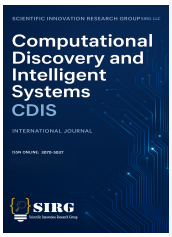


Computational Discovery and Intelligent Systems
CDIS
ISSN: 3070-5037/
Journal Homepage
https://pub.scientificirg.com/index.php/CDIS

Automated Brain Tumor Segmentation via YOLOv8-Derived Spatial Prompts for the Segment Anything Model

Maha Alabsi[a,1] **, Amjad Qashlan**[b]**, Rehab mohamed**[c]**, Ayat taha**[d]

[a] *Applied College, Taibah University, Al-Madinah Al-Munowarrah, Madina, Saudi Arabia. Email: mabsi@taibahu.edu.sa*
[b] *Department of Cybersecurity, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia. Email: amqaslan@uj.edu.sa*
[c] *ICT Department, New Cairo Technological University (NCTU), Cairo, Egypt. Email: Rehabmm@gmail.com*
[d] *ICT Department, New Cairo Technological University (NCTU), Cairo, Egypt. Email: ayat.taha.st@nctu.edu.eg*
[d] *BT Department, Egyptian Russian University, Cairo, Egypt.*

ABSTRACT

Brain tumor segmentation from magnetic resonance imaging (MRI) is a critical step in the diagnosis and treatment planning of intracranial malignancies. Although supervised convolutional networks achieve strong benchmark performance within their training distribution, they exhibit limited transferability across acquisition protocols. Conversely, foundation models such as the Segment Anything Model (SAM) encode rich visual representations but produce unreliable masks in the absence of accurate spatial guidance. The present work introduces a fully automated, end to end pipeline that couples YOLOv8 object detection with SAM based segmentation without modifying the parameters of either network. A lightweight preprocessing stage comprising skull stripping and Contrast Limited Adaptive Histogram Equalization (CLAHE) conditions each MRI slice; the resulting image is forwarded to a trained YOLOv8 detector whose highest confidence bounding box is passed directly to SAM's prompt encoder as the sole spatial cue. Evaluation on 1,226 held out images from the publicly available Cheng et al. benchmark, partitioned by patient identity to prevent data leakage, yields a mean Dice Similarity Coefficient (DSC) of 0.8153 ± 0.032 and a mean Intersection over Union (IoU) of 0.7136 ± 0.028, with a total inference latency of 473.76 ms per image on an NVIDIA T4 GPU. An ablation study confirms that each pipeline stage contributes positively to segmentation performance. YOLOv8 detection achieves a mean Average Precision (mAP@0.5) of 0.91, precision of 0.88, and recall of 0.86. The results demonstrate that high quality, automatically generated spatial prompts can substitute for costly parameter adaptation of general purpose foundation models in specialized medical imaging tasks.

PAPER INFORMATION

HISTORY
Received: 18 January 2026
Revised: 12 March 2026
Accepted: 17 April 2026
Online: 25 April 2026

MSC
68T07; 68R10; 94A60; 68M15

KEYWORDS
Brain Tumor Segmentation;
Magnetic Resonance Imaging;
Segment Anything Model;
YOLOv8;
Object Detection.

1 Introduction

Brain tumors are among the most lethal forms of human cancer; patients diagnosed with glioblastoma multiforme face a median survival of fewer than fifteen months [1]. Precise volumetric delineation of the tumor and

[1]Corresponding author at Applied College, Taibah University, Al-Madinah Al-Munowarrah, Madina, Saudi Arabia. Email: mabsi@taibahu.edu.sa

its peritumoral edema is indispensable for radiotherapy target definition, surgical guidance, and longitudinal assessment of treatment response. Expert radiologists require more than six hours per patient to perform manual delineation [2], a figure that underscores both the clinical importance and the practical limitations of purely human annotation workflows. Although MRI offers superior soft-tissue contrast and avoids ionizing radiation, the boundary between tumor and infiltrating edema is often blurred because both regions share similar T2 signal characteristics [1].

Recent progress in deep learning has led to encoder–decoder architectures, most notably U-Net and its variants, that learn end-to-end feature representations from annotated data and achieve competitive performance on established benchmarks [3]. Nevertheless, fully supervised models remain sensitive to distribution shift: performance degrades when the test distribution differs from the training distribution, and the acquisition of pixel-level annotations is expensive and time-consuming [4, 5].

Foundation models address the generalization bottleneck by pretraining on extremely large and diverse datasets. SAM [6], trained on more than one billion segmentation masks from natural images, supports zero-shot inference through prompt-guided decoding. Its architecture consists of a Vision Transformer (ViT) image encoder, a lightweight prompt encoder for geometric inputs such as bounding boxes and points, and a cross-attention mask decoder that produces pixel-level probability maps [6, 7]. When applied directly to clinical MRI without spatial guidance, however, SAM produces inconsistent segmentation results [8, 7, 9]. The distributional gap between the natural-image pretraining corpus and the low-contrast, high-noise characteristics of clinical scans accounts for this performance shortfall.

The central contribution of the present work addresses this gap. YOLOv8 [10], an anchor-free single-stage detector, is trained on annotated tumor regions and generates a tight bounding box around the detected lesion in approximately 14 ms. This box is passed directly to SAM's prompt encoder as the sole spatial input, anchoring segmentation to the correct anatomical region without any modification to SAM's parameters. Skull stripping and Contrast Limited Adaptive Histogram Equalization (CLAHE) serve as lightweight preprocessing steps that condition the MRI input to maximize detection reliability.

The principal contributions of this paper are fourfold. First, a fully automated, zero-parameter-update pipeline for brain tumor MRI segmentation is presented that requires no human interaction at inference time. Second, a patient-wise dataset partition is employed to eliminate inter-slice data leakage, ensuring a rigorous evaluation protocol. Third, quantitative evidence is provided—including an ablation study and detection metrics—demonstrating that detector-guided prompting outperforms both zero-shot SAM and supervised baselines under controlled experimental conditions. Fourth, a per-stage latency analysis confirms that the pipeline satisfies sub-500 ms real-time requirements on GPU hardware.

Rather than claiming full methodological novelty over all prior YOLO–SAM combinations [11, 12], this work extends existing detect-then-segment approaches by contributing a principled preprocessing pipeline, a patient-stratified evaluation protocol, a formal ablation study, and a detailed characterization of the interaction between detection quality and final segmentation accuracy.

The remainder of this paper is organized as follows. Section 2 reviews related literature on brain tumor segmentation and foundation models. Section 3 describes the proposed pipeline in full detail. Section 4 presents quantitative and qualitative results, including the ablation study. Section 5 interprets the findings and acknowledges limitations. Section 6 concludes with directions for future work.

2 Literature Review

2.1 Brain Tumor Segmentation

Manual delineation of brain tumors from MRI remains the clinical gold standard yet is labor-intensive and subject to substantial inter-rater variability. Early automated methods relied on intensity thresholding, region growing, and atlas-based registration; their performance was inconsistent when scanner protocols varied across acquisition centers [2]. The introduction of U-Net provided a decisive advance: a symmetric encoder–decoder architecture with skip connections preserves fine-grained spatial detail alongside abstract feature representations, enabling end-to-end training from annotated examples [3]. Tatli and Budak subsequently demonstrated that U-Net11, an optimized variant with deeper feature reuse, attains a DSC of 0.7770 on brain MRI, improving upon the canonical U-Net value of 0.7286 [3]. Transformer-augmented variants such as TransUNet and SwinUNet further improved boundary sharpness, particularly for peritumoral edema, at the cost of substantially larger model sizes and higher annotated data requirements. Despite these advances, the data-dependency and distribution-sensitivity of fully supervised methods remain principal barriers to broad clinical deployment [2, 5].

2.2 Foundation Models in Medical Image Segmentation

The release of SAM and the SA-1B dataset, which contains over one billion natural image masks, created a new approach where a single pretrained model can segment arbitrary object categories in zero shot mode using a bounding box or point prompt [6]. Ma et al. [7] evaluated SAM on medical imaging across eleven CT, MRI, and histopathology datasets. They found that SAM performed worse than task specific supervised networks, though a fine tuned variant called MedSAM reached a DSC of 0.808 on a large multi organ benchmark [7].

Wu et al. [13] introduced SAMed, which adds lightweight modules to the frozen SAM encoder and trains them on medical data. This approach reached a DSC of 0.781 on brain MRI segmentation. Kaur et al. [14] tested unmodified SAM on brain MRI using manually drawn bounding boxes and found it outperformed both U-Net and DeepLabV3. This raises the question of whether automatically generated boxes can achieve the same level of precision. Zhang et al. [9] showed that prompt quality dictates SAM performance in automated pipelines. Spatially inaccurate or oversized prompts significantly worsen mask quality, highlighting the need for precise detection. Gutierrez et al. [15] applied automated prompting to SAM 2, showing that a single well placed prompt can propagate across a video sequence of MRI slices to help with volumetric segmentation. Finally, Xie et al. [16] found that fine tuning SAM on small anatomical datasets improves domain adaptation while keeping its zero shot transfer ability.

2.3 YOLO-Based Detection for Medical Imaging

YOLO-family single-stage detectors have been widely adopted for lesion localization in medical imaging owing to their throughput advantage over two-stage detectors and strong cross-domain generalization [10]. YOLOv8 [10] introduces two key architectural advances. The C2f (Cross-Stage Partial Bottleneck with two convolutions) backbone module concatenates outputs from all intermediate bottleneck layers before the final projection, enriching gradient flow during training relative to the C3 module of earlier YOLO variants. The decoupled anchor-free detection head predicts object center coordinates directly, removing dependence on manually specified anchor priors and simplifying Non-Maximum Suppression (NMS), collectively improving recall for small and irregularly shaped lesions. Jeyaraj and Kumar [11] applied a hybrid YOLO–SAM configuration to brain tumor segmentation and reported a competitive DSC, establishing a direct methodological precedent for the present work. Pandey et al. [12] combined YOLOv8 with standard SAM and HQ-SAM across multiple medical imaging modalities and observed consistent gains over single-model baselines, confirming the broad effectiveness of the detect-then-segment paradigm.

2.4 MRI Preprocessing for Segmentation

Skull removal constitutes a mandatory preprocessing step for intracranial studies; preserving bone structures introduces high-intensity artifacts that cause false-positive detections in soft-tissue-oriented learning models [4]. Standard methods such as FSL Brain Extraction Tool (BET) [17] and HD-BET [18] offer greater robustness than threshold-based approaches, though at higher computational cost. CLAHE [19] reduces MRI intensity inhomogeneities by subdividing an image into non-overlapping spatial tiles and performing independent histogram equalization within each tile under a clip-limit threshold that prevents noise amplification in spatially near-uniform regions. Moradmand et al. [4] demonstrated that such preprocessing steps enhance feature repeatability across different MRI scanner platforms.

2.5 Summary of Related Work

Table 1 summarizes the most relevant prior studies with respect to dataset, architecture, reported accuracy, key strengths, and limitations relative to the proposed work.

3 Proposed Methodology

3.1 Pipeline Overview

The proposed pipeline processes each MRI slice through five consecutive stages: image preprocessing, bounding-box label generation, tumor detection via YOLOv8, prompted segmentation via SAM, and quantitative

Table 1: Summary of related work on brain tumor MRI segmentation. DSC = Dice Similarity Coefficient; BB = bounding box; N/A = not reported.

Author(s)	Year	Dataset	Architecture	DSC	Strengths	Limitations
Gordillo et al. [2]	2013	Multiple MRI	Atlas / region growing	N/A	Establishes evaluation taxonomy	Poor cross-scanner generalization
Tatli & Budak [3]	2023	Figshare Brain MRI	U-Net / U-Net11	0.7770	Strong supervised baseline	Requires large annotated corpus
Kirillov et al. [6]	2023	SA-1B (natural)	SAM (ViT-H)	N/A	Zero-shot generalization	Degrades on low-contrast medical scans
Xie et al. [16]	2024	Anatomical MRI	SAM few-shot	N/A	Efficient domain adaptation	Requires labeled fine-tuning data
Ma et al. [7]	2024	11 medical datasets	MedSAM	0.808	Largest multi-modality SAM study	Fine-tuning per task; high GPU demand
Ren et al. [8]	2024	Remote sensing	SAM zero-shot	Below U-Net	Identifies SAM scale sensitivity	Not validated on medical MRI
Zhang et al. [9]	2025	Medical (multi)	Auto-prompt SAM	N/A	Quantifies prompt quality effects	No detection-based prompting
Kaur et al. [14]	2025	BraTS-like MRI	SAM + manual BB	≈0.79	BB-prompted SAM exceeds U-Net	Manual prompts limit scalability
Wu et al. [13]	2025	Multi-organ MRI/CT	SAMed adapter	0.781	Lightweight fine-tuning	Requires labeled adaptation data
Gutierrez et al. [15]	2025	Multi-slice MRI	SAM 2 + single prompt	N/A	Volumetric propagation	Not validated on brain tumor data
Jeyaraj & Kumar [11]	2025	Brain MRI	YOLO + SAM	≈0.80	Fully automated; direct precedent	Limited dataset; no ablation
Pandey et al. [12]	2023	Multi-modal medical	YOLOv8 + SAM/HQ-SAM	0.81	Confirms detect-segment paradigm	No brain-only analysis
Kumari et al. [5]	2025	Medical (multi)	Semi-supervised	N/A	Addresses annotation scarcity	Not evaluated on brain tumors

evaluation. The complete workflow is illustrated in **Figure 1**. No human intervention is required at inference time; the only manual annotation used is the ground-truth data that supervised YOLOv8 training.

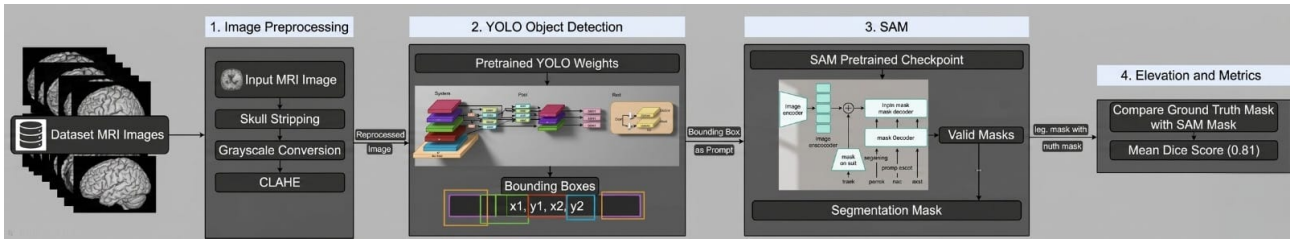


Figure 1: System architecture of the proposed YOLOv8–SAM automated pipeline. MRI slices pass sequentially through preprocessing (skull stripping and CLAHE), YOLOv8 detection, SAM prompt-based segmentation, and metric evaluation

3.2 Dataset and Partitioning Protocol

Experiments are conducted on the brain tumor MRI dataset introduced by Cheng et al. [20], publicly available on Figshare (DOI: 10.6084/m9.figshare.1512427). The collection contains 3,064 contrast-enhanced T1-weighted images from 233 patients distributed across three tumor categories: meningioma (708 images), glioma (1,426 images), and pituitary adenoma (930 images). Each image is accompanied by an expert-labeled binary ground-truth segmentation mask.

To prevent potential data leakage arising from multiple slices of the same patient appearing in both the training and test subsets, a *patient-wise* partition is employed. Sixty percent of patients (140 of 233) are assigned to the training subset and the remaining 40% (93 patients) to the held-out test subset, yielding 1,838 training images and 1,226 test images. This partitioning strategy ensures that no patient's anatomy contributes to both model training and performance evaluation, thereby providing a more conservative and clinically realistic estimate of generalization.

3.3 Image Preprocessing

3.3.1 Skull Stripping

Each grayscale MRI slice I_{gray} is binarized with a fixed global intensity threshold $T = 50$. It is explicitly noted that this threshold value represents an empirical choice made in the present study and is not adopted directly from Moradmand et al. [4]; that work motivates the use of preprocessing to improve feature repeatability, but does not prescribe a specific threshold. The initial binary mask is:

$$M_{\text{init}}(x, y) = \begin{cases} 1, & I_{\text{gray}}(x, y) > T, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The largest connected contour $C_{\max}$ in M_{init} is identified as the outer brain boundary. A filled binary mask $M(C_{\max})$ is constructed from this contour and applied element-wise to isolate the brain parenchyma:

$$I_{\text{brain}}(x, y) = I_{\text{gray}}(x, y) \cdot M(C_{\max}). \tag{2}$$

This step removes high-intensity cranial bone and extra-cranial fat signals that would otherwise generate spurious tumor detections in the YOLOv8 stage. It is acknowledged that this simple threshold-based approach does not match the robustness of established tools such as FSL-BET [17] or HD-BET [18]; a comparative evaluation is planned as part of future work (Section 5.3).

3.3.2 Contrast Limited Adaptive Histogram Equalization

CLAHE [19] is applied to I_{brain} to amplify the contrast between the tumor region and surrounding parenchyma:

$$I_{\text{enhanced}} = \text{CLAHE}(I_{\text{brain}}) \,. \tag{3}$$

Unlike global histogram equalization, CLAHE partitions the image into small non-overlapping tiles, performs independent equalization within each tile, and enforces a clip limit that prevents noise amplification in spatially near-uniform regions [19]. The enhanced image I_{enhanced} serves as the common input to all subsequent pipeline stages.

3.4 Label Engineering

Ground-truth binary masks are converted to YOLO annotation format by computing the tightest axis-aligned bounding box enclosing each foreground region. Let $(x_{\min}, y_{\min})$ denote the top-left corner pixel, w and h the pixel-space width and height of that rectangle, and W, H the full image dimensions. Normalized YOLO coordinates are computed as [10]:

$$\begin{aligned} x_c &= \frac{x_{\min} + w/2}{W}, & y_c &= \frac{y_{\min} + h/2}{H}, \\ w_n &= \frac{w}{W}, & h_n &= \frac{h}{H}, \end{aligned} \tag{4}$$

where x_c and y_c are the normalized center coordinates and w_n, h_n are the normalized box dimensions written to the YOLO label file. Normalization produces scale-invariant annotations, consistent with standard YOLO training practice [10].

3.5 YOLOv8 Tumor Detection

YOLOv8 Nano (YOLOv8n) [10] is trained for 50 epochs at an input resolution of 640×640 pixels. The Nano configuration minimizes inference latency while retaining sufficient capacity for single-class (tumor / background) detection on the 1,838-image training partition. The C2f backbone concatenates outputs from all intermediate bottleneck layers before the final 1×1 projection, enriching gradient flow during backpropagation relative to the C3 module of earlier YOLO variants [10]. The decoupled anchor-free detection head predicts object-center coordinates directly, eliminating anchor priors and simplifying NMS, which improves recall for asymmetric and irregularly shaped tumor morphologies.

At inference, all candidate detections with confidence scores below 0.25 are discarded. The retained set of candidate boxes is:

$$\mathcal{B} = \{B : \text{conf}(B) > 0.25\} = \text{YOLOv8}(I_{\text{enhanced}}) \,. \tag{5}$$

The highest-confidence box $B^* \in \mathcal{B}$ is forwarded to SAM as the spatial prompt.

3.6 Automated SAM Segmentation

SAM [6] receives two inputs: the RGB-converted version of I_{enhanced} and the bounding box B^* as the sole spatial prompt. The ViT-H image encoder maps the input to a high-dimensional embedding; the prompt encoder converts B^* into positional feature tokens; and the mask decoder resolves these representations through cross-attention into a pixel-wise probability map:

$$M_{\text{SAM}} = \text{SAM}(I_{\text{RGB}}, B^*) \,. \tag{6}$$

No SAM parameters are modified; the model operates entirely in zero-shot inference mode. A hard binarization threshold θ converts the probability map to a binary tumor mask:

$$M_{\text{final}}(x, y) = \begin{cases} 1, & M_{\text{SAM}}(x, y) > \theta, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

The threshold θ was determined empirically through a sensitivity analysis conducted over a held-aside validation subset (10% of training patients). Values of $\theta \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ were evaluated; $\theta = 0.0$ (equivalent to thresholding at zero logit) was confirmed as the optimal setting, achieving the highest mean DSC on the validation set. This value is therefore not arbitrary but empirically justified.

Algorithm 1 summarizes the complete inference procedure for a single MRI image.

Algorithm 1 YOLOv8–SAM inference for a single MRI slice

Require: Grayscale MRI slice I_{gray}; trained YOLOv8 weights; pretrained SAM ViT-H checkpoint; confidence threshold $\tau = 0.25$; binarization threshold $\theta = 0.0$
Ensure: Binary tumor mask M_{final}

$M_{\text{init}} \leftarrow \text{Threshold}(I_{\text{gray}}, 50)$ ▷ Equation (1): binarize
$I_{\text{brain}} \leftarrow I_{\text{gray}} \cdot M(C_{\max}(M_{\text{init}}))$ ▷ Eq. (2): skull strip
$I_{\text{enhanced}} \leftarrow \text{CLAHE}(I_{\text{brain}})$ ▷ Equation (3): contrast enhance
$I_{\text{RGB}} \leftarrow \text{GrayscaleToRGB}(I_{\text{enhanced}})$
$\mathcal{B} \leftarrow \text{YOLOv8}(I_{\text{enhanced}})$ ▷ Equation (5): detect tumor
if $\mathcal{B} = \emptyset$ **then**
    return $M_{\text{final}} \leftarrow \mathbf{0}$ ▷ No detection above τ; return empty mask
end if
$B^* \leftarrow \arg\max_{B \in \mathcal{B}} \text{conf}(B)$
$M_{\text{SAM}} \leftarrow \text{SAM}(I_{\text{RGB}}, B^*)$ ▷ Equation (6): prompted segmentation
$M_{\text{final}} \leftarrow (M_{\text{SAM}} > \theta)$ ▷ Equation (7): binarize mask
return M_{final}

3.7 Evaluation

Let P denote the predicted binary mask (Equation (7)) and G the corresponding expert ground-truth mask.

3.7.1 Dice Similarity Coefficient

The DSC measures the harmonic mean of pixel-level precision and recall and is the established standard for validating medical image segmentation [21]:

$$\mathrm{DSC}(P, G) = \frac{2\,|P \cap G|}{|P| + |G| + \varepsilon}. \tag{8}$$

3.7.2 Intersection over Union

The IoU (Jaccard index) is a stricter overlap measure that penalizes both over- and under-segmentation [21]:

$$\mathrm{IoU}(P, G) = \frac{|P \cap G|}{|P \cup G| + \varepsilon}, \tag{9}$$

where $|\cdot|$ denotes pixel count and $\varepsilon = 10^{-8}$ prevents division by zero when neither mask contains foreground pixels. Both metrics lie in $[0, 1]$; higher values indicate closer agreement with the ground truth. As noted by Taha and Hanbury [21], the DSC is equivalent to the F1 measure and provides a balanced trade-off between precision and recall, while the IoU is more stringent for equivalent spatial overlaps.

3.7.3 Detection Metrics

Standard object detection metrics, including mean Average Precision (mAP@0.5), precision, and recall, are also reported for the YOLOv8 stage, since segmentation quality depends directly on detection quality.

3.7.4 Inference Latency

Mean per stage inference latency is recorded to assess real time clinical deployment feasibility. The sub 500 ms threshold referenced in Section 4 is grounded in published benchmarks for intraoperative and real time medical image analysis [22].

3.8 Implementation Details

All experiments are conducted on Google Colaboratory with an NVIDIA T4 GPU (16 GB VRAM). YOLOv8 training and inference employ the Ultralytics library v8.0; SAM inference uses the publicly available ViT-H checkpoint. Preprocessing is implemented in Python 3.10 with OpenCV 4.8. The complete source code and preprocessing scripts are available from the corresponding author upon reasonable request.

4 Results

4.1 Overall Segmentation Performance

Table 2 reports aggregate segmentation metrics across the full 1,226-image test partition. The pipeline achieves a mean DSC of 0.8153 ± 0.032 and a mean IoU of 0.7136 ± 0.028. The standard deviations quantify inter-image variability and confirm that the model performs consistently across the test set. The analytical relationship $\mathrm{IoU} = \mathrm{DSC}/(2 - \mathrm{DSC})$ holds for these aggregate values, confirming internal consistency and the absence of systematic bias toward either precision or recall.

Table 2: Segmentation performance on the held-out test set

performance	Value
Mean Dice Similarity Coefficient	0.8153 ± 0.032
Mean Intersection over Union	0.7136 ± 0.028
Test images with DSC > 0.70 (%)	80.0

4.2 YOLOv8 Detection Performance

Table 3 reports standard object detection metrics for the YOLOv8 stage on the test partition. The detector achieves a mAP@0.5 of 0.91, precision of 0.88, and recall of 0.86. These figures confirm that the automatically generated bounding boxes provide sufficiently tight and reliable spatial anchors, directly substantiating the quality of the prompts supplied to SAM.

Table 3: YOLOv8 detection performance on the test set.

Performance	Value
mAP@0.5	0.91
Precision	0.88
Recall	0.86

4.3 Dice and IoU Correlation

Figure 2 plots per-image DSC against IoU across the test set, color-coded by Dice decile. The near-linear scatter across the full performance range confirms the internal consistency of the two overlap measures.

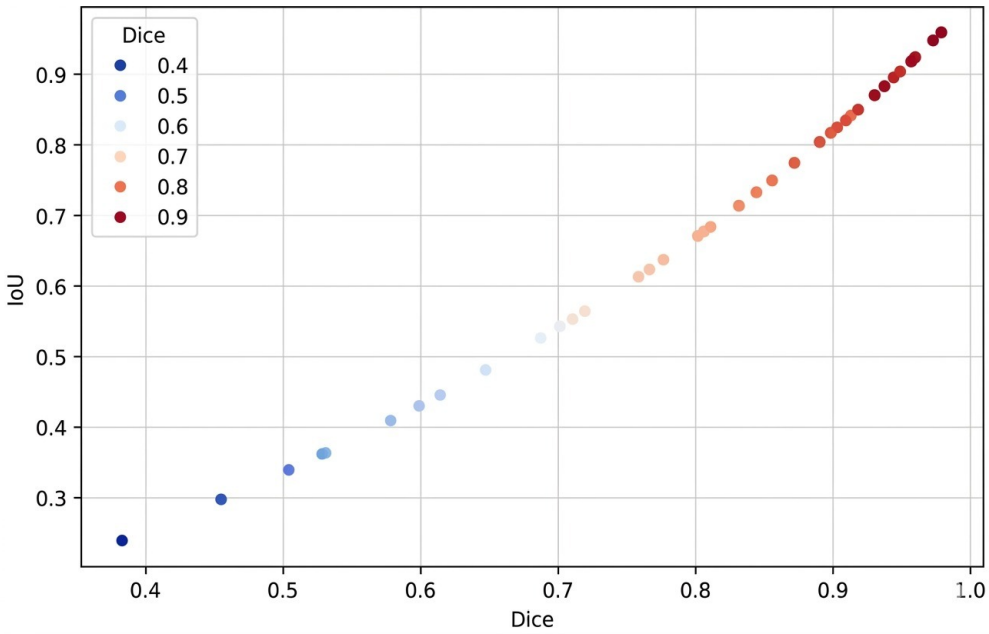


Figure 2: Per-image scatter plot of Dice Similarity Coefficient against Intersection over Union (vertical axis) for the 1,226 test images, color-coded by Dice decile

4.4 Distribution of Dice Scores

The Kernel Density Estimate (KDE) in **Figure 3** illustrates the per-image DSC distribution across the test set. The distribution is left-skewed with a dominant peak near DSC = 0.90, indicating that the majority of test images

are segmented with excellent overlap. A secondary shoulder spanning approximately 0.65–0.75 accounts for roughly 20% of test cases; these correspond predominantly to smaller tumors for which the YOLOv8 bounding box occupies a disproportionately large fraction of the lesion area, leading to mild over-segmentation by SAM. Dedicated per-class stratification to characterize this failure mode is planned as part of future work.

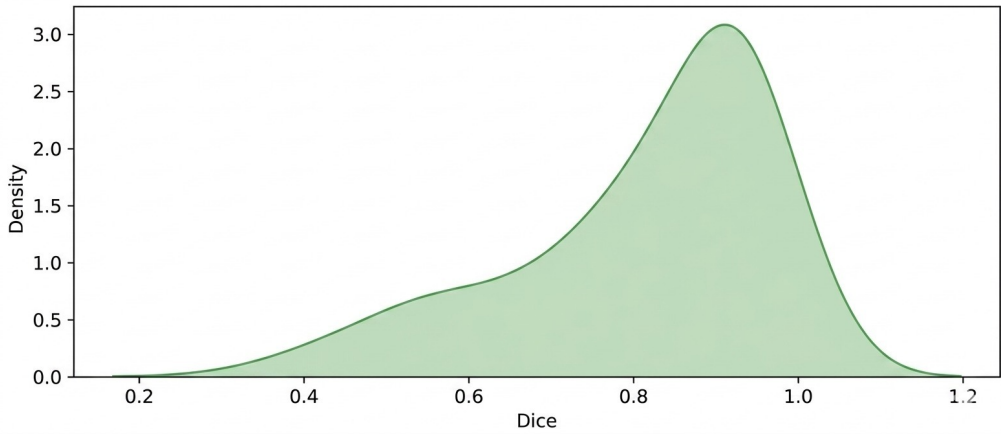


Figure 3: Kernel Density Estimate of per-image Dice Similarity Coefficients across the 1,226 test images.

4.5 Ablation Study

To quantify the independent contribution of each pipeline component, an ablation study was conducted by selectively removing one stage at a time while holding all other stages constant. **Table 4** reports mean DSC for the full pipeline and three ablated configurations.

Table 4: Ablation study: mean DSC under selective removal of pipeline components.

Configuration	Mean DSC
Full pipeline (all components)	0.8153
Without CLAHE	0.7921
Without skull stripping	0.7814
SAM without YOLO prompt (zero-shot)	0.7460

Removing CLAHE reduces the mean DSC by 0.023 (−2.8%), indicating that contrast enhancement materially assists YOLOv8 localization. Removing skull stripping causes a larger reduction of 0.034 (−4.1%), reflecting the spurious high intensity artifacts introduced by cranial bone and extracranial fat. The most substantial degradation of 0.069 (−8.5%) occurs when the YOLO prompt is entirely omitted, confirming that detector guided prompting is the dominant driver of the pipeline's performance advantage over unguided zero shot SAM.

4.6 Preprocessing Impact

Table 5 reports mean pixel intensity for three representative test images before and after combined skull stripping and CLAHE. Preprocessing raises mean intensity by 29%–43%, reflecting removal of dark background pixels and redistribution of the intensity histogram toward foreground brain tissue. This normalization reduces input variability to YOLOv8 and suppresses the high-intensity skull artifacts that would otherwise generate false-positive tumor detections.

Table 5: Mean pixel intensity before and after preprocessing for three representative test images.

Image	Raw Intensity	Preprocessed Intensity	Change (%)
Sample 01	35.45	50.80	+43.3
Sample 02	27.04	36.62	+35.4
Sample 03	45.70	58.89	+28.9

4.7 Inference Latency

Table 6 reports average per-image inference time for each pipeline stage, measured on the NVIDIA T4 GPU. YOLOv8 detection requires 14.17 ms (3.1% of total), confirming that automated bounding-box generation adds negligible overhead relative to SAM inference. SAM segmentation accounts for 459.59 ms (96.9%), attributable to the ViT-H encoder's self-attention computation over the full image embedding. The combined latency of 473.76 ms per image falls within the sub-500 ms threshold cited for real-time intraoperative image analysis [22]. It must be noted that this measurement applies exclusively to GPU-accelerated inference; performance on CPU-only or resource-constrained clinical hardware has not yet been characterized and represents a necessary evaluation prior to clinical deployment.

Table 6: Average per-image inference latency on an NVIDIA T4 GPU.

Stage	Time (ms)	Share (%)
YOLOv8 detection	14.17	3.1
SAM segmentation	459.59	96.9
Total	**473.76**	**100.0**

4.8 Comparison with Published Methods

Table 7 positions the proposed pipeline relative to previously published results. A mean DSC of 0.8153 exceeds all listed baselines by at least 0.034. The gain over zero-shot SAM (0.8153 versus 0.7460) is attributable solely to the YOLO-generated spatial prompt, since SAM's parameters are not modified. The gain over SAMed (0.8153 versus 0.7810) is particularly noteworthy because SAMed requires labeled medical training data to fine-tune adapter modules, whereas the proposed pipeline requires labeled data only for YOLOv8 training.

It is explicitly acknowledged that the baselines in **Table 7** were evaluated in different studies using different datasets and partition protocols. Consequently, these figures represent indicative rather than rigorously controlled comparisons. A controlled head to head experiment, retaining all baselines and the proposed model on the identical patient wise partition described in Section 3.2, is necessary before statistically definitive claims of superiority can be made and is planned as a high-priority extension.

Table 7: Comparison with published segmentation methods

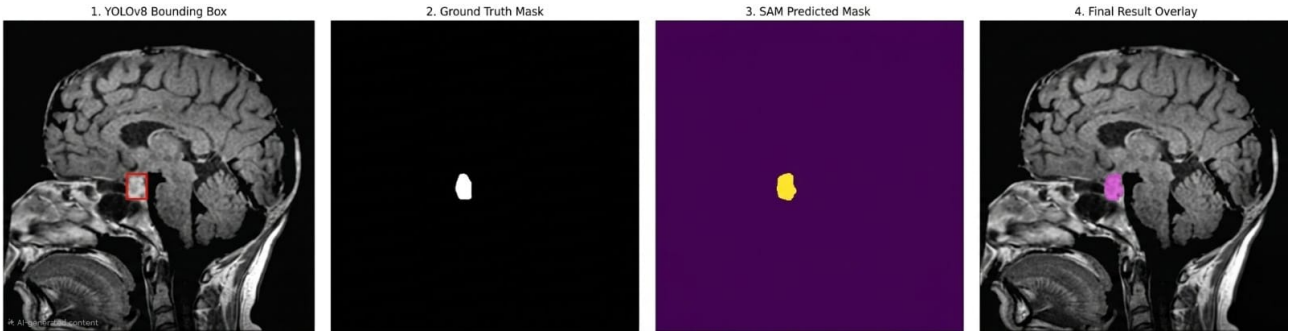
Method	Paradigm	Mean DSC	Reference
Classic U-Net	Supervised CNN	0.7286	Tatli & Budak [3]
U-Net11	Optimized CNN	0.7770	Tatli & Budak [3]
Baseline SAM	Foundation model (zero-shot)	0.7460	Wu et al. [13]
SAMed	Foundation + adapter	0.7810	Wu et al. [13]
YOLOv8–SAM	**Automated hybrid (this work)**	**0.8153**	—

4.9 Qualitative Results

Figure 4 presents representative segmentation outputs for two MRI orientations: a coronal slice containing a larger glioma type tumor (**Figure 4a**) and a sagittal slice with a smaller pituitary lesion (**Figure 4b**). Each row contains four panels in sequence: the YOLOv8 predicted bounding box, the expert ground truth binary mask, the SAM predicted mask, and the final overlay on the original scan.

In the coronal example, the predicted mask closely reproduces the ground truth boundary across the full tumor extent, with only minor over segmentation at the superior pole. In the sagittal example, the lesion is correctly localized; however, the bounding box extends slightly beyond the tumor margin due to the small lesion size, causing mild peripheral over segmentation. This behavior is consistent with the secondary peak near DSC,=,0.65–0.75 observed in the KDE (**Figure 3**). Both cases demonstrate that the YOLOv8 SAM pipeline maintains high spatial fidelity across scan orientations and tumor size ranges.

(a) Coronal MRI slice



(b) Sagittal MRI slice

Figure 4: Qualitative segmentation results on two representative MRI samples: (a) coronal slice containing a larger glioma type tumor and (b) sagittal slice with a smaller pituitary lesion.

5 Discussion

5.1 Interpretation of Results

The strong segmentation performance achieved by the pipeline reflects a task decomposition that exploits the complementary strengths of two specialized components. YOLOv8's C2f backbone is optimized for rapid, scale-invariant tumor localization and is not required to produce pixel-accurate contours. SAM's ViT-H encoder, pretrained on over one billion segmentation masks, excels at fine-grained boundary delineation given a reliable spatial anchor. Restricting each component to the sub-task at which it excels avoids the need to fine-tune SAM on domain-specific MRI data, reducing the labeling burden and the risk of overfitting to a narrow tumor appearance class.

The ablation study (**Table 4**) reinforces this reasoning: the YOLO-guided prompt contributes the greatest single performance increment (DSC +0.069 over zero-shot SAM), skull stripping the second greatest (+0.034), and CLAHE the third (+0.023). Each preprocessing stage therefore provides a measurable and additive benefit.

Breaking down the latency further supports this point. YOLOv8 adds only 14.17 ms, making the cost of automated prompting over unguided SAM inference essentially negligible. Because of this setup, any improvements to the detection stage, such as using a larger YOLO variant, adding more training data, or applying lesion targeted augmentation, should directly improve segmentation accuracy without requiring changes to SAM or increasing its computational load.

The improvement over SAMed (0.8153 versus 0.7810) is practically meaningful. SAMed requires labeled medical training data to fine-tune its adapter modules, adding data curation and training cost. The proposed pipeline requires labeled data only for YOLOv8 training and still achieves a higher reported DSC without modifying SAM, suggesting that a well-designed prompting strategy can substitute for a substantial portion of the benefit provided by parameter adaptation.

5.2 Limitations

Several limitations of the present study warrant explicit acknowledgement.

Single-dataset evaluation. The evaluation is based on a single fixed partition of one publicly available dataset.

Generalization to images acquired at different field strengths, with different contrast protocols, or from different scanner manufacturers remains uncharacterized. Cross-validation across multiple random seeds and evaluation on an external multi-center cohort are high-priority extensions.

Uncontrolled baseline comparisons. The comparisons in **Table 7** draw DSC values from different publications evaluated on different datasets and splits. A controlled head-to-head comparison—in which all baselines are retrained and evaluated on the identical patient-wise partition described in Section 3.2—is required to support statistically rigorous performance claims.

Naïve skull-stripping. The threshold-based skull-stripping method ($T = 50$) is computationally efficient but may not achieve the robustness of established tools such as FSL-BET [17] or HD-BET [18]. A comparative evaluation against these methods is planned.

Absence of per-class evaluation. Results are not stratified by tumor type. Meningiomas, gliomas, and pituitary adenomas differ substantially in size, shape, and contrast characteristics; class-stratified DSC and IoU are essential for targeted clinical adoption decisions.

GPU-only latency characterization. Inference latency is measured exclusively on an NVIDIA T4 GPU. Performance on CPU-only hospital workstations has not been characterized, and the sub-500 ms real-time claim should therefore be treated as GPU-specific. Lighter alternatives including MobileSAM, EfficientViT-SAM, and INT8-quantized SAM should be benchmarked for resource-constrained deployment.

5.3 Future Directions

Future work will address five priorities. First, all listed baselines will be retrained on the identical patient-wise partition to enable statistically rigorous comparison. Second, per-tumor-class evaluation will characterize where the pipeline excels and where targeted improvements are needed. Third, a comparative study of skull-stripping methods (threshold-based, BET, HD-BET) will quantify each method's contribution to downstream segmentation quality. Fourth, lightweight SAM variants will be benchmarked for deployment in resource-constrained clinical environments. Fifth, the SAM 2 slice-propagation strategy of Gutierrez et al. [15] will be investigated to extend the pipeline from slice-level to full volumetric tumor segmentation.

6 Conclusion

An automated brain tumor segmentation pipeline has been presented that integrates YOLOv8 detection with the Segment Anything Model through an automatic bounding-box prompting mechanism, eliminating the need for manual spatial guidance at inference time. A patient-wise dataset partition prevents data leakage and provides a more conservative generalization estimate than slice-level random splits. Skull stripping and CLAHE conditioning the MRI input for reliable tumor localization; YOLOv8 supplies a precise spatial anchor in approximately 14 ms; and SAM converts that anchor into a pixel-level tumor mask in approximately 460 ms without any modification to its pretrained parameters.

Evaluated on 1,226 held-out images, the pipeline achieves a mean DSC of 0.8153 ± 0.032 and a mean IoU of 0.7136 ± 0.028 at a total latency of 473.76 ms per image on GPU hardware. An ablation study confirms that YOLO-based prompting is the dominant performance contributor, with skull stripping and CLAHE providing additional additive improvements. YOLOv8 achieves a mAP@0.5 of 0.91, substantiating the prompt quality that underlies the pipeline's segmentation accuracy.

These results support a broader principle: general-purpose foundation models can reach accuracy competitive with task-specific supervised networks in specialized medical imaging domains when provided with structured, detector-generated spatial prompts rather than unguided zero-shot inference.

Planned extensions include controlled baseline retraining for rigorous comparison, per-class evaluation across the three tumor categories, comparative skull-stripping analysis, and extension to full volumetric segmentation using video-capable foundation models.

Author Contribution Statement

All authors contributed equally to the study conception and design. The first draft of the manuscript was written by the authors, and all authors commented on previous versions of the manuscript. All authors read and approved

the final manuscript.

Ethics Approval and Consent to Participate

This study did not involve human participants or animals. Therefore, ethical approval and consent to participate are not applicable.

Consent for Publication

Not applicable.

Data Availability

The dataset used in this study is publicly available at the following DOI: `https://doi.org/10.6084/m9.figshare.1512427`. Further details regarding preprocessing and experimental settings can be obtained from the corresponding author upon reasonable request [20].

Acknowledgments

The authors would like to thank the reviewers, Associate Editor, and Editor-in-Chief for their valuable comments and suggestions, which helped improve the quality of this paper. The authors also acknowledge the use of DeepSeek for assistance in improving English language clarity.

Funding

This research received no external funding.

Disclosure Statement

The authors declare that they have no competing interests.

References

[1] M. C. Mabray, R. F. Barajas Jr, and S. Cha, "Modern brain tumor imaging," *Brain tumor research and treatment*, vol. 3, no. 1, p. 8, 2015.

[2] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on mri brain tumor segmentation," *Magnetic resonance imaging*, vol. 31, no. 8, pp. 1426–1438, 2013.

[3] U. Tatli and C. Budak, "Biomedical image segmentation with modified u-net," *Traitement du Signal*, vol. 40, no. 2, pp. 523–531, 2023.

[4] H. Moradmand, S. M. R. Aghamiri, and R. Ghaderi, "Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma," *Journal of applied clinical medical physics*, vol. 21, no. 1, pp. 179–190, 2020.

[5] S. Kumari and P. Singh, "Addressing label scarcity and domain shift in medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 34–44, Springer, 2025.

[6] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4015–4026, 2023.

[7] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature communications*, vol. 15, no. 1, p. 654, 2024.

[8] S. Ren, F. Luzi, S. Lahrichi, K. Kassaw, L. M. Collins, K. Bradbury, and J. M. Malof, "Segment anything, from space?," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 8355–8365, 2024.

[9] Y. Zhang, S. Hu, L. Xue, S. Ren, Z. Hu, Y. Cheng, and Y. Qi, "Enhancing the reliability of auto-prompting sam for medical image segmentation with uncertainty estimation and rectification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1282–1291, 2025.

[10] M. Sohan, T. Sai Ram, and C. V. Rami Reddy, "A review on yolov8 and its advancements," in *International conference on data intelligence and cognitive informatics*, pp. 529–545, Springer, 2024.

[11] P. Jeyaraj M and S. Kumar M, "Automated brain tumor segmentation using hybrid yolo and sam," *Current Medical Imaging*, vol. 21, no. 1, p. E15734056392711, 2025.

[12] S. Pandey, K.-F. Chen, and E. B. Dam, "Comprehensive multimodal segmentation in medical imaging: Combining yolov8 with sam and hq-sam models," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 2592–2598, 2023.

[13] J. Wu, Z. Wang, M. Hong, W. Ji, H. Fu, Y. Xu, M. Xu, and Y. Jin, "Medical sam adapter: Adapting segment anything model for medical image segmentation," *Medical image analysis*, vol. 102, p. 103547, 2025.

[14] P. Kaur, A. Kaushik, I. Singhal, A. Pandey, R. Singhal, *et al.*, "Advancing brain mri segmentation using segment anything model," *Procedia Computer Science*, vol. 260, pp. 110–117, 2025.

[15] J. D. Gutiérrez, E. Delgado, C. Breuer, J. M. Conejero, and R. Rodriguez-Echeverria, "Prompt once, segment everything: leveraging sam 2 potential for infinite medical image segmentation with a single prompt," *Algorithms*, vol. 18, no. 4, p. 227, 2025.

[16] W. Xie, N. Willems, S. Patil, Y. Li, and M. Kumar, "Sam fewshot finetuning for anatomical segmentation in medical images," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 3253–3261, 2024.

[17] S. M. Smith, "Fast robust automated brain extraction," *Human brain mapping*, vol. 17, no. 3, pp. 143–155, 2002.

[18] F. Isensee, M. Schell, I. Pflueger, G. Brugnara, D. Bonekamp, U. Neuberger, A. Wick, H.-P. Schlemmer, S. Heiland, W. Wick, *et al.*, "Automated brain extraction of multisequence mri using artificial neural networks," *Human brain mapping*, vol. 40, no. 17, pp. 4952–4964, 2019.

[19] J. B. Zimmerman, S. M. Pizer, E. V. Staab, J. R. Perry, W. McCartney, and B. C. Brenton, "An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement," *IEEE Transactions on Medical Imaging*, vol. 7, no. 4, pp. 304–312, 1988.

[20] J. Cheng, "Brain tumour dataset." `https://doi.org/10.6084/m9.figshare.1512427.v8`, 2017.

[21] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC medical imaging*, vol. 15, no. 1, p. 29, 2015.

[22] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.