

A Multi-Modal Hybrid U-Net and Mask R-CNN Approach to Brain Tumor Segmentation in MRI: Baseline Analysis and Proposed Framework

Heli Dave¹, Dr. Lokesh Gagnani², Dr. Mehul Barot³

¹Department of Computer Engineering, Kadi Sarva Vishwavidyalaya, Gandhinagar, Gujarat, India.

Email: helidave06@gmail.com

²Associate Professor, LDRP-ITR, Kadi Sarva Vishwavidyalaya, Gandhinagar, Gujarat, India. Email: gagnani.lokesh@gmail.com

³Professor, LDRP-ITR, Kadi Sarva Vishwavidyalaya, Gandhinagar, Gujarat, India. Email: m.p.barot@gmail.com

Abstract: Detecting brain tumors early is still hard to get right, and it matters a lot because it feeds directly into how a patient gets treated. MRI is the modality most people trust for this because it shows soft tissue clearly, but drawing tumor boundaries by hand takes forever, and two radiologists rarely agree exactly. That disagreement problem is a big part of why automated segmentation has become such an active research area recently. In this paper, I look at where current deep learning methods for MRI tumor segmentation are still weak: they usually depend on just one MRI sequence, false positives don't go away easily, tumor boundaries are often imprecise (HD95 numbers make this obvious), and diffusion-based models in particular cost too much compute to run quickly. I implemented two baseline pipelines on BraTS 2020 to check where things actually stand - a GGMM-enhanced U-Net and a diffusion probabilistic model - and used what came out of that to shape a hybrid framework. The framework has four parts: it fuses T1, T2, FLAIR, and T1CE features instead of using just one sequence; it uses a Dice-BCE loss so class imbalance doesn't dominate training; it applies a threshold that adjusts itself based on the mean and standard deviation of each prediction rather than sticking to a fixed 0.5 cutoff; and it removes tiny, implausible regions in a post-processing step. The GGMM baseline landed around a 0.81-0.82 Dice score; the diffusion model got closer to 0.89 but took far longer per case to run. Based on that gap, four hypotheses are proposed for testing once the full model is trained: Dice of 0.90 or higher, IoU above 0.85, at least 15% fewer false positives than a plain U-Net, and no more than a 5% performance drop under noisy input. The hope is that this ends up being a reasonable middle ground - accuracy near what diffusion models get, minus the runtime problem that keeps them out of actual clinics.

Keywords: Brain tumor segmentation; Multi-modal MRI; U-Net; Mask R-CNN; BraTS; Deep learning; Dice coefficient; False-positive reduction; Dynamic thresholding.

1. Introduction

A brain tumor is essentially abnormal cell growth inside brain tissue. Where it grows and how fast it grows decide a lot about how serious it turns out to be for the patient - sometimes it's manageable, sometimes it isn't. Prognosis depends heavily on catching it early. Every year, a considerable number of people worldwide get diagnosed with brain tumors, and among children and young adults, it remains one of the more common causes of cancer death.



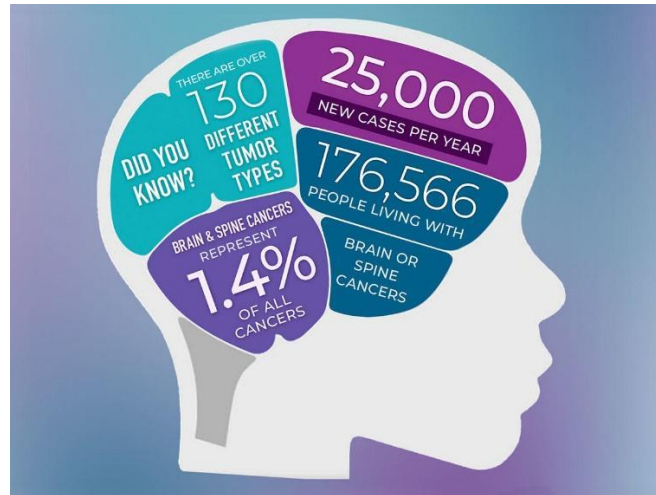


Figure 1. Brain tumor Statistics [22]

The MRI scan has long been the imaging modality for brain tumors due to its excellent separation of soft tissue types, and the fact that one MRI exam already provides several sequences to work with: T1, contrast-enhanced T1 (T1CE), T2, and FLAIR. Every sequence tends to reveal something a bit different about the tumor. Though T1CE is better at lighting up the enhancing tumor core, FLAIR is more sensitive to the edema surrounding it. Normally, radiologists check out all these things together when they read a scan. However, they must manually segment the tumor region across every sequence, slice by slice. This takes a long time, and different readers often do not land on the same boundary. This gap pretty much constitutes the entire rationale behind the construction of automated segmentation pipelines that yield reproducible results.

In recent years, deep learning has become the default for this kind of segmentation problem. U-Net, Attention U-Net, and 3D U-Net perform well on benchmarks like BraTS, among others. In recent times, accuracy limits have stretched further by transformer and diffusion-based models. Patterns appear to be observed in information articles from 2024 and 2025. Improving the accuracy, boundary sharpness, or robustness of a product significantly increases its computational costs. A clear example of this would be diffusion models. While they get competitive Dice scores, refinement through multiple iterations takes tens of minutes per patient for inference. A busy clinical workflow can't absorb that.



Figure 2. Image processing [23]

There's additionally a metric problem worth mentioning early on, as it shapes a lot of the manner in which this paper reads existing results. It is not reliable to use pixel-wise accuracy alone to judge segmentation quality in this case. Any scan will have a small piece of tissue where the tumor is located within the total volume of the brain. So, a model that predicts no tumor everywhere will still score very close to 98% accuracy. That's not doing anything useful. The Dice score, Intersection-over-Union (IoU), and the 95th percentile Hausdorff Distance (HD95) do a much better job of showing what's actually happening at the boundary and overlap level. With this in mind, the study scrutinizes contemporary literature, executes two baseline approaches to establish a point of reference, examines the particular

open gaps in the industry, and presents a hybrid U-Net & Mask R-CNN framework that aims to advance accuracy, reliability, generalizability, and efficiency all at once, instead of sacrificing one for the other.

2. Literature Review

The recent literature extensively covers various aspects of work related to brain tumor detection, classification, and segmentation. According to reports, YOLO-NeuroBoost pairs together an attention-enhanced YOLOv8 backbone and a custom loss and achieves very high mAP. Although it was created for bounding-box detection rather than pixel-level segmentation, it is not yet known how efficacious it is clinically [1]. A rotation-invariant Vision Transformer approaches the problem differently and achieves 98% classification accuracy with less pre-processing required. However, it underperforms on scans from non-standard protocols and on small or irregular-shaped tumors [2].

Certain CNN pipelines utilize a combination of a pre-trained backbone, feature extractor, and ensemble classifier. These algorithms also have good classification results, but they require large labeled datasets to be able to do so. Moreover, they do not transfer well across clinical sites [3]-[4]. Earlier pipelines for image processing that merged histogram equalization with segmentation did a decent job of coping with contrast problems. However, they still do not help with low-contrast scans effectively [5]. Another option is EfficientNetB2 in tandem with homomorphic filtering, which improves the signal-to-noise ratio and downstream classification accuracy is also an AI-based option. However, it too was designed for classification and not segmentation [6].

Segmentation-focused architectures have considerably advanced as well. The merging of Faster R-CNN and YOLOv5 to enhance segmentation motivated by detection creates false positives, as the boundaries of the tumor area are ambiguous. Neural networks made locally, employing histogram equalization and also manual feature extraction, work fine at usual resolution but drop off rather suddenly at low resolution [10]. According to [11], [16], ensemble methods improve the average Dice and IoU scores and generalise better across datasets, but reduce inference speed and interpretability. The implementation of recurrent Atrous Spatial Pyramid Pooling on the Attention U-Net showed competitive results with a Dice score of 0.88-0.89 on multi-scale and irregular tumours. Nevertheless, the increase in architectural complexity renders it more sensitive to hyperparameters.

Two of these are more directly relevant here because they were actually implemented as baselines for this work. The first is a GGMM-enhanced U-Net, which is compared against plain U-Net and U-Net + GMM versions. The model GGMM+U-Net has produced the best results in Dice, recall, and IoU. It is noteworthy that training took place only for a few epochs, the images were downsampled at the start, T1CE was used for preprocessing only, and due to hardware constraints, enhancements to the setup were not implemented. This baseline is a diffusion probabilistic segmentation model equipped with Fourier feature processing, uncertainty estimation, and multi-step refinement. It gives a Dice score of 0.89. Moreover, it performs good boundary detection. This approach mainly suffers from being slow; the repeated denoising pushes the runtime into the thousands of seconds per case. Furthermore, as it is a 2D algorithm that works slice by slice, it never really captures the 3D shape of the tumor.

Reviews across the board keep flagging the same issues: frequent false positives, poor recall of small lesions, generalization that fails across institutions, and inconsistent robustness to noise [26]. It also spills over to the multi-modal MRI benchmark in evaluators looking for standardized evaluation, so that they make a fair comparison range on the doctors' note. Additionally, crossover center generalization largely remains unsolved [25]. An ideal evolutionary design of a lightweight 3D attention U-Net, such as LATUP-Net, would attempt to minimize compute cost without significantly losing important accuracy features as compared to heavier 3D models. While the results are promising, clinical validation is lacking. Table 1 provides an overview of the aforementioned methods, detailing their reporting capabilities and highlighting the shortcomings of each method compared to the present approach utilized here.

Table 1. Summary of representative methods from the literature

Method / Model	Dataset(s)	Key Metric(s)	Gap vs. Proposed Work
----------------	------------	---------------	-----------------------

YOLO-NeuroBoost [1]	Br35H, RoboFlow MRI	mAP $\approx$ 97–99%	Strong detection but limited pixel-level segmentation
Rotation-Invariant ViT [2]	Kaggle Brain MRI	Accuracy $\approx$ 98%, F1 $\approx$ 0.98	Weak for small/irregular tumors
EfficientNetB2 + Preproc. [6]	Public MRI, BraTS	Higher classification accuracy	Classification-only, no segmentation
Attention U-Net + ASPP [21],[30]	BraTS, clinical MRI	Dice $\approx$ 0.88–0.89, IoU $\approx$ 0.82–0.84	High compute cost, cross-center issues
Diffusion-based segmentation [33]	BraTS multi-modal MRI	Dice $\approx$ 0.89, IoU $\approx$ 0.83	Slow multi-step inference, high memory demand
GGMM + U-Net (baseline, this study)	BraTS 2020	Dice $\approx$ 0.81–0.82, HD95 $\approx$ 15	Single-modality input, limited boundary precision
LATUP-Net [29]	3D BraTS-like MRI	Comparable to heavy 3D models	Needs further real-world validation

3. Research Gap

Based on the literature review and the results of the baseline runs, a set of gaps kept coming up. These were truly what started to push the direction of the proposed framework.

- Dependence on One Mode: Many of the most successful approaches only utilize a single MRI sequence, typically T1CE, and end up discarding useful information that can be found in T1, T2, and FLAIR.
- False positives remain on the rise. The fixed-threshold outcomes mark uncertain and low-confidence pixels as tumor and hurt precision.
- Boundaries are not clearly defined. The HD95 (Hausdorff) metric may frequently be elevated even when Dice looks good, indicating the predicted boundary is noticeably deviated from the ground truth.
- Models are too costly to operate. Competitive accuracy is reached by diffusion-based and heavier attention architectures; however, they take too much time for clinical (real-time or resource-limited) applications.
- Three-dimensional slice processing: Many pipelines function using a single slice at a time, which leads to inconsistency across the third spatial dimension.
- The accuracy number will mislead you here. Since the tumor tissue represents only a small fraction of the total brain volume, the accuracy level is high and close to 98%. This means that a model could yield a lot of false positives and false negatives, yet still fall within that accuracy measure. Therefore, it's important to report on Dice, IoU, and HD95 as the actual metrics that need to be noted.
- Most frameworks will work on improving the accuracy of the framework, but hardly any work on improving the reliability, cross-modality generalization, and noise robustness together in one model.

4. Proposed Methodology

Based on these gaps, the study proposes a hybrid 3D U-Net and Mask R-CNN framework around multi-modal preprocessing, an adaptive loss function, dynamic thresholding, and a post-processing refinement phase. Our pipeline accepts T1, T2, FLAIR, and T1CE MRI volumes as input and outputs a better tumor segmentation mask. All the

experiments use BraTS 2020: 369 patient volumes for training, each patient volume consisting of the four sequences and a ground-truth mask, and 125 volumes held out for validation.

4.1 Multi-Modal Feature Fusion

Rather than depending on a single sequence the way the GGMM+U-Net baseline does, feature maps from all four modalities get averaged together. That way, the network has access to whatever complementary tumor information each sequence carries:

$$F = (f_{T1} + f_{T2} + f_{FLAIR} + f_{TICE}) / 4$$

Richer input features should mean better tumor visibility than a single-modality baseline can offer. That's the basic idea behind this step.

4.2 Enhanced Segmentation Backbone

A U-Net-style encoder-decoder pulls hierarchical texture, shape, and boundary features out of the fused input, following the standard convolutional feature-mapping form:

$$F(x) = \sigma(W \cdot X + b)$$

The encoder-decoder output then goes through an instance-level Mask R-CNN stage. Combining dense semantic segmentation with instance-aware boundary refinement this way is aimed specifically at the residual false positives and boundary imprecision reported for standalone attention U-Net and Mask R-CNN variants elsewhere.

4.3 Adaptive Loss Function

Tumor and background voxels are heavily imbalanced in count, so a combined Dice and Binary Cross-Entropy (BCE) loss is used instead of relying on either alone:

$$Dice = 2 \cdot TP / (2 \cdot TP + FP + FN)$$

$$Loss = \lambda_1 \cdot Dice_Loss + \lambda_2 \cdot BCE$$

Here $Dice_Loss = 1 - Dice$, and λ_1 , λ_2 weight the two terms. In practice, this makes the model pay more attention to small tumor regions, which a cross-entropy-only loss tends to overlook.

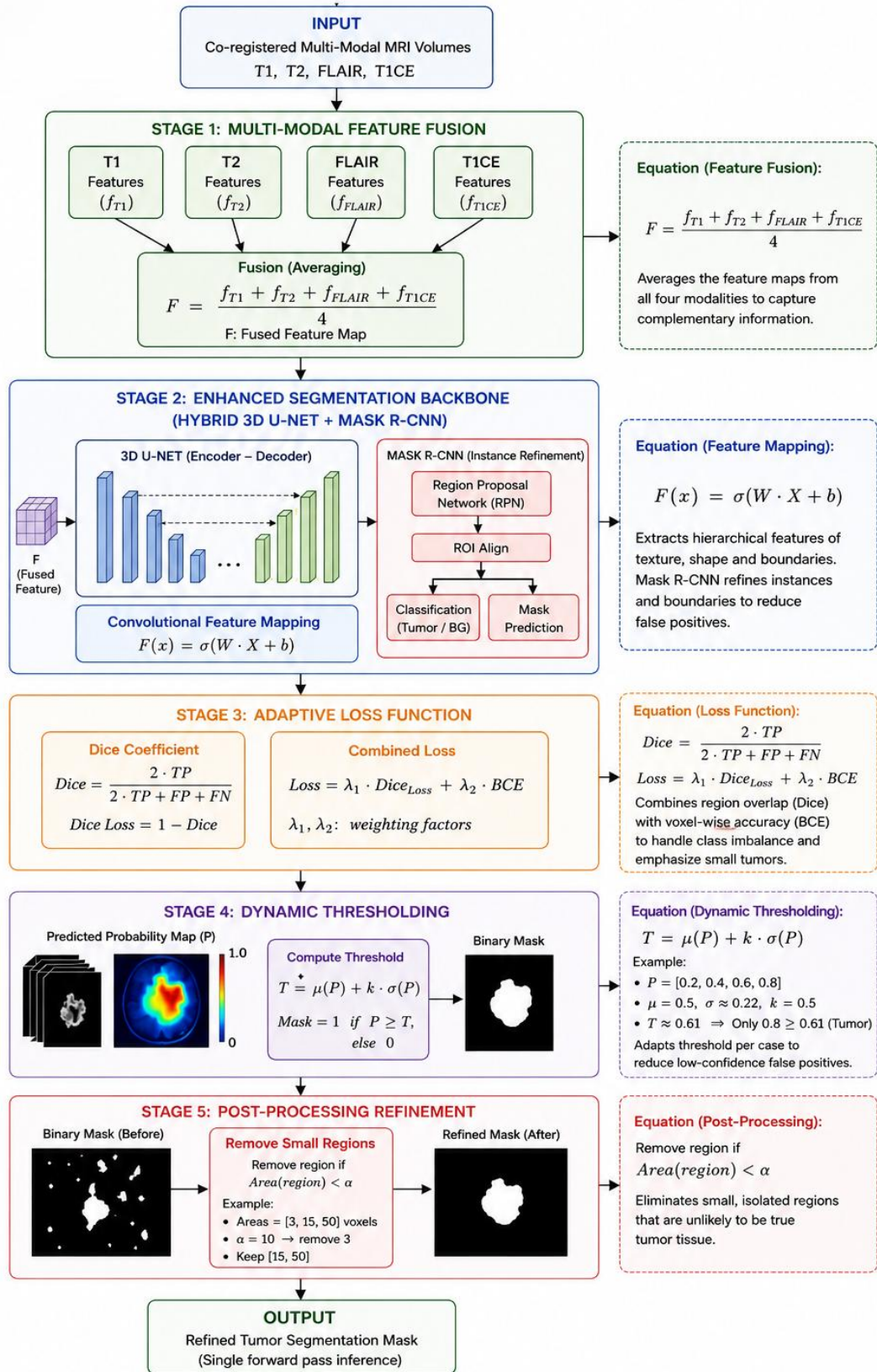


Figure 3. Proposed Methodology

4.4 Dynamic Thresholding

A fixed 0.5 threshold lets low-confidence, noisy predictions through as "tumor" more often than it should. So instead, a threshold gets computed from the statistics of the predicted probability map P for each case:

$$T = \mu(P) + k \cdot \sigma(P)$$

Mask = 1 if $P \geq T$, else 0

4.5 A quick example: with $P = [0.2, 0.4, 0.6, 0.8]$ and $k = 0.5$, we get $\mu = 0.5$, $\sigma \approx 0.22$, so $T \approx 0.61$. Only the 0.8 entry counts as tumor here. A fixed 0.5 cutoff would have kept the 0.6 entry too, which is exactly the kind of false positive this method is meant to catch

4.5 Post-Processing Refinement

A last morphological step clears out small, isolated regions unlikely to be actual tumor tissue:

Remove region if $\text{Area}(\text{region}) < \alpha$

Say the candidate regions come out with areas [3, 15, 50] voxels and $\alpha = 10$. The area-3 region gets dropped, leaving [15, 50] as the final output. Noise gone, the clinically relevant regions stay intact.

4.6 Summary of the Proposed Pipeline

- Input: co-registered T1, T2, FLAIR, and T1CE MRI volumes.
- Stage 1: multi-modal feature fusion across the four sequences.
- Stage 2: hybrid U-Net + Mask R-CNN feature learning with instance-aware refinement.
- Stage 3: adaptive Dice + BCE loss optimization for class imbalance.
- Stage 4: dynamic mean/standard-deviation thresholding to cut false positives.
- Stage 5: morphological post-processing to clear out small, noisy regions.
- Output: a refined tumor segmentation mask, produced in one forward pass instead of the iterative inference diffusion-based baselines require.

4.7 Hypotheses

Four hypotheses follow from the four objectives above, each tied to a specific metric that the proposed model will be tested against once training is complete:

- H1 (Accuracy): the hybrid U-Net + Mask R-CNN framework reaches $\text{Dice} \geq 0.90$ and $\text{IoU} \geq 0.85$, using $\text{Dice} = 2\text{TP}/(2\text{TP} + \text{FP} + \text{FN})$ and $\text{IoU} = \text{TP}/(\text{TP} + \text{FP} + \text{FN})$.
- H2 (Reliability): boundary-aware loss and post-processing together cut false positives by $\geq 15\%$ versus a baseline U-Net, using $\text{FPR} = \text{FP}/(\text{FP} + \text{TN})$ and $\text{reduction} = (\text{FP}_{\text{baseline}} - \text{FP}_{\text{proposed}})/\text{FP}_{\text{baseline}} \times 100$.
- H3 (Generalization): transfer learning keeps average $\text{Dice} \geq 0.85$ across multiple modalities (MRI and CT), measured as mean Dice across sequences.
- H4 (Robustness): performance drop under noisy MRI conditions stays $\leq 5\%$, measured as $(\text{Dice}_{\text{clean}} - \text{Dice}_{\text{noisy}})/\text{Dice}_{\text{clean}} \times 100$.

5. Results and Discussion

Before the deployment of the complete hybrid model, two baseline methods were implemented on BraTS 2020 for a solid point of comparison. The following findings are preliminary results from those two. The following stage of the project will test these two properly with the proposed hybrid framework, including an ablation of each added component.

5.1 Baseline 1: U-Net / GGMM-Enhanced Segmentation

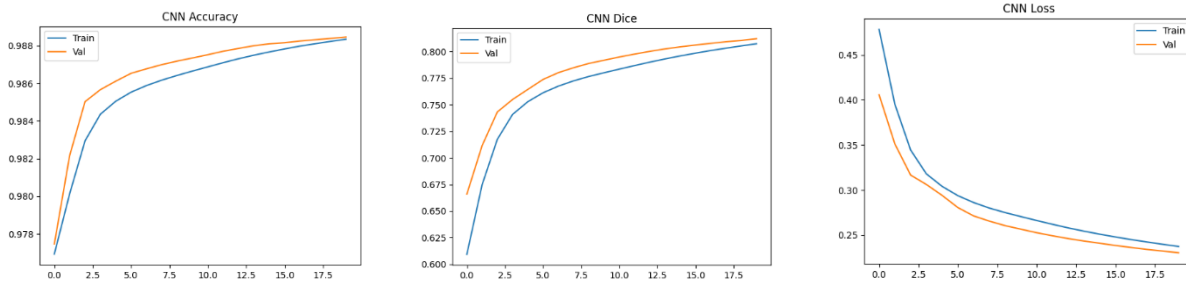


Figure 4. Baseline Output

The baseline achieved approximately 98% pixel-wise accuracy, with a Dice coefficient of around 0.81-0.82, IoU around 0.70, and HD95 of around 15. The accuracy numbers look good. However, looking closer at the actual outputs, there were clear false positives, and boundaries were not clean, with no clear delineation. The precision issue applies here too, where accuracy alone falsely overstates how good segmentation is once tumour tissue only accounts for a small portion of total brain volume.

5.2 Baseline 2: Diffusion-Based Segmentation

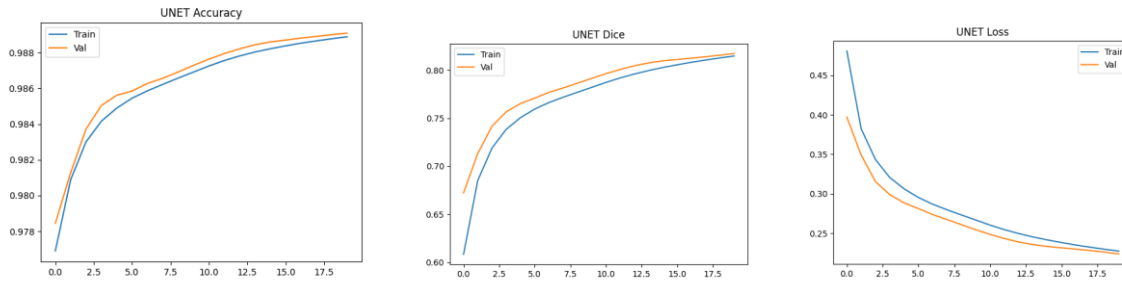


Figure 5. Baseline 2 output

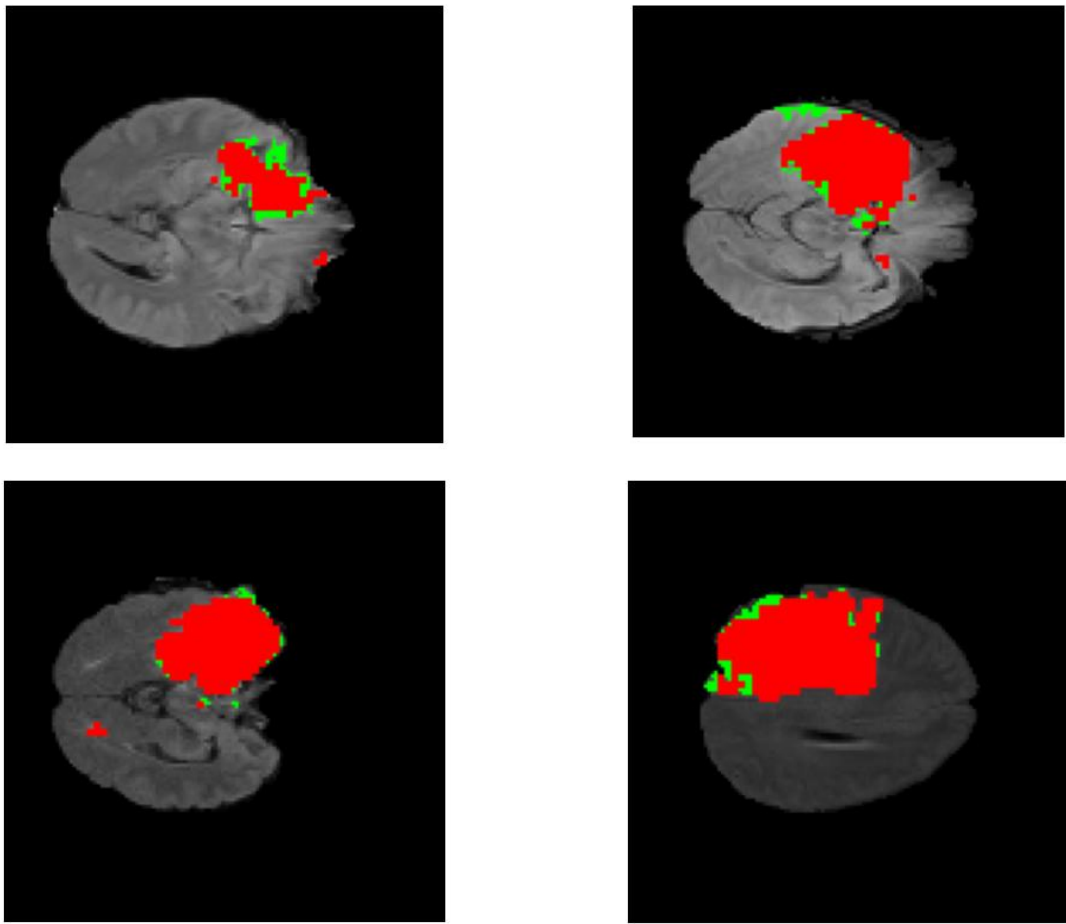
This baseline improves on the first for both Dice and boundary quality at great computational cost. Running Inference involves a multi-step iterative refinement process, and the observed runtime came out in thousands of seconds per case. The quality of the boundary improved, yet it was still not accurate enough for the irregular shape of the tumor. In simpler terms, it's not ready for real-time clinical applications as it is now.

Table 2. Preliminary baseline results vs. proposed targets

Metric	Baseline 1 (U-Net/GGMM)	Baseline 2 (Diffusion)	Proposed Target
Dice Coefficient	0.81 – 0.82	≈ 0.89	≥ 0.90
IoU	≈ 0.70	≈ 0.83	≥ 0.85

HD95	≈ 15	≈ 16	Reduced (improved boundary precision)
False Positives	Present, moderate-high	Reduced but still present	$\geq 15\%$ reduction vs. baseline U-Net
Inference Time	Low-moderate	Very high (≈ 2739 s)	Single-pass, real-time capable

A couple of patterns stand out here. For one, the two baselines sit at opposite ends of an accuracy-versus-efficiency trade-off - the U-Net/GGMM baseline is cheap to run but leaves accuracy and boundary precision below where they need to be, while the diffusion baseline closes some of that gap but at a computational cost that rules out routine clinical use.



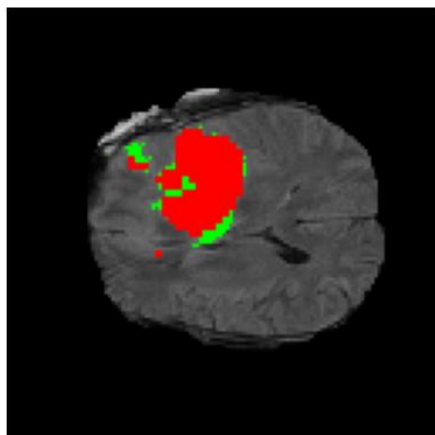
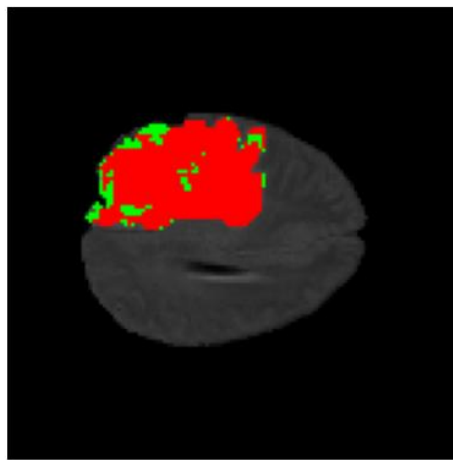
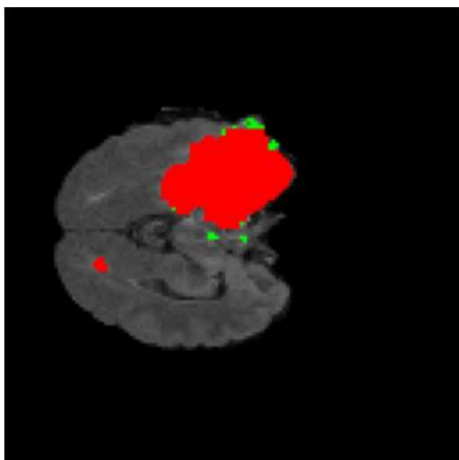
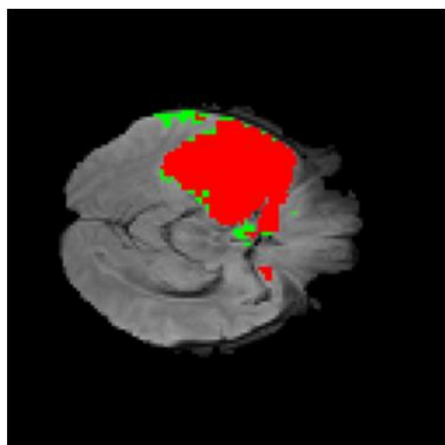
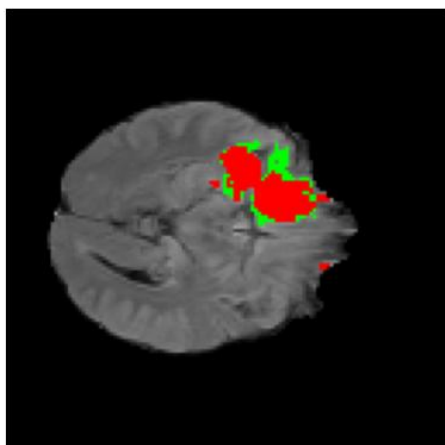


Figure 6. Tumor Detection (Baseline 1)

That trade-off is basically the reason the proposed framework goes with a single-pass hybrid design aiming for diffusion-level accuracy without the iterative overhead. Second point: both baselines back up the earlier claim that headline accuracy isn't worth much here, and Dice, IoU, and HD95 stay the metrics that matter when the proposed model gets evaluated.



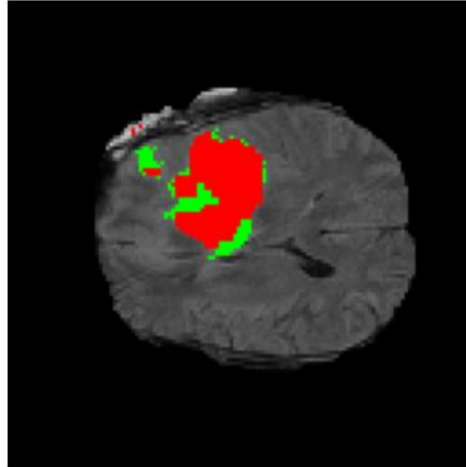


Figure 7. Tumor Detection (Baseline 2)

Dynamic thresholding and post-processing go after the false-positive and boundary-precision problems seen in Baseline 1 specifically, while the single-pass hybrid design goes after the inference-time problem from Baseline 2. Actually, testing Hypotheses H1 through H4 - comparative Dice/IoU/HD95 results, false-positive reduction, cross-modality generalization, noise robustness - is what comes next in this research.

There are several applications of the proposed system for brain tumor diagnosis and treatment in medical applications. It may assist with early diagnosis and screening of brain tumors, allowing clinicians to flag suspicious cases early. The system can also support treatment and radiotherapy planning by providing clinically relevant imaging-based information. Furthermore, it may aid surgical assistance and pre-operative planning, allowing medical professionals to better understand the nature and location of tumours before the operation. The system can serve as a decision-support tool in the hands of a radiologist, especially in high-throughput clinical areas requiring fast analysis of medical imaging. In addition, due to fast and low-compute inference ability, it can lead to important work for telemedicine and other remote healthcare delivery in low-compute healthcare environments or with specialized medical expertise.

6. Conclusion and Future Work

The objective of brain tumor segmentation from multi-modal MRI is to achieve a good balance between precision and computational efficiency. Existing diffusion-based and attention-based models fail to achieve both goals. This paper examined more recent literature on brain tumor detection and segmentation from 2024 to 2026, implemented on BraTS 2020 two representative baseline methods, and pulled out the repeated gaps: single-modality dependence, false positives, imprecise boundaries, high computational overhead, and pixel-wise accuracy being an unreliable metric. Based on those findings, a hybrid 3D U-Net and Mask R-CNN framework with multi-modal feature fusion, adaptive Dice-BCE loss, dynamic thresholding, and morphological post-processing is proposed to achieve a Dice coefficient of at least 0.90 and an IOU of at least 0.85, running sufficiently fast in a single pass for real-time application. The only remaining steps necessary to finalize the work are to implement and train the proposed model, run it against both baselines and other methods in the literature, test all four hypotheses (H1-H4), including cross-modality generalization and noise robustness, and assemble the results for a round journal submission.

REFERENCES

1. Chen A., Lin D., Gao Q. "Enhancing brain tumor detection in MRI images using YOLO-NeuroBoost model." *Front Neurol.* 2024 Aug 22;15:1445882. doi:10.3389/fneur.2024.1445882.
2. Krishnan P.T., Krishnadoss P., Khandelwal M., Gupta D., Nihaal A., Kumar T.S. "Enhancing brain tumor detection in MRI with a rotation invariant Vision Transformer." *Front Neuroinform.* 2024 Jun 18;18:1414925. doi:10.3389/fninf.2024.1414925.

3. Chattopadhyay A., Maitra M. "MRI-based brain tumour image detection using CNN based deep learning method." *Neuroscience Informatics*, Vol. 2, Issue 4, 2022.
4. Ranganathan V. et al. "Brain Tumor Classification From MRI Images Using Machine Learning." arXiv:2407.10630, 2024.
5. S. C. K. A, S. J. "Brain Tumor Detection Using Image Processing." 8th ICACCS, Coimbatore, India, 2022, pp. 582–587. doi:10.1109/ICACCS54159.2022.9785298.
6. Zubair Rahman A.M.J., Gupta M., Aarathi S. et al. "Advanced AI-driven approach for enhanced brain tumor detection from MRI images utilizing EfficientNetB2 with equalization and homomorphic filtering." *BMC Med Inform Decis Mak* 24, 113 (2024).
7. Bhagyalaxmi K., Dwarakanath B., Reddy P.V.P. "Deep learning for multi-grade brain tumor detection and classification: a prospective survey." *Multimed Tools Appl* 83, 65889–65911 (2024).
8. Hashemi S.M.H., Safari L., Dadashzadeh Taromi A. "Realism in action: anomaly-aware diagnosis of brain tumors from medical images using YOLOv8 and DeiT." arXiv:2401.03302, 2024.
9. Ahsan R., Shahzadi I., Najeeb F. et al. "Brain tumor detection and segmentation using deep learning." *Magn Reson Mater Phy* (2024).
10. Ullah F., Nadeem M., Abrar M., Al-Razgan M., Alfakih T., Amin F., Salam A. "Brain Tumor Segmentation from MRI Images Using Handcrafted Convolutional Neural Network." *Diagnostics* 2023, 13, 2650.
11. Kiran L., Zeb A., Rehman Q.N.U. et al. "An enhanced pattern detection and segmentation of brain tumors in MRI images using deep learning technique." *Front Comput Neurosci*. 2024 Jun 26;18:1418280.
12. Phan T., Hieu L. "Tumor Segmentation and Classification Using Machine Learning Approaches." *Intl. J. Data Informatics and Intelligent Computing*. 3, 1–11 (2024).
13. Xu X., Zhang Q., Sun P. "Automated brain tumor segmentation using a U-Net with hybrid training strategies." *IEEE Trans. Neural Networks and Learning Systems*, 35(6), 813–827, 2023.
14. Yang L., Ma H., Zhu M. "Brain tumor analysis using 3D CNN models on MRI datasets." *IEEE Access*, 11, 66899–66911, 2023.
15. Rathore S., Bhatt R. "Hybrid approach combining deep learning and image enhancement for brain tumor segmentation." *IEEE Trans. Biomedical Circuits and Systems*, 17(1), 86–95, 2024.
16. Chan C., Li J., Wei T. "A novel ensemble learning model for brain tumor classification." *IEEE Trans. Machine Learning in Biomedical Imaging*, 8(3), 52–62, 2023.
17. Zhang K., Zhao P. "Attention-guided deep learning for MRI-based brain tumor segmentation." *IEEE Trans. Biomedical Engineering*, 71(5), 840–850, 2024.
18. Patel A., Joshi D. "Enhanced segmentation of brain tumors using deep learning." *IEEE Trans. Cognitive Computing*, 12(2), 161–173, 2023.
19. Liu S., Cheng Y., Wang H. "Fusion of deep learning models for robust brain tumor classification." *IEEE Trans. Medical Imaging*, 43(3), 124–135, 2024.
20. Bhagyalaxmi K., Dwarakanath B., Reddy P.V.P. "Deep learning for multi-grade brain tumor detection and classification: a prospective survey." *Multimed Tools Appl* 83, 65889–65911 (2024).
21. Yuan J. "Brain tumor image segmentation method using hybrid attention module and improved mask RCNN." *Sci Rep* 14, 20615 (2024). doi:10.1038/s41598-024-71250-4.
22. de Verdier M.C., Saluja R., Gagnon L., LaBella D., Baid U., Tahon N.H. et al. "The 2024 brain tumor segmentation (BraTS) challenge: Glioma segmentation on post-treatment MRI." arXiv:2405.18368, 2024.
23. Bouhafra S., El Bahi H. "Deep Learning Approaches for Brain Tumor Detection and Classification Using MRI Images (2020 to 2024): A Systematic Review." *J Imaging Inform Med*. 2025 Jun;38(3):1403–1433.
24. Rao K.N., Khalaf O.I., Krishnasree V., Kumar A.S., Alsekait D.M., Priyanka S.S., Alattas A.S., Abdelminaam D.S. "An efficient brain tumor detection and classification using pre-trained convolutional neural network models." *Heliyon*, 10(17), e36773, Sep. 2024.
25. Pourmahboubi A., Arsalani Saeed N., Tabrizchi H. "A brain tumor segmentation enhancement in MRI images using U-Net and transfer learning." *BMC Med Imaging* 25, 307 (2025).
26. Alwadee E.J., Sun X., Qin Y., Langbein F.C. "LATUP-Net: A lightweight 3D attention U-Net with parallel convolutions for brain tumor segmentation." *Computers in Biology and Medicine*, 184, 109353, Jan. 2025.
27. Chowdhury S.R., Mirzaei G. "Hybridization of Attention UNet with Repeated Atrous Spatial Pyramid Pooling for Improved Brain Tumour Segmentation." arXiv:2501.13129, 2025.
28. R. P., M J.P.P., J.S. N. "Brain tumor segmentation using multi-scale attention U-Net with EfficientNetB4 encoder for enhanced MRI analysis." *Sci Rep* 15, 9914 (2025).
29. Wang R., Liu H., Wang G. "CMT-Unet: leveraging stage-wise hybrid framework for enhanced accuracy and efficiency in medical image segmentation." *Sci Rep* 16, 10079 (2026).
30. Yaseen M., Ali M., Kim H.-C. "Diffusion-Based Approaches for Medical Image Segmentation: An In-Depth Review." *Electronics*, 15(7), 1400 (2026).
31. BraTS 2020 Dataset. Kaggle: awsaf49/brats20-dataset-training-validation.