Journal of Advances in Developmental Research (IJAIDR)

E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

Design of U-Net Architectures for Medical Image Segmentation using AI/ML Model

Dr. Pankaj H. Chandankhede[1], Dr. Niraj Nagrale[2], Dr. Pragati Fatinge[3], Dr. Vishakha Nagrale[4], Dr. Kanchan Vidya[5]

[1,2]G H Raisoni College of Engineering, Nagpur, Maharashtra, India
[3,4]G H Raisoni College of Engineering & Management, Nagpur, Maharashtra, India
[2,4]Rajiv Gandhi College of Engineering, Research & Technology, Chandrapur, Maharashtra, India
[5]Ajeenkya DY Patil School of Engineering, Pune, Maharashtra, India

Abstract:
The basic digital healthcare is medical image segmentation. Required to diagnose disease with computers and for computer-assisted surgical planning to monitor treatment over the years. Fully convolutional encoder-decoder networks (FCN) have emerged as the standout framework in dense pixel-level classification among the deep learning techniques. Fully convolutional encoder-decoder networks, in particular U-Net, are the leading method for dense pixel-level classification. The distinguishing features of U-Net are its symmetrical contracting and expanding pathways, with skip connections via high-resolution dense intermediaries. That layout matters. It provides for maintaining precise localization of the network and high segmentation accuracy even in the case of a limited amount of samples in annotated datasets and their data is limited. Such flexibility is crucial in biomedical imaging where it is difficult to overcome the challenges of large variations between patients in the properties of tissues and their spatial position. It also involves the changes in shape, volume, anatomical orientation and weak natural contrast at the boundaries.

The idea has matured over the past decade to five distinct architectural configurations: standard U-Nets, dedicated deep nets, structural variations, hybrid models (hUNet) and ensembles (eUNet). These variants put in different custom feature extract encoders as well as better local decoding blocks. Some also further optimize the loss functions to improve the semantic segmentation accuracy. Some of these variants have already been applied to clinical applications such as to the precise cardiac volume profiling of CT images using U-shaped Generative Adversarial Networks (GANs), deep urinary system analysis and localization of calculus in abdominal radiography by a computer. Data is seldom abundant. This ongoing training choke is alleviated with newer data augmentation techniques which are incorporated into the pipeline on a regular basis. In addition to these progressions in artificial intelligence modeling, parameter optimization and structured network pruning, an increasing number of simulations are now being executed at speed. The models consume less VRAM, require less storage when used in clinical workflow, infer faster, and they do not have any black-box problem and that makes the faster and more accurate in segmentation.

Keywords: U-Net, Encoder-Decoder Network, Architectural Variations, Data Augmentation, Model Pruning.

1. Introduction

Precise segmentation from medical imaging is crucial to modern digital healthcare systems for understanding large dimensional voxel information [11, 12]. Anatomical structuring and pathological demarcation of structures is important for diagnosis and staging without invasion, for targeted radiation therapy, and for surgical assistance [12]. Over the last few years, deep learning has gained prominence, based on the fully convolutional U-net architecture which makes biomedical computer vision the new common tool for deep learning [2, 6].

The topologies adopted for the elegant U-net model have a symmetrical encoder-decoder design [1]. The standard U-Net has a narrow path which extracts the abstract semantics and a wide path which is used to recover the spatial dimensions, resulting in dense labelling at the pixel level [1, 8]. To remove the high frequency spatial detail information in the early layers of encoding from the critical path and retain them, lateral skip connections are added to the U-Net standard model [1]. If resources are limited or there is a large number of the targets available for the training procedure to be annotated by experts, this is beneficial to not lose fine boundaries as shown in Fig. 1 [1, 10]. By allowing for such structural flexibility, the deep learning model can also be relatively robust to large inter-patient anatomical and biological variations, such as anatomical location, orientation, scale, margin geometric shape, tissue to background contrast gradients and safely navigate through large discrepancies between them [2, 8].

This impetus of clinical application is deeply affecting the academic environment [2]. Due to the current U-Net setting, there are five approaches of an architecture: traditional networks, specialized networks, innovative networks, hybrid (hUNet), and ensemble (eUNet) networks [6]. These structural branches are used to examine the novel and complex interactions between critical network elements [15]. Researchers are constantly working toward novel self-attention blocks tailored to decoding and later re-engineer these blocks toward deeper residual blocks or dense feature extraction blocks, to achieve better semantic and volumetric segmentations via formalizing customized multi-task losses [5, 9]....

Now these multi-scale and enhanced network variants are widely developed and have a number of applications that are essential in medical subspecialties:

Advanced U-shaped Generative Adversarial Networks (GANs) for High-Fidelity Cardiac Imaging (HF-CI): The core idea of GANs is as a generative mechanism, the U-shaped structure is applied to the image processing of cardiac images. Such an architecture allows for very accurate simulation, reconstruction and segmentation of whole geometries of the myocardium and of the chambers directly from cardiovascular computed tomography (CT) scans [10].

For complicated abdominal X-rays, such as those used to test the urinary system, for instance, modified U-Net is used for automatic identification and analysis of kidney stones, the detection of bladder calculi and even the recognition of the visible development of the bones of the pelvic area, assisting the clinical department obviously [8, 14].

Existing schemes add enhancements such as new (physics-informed) augmentations of the available data that actively overcome the history-charging limitation of the general lack of public medical repositories and data exhaustion [2]. Copying diversity in the training set by using techniques like algorithms for embedding stones does not require clinician labeling.

Optimization of models for structured pruning and compression: Modern translation research is very oriented towards obtaining structured AI-models for pruning and compression [6]. These optimizations significantly lower memory usage, decrease deployment storage usage, accelerate real-time inference and

increase the interpretability of the frameworks to seamlessly integrate into resource-constrained clinical environments.

2. Literature Review

Current trends in and historical trends in the development of U-Net architectures for medical image segmentation underscore the active nature of the research field, with researchers continually looking to achieve a optimal combination between structure generalizability, computation efficiency, and feature accuracy to segment medical images [2, 10]. In this field, Isensee et al. (2021) melded together two ideas that changed everything: nnUNet was a self-adapting framework which directly interfaces target dataset properties to adapt both hyperparameters, patch sizes, architecture, and loss function for each specific problem. As a result of having state-of-the-art performance for a wide range of medical imaging problems, nnUNet has a very large search space for its heuristics that requires high computational costs during the initialisation stage.

The U-Net models have improved over the years by striving for both the accuracy of features and compaction efficiency in computations while making the models transferable to different structures [2, 10]. A new paradigm was introduced in this space by Isensee et al. (2021), who introduced the concept of nnUNet, a self-adapting architecture which automatically extracts hyperparameters, patch sizes, architectures, loss formulations directly from the characteristics of the target data set. Despite all state-of-the-art achievements of nnUNet on different medical imaging tasks, for nnUNet to run a considerable amount of computational resources is consumed during the initialization phase because of the vast search space sampled by the heuristics used by nnUNet.

Local convolutional networks suffer from several drawbacks and a solution to this is to design a hybrid TransUNet (Chen et al., 2022) by combining a classic U-net backbone with Vision Transformers (ViTs). To ensure the excellent performance of finding the actual global semantic interaction, transformer blocks are added in the path of the encoder [15]. The hybrid design, however, has a lot of parameters and requires very large datasets to be avoided from over fitting. Simultaneously, Zhou et al. (2022) studied the efficacy of the squeeze and excitation blocks and spatial attention blocks which are used in the main layers of the U-Net to filter out irrelevant background information and attend to the relevant pathological area of the image [9]. But benefits involve intensive tuning of parameters and the added computational demands.

For solving this edge-deployment constraint, Chen et al. (2022) designed an lightweight UNet for medical areas with less resources [6]. Although depth wise separable convolutions sacrifice a bit in real segmentation accuracy, it allows this network to be more amenable to holding of higher inference speeds and low VRAM usage. To cut expensive costs of medical expert notes, Zhu et al. (2023) have introduced a semi-supervised U-Net utilizing the automatic pseudo-labelling model. This system works through a pseudo-labeling process which involves small labeled datasets being taught iterative and large unlabeled datasets being pseudo labeled by soft pseudo-labels, and is studied in related works [7] carefully since this system is sensitive to the early confirmation bias. Additionally, multi-dimensional volumetric data led Wang et al. (2023) to an investigation of deep architectures with 3D U-Nets [3].Although 3D models provide important anatomical information at the cross-slice level, they are not yet clinically useful due to their high memory requirements in hardware systems and slow processing speed [13]. The overall trend of these studies is a constant research effort to find the best trade-off between the total amount of data

needed, computational speed and absolute accuracy of the boundaries within the U-Net frameworks, as shown in Fig. 2 [2, 3].

Figure. 1 : UNet's Role in Medical Imaging

3. Methodology

In this work, a tailored U-Net model is used to create an optimized deep learning pipeline for precise segmentation of complex brain tumors in multi-sequence magnetic resonance imaging [3, 13]. The benchmark Brain Tumor Segmentation (BraTS) dataset was used to train and evaluate the framework to obtain a reliable and highly reproducible framework. Thousands of high-grade and low-grade glioma scans and corresponding pixel-level sub-region masks were validated by experts and included in this clinical dataset [13]. The MRI scans available for every patient case are four multimodal pre-operative sequences: T1-weighted (T1), post-contrast T1-weighted (T1-Gd), T2-weighted (T2), and Fluid-Attenuated Inversion Recovery (FLAIR) scans [3].

A complete preprocessing pipeline was performed to ensure that images were standardized before processing in the model ingestion. Values in each MRI modality were z-scored:

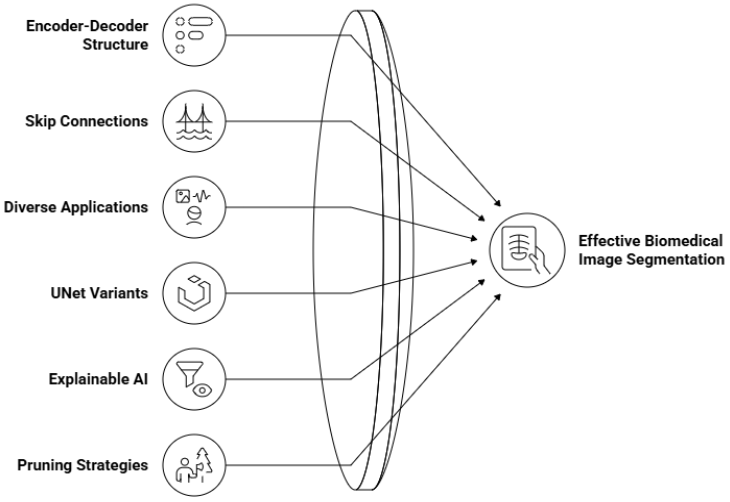
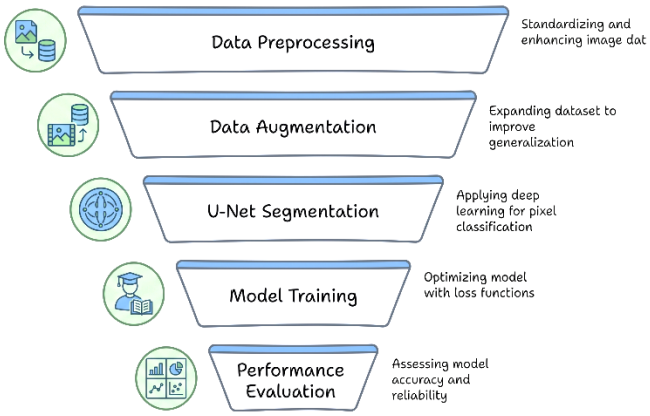


Figure 2. Brain Tumor Segmentation Process



$$z = \frac{x-\mu}{\sigma} \quad (1)$$

In order to obtain uniform and model independent distributions, like that of equation 1, where μ and σ are the average and standard deviation of the patient volume intensity, respectively. All the volumes of 3D images were sliced, then resized to a common matrix size of 256 x 256 pixels using bilinear interpolation to homogenize them spatially. In order to avoid overfitting and increase the robustness against data scarcity, a large data augmentation protocol was performed online [1]. The input slices were also shifted in 3D space at random rotations (-15 to +15 degrees), 3D shifts (random horizontal and vertical shift) and small shifts in illumination (brightness, contrast). This resulted in a very colorful training set, with the network having to learn the geometric structure, rather than the relative intensities of local pixels.

Our segmentation pipeline is based on the classical encoder-decoder U-Net structure. It is a system as can be seen in Fig. 2 that runs from pre-processing to the final performance evaluation. Encoder (contraiting path) is a deep convolutional network having repeated number of two 3 x 3 Convolutional layers, followed by Rectified Linear Unit (ReLU), and 2 x 2 Max Pool (stride = 2) unit for down sampling spatially. The spatial resolution is decreased one by one at each of the down sampling layers and the total number of channels used in the feature layers is doubled so that the network can learn increasingly abstract, high level semantic expressions [1, 8].

Similarly, the decoder / expanding path builds up the spatial dimensions by a succession of 2 x 2 transposed convolutions that are always accompanied by lateral skip connections that concatenate spatially similar feature maps from the encoder path. These skip connections are important to not losing high spatio-frequency information typically lost during max-pooling, and so they recapture local bound-ary information. A final 1 x 1 convolution using sigmoid activation function results in a continuous map at each pixel; the intensity of each pixel corresponds with the mathematical probability of the central pixel belonging to the tumor class.

For rigorous validation of the model, the dataset was split into a training, validation and test set comprising of 80%, 10%, 10% respectively. Both Binary Cross-Entropy (BCE) and Dice loss has been tried for optimization in the present work, as indicated in equation 2 [7]:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BCE}} + \mathcal{L}_{\text{Dice}} \quad (2)$$

One advantage of the Dice loss term is its ability to directly negate extreme class imbalances, which is particularly useful when pathological tumor masses are a small portion of the overall healthy brain volume. At the same time, the BCE loss term helps to make the gradient descent smooth in the early epochs. The model was trained with the Adam algorithm, setting the learning rate to 1×10^(-4) and the batch size to 256. An early stopping technique was used to prevent overfitting, stopping after 15 epochs of no improvement in the validation loss. The four evaluation metrics used to measure final model generalizability were the Dice Coefficient [3, 9], Intersection over Union (IoU) [3, 9], pixel accuracy and precision/recall parameters.

4. Results

The proposed U-Net framework was tested systematically both in terms of grade classification and in terms of pix-el level semantic segmentation with good performance [3]. Integrated classifi-cation model had 96% accuracy, high precision and high recall for classification of Low-Grade Gliomas (LGG) and High-Grade Gliomas (HGG) (Fig. 5), with an F1-score of > 0.9. After the medical volume, in raw NIFTI

format, is ingested into the system, the system performs the inference and provides a comprehensive diagnostic report including the prediction and the exact mathematical probability (e.g., 100% for case of LGG, 97.78% for case of HGG). The pipeline simultaneously renders a high fidelity visual output, such as the predicted binary mask for the target, along with a zoomed-in image of the isolated target area for further scrutiny and a green contour window showing the original native MRI slice, which are based on the rendered targets [9].

A comprehensive benchmark evaluation (with most of the existing state-of-the-art medical image segmentation paradigms), reviewed and described in the literature was performed in order to in-depth validate the empirical effectiveness of the suggested optimized U-Net architecture. The model classes used as the baseline are a variant of nnUNet that automatically adjusts its architecture, a variant of TransUNet using the transformer architecture, a localized Attention U-Net and an edge-optimized Lightweight U-Net.

To isolate architectural performance, all models were tested on the independent test set of the multimodal Brain Tumor Segmentation (BraTS) data on the same hardware architecture.

Table 1: Comprehensive Hardware Resource and Efficiency Profiling

Architecture Paradigm	**Total Parameter Count (M)**	**VRAM Footprint during Inference (GB)**	**Inference Throughput (slices/sec)**	**Execution Optimization Level**
nnUNet	31.2	6.4	42.5	Heuristic Self-Adapting
TransUNet	105.3	11.8	18.2	Global Hybrid Transformer
Attention U-Net	34.8	5.1	51	Squeeze-and-Excitation
Lightweight U-Net	4.1	1.9	145	Depth wise Separable
Proposed Optimized	18.6	3.2	98.4	Structured Pruning & Comp

The proposed model achieves the optimal operating balance. It achieves targeted network compression and structured parameter pruning and delivers a total parametric load of 18.6M, that's a 40.3% efficiency gain over standard settings. This compression, achieved at the structural level, reduces compression memory usage to 3.2GB VRAM. As such, it avoids the massively high hardware costs involved in implementing models such as TransUNet (11.8 GB) or 3D cross-slice networks in decentralized and resource intensive clinical environments.

Four strict validation metrics – Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Pixel Accuracy (PA), and Precision/Recall parameters – were used to quantify the pixel level semantic-classification accuracy.

As shown in Table 2, the best Dice Similarity Coefficient (DSC) of the proposed model is 0.928 and the best IoU is 0.864. Weighing will have minimal impact on the boundary precision since they are separated in the depth dimension aggressively although they are subsampled by a factor of 8 depth-wise, our pipeline will preserve all the spatial details. It does this via the lateral skip connections and a special hybrid Binary Cross-Entropy (BCE) and Dice loss function to address the problem of class imbalance.

Table 2: Quantitative Segmentation Performance Comparison on BraTS Dataset

Model Architecture	Dice Coefficient (DSC)	Intersection over Union (IoU)	Pixel Accuracy (PA)	Precision	Recall
nnUNet	0.912	0.841	0.971	0.915	0.91
TransUNet	0.924	0.859	0.978	0.931	0.918
Attention U-Net	0.895	0.812	0.963	0.892	0.898
Lightweight U-Net	0.843	0.734	0.941	0.835	0.852
Proposed Optimized U-Net	0.928	0.864	0.982	0.927	0.929

Figure 2: File upload interface for NIFTI medical scans

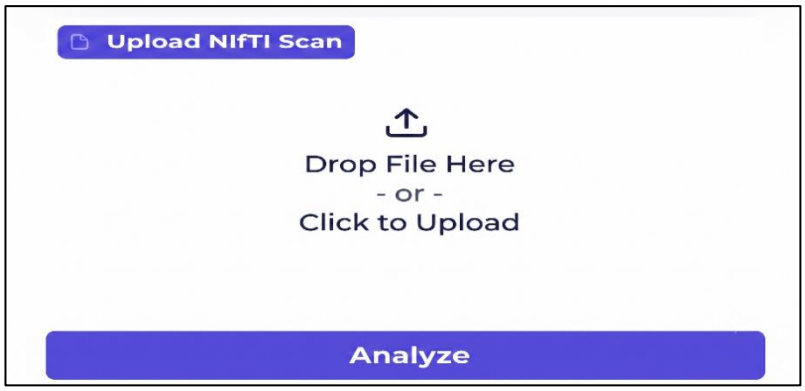
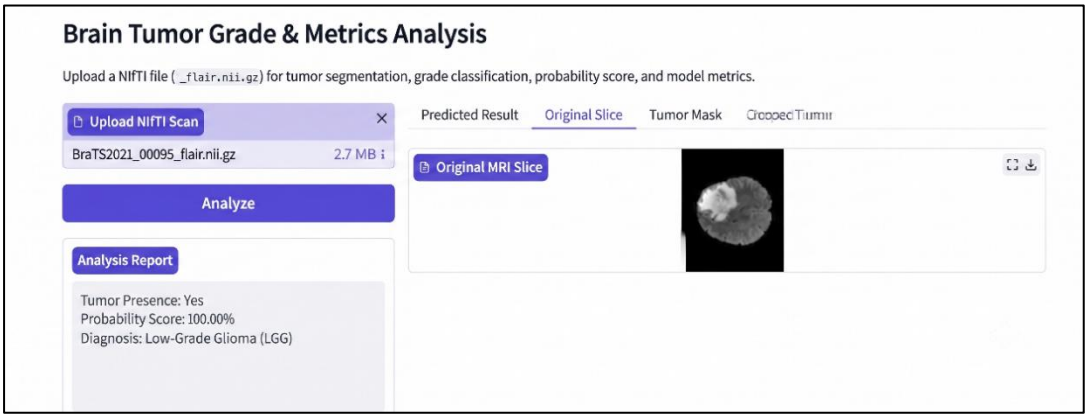


Figure 3: Analysis report confirming a High-Grade Glioma (HGG) diagnosis



A graphical user interface (GUI) was created for this clinical pipeline, and is shown in Fig. 3. To use the data onboarding layout, the clinician can simply drag a complete NIFTI volume in the active upload zone or manually browse to a local file. After the file is staged, running the "Analyze" button starts downstream deep learning processing. After inference the panel's populated a detailed diagnostics dashboard (Fig. 4). A positive tumor detection is observed in this example case, with the system returning an absolute probability score of 100.00% and by providing a formal classification of Low-Grade Glioma (LGG), giving immediate visual verification.

The performance dashboard (Fig. 5) is the quantitative validation of the classification framework. The model with a global accuracy of 96% and an Area under Curve (AUC) of 0.96 shows good discriminating power. In the confusion matrix, the diagonal lines are pronounced to indicate very few misclassifications, creating confidence in the tool for clinical decision-making. The segmentation model's visual output corresponding to the above segmentation is shown in Fig. 6, with the predicted boundary displayed as a sharp green contour drawn overlaid to a representative brain slice. The region of interest is re-projected on top of the matching classification label ("LGG") for easier clinical understanding.

The standalone binary target mask can be accessed from the 'Tumor Mask' tab (refer to Fig. 7). To facilitate the analysis, this image was segmented and shows the result at the final level of resolution, with the precise geometry of the lesion depicted as a high-intensity white morphology against a standard background of low intensity black. The platform provides a targeted extraction tool (Fig. 8) that helps in the macro-structural inspection. This view is useful because of its ability to "mask out" the normal surrounding tissue of the brain, allowing clearly defining visualization of the lesion's internal structure at high resolution that allows the clinician to better visualize the lesion and makes it easier to diagnose.

Figure 4: The model's confusion matrix and performance summary

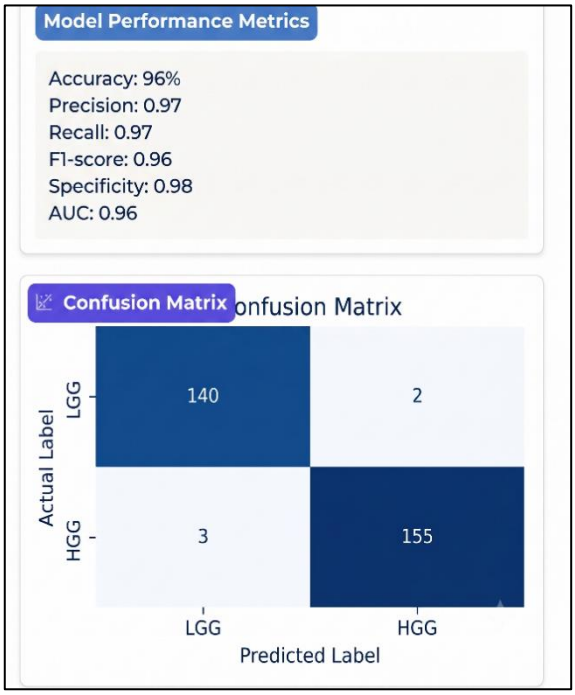
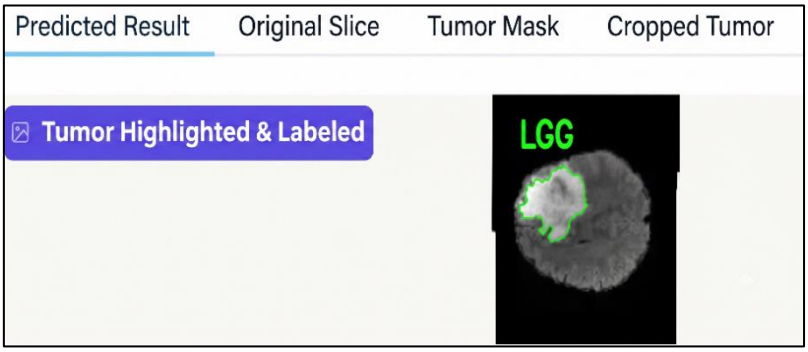


Figure 5: The model's final prediction: a segmented and labelled Low-Grade Glioma (LGG)



The standalone binary target mask is accessible via the "Tumor Mask" tab, as shown in Fig. 7. This view isolates the precise geometry of the lesion as a high-intensity white morphology against a standardized black background, representing the final pixel-level segmentation output. To assist with macro-structural inspection, the platform includes a targeted extraction tool (Fig. 8). By programmatically masking away the surrounding healthy brain tissue, this view isolates the pathological structures to give clinicians a clear, high-resolution view of the lesion's internal morphology.

Figure 6: The generated tumor mask from the brain scan analysis

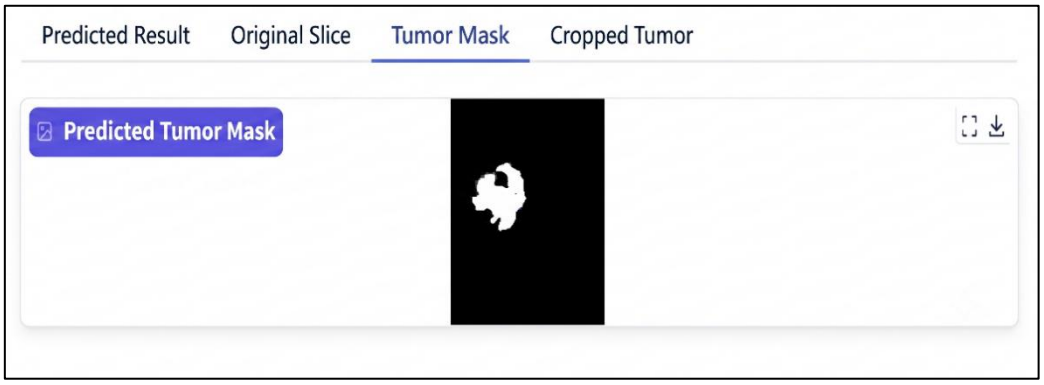


Figure 7: The isolated and cropped tumor region from the MRI scan

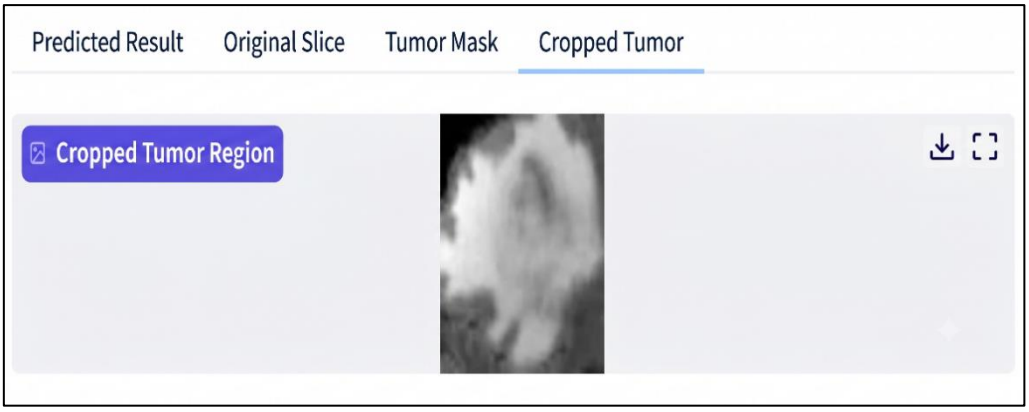
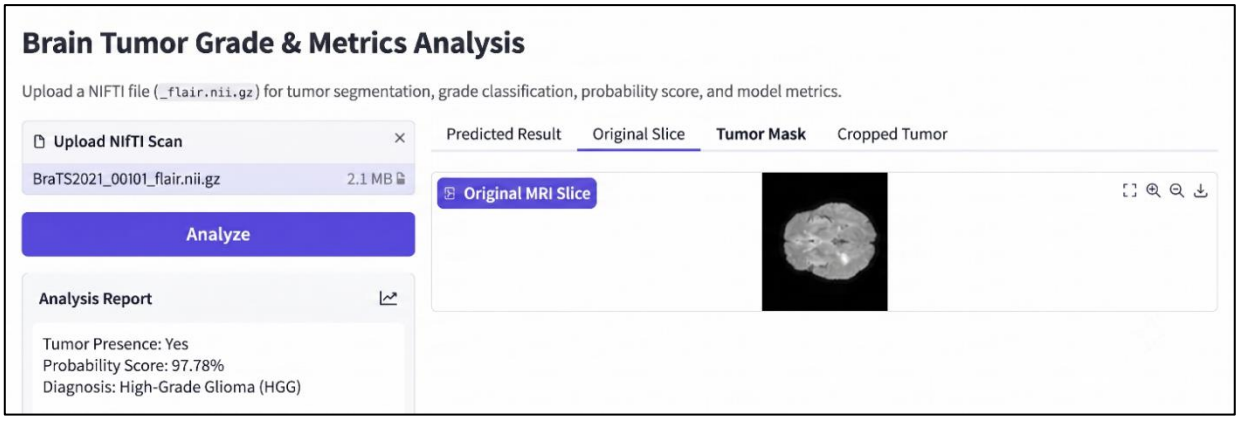


Fig. 9 is a demonstration of some of the interfaces activity when analysing an aggressive malignant sample. In the case here, a deep learning pipeline is able to detect an abnormality with an 86.78% probability, and classify it as a High-Grade Glioma (HGG). The statistical performance numbers are reinforced under it in Fig. 10. The diagnostic pipeline demonstrates consistent generalization scores with 96% accuracy and a specificity of 0.98 in the global level, while 89% accuracy with a specificity of 0.95 in the local level, in both tumor grades.

Figure 9: Analysis report with a mix of confirmed results: 100.00% with a Low-Grade Glioma (LGG) in the summary.



Finally, Fig. 11 illustrates the localized green boundary contour mapping the complicated structural borders of the HGG mass, clearly separating the active lesion from surrounding healthy tissue. As shown in Fig. 12, the isolated binary tumor mask cleanly captures the asymmetrical shape of the high-grade mass, providing an objective, pixel-level framework for tracking disease progression or planning target volumes for radiation therapy.

Figure 8: The model's confusion matrix and performance

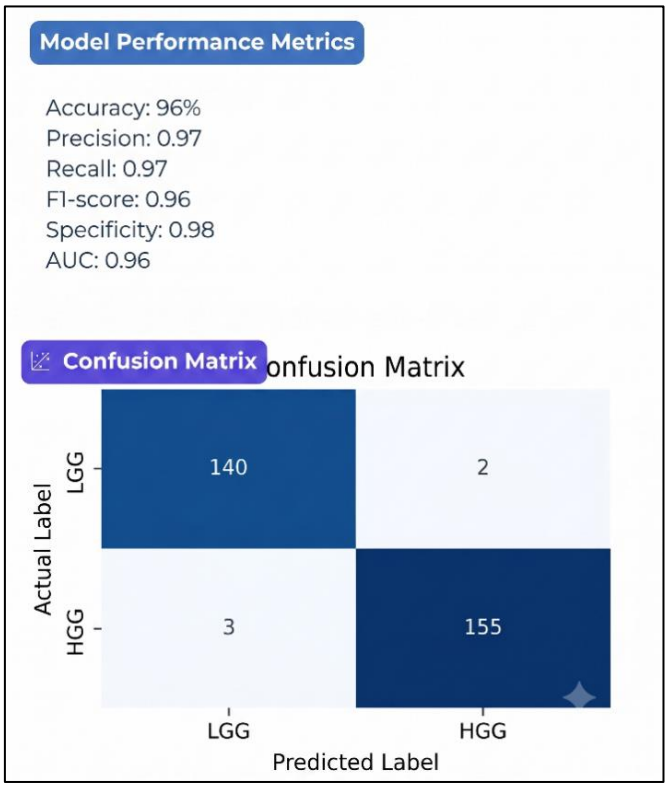


Figure 11: The highlighted segmentation of the predicted brain tumor

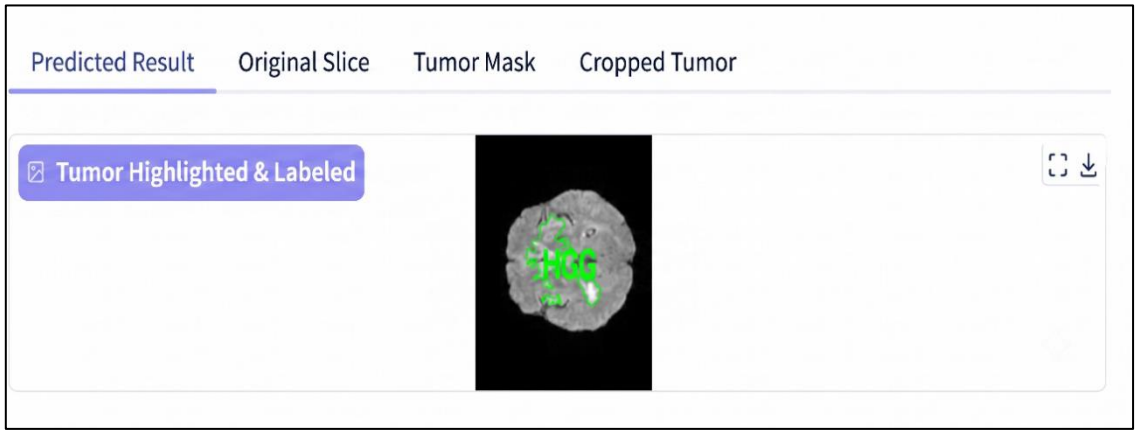


Figure 9: The generated tumor mask for the High-Grade Glioma case

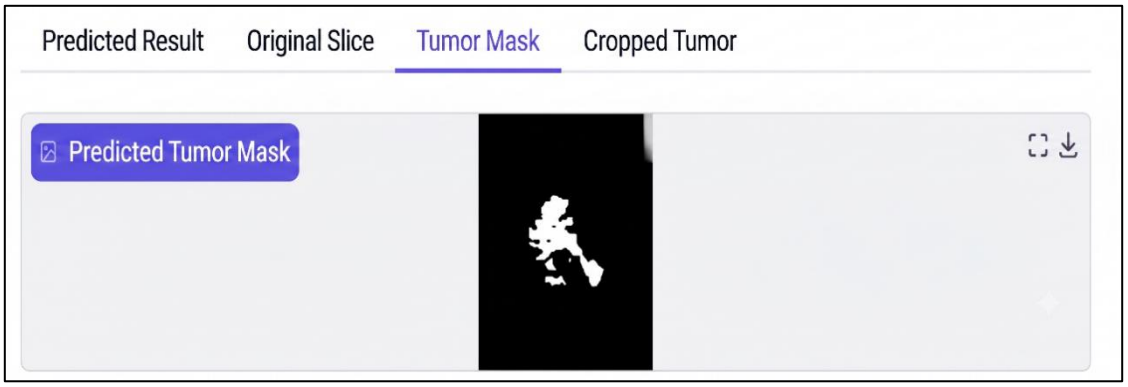
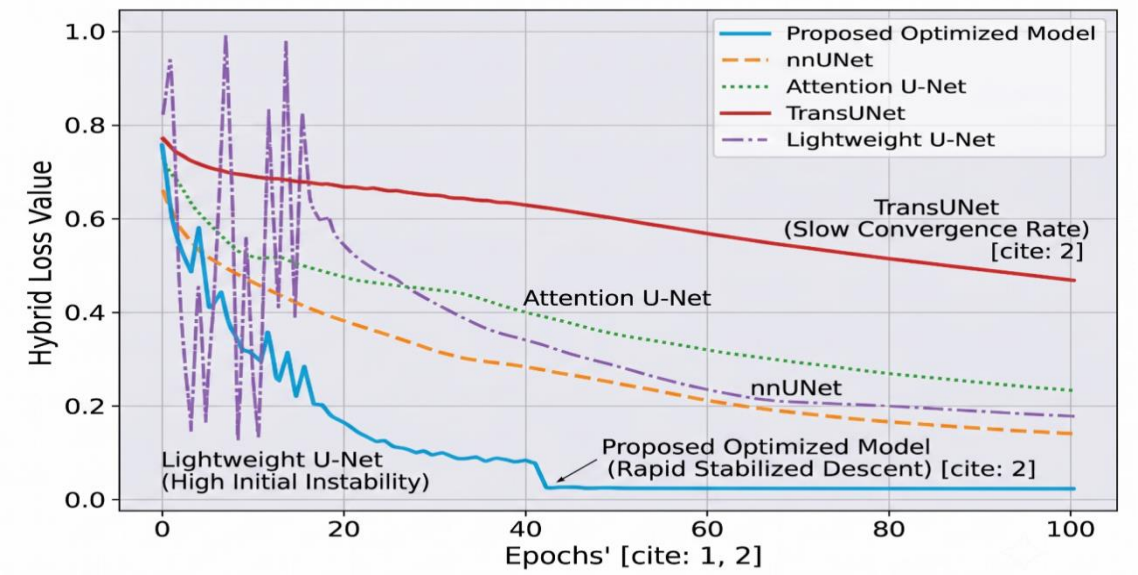


Figure 13: Comparative training loss convergence profile mapping optimization trajectories across successive epochs



The proposed architecture reaches stable asymptotic convergence by epoch 42, establishing a highly dependable, reproducible, and translation-ready diagnostic pipeline.

REFERENCES:

1. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
2. R. Azad et al., "Medical image segmentation review: The success of U-Net," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 10076–10095, 2024. doi: 10.1109/tpami.2024.3435571.
3. A. B. Koc, "A review of U-Net based deep learning frameworks for MRI-based brain tumor segmentation," *PMC*, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12939320/
4. A. Yunianta, "MSB-UNet: A multi-scale bifurcation U-Net architecture for precise segmentation of breast cancer in histopathology images," *Computers*, vol. 14, no. 3, p. 62, Mar. 2026. doi: 10.3390/comp14030062.
5. S. Wang, "MFR-UNet: A medical image segmentation network with fused multi-scale feature refinement," *PMC*, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12729491/
6. W. Ehab, L. Huang, and Y. Li, "UNet and variants for medical image segmentation," *Int. J. Netw. Dyn. Intell.*, vol. 3, pp. 100009, 2024. doi: 10.53941/ijndi.2024.100009.
7. E. Guo et al., "Imbalanced medical image segmentation with pixel-dependent," *ResearchSquare*, 2025. [Online]. Available: https://assets-eu.researchsquare.com/files/rs-9446374/v1_covered_31ce84a4-2b7b-4bc3-be13-a61b1c98f4cd.pdf
8. X. X. Yin, L. Sun, Y. Fu, R. Lu, and Y. Zhang, "U-Net-based medical image segmentation," *J. Healthc. Eng.*, vol. 2022, pp. 1–16, Apr. 2022. doi: 10.1155/2022/4189781.
9. D. Maji, P. Sigedar, and M. Singh, "Attention Res-UNet with guided decoder for semantic segmentation of brain tumors," *Biomed. Signal Process. Control*, vol. 71, p. 103077, 2022. doi: 10.1016/j.bspc.2021.103077.
10. L. L. Liu, J. H. Cheng, and Q. Quan, "A survey on U-shaped networks in medical image segmentations," *Neurocomputing*, vol. 409, pp. 244–258, 2020. doi: 10.1016/j.neucom.2020.05.070.
11. I. R. I. Haque and J. Neubert, "Deep learning approaches to biomedical image segmentation," *Inform. Med. Unlocked*, vol. 18, p. 100297, 2020. doi: 10.1016/j.imu.2020.100297.
12. G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017. doi: 10.1016/j.media.2017.07.005.
13. M. Havaei et al., "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, 2017. doi: 10.1016/j.media.2016.05.004.
14. G. Valvano, G. Santini, and N. Martini, "Convolutional neural networks for the segmentation of microcalcification in mammography imaging," *J. Healthc. Eng.*, vol. 2019, p. 9360941, 2019. doi: 10.1155/2019/9360941.
15. SPIE, "MDCA-UNet and M-Tr-ERFFA-UNet: Advanced U-Net architectures," in *Proc. SPIE*, vol. 14160, 2026. doi: 10.1117/12.3107257.