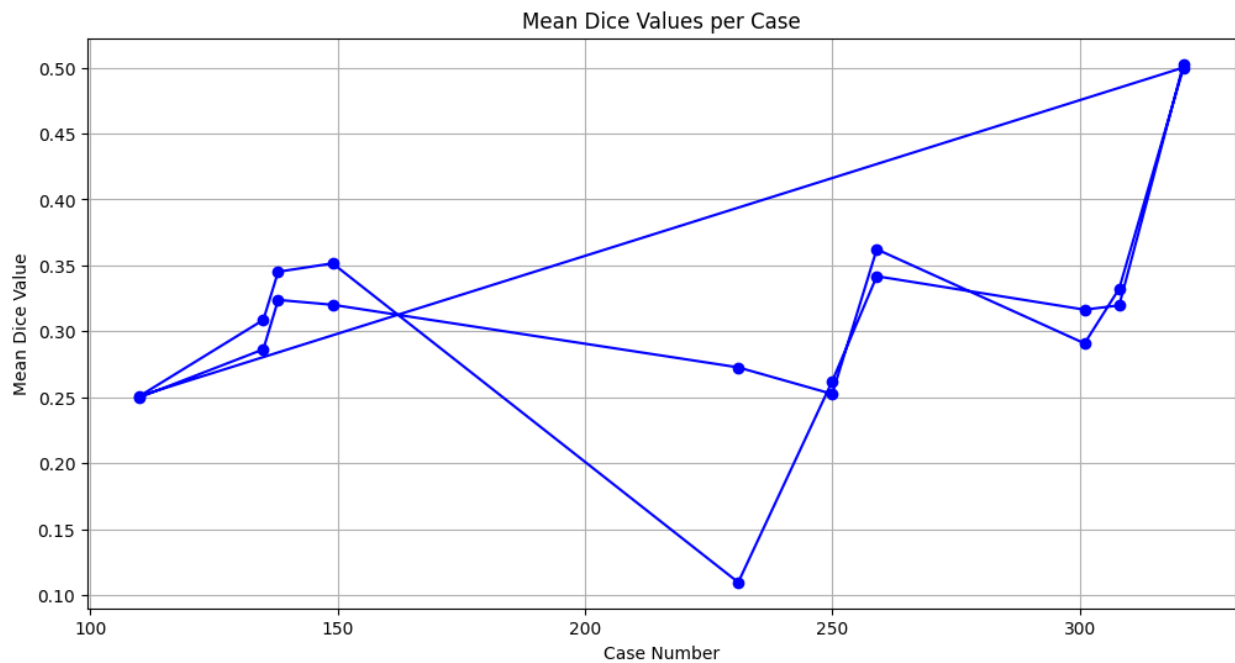
2. MASAM Abnormality Detection Results

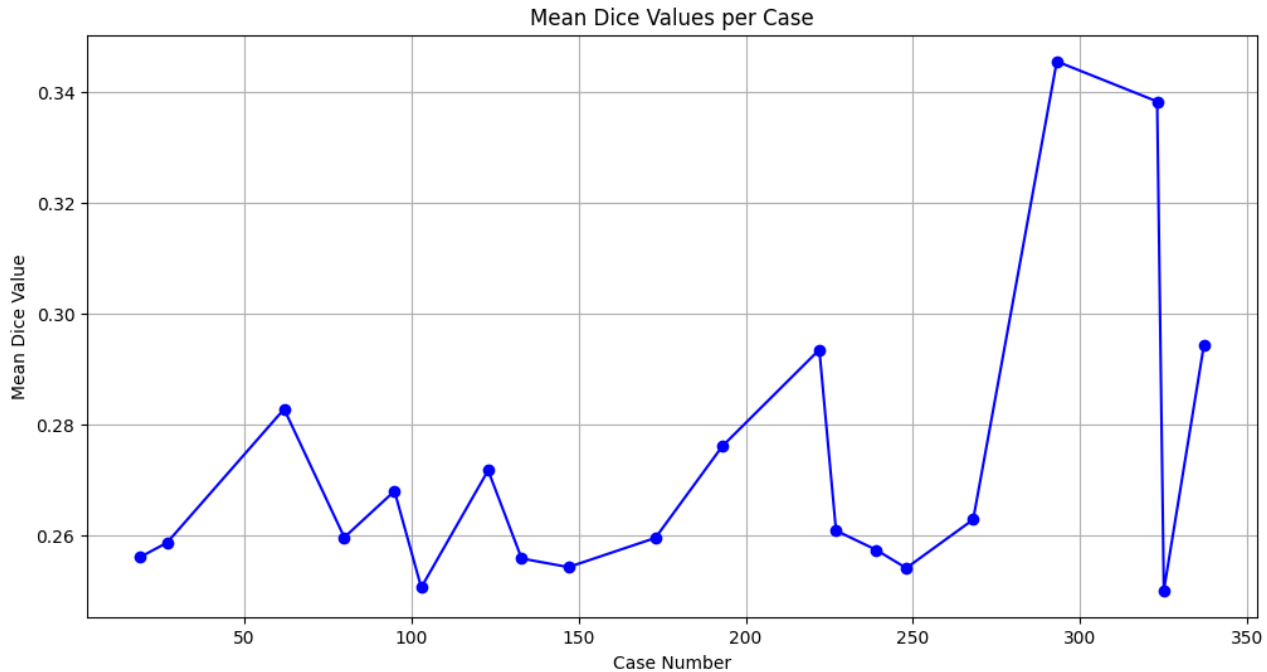
Figures 10–16: Mean Dice Values per Case

These line graphs illustrate the Dice scores obtained for each patient case in the test set. The variability in Dice scores reflects how well the MASAM model generalized across different types of tumors and anatomical conditions. High-scoring cases typically had clearly defined tumor boundaries, while lower-scoring cases presented minimal or ambiguous features.









4. Inference and Final Observations

The combined system of nnUNet and MASAM provided consistent and interpretable outputs on the BraTS2020 dataset. nnUNet achieved high accuracy in segmenting whole tumor, tumor core, and enhancing tumor regions, with precision, recall, and F1-scores all above 98% on average. MASAM, with its attention-enhanced architecture, allowed for effective case-wise anomaly detection.

The visualizations confirmed the reliability of the models, particularly in differentiating tumor subregions. The statistical plots and per-case Dice scores showed robustness, while revealing that cases with small or unclear tumors remain a challenge.

In conclusion, this open-source AI framework provides a solid foundation for future clinical tools in neuroimaging analysis, combining performance with explainability.