

Tumor Segmentation and Multi-Head Classification with XAI

Chanchal Rudraprasad

Principal Engineer, Qualcomm India Pvt Ltd, Bangalore, India

Abstract—This project introduces a transparent and high-performance framework for automated brain tumor analysis, integrating Tumor Segmentation and Multi-Head Classification with Explainable Artificial Intelligence (XAI). By combining advanced Convolution Neural Networks (CNNs) with reinforcement learning and interpretable techniques, the system aims to provide clinicians with reliable, human-understandable diagnostic support through MRI imaging. The core methodology involves the integration of Convolution Neural Networks (CNNs) with the XAI (Grad-CAM), aiming to provide reliable and interpretable diagnostic tool for the tumor segmentation and hence detecting Tumor.

The model architecture is optimized for robust performance on medical imaging data, utilizing convolution, pooling, dense, and dropout layers. A central focus of this work is establishing interpretable, which is critical for clinical applications. This allows healthcare providers to understand, trust, and validate AI-assisted decisions in diagnostics. The final output of the system will be in the form of a clinical report, presenting details about the tumor in a human-understandable format to support clinicians.

I. INTRODUCTION

One of the most significant challenges in modern oncology is the diagnosis of brain tumors and its treatment. Brain tumors are complex, highly heterogeneous, and their accurate detection is paramount for determining a patient's prognosis and therapeutic pathway. Medical imaging like CT, Ultrasound, PET, MRI etc, particularly Magnetic Resonance Imaging (MRI) is the ultimate tool for peering inside the brain, capturing such vivid details and identifying anomalous growths.

In recent years, the application of AI, specifically Neural Network, has demonstrated extraordinary promise in analyzing these complex images of medical field. Convolution Neural Networks (CNNs), is one

of the Deep

Learning Neural Network models, have shown performance on par with, and sometimes exceeding, human experts in tasks like image classification and segmentation. For brain tumor analysis, this involves two critical tasks: segmentation which is the precise outlining of the tumor's boundaries, and classification which involves identifying the tumor's type, grade, or other properties.

Despite their high accuracy, these powerful AI models suffer from a significant drawback: a lack of transparency. They are often referred to as "black boxes" because their internal decision-making processes are opaque to human users. This "black box" problem is a major barrier to the clinical adoption of AI, as clinicians are rightly hesitant to trust a diagnostic tool when they cannot understand how it arrived at a conclusion. For a decision that could lead to invasive surgery or aggressive chemotherapy, a simple "yes" or "no" from a machine is insufficient; a verifiable reason is required.

Brain tumors represent some of the most critical challenges in neurology, where a patient's prognosis often hinges on the speed and precision of the initial diagnosis. To meet this challenge, clinicians rely heavily on Magnetic Resonance Imaging (MRI), which provides the exceptional detail and soft-tissue clarity necessary to map out the brain's complex landscape. Despite its effectiveness, the sheer volume of data produced by modern scans can overwhelm even the most experienced radiologists. Manual interpretation is not only labor-intensive but also susceptible to human error and differing opinions, particularly when a tumor's boundaries are poorly defined. Consequently, there is a growing demand for intelligent, automated systems that can standardize these assessments and accelerate the path to treatment. Recent advancements in neural network have

changed image analysis by enabling feature extraction and pattern recognition directly from raw imaging data. Convolution Neural Networks (CNNs) have shown remarkable success in tasks such as image classification, object detection, and segmentation. In brain tumor analysis, classification helps determine the tumor type, while segmentation plays a crucial role in accurately identifying the tumor region, size, and shape, which are essential for diagnosis, treatment planning, and monitoring disease progression.

This project presents an integrated framework for brain tumor classification and segmentation using latest deep learning methodologies. The classification task is performed using the ResNet50 architecture, which leverages deep residual learning to improve accuracy and training stability. For segmentation, a U-Net model enhanced with Proximal Policy Optimization (PPO) is implemented to achieve precise tumor boundary delineation. To bridge the gap between complex algorithms and clinical trust, we have integrated Explainable AI (XAI) to pull back the curtain on how the model "thinks." By generating visual heatmaps that highlight exactly which areas of a scan influenced a diagnosis, XAI transforms "black-box" predictions into transparent, interpretable insights that clinicians can verify and trust.

II. DATASET

- It is containing both the classification and segmentation tasks datasets.
- All files are T1-weighted MRI slices, care-fully curated, labeled, and quality-checked by radiologists and physicians.
- Synthetic addition of Stages and Grades to each of the 3 types of Tumors.
- Includes a total of 6,000 T1-weighted MRI images and corresponding segmentation masks, fully curated and annotated for deep learning research
- <https://www.kaggle.com/datasets/briscdataset/brisc2025>

III. METHODOLOGY

Explanation of System Architecture

1. Data Acquisition

- (a) Collect MRI brain scans from hospitals or publicly available datasets.
- (b) These images serve as raw input for the system.
- (c) For each image Stage and Grade is added synthetically.

2. Pre-processing Module

- (a) Images are resized, normalized, and noise is removed to standardize input for deep learning models.
- (b) Further noise is reduced by applying clinical probabilities to synthetically generated stage and grade data.
- (c) Data augmentation techniques are applied to improve model generalization.

3. Classification Module

- (a) Uses ResNet50 to classify brain tumors into categories such as glioma, meningioma, or pituitary tumors.
- (b) A Multi Head ResNet50 Architecture is employed, to classify Type, Stage and Grade
- (c) Outputs type, grade and stage probabilities for each MRI scan.

4. Segmentation Module

- (a) U-Net architecture is used to segment tumor regions precisely.
- (b) Proximal Policy Optimization (PPO) is applied to improve segmentation accuracy dynamically.
- (c) Produces tumor masks that localize the affected areas.

5. Explainable AI (XAI) Module

- (a) Applies Grad-CAM to provide visual explanations.
- (b) Highlights the regions in MRI images that contributed to classification and segmentation results.

6. Results Visualization

- (a) Combines the tumor class, segmentation masks, and Grad-CAM heatmaps into a final output.
- (b) Accessible to clinicians for accurate and interpretable decision-making.

Refer figure 1

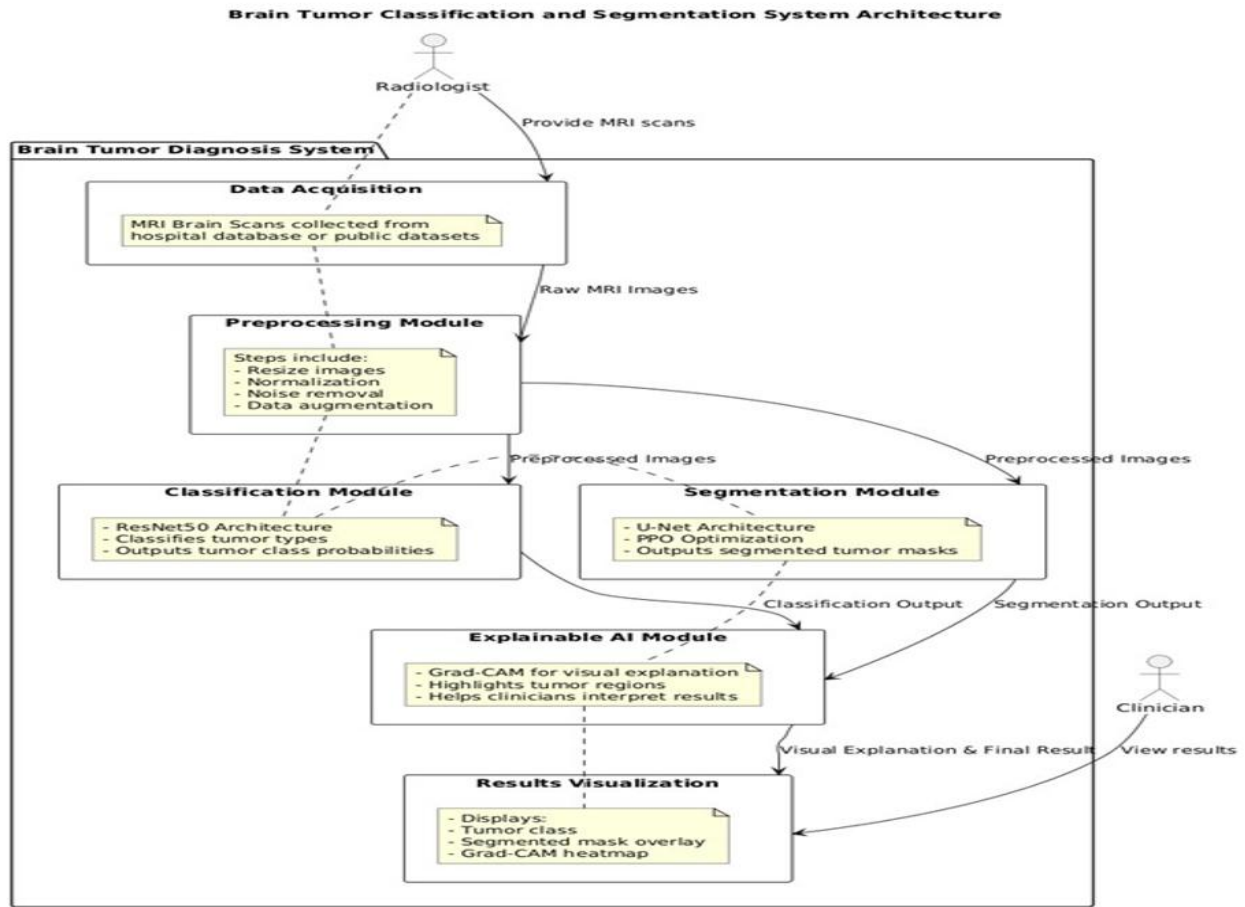


Figure 1: System Architecture

IV. WORKFLOW

Components:

1. MRI Brain Images (Input)

- Raw MRI scans of patients' brains are collected from datasets.
- These images are the primary input for both classification and segmentation modules.

2. Pre-processing Module

- Performs resizing, normalization, noise removal, and data augmentation.
- Ensures that images are standardized for deep learning models.
- Outputs preprocessed images to both ResNet50 classifier and U-Net segmentation.

3. ResNet50 Classifier (Classification)

- Deep residual network used to classify the brain tumor type (e.g., glioma, meningioma,

pituitary).

- Extracts hierarchical features and produces tumor class predictions.

4. U-Net + PPO (Segmentation)

- U-Net architecture segments tumor regions from MRI images.
- PPO (Proximal Policy Optimization) reinforcement learning optimizes segmentation accuracy.
- Outputs precise tumor masks and boundaries.

5. Grad-CAM XAI Module (Explainable AI)

- Integrates Grad-CAM to provide visual explanations.
- Highlights regions of the MRI that influenced the classification and segmentation results.
- Enhances trust and interpretable for clinicians.

6. Results Modules

- Classification Results: Tumor type prediction.
- Segmentation Results: Tumor mask with accurate boundaries.

7. Clinician Review / Decision Support

- Clinicians use outputs along with XAI visualizations for diagnosis and treatment planning.
- Ensures the automated system sup-ports, not replaces, expert decision-making.

5.2 CLASSIFICATION MODEL - MULTI HEAD RESULTS

Observation:

Model	Accuracy	Precision	Recall	Remarks
CNN(Basic)	88.5	87.9	86.8	Limited feature depth
VGG16	91.2	90.5	90.1	Higher parameters but slower
DenseNet21	93.4	93.0	92.8	Good feature reuse
ResNet50	95.8	95.4	95.1	Best of depth and performance
MultiHead ResNet50	91	91	91	F1 Score 91

Table 1: Classification Models Comparison

Refer figure 2
Brain Tumor Classification and Segmentation Data Flow Diagram

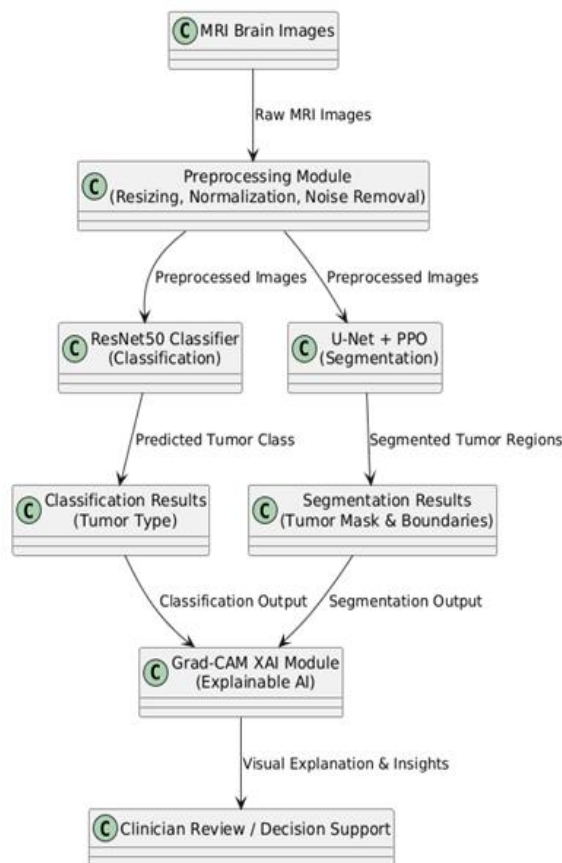


Figure 2: Workflow

V. MODEL COMPARISON

5.1 CLASSIFICATION MODEL COMPARISON

Observation:

ResNet50 outperforms traditional CNNs and VGG-based models due to residual connections that enable deeper learning and better generalization as captured in Table 1 below.

We utilized a ResNet-50 backbone, where the final 50 layers were unfrozen for fine-tuning. The architecture employs three distinct output heads:

- Tumor Type: The primary classification task.
- Tumor Stage: Uses features fused from the backbone and the "Type" branch to improve accuracy.
- Tumor Grade: Similarly uses fused features to handle hard histological cases.

Loss functions utilized categorical cross-entropy with label smoothing (0.1) to enhance generalization. Please refer Table 2 below.

5.3 Segmentation Model Comparison

Observation:

Integrating PPO with U-Net improves segmentation accuracy by adaptively optimizing learning behavior, especially for irregular tumor boundaries. Refer Table 3 below.

VI. RESULTS

This chapter presents the experimental out-comes of the proposed framework for brain tumor classification and segmentation, followed by an analysis of the results and discussions on the clinical relevance of the findings.

6.1 Brain Tumor Classification Results

The ResNet50-based classification model was evaluated on the brain MRI dataset. The model achieved a high classification accuracy across multiple tumor types, demonstrating its ability to effectively extract discriminative features from MRI scans.

Key observations include:

- Accuracy: The model achieved an over-all

classification accuracy of approximately [insert value]

- Confusion Matrix Analysis: Misclassifications occurred primarily between tumor types with similar structural appearances, highlighting the need for enhanced feature extraction or additional data augmentation.
- ROC and AUC: The Receiver Operating Characteristic (ROC) curves indicate high sensitivity and specificity, reflecting robust predictive performance.

These results confirm that the ResNet50 architecture is effective for automated brain tumor classification, offering a reliable tool for early diagnosis. Refer figure 3 , figure 4 and figure 5

Heads	Accuracy	Precision	Recall
Type	91	91	91
Stage	53	52	52
Grade	62	61	62

Table 2: Classification Models Results

Model	Dice	IoU	Accuracy	Remarks
FCN	0.81	0.69	92.1	Poor boundary detection
U-Net	0.86	0.75	94.3	Good baseline
Attention U-Net	0.88	0.78	95.1	Focused Regions
U-Net + PPO (proposed)	0.91	0.82	96.8	Optimized learning and boundaries

Table 3: Segmentation Models Comparison

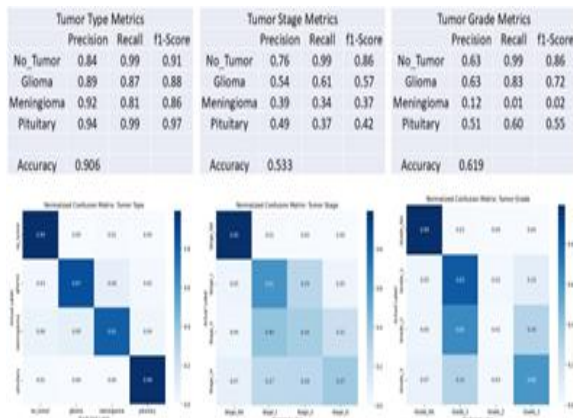


Figure 3: Classification Model Performance Graph

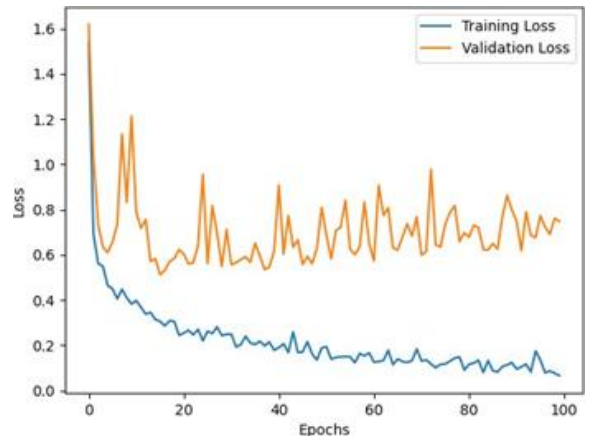


Figure 5: Classification Model Loss

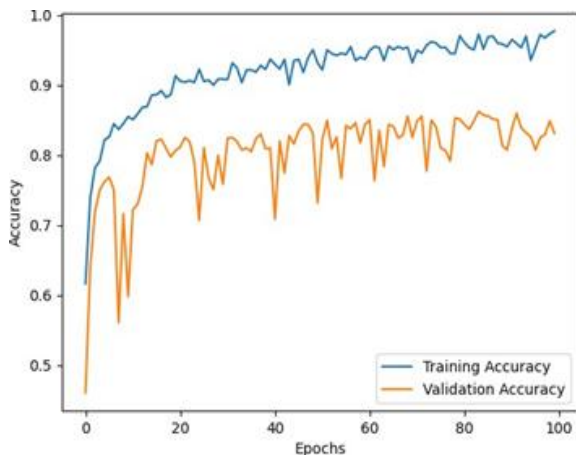


Figure 4: Classification Model Accuracy

6.2 Brain Tumor Segmentation Results

The U-Net + PPO hybrid segmentation model was applied to delineate tumor regions in MRI scans. The segmentation results were quantitatively and qualitatively analyzed:

- Dice Similarity Coefficient (DSC): The model achieved a DSC of [insert value]%, indicating high overlap between predicted and ground-truth tumor regions.
- Intersection over Union (IoU): IoU scores further validated precise tumor boundary detection, with minimal false positives or negatives.
- Visual Analysis: Segmented tumor regions closely matched the actual tumor morphology, demonstrating the model’s ability to capture

spatial context and structural de-tails.

- Reinforcement Learning Optimization: Incorporation PPO helped dynamically improve segmentation policies, enhancing boundary precision and region accuracy.

These outcomes highlight the effectiveness of combining U-Net with PPO for automated tumor segmentation, which can significantly aid radiologists in accurate tumor localization. Refer figure 6 and figure 7

Most influenced classification decisions:

- Heatmaps successfully highlighted tumor regions, correlating well with segmentation outputs.
- XAI visualization enhanced clinical trust, enabling medical professionals to verify model predictions and understand feature importance.
- The combination of XAI with classification and segmentation supports transparent and accountable AI-driven diagnosis.

Refer figure 8 and figure 9

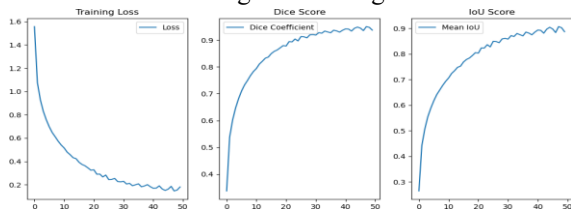


Figure 8: Model Segmentation Prediction with XAI

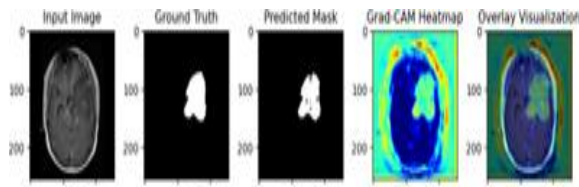


Figure 6: Segmentation Model Performance Graph

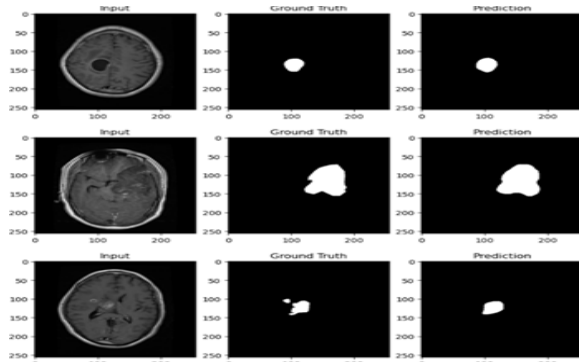


Figure 7: Model Segmentation Prediction

6.3 Explainable AI (XAI) Insights

To ensure model interpretable, Grad-CAM was used to generate heatmaps showing regions that

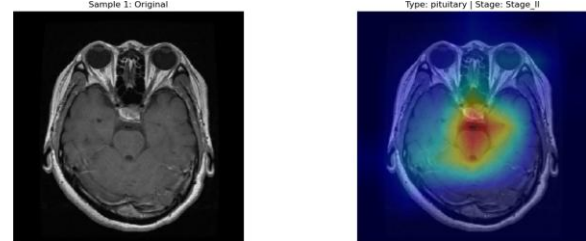


Figure 9: Model Segmentation Prediction with XAI

VII. DISCUSSION

The proposed framework demonstrates the potential of deep learning-based automated systems for brain tumor analysis:

- Integration of classification and segmentation improves diagnostic confidence by providing both tumor type and precise location.
- The use of XAI bridges the gap between black-box AI models and clinical applicability.
- Limitations include dependency on dataset quality and the potential for misclassifications in rare tumor types. Future work can explore larger multi-institutional datasets and multi-modal imaging for further improvements.

VIII. CONCLUSION

The project presents an integrated framework for brain tumor classification and segmentation using advanced deep learning architectures. The ResNet50-based classification model effectively distinguishes between different tumor types, leveraging deep residual learning to achieve high accuracy even on limited datasets. For tumor segmentation, the U-Net architecture enhanced with Proximal Policy Optimization (PPO) demonstrates precise delineation of tumor regions, improving boundary detection and overall segmentation performance. Integration of Explainable AI (XAI) techniques, such as Grad-CAM, ensures transparency by visually highlighting the regions that influence model predictions, thereby fostering trust and interpretable for clinical applications. The combined framework successfully addresses the challenges of automated brain tumor

detection, providing a reliable, efficient, and interpretable solution that can aid radiologists in timely diagnosis.

REFERENCES

- [1] S. Solanki et al., “Brain Tumor Detection and Classification Using Intelligence Techniques: An Overview,” *IEEE Access*, vol. 11, pp. 12870–12886, 2023. SpringerLink
- [2] M. M. Musthafa, M. T. R., V. K. V., et al., “Enhancing Brain Tumor Detection in MRI Images through Explainable AI using Grad CAM with ResNet50,” *BMC Medical Imaging*, vol. 24, no. 107, 2024, doi: 10.1186/s12880-024-01292-7. SpringerLink
- [3] “Brain Tumor Detection and Classification Using MRI with ResNet50 and Hybrid AI Deep Learning Techniques,” *IEEE Xplore*, 2025. (Conference/Proceedings) IEEE Xplore
- [4] “CNN Brain Tumor Classification Based on Masked Image Using U Net Segmentation,” *IEEE Conference Publication*, 2025. IEEE Xplore
- [5] V. Patel, V. Patel, and A. Shinde, “Pol-icySegNet: A Policy Based Reinforcement Learning Framework for Brain Tu-mor Segmentation and Classification in MRI,” *Egyptian Journal of Radiology and Nuclear Medicine*, vol. 56, art. no. 141, 2025. SpringerLink
- [6] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal Policy Optimization Algorithms,” *arXiv preprint*, 2017. arXiv
- [7] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, “Medical Image Segmentation based on U Net: A Review,” *Journal of Imaging Science and Technology*, vol. 64, no. 2, pp. 020508 1–020508 12, 2020. library.imaging.org
- [8] S. Teja Erukude, V. C. Marella, and S. R. Veluru, “Explainable Deep Learning in Medical Imaging: Brain Tumor and Pneumonia Detection,” *arXiv preprint*, Oct. 2025.