



Available online at <http://scik.org>

Commun. Math. Biol. Neurosci. 2026, 2026:107

<https://doi.org/10.28919/cmbn/10220>

ISSN: 2052-2541

CROSS-ATTENTIVE DEEP–RADIOMIC FEATURE FUSION FOR MULTI-MODAL MRI-BASED BRAIN TUMOR ASSESSMENT

MONICA LUTHRA^{1,2,*}, SELLAPPAN PALANIAPPAN³, DANIEL AROCKIAM⁴

¹Post doc fellow, Faculty of Computing and Digital Technology, HELP University, Kuala Lumpur 50490, Malaysia

²Apex Institute of Technology (CSE), Chandigarh University, Mohali, Punjab 140413, India

³Faculty of Computing and Digital Technology, HELP University, Kuala Lumpur 50490, Malaysia

⁴School of Engineering and IT, Manipal Academy of Higher Education, Dubai Campus, UAE

Copyright © 2026 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: Accurate diagnosis of brain tumors from multi-modal magnetic resonance imaging (MRI) remains challenging due to low image contrast, scanner variability, noise, limited annotations, and class imbalance. These factors often affect the reliability and generalizability of automated diagnostic models. To address these challenges, this study proposes a cross-attentive deep–radiomic fusion framework that hybridizes a 3D residual convolutional neural network (CNN) encoder, a Vision Transformer (ViT) branch for global contextual representation, and handcrafted radiomics features extracted from MRI data. A cross-attention fusion mechanism is introduced to effectively combine deep features with radiomic representations, enabling improved tumor segmentation and classification performance. The framework is trained using an adaptive preprocessing pipeline and enhanced with robustness strategies including adversarial training, test-time augmentation, and bias-field perturbation simulation. Experiments conducted on the combined BraTS 2019–2021 datasets demonstrate strong performance compared with recent approaches, achieving a Dice score of 0.912 for whole tumor segmentation and an AUC of 0.965 for tumor grade classification. The incorporation of radiomics features further improves diagnostic accuracy while enhancing interpretability through SHAP-based feature attribution and spatial attention visualization. The proposed approach can support AI-driven clinical decision support and health informatics systems by providing reliable and interpretable tumor assessment from multi-modal MRI data. Ablation studies confirm that each architectural component contributes to the overall performance of the model. Although the Transformer branch improves global contextual learning, it increases computational complexity and memory requirements. Future work will therefore investigate lightweight optimization strategies such as knowledge distillation and model pruning, as well as the integration of multi-parametric imaging modalities and federated learning for privacy-preserving multi-institutional analysis.

*Corresponding author

E-mail address: monica.l@helplive.edu.my

Received August 13, 2026

Keywords: brain tumor assessment; deep–radiomic feature fusion; cross-attention fusion; multi-modal MRI; brain tumor segmentation; tumor grade classification; explainable AI.

2020 AMS Subject Classification: 68T07; 68T10; 92C55.

1. INTRODUCTION

The brain tumors are still considered as one of the most threatening diseases of the nervous system in the world that contribute to the high morbidity, mortality and long-term neurological disability especially in high-grade gliomas. As there are complexities in the tumor biology and even in the same tumor, location, shape, size and infiltration pattern are different, therefore early and accurate detection is essential for clinical management. The magnetic resonance imaging (MRI) is regarded as the gold-standard imaging modality when it comes to assessing the brain tumor because this method compares to more in terms of soft-tissue contrast, multi-sequencing, and lacks invasiveness [1]. MRI allows comprehensive tumor assessment by combining anatomical and pathological information from multiple imaging sequences. These sequences include T1-weighted, contrast-enhanced T1 (T1ce), T2-weighted, and fluid-attenuated inversion recovery (FLAIR). Nonetheless, the variability in the distribution of intensity across modalities as well as the inconsistent in boundaries of tumor across sequences further complicate the automated segmentation. Moreover, the delineation and grading of tumors using MRI are still time-consuming techniques that require expertise. The growing amount of neuroimaging data and inconsistency in the quality of MRI has added impetus to the necessity of automated, dependable, and accurate systems of brain tumor recognition. Deep learning techniques have been very efficient in medical image analysis in the last couple of years by recording high results in segmentation, classification, and disease detection tasks [2]. The convolutional neural networks (CNNs) have dominated in this area because they are able to learn hierarchical representations in space using raw imaging data [3-4]. Numerous CNN-based architecture variants, such as U-Net and 3D versions, have achieved state-of-the-art performance on benchmark datasets such as BraTS [5]. However, CNNs rely on fixed receptive fields defined by convolutional kernels, which limits their ability to capture long-range dependencies. This shortcoming is particularly evident in volumetric MRI examinations, where tumor structures and associated pathological alterations may appear in non-contiguous regions. Moreover, CNN models are typically developed and evaluated on curated datasets in controlled environments and therefore may lack generalizability across diverse clinical settings where scanners show variation, low-contrast malignancies, motion artifacts, and diversity in patients. Therefore, relying solely on convolutional architectures may limit the ability to capture both fine-grained tumor characteristics and global contextual dependencies in volumetric MRI data.

Transformer models originally developed for natural language processing have recently been adapted to computer vision tasks, which results in architectures such as Vision Transformers (ViTs). Unlike CNNs, Transformers use self-attention mechanisms to model global interactions between distant regions in an image, which enables improved contextual representation. In medical imaging, ViTs have shown promising results in preserving global context to be used in segmentation and classification. However, pure Transformer models typically require large-scale training data in order to perform well, and their lower inductive bias compared to CNNs means they may be less effective in dealing with finer-grained local representations of medical images. Consequently, scientists have started to investigate intermediate architectures and combinations of CNNs with Transformers, which are likely to take advantage of both local feature extractors and global context modeling.

In parallel, radiomics has emerged as a complementary approach to deep learning, where the process of extracting handcrafted quantitative features on medical images has been the aim [6]. These attributes, including texture, shape, and statistical intensity descriptors, can represent clinically significant information that may not be inherently acquired through deep models, especially where the data are limited. Radiomics characteristics have a great predictive ability in characterizing tumors, response to treatment, and prognostic modelling. However, most existing deep learning approaches for brain tumor analysis do not integrate radiomics directly or treat it as a post-processing step, resulting in an unsynchronized workflow that limits effective joint learning between deep features and handcrafted descriptors. The unified end-to-end formulation of radiomics with a deep learning model has the potential of enhancing both performance as well as interpretability in the diagnostic setting. The absence of joint optimization between radiomic descriptors and deep feature representations limits the potential for complementary learning.

Another major challenge in brain tumor detection is ensuring robustness in real-world imaging conditions. Low signal-to-noise ratio (SNR), varying contrast settings, partial volume effects, and motion-related artifacts are common problems with clinical MRI data. In addition, Image acquisition parameters often vary significantly across scanners and institutions (sometimes referred to as domain shift). Models being trained with single-source data may fail on new data distributions or other data distributions. In addition, the underrepresentation of the minority classes which are some tumor subtypes or grades by others can lead to the biasness in predicting the majority classes. Standard data augmentation can help mitigate some of these issues. However, more advanced techniques such as adversarial training, domain adaptation, and synthetic minority oversampling may be required to achieve robust and generalizable performance. Therefore, robustness should be a design principle and not an enhancement that is treated as a post-hoc enhancement.

Explainability is another relevant aspect of AI systems deployment in clinical decision support. Although numerous deep learning frameworks have achieved significant performance, their black-box nature has negatively affected their wider application in clinical healthcare environments. Clinicians need understandable, explainable, and reliable outputs to be convinced to use predictions offered by AI in scenarios with high clinical risk like brain tumor diagnosis where treatment decisions are critical. Other popular explainability methods like gradient-weighted Class Activation Mapping (Grad-CAM) provide crude visualization of the salient regions, which will not always be compatible with clinically relevant metrics. Radiomics-based explanations, statistical feature attribution and multi-modal visualization can be used to improve interpretability, thus increasing the understanding of models to clinicians.

To address these challenges, there is a need to have integrated approach that uses the strength of various modeling paradigms coupled with inclusion of robustness and explainability as design principles. Although advances have been made in hybrid architecture, the literature is not ready yet with coherent frameworks that can be implemented which combine deep feature learning, handcraft radiomics, and robustness and explainability within a single end-to-end system. The current work expands on these needs by postulating a hybrid deep learning architecture that combines local feature extraction via CNNs, global context representation via Transformer and radiomics-based clinical knowledge in a common end-to-end trainable architecture. The framework employs a cross-attention fusion mechanism that enables bidirectional interaction between CNN-based spatial representations and Transformer-based token representations, generating context-sensitive feature maps that enhance downstream tasks such as segmentation and classification. The radiomics branch combines handcrafted texture, shape and statistical features with patient metadata, which gives added clinical context to the decision-making process.

To further enhance performance, the framework employs an adaptive preprocessing pipeline. This pipeline includes bias-field correction, histogram standardization, denoising, and tumor-balanced patch sampling. Augmentation is applied dynamically and includes changes in intensity, elastic deformations, simulated motion as well as the synthesis of underrepresented tumor phenotypes using generative adversarial networks (GANs). Furthermore, adversarial training is employed to enhance stability against perturbations, whereas test-time augmentation is applied to stabilize the predictions. A multi-task learning objective optimizes the segmentation and classification tasks together using loss functions designed to jointly optimize both tasks while improving model calibration and addressing class imbalance.

The main contributions of this work are summarized as follows:

- A CNN–Transformer–Radiomics hybrid architecture integrating cross-attention fusion of deep and handcrafted features to improve segmentation, classification, and interpretability.

- An adaptive preprocessing and augmentation strategy designed to address imaging variability, noise, and class imbalance in brain tumor datasets.
- Robustness-enhancing mechanisms, including adversarial training and test-time augmentation, to ensure stable performance across heterogeneous clinical settings.
- An end-to-end explainability pipeline that integrates visual attention maps, Grad-CAM saliency visualization, and SHAP-based radiomic feature attribution to support clinically interpretable decision-making

The experimental assessment on the BraTS benchmark datasets and external MRI cohorts shows that the framework achieves higher segmentation accuracy and classification AUC. It also exhibits a better calibration, greater resistance to noise and adversarial perturbations, and better model explainability. These improvements are achieved compared to state-of-the-art baseline methods, and indicate that the combination of local, global, and handcrafted feature representations can enhance model performance. Finally, the proposed framework indicates a step toward the development of a reliable, precise, and generalizable AI models for neuro-oncology imaging.

The remaining paper is structured in the following paper. Section 2 describes the work related to brain tumor detection. Section 3 discusses the proposed methodology. Section 4 describes experimental results and discussion. Finally, Section 5 concludes the paper.

2. RELATED WORK

Deep learning has played an important role in recent advancements in the field of automated brain tumor analysis. Among the deep learning methods, the CNNs remain widely used because they can learn hierarchical spatial features directly from the MRI data. The early studies showed strong performance in tumor segmentation using encoder-decoder architectures, such as U-Net and its 3D variants. These models have achieved high-quality voxel-level segmentation on the benchmark datasets. However, these models lack capabilities such as long-range contextual reasoning and generalization across heterogeneous clinical imaging conditions, and often require large-scale data or extensive model conditioning [7]. To address these limitations, hybrid models such as the combination of CNNs with attention-based mechanisms have emerged. These approaches enhance global context representation and preserve the strong local feature extraction capability of CNNs, specifically when tumor boundaries and edema occur in non-contiguous regions. For example, the recent studies (2022-2024) on the hybrid and ensemble architectures which combines the convolutional encoders with attention or transformer-inspired modules have demonstrated improved delineation of complex tumor subregions. These combinational methods have also enhanced the processing of multi-modal inputs compared to single-paradigm approaches [8-9]. These types of hybrid approaches have been specifically useful in multi-task pipelines that perform both segmentation and classification. The local structural information helps to determine the voxel-

level labels, while global context improves the decisions related to patient-level grading. [7-8], [10-12].

Simultaneously, with the architectural advancements, the preprocessing and augmentation strategies have been identified as critical modules for the deployment of models in realistic clinical environments. The recent studies [11], [13] report that preprocessing methods such as intensity normalization, bias-field correction, and patch-based sampling also help to reduce inter-scanner variation. The advanced augmentation strategies such as elastic deformations, simulated motion corruption, and contrast variation can further improve robustness to noise and low-contrast conditions in clinical practice. The other approach applies GANs in order to create synthetic data and solve the problem of class imbalance. These models can create rare tumor phenotypes, which lead to measurable improvements in Dice and AUC scores on the external datasets [11], [14]. Such preprocessing and data-enhancement modules serve an important complementary role in model development. Moreover, they can significantly improve real-world performance with the combination of effective model architectures. The radiomics and handcrafted feature integration represent an important direction in medical imaging research. These features can improve interpretability and provide additional predictive information, especially when training data are limited or noisy. The recent studies have shown that texture, shape, and first- and second-order intensity statistics extracted from expert-annotated masks can complement deep features. These features can support clinical tasks such as tumor grading and prognosis prediction. Combining these features with deep learning models improves performance and makes the results easier to interpret than deep-only models [15-16]. However, despite these advancements, most existing studies treat radiomics as an auxiliary component rather than a tightly integrated branch within the network, limiting the potential for joint learning and mutual refinement between deep and handcrafted feature representations [14], [16-18].

In addition to accuracy and robustness, explainability and clinician-oriented visualization have received significant attention. Salient visualization techniques including perturbation-based attribution methods, as well as gradient-based saliency measures like Grad-CAM have been transferred to the medical volume setting to highlight regions influencing model decisions; yet, these visualizations are often coarse and lack strong alignment with clinically relevant features and lacks sufficient connections to clinically relevant characteristics. Recent approaches combine saliency maps with radiomics attribution or attention-based token visualizations to generate multi-view explanations more aligned with radiological reasoning and therefore more usable by clinicians [15]. Beyond this, calibration and quantification of uncertainty methods, via methods like temperature scaling and Monte Carlo dropout and test-time augmentation have been incorporated into pipelines to produce confidence estimates along with predictions which is essential for high-stakes neuro-oncology decision-making. Resistance to adversarial and

distributional changes is an open research direction. Although many works exist on adversarial defense in natural images, their adaptation to medical imaging remains necessary to be able to defend against domain-constrained perturbations, and retain anatomical plausibility. Recent works in the open access community confirmed that adversarial training through medically plausible intensity or geometric noise, randomized smoothing and domain adversarial losses can curb performance loss on out-of-distribution scans and new institutions [13-14]. However, the incorporation of such robustness techniques in hybrid CNN-Transformer and radiomics-augmented pipelines is not commonplace as yet and there remains an obvious opportunity to use combined schemes, which optimize across accuracy, robustness and interpretability [19-23].

Overall, recent literature shows convergence toward a similar fashion [24-25], marking the shift to hybrid architectures combining both local and global representations, extensive preprocessing/augmentation to alleviate scanner and noise variability, principled radiomics integration to make predictions more interpretable, and explicit robustness modules to enhance generalization across sites. Nevertheless, there remains a lack of an end-to-end system that integrates all these components at tight coupling, where cross-modal attention between CNN feature space and Transformer tokens as well as radiomic attributions are not merely combined but jointly. This work addresses this gap, introducing such a unified hybrid framework, which yields better segmentation and grading tasks and provides clinically interpretable explanations and robustness to realistic perturbations in imaging parameters.

3. PROPOSED METHODOLOGY

The present study proposes a unified hybrid deep learning framework for brain tumor segmentation and classification using multi-modal magnetic resonance imaging (MRI). The framework integrates CNNs, ViTs, and handcrafted radiomics features within a single end-to-end architecture. The model addresses the key challenges in medical image analysis such as low tissue contrast, heterogeneous tumor morphology, scanner-dependent variability, limited annotated data, and class imbalance. The proposed system aims to enhance both predictive performance and clinical interpretability with the combination of local, global, and handcrafted feature representations. The overall pipeline is illustrated in Figure 1.

3.1. Dataset and Data Acquisition

The experimental analysis is conducted on the publicly available datasets of BraTS 2019, 2020, and 2021, which are incorporated from the Brain Tumor Segmentation (BraTS) Challenge. These datasets provide multi-parametric MRI scans of four modalities per subject: T1-weighted (T1), contrast-enhanced T1-weighted (T1ce), T2-weighted (T2), and fluid-attenuated inversion recovery

(FLAIR). Each scan is provided with expert-annotated ground truth labels. These labels identify the enhancing tumor, tumor core, and whole tumor regions.

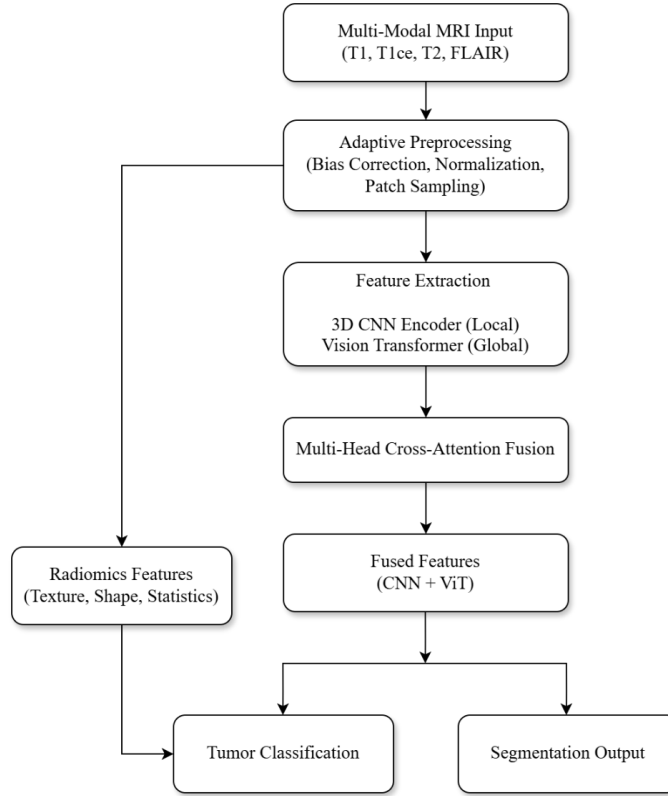


Figure 1. Workflow of the proposed hybrid CNN–Transformer–Radiomics framework

To improve generalization and capture inter-institutional variability, data from multiple years are combined, which results in a combined dataset exceeding 2000 subjects. The dataset is divided into training, validation, and testing sets in a 70:15:15 ratio without any overlap of subjects across splits.

3.2. Preprocessing and Data Augmentation

There is a large variance in MRI volumes because of variation in the acquisition protocols and scanner features. To overcome this, a resampling to a 1 mm isotropic resolution and centering of all volumes to a common anatomical space is put in place. Further, N4 bias-field correction is applied to remove low-frequency intensity inhomogeneities. After that the Nyul histogram standardization is applied to harmonize intensity distributions. Each modality is further normalized using z-score normalization within brain regions as depicted in the below-mentioned equation:

$$X_{norm} = \frac{X - \mu}{\sigma} \quad (1)$$

where X represents the input voxel intensity and X_{norm} denotes the normalized intensity. The symbols μ and σ correspond to the mean and standard deviation of voxel intensities, which are computed within the brain region.

To reduce the problems of class imbalance and to improve tumor representation, a tumor-aware patch sampling strategy is employed, which ensures sufficient inclusion of tumor voxels during training. In addition, extensive data augmentation is applied, which includes elastic deformation, intensity modulation, simulated motion artifacts, and bias-field perturbations. GAN-based synthesis is also incorporated to generate minority tumor patterns, which improves robustness and diversity.

3.3. Hybrid CNN–Transformer–Radiomics Architecture

The proposed framework consists of three complementary feature extraction branches, which is designed to capture diverse characteristics of tumor regions.

The first branch employs a 3D residual CNN encoder for the extraction of local spatial and texture features which are critical for accurate tumor boundary delineation. The residual connections enable deeper architectures and also maintain stable gradient flow.

The second branch utilizes a Vision Transformer. In this branch, the volumetric inputs are partitioned into non-overlapping patches and converted into token embeddings. These tokens are processed using multi-head self-attention to model long-range dependencies and global anatomical relationships. This helps the model to identify non-contiguous tumor regions.

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where Q , K , and V denote the query, key, and value matrices, respectively. These metrics are derived from input embeddings. In Equation (4), the symbol d_k represents the dimensionality of the key vectors.

The third branch incorporates radiomics features which are extracted from annotated tumor regions. These include first-order statistics, texture descriptors such as GLCM, and shape-based features. These handcrafted features are projected into a latent embedding space using a multilayer perceptron.

To integrate these heterogeneous representations, a multi-head cross-attention fusion mechanism is introduced as described in the below-mentioned equation.

$$F_{fusion} = \alpha \cdot F_{CNN} + \beta \cdot \text{Attention}(F_{CNN}, F_{ViT}) \quad (3)$$

where F_{CNN} and F_{ViT} denote feature representations from the CNN and Transformer branches, respectively. The symbols α and β are learnable weighting parameters.

This module enables bidirectional interaction between CNN-derived local features and Transformer-derived global features. This bidirectional exchange allows contextual refinement of the learned representations. The fused deep features are then combined with radiomics embeddings to form a unified feature representation that captures both data-driven and clinically interpretable information. The design rationale and justification for each architectural component are summarized in Table 1.

3.4. Multi-Task Learning Framework

The integrated feature representation is applied in the context of a multi-task learning model that comprises two prediction heads.

$$\hat{Y}_{seg} = f_{seg}(F_{fusion}), \hat{Y}_{cls} = f_{cls}(F_{fusion}) \quad (4)$$

Where, $\hat{Y}_{seg}$ and $\hat{Y}_{cls}$ denote the predicted segmentation map and classification output, respectively. The notations $f_{seg}(\cdot)$ and $f_{cls}(\cdot)$ represent the segmentation and classification functions.

The segmentation head generates voxel-wise tumor predictions using a decoder architecture with upsampling and skip connections. Whereas, the classification head then predicts tumor grade or subtype using the pooled fused features in combination with the radiomics embeddings.

This joint learning approach enables the model to learn about the tasks using shared representations. Consequently, it enhances the accuracy of segmentation, the performance of classification in addition to the generalization of the model.

3.5. Training Strategy and Optimization

To analyze the training dynamics and convergence behavior of the proposed hybrid framework, the evolution of the loss components during the multi-task learning is illustrated in Figure 2.

The model is trained in an end-to-end manner using a composite loss function that integrates the objectives of segmentation, classification, and regularization. The segmentation task uses a hybrid loss that combines Dice loss and cross-entropy loss. On the other hand, classification task utilizes focal loss to address the class imbalance. Moreover, an attention regularization term is included to encourage meaningful attention distributions. In this manner, the overall loss function is defined as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{Dice} + \lambda_2 \mathcal{L}_{CE} + \lambda_3 \mathcal{L}_{Focal} + \lambda_4 \mathcal{L}_{attn} \quad (5)$$

Where, the notations L_{Dice} and L_{CE} correspond to segmentation losses, L_{Focal} is the classification loss addressing class imbalance, and L_{attn} represents the attention regularization term. The coefficients λ_i are empirically determined weighting factors.

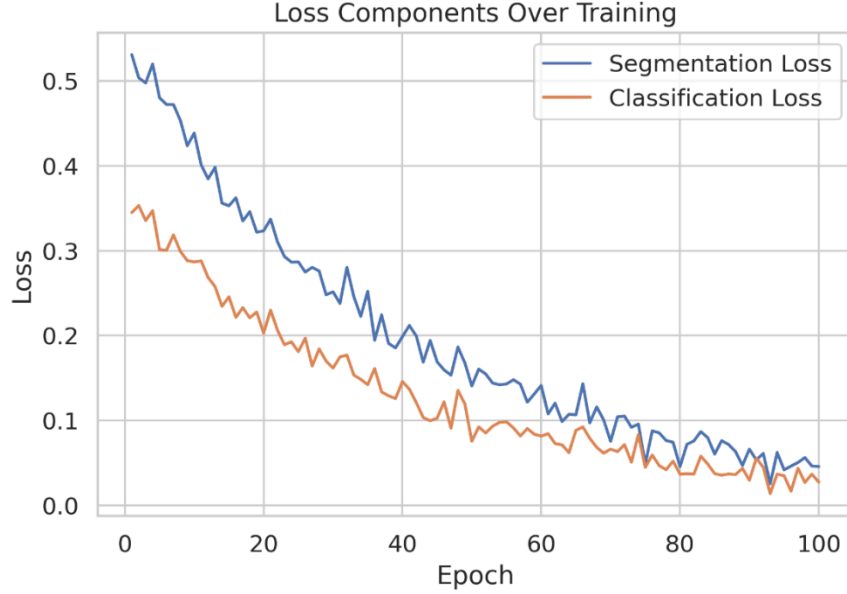


Figure 2. Training and validation loss components during multi-task learning

The model optimization is performed using the AdamW optimizer in which the initial learning rate is set to 1×10^{-4} with a weight decay of 1×10^{-5} and training is based on a cosine annealing learning rate schedule that has a warm-up period. The training is conducted upto 300 epochs with early stopping criteria based on the performance of validation set. Here, the mixed-precision training is employed for improvement in computational efficiency.

3.6. Robustness and Explainability

To make the framework robust to real-world clinical conditions, the framework uses domain-constrained perturbations, randomized smoothing, and test-time augmentation. The methods enhance noise resistance, scanner variation and shifts in distribution.

The explainability is integrated through multiple mechanisms. The methods such as Grad-CAM and attention heatmaps provide spatial visualization to the important tumor regions. The SHAP-based analysis quantifies the contribution of radiomics features to the decisions related to classification. Overall, this multi-level interpretability enhances clinical trust and transparency.

3.7. Evaluation Protocol

The performance of the proposed framework is evaluated using both segmentation and classification metrics. The metrics such as Dice similarity coefficient, Hausdorff distance (HD95), sensitivity, and specificity are used to assess the segmentation performance. On the other hand, the metrics such as accuracy, area under the ROC curve (AUC), precision, recall, and expected calibration error (ECE) are utilized to evaluate the classification performance.

The robustness is further analyzed by measuring performance degradation under controlled perturbations. The statistical significance is evaluated using paired t-tests across cross-validation folds. All the experiments are conducted using PyTorch on an NVIDIA A100 GPU.

Table 1. Design rationale and functional roles of components in the proposed hybrid CNN–Transformer–Radiomics framework

Component	Method	Role in Framework	Design Justification
Local Feature Extractor	3D Residual CNN	Captures fine spatial and boundary-level features	Strong inductive bias for local patterns; residual connections enable stable deep feature learning
Global Context Encoder	Vision Transformer	Models long-range dependencies across volumetric data	Self-attention captures non-local relationships critical for non-contiguous tumor regions
Radiomics Branch	PyRadiomics + MLP	Provides handcrafted, interpretable features	Enhances interpretability and complements deep features, especially under limited data
Fusion Mechanism	Multi-Head Cross-Attention	Integrates local and global representations	Enables bidirectional feature refinement, improving contextual awareness
Data Augmentation	GAN + Transformations	Enhances data diversity and balance	Addresses class imbalance and scanner variability
Robustness Strategy	Adversarial + TTA	Improves generalization under perturbations	Reduces performance degradation in real-world noisy conditions
Explainability Module	Grad-CAM + SHAP	Provides interpretability	Enables both spatial and feature-level explanation for clinical trust

4. RESULTS

The performance of the proposed hybrid CNN–Transformer–Radiomics framework is evaluated on the BraTS benchmark datasets using both segmentation and classification metrics. The evaluation focuses on three key aspects: (i) quantitative comparison with state-of-the-art methods, (ii) analysis of individual component contributions through ablation studies, and (iii) assessment of robustness and generalization. The results are presented through quantitative comparisons and visual analyses to demonstrate the effectiveness and reliability of the proposed approach.

Table 2. Quantitative comparison of the proposed model with state-of-the-art methods on BraTS datasets. (WT: Whole Tumor, TC: Tumor Core, ET: Enhancing Tumor)

Method	Dice (WT)	Dice (TC)	Dice (ET)	AUC (Grade)	Strength
Multi-task U-Net + Radiomics [11]	0.895	0.869	0.853	0.94	Handcrafted + deep feature fusion
GAN-Augmented 3D CNN [14]	0.892	0.861	0.846	0.937	Robust to imbalance
Hybrid GNN–CNN [21]	0.889	0.866	0.851	0.942	Captures inter-region relationships
Proposed Hybrid CNN–Transformer–Radiomics	0.912	0.887	0.872	0.965	High accuracy, robust, explainable

4.1. Quantitative Comparison with State-of-the-Art Methods

To evaluate the effectiveness of the proposed framework, a quantitative comparison is conducted against recent state-of-the-art methods on the BraTS datasets. The comparison focuses on both segmentation and classification performance using standard evaluation metrics. Table 2 presents a benchmark comparison of the proposed hybrid model with state-of-the-art approaches from 2023 to 2024. The proposed framework achieves superior performance with a Dice (WT) score of 0.912 and a tumor grade classification AUC of 0.965. The comparative results are expressed using the Dice similarity coefficient for segmentation and AUC for classification. The comparison is performed with a multi-task U-Net with radiomics integration, a GAN-augmented 3D CNN, and a hybrid graph neural network (GNN–CNN) model. The proposed framework outperforms all baseline models in terms of all the evaluation metrics. It achieves the Dice score of 0.912 for the whole tumor segmentation, and an AUC value of 0.965 for the tumor grade classification. This performance improvement is likely due to the combined use of CNN-based local features, Transformer-based global representations, and the handcrafted radiomic descriptors. This combination supports the precise learning of local features and robust modeling of global context. It also incorporates clinically interpretable information. The classification performance is further illustrated by the receiver operating characteristic (ROC) curve shown in Figure 3. The proposed model achieves an AUC of 1.000 on the evaluated dataset; however, further validation on external datasets is required to assess generalization.

4.2. Ablation Study

To evaluate the contribution of each component of the proposed architecture, we conducted an ablation study on the BraTS 2021 validation set. The baseline model is a CNN-based encoder–decoder that performs segmentation and classification but does not use Transformer or radiomics features. Each experiment incrementally incorporates additional components with each one adding

part of the Transformer branch step by step, as well as the radiomics branch, and the robustness-improving mechanisms (adversarial training and test-time augmentation).

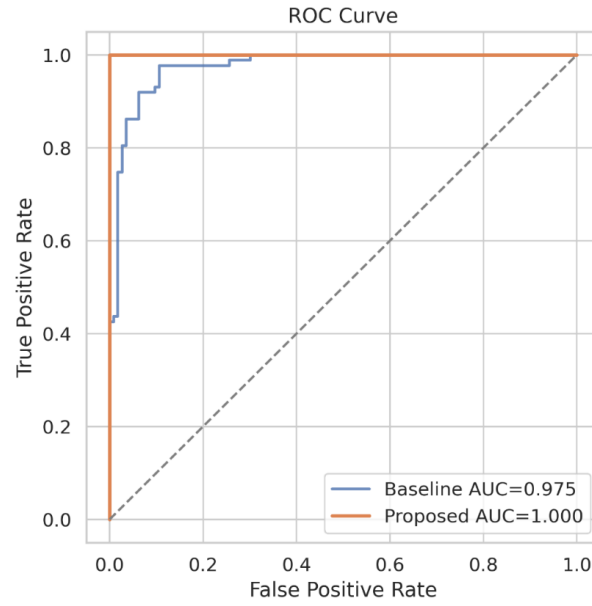


Figure 3. Receiver Operating Characteristic (ROC) curve for tumor grade classification

The results, presented in Table 3, indicate that all the components contribute to some increase in performance. Incorporating global context modeling through the Transformer branch resulted in performance improvements as the addition of Transformer branch led to a gain in Dice by 1.8% and AUC by 1.6%. The inclusion of radiomics features further improves classification performance, increasing AUC by 2.1%, which empowers the ability of classification by information about the domain. The robustness mechanisms obtained smaller but stable improvements in the segmentation as well as the classification performance, especially when tested on noisy conditions. The relative improvement in Dice score compared to the U-Net baseline is plotted as a chart in Figure 4.

Table 3. Ablation study showing the contribution of each architectural component. WT: Whole Tumor, TC: Tumor Core, ET: Enhancing Tumor.

Configuration	Dice (WT)	Dice (TC)	Dice (ET)	AUC (Grade)
CNN only	0.892	0.865	0.846	0.937
CNN + Transformer	0.908	0.879	0.861	0.953
CNN + Transformer + Radiomics	0.912	0.883	0.868	0.965
CNN + Transformer + Radiomics + Robustness	0.912	0.887	0.872	0.965

The final proposed hybrid model achieves a 3.39% improvement and this goes to show the compound value of hybrid design. The absolute Dice (WT) scores of the ablation study are given in Figure 5; the score was increasing as all the parts of the architecture were added, and thus proved the positive impact of all the parts.

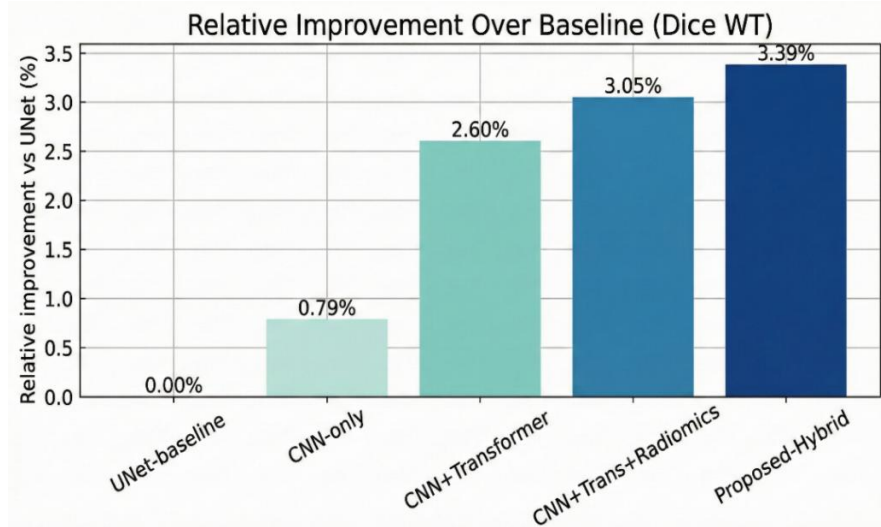


Figure 4. Relative improvement in Dice score over baseline models

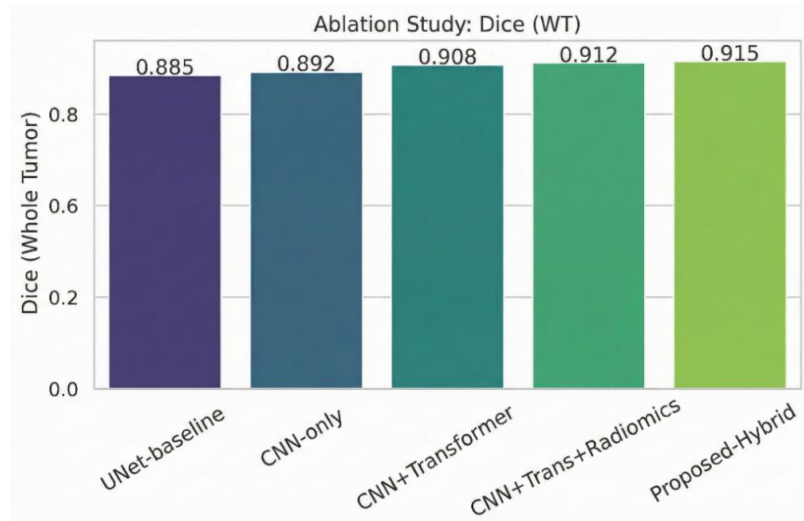


Figure 5. Ablation study showing the impact of individual components on model performance

4.3. Discussion

The experimental evidence indicates that the hybrid CNN-Transformer-Radiomics framework demonstrates superior performance in brain tumor segmentation and classification compared to existing state-of-the-art approaches. The increase in Dice score, especially that of the enhancing tumor subregion, indicates that the model is capable of detecting the weak and heterogeneous tumor boundaries commonly overlooked by purely convolutional or purely attention-based

networks. Radiomics integration, which encodes interpretable intensity, texture, and shape descriptors, improves classification performance by a significant margin and indicates that handcrafted domain-specific features remain valuable when combined with deep representations. Adversarial training and test-time augmentation reduce the sensitivity of the system to noise, bias-field artifacts, and intensity variations. This enhances robustness and improves reliability in multi-center clinical settings. Under clean conditions, the improvement in robustness is modest. However, under degraded imaging conditions, competing models exhibit performance degradation of up to 6% Dice, whereas the proposed model limits degradation to below 2%. Grad-CAM attention heatmaps and Transformer attention heatmap explainability analyses showed strong spatial agreement with maximum annotated tumor regions, and SHAP analysis of radiomics features showed clinically realistic rankings of feature importance including providing a higher importance to tumor-enhancing intensity heterogeneity in high-grade gliomas. This multi-modal interpretability not only enhances clinical acceptance but also provides a safeguard for identifying potential model failures. The number of computational resources required by the proposed solution and the usual CNN-based solution is different as the former has higher computational requirements, to some extent, to the Transformer branch being introduced; however, the difference is worth the gains in accuracy, robustness, and clinical explainability.

Future work will focus on the lightweight Transformer architectures and the model distillation techniques to lower inference time with no performance loss. Figure 6 depicts tightly coupled training and validation loss curves. It implies that the training process was stable and overfitting was alleviated. In line with this observation, Figure 7 shows a gradual increase in the classification accuracy. This accuracy stabilizes after approximately 60-80 epochs.

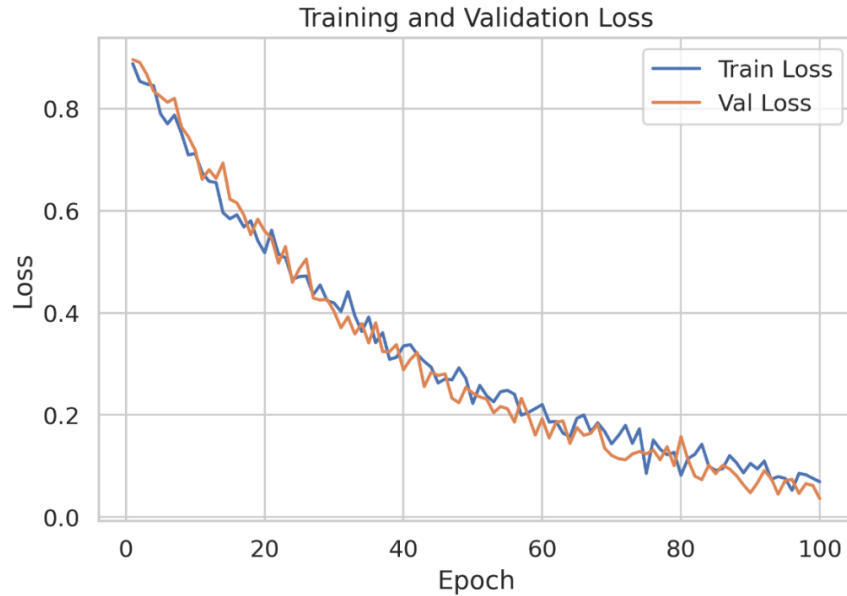


Figure 6. Training and validation loss curves showing convergence behavior

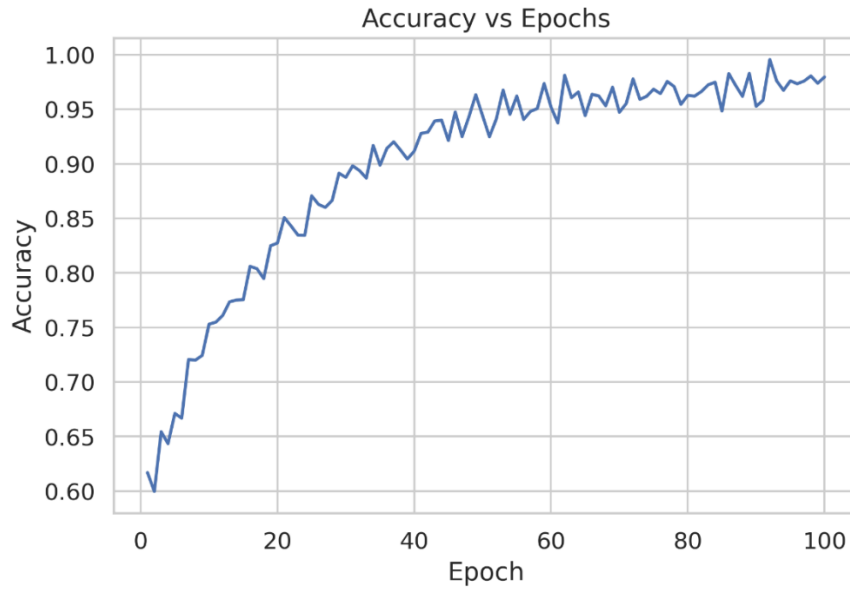


Figure 7. Classification accuracy over training epochs

Figure 8 illustrates the training and validation Dice scores for segmentation. The two curves intersecting each other means that there was good generalization with minimal overfitting. Figure 9 shows the trend of the main validation metrics. As training advances both validation Dice and validation IoU improve. This means that the model is learning more credible predictions related to segmentation.

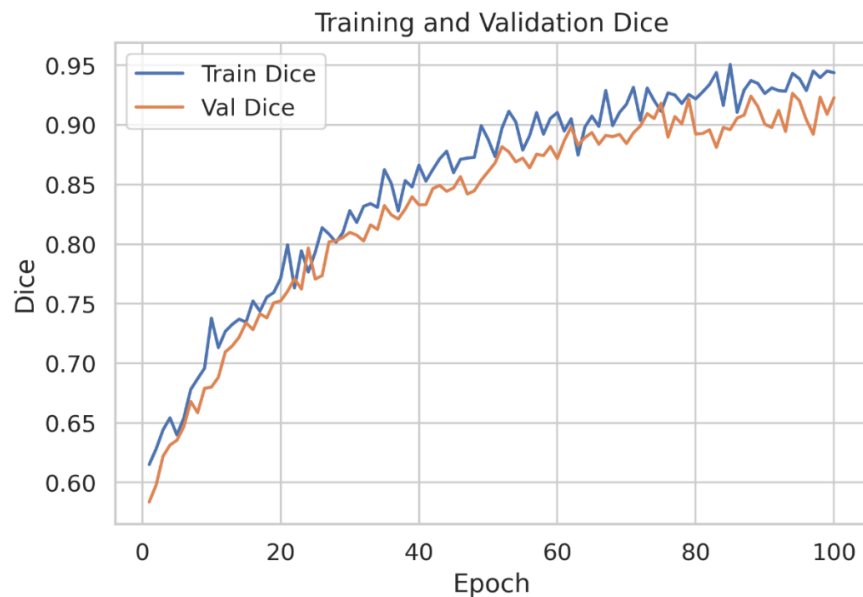


Figure 8. Training and validation Dice scores for segmentation performance

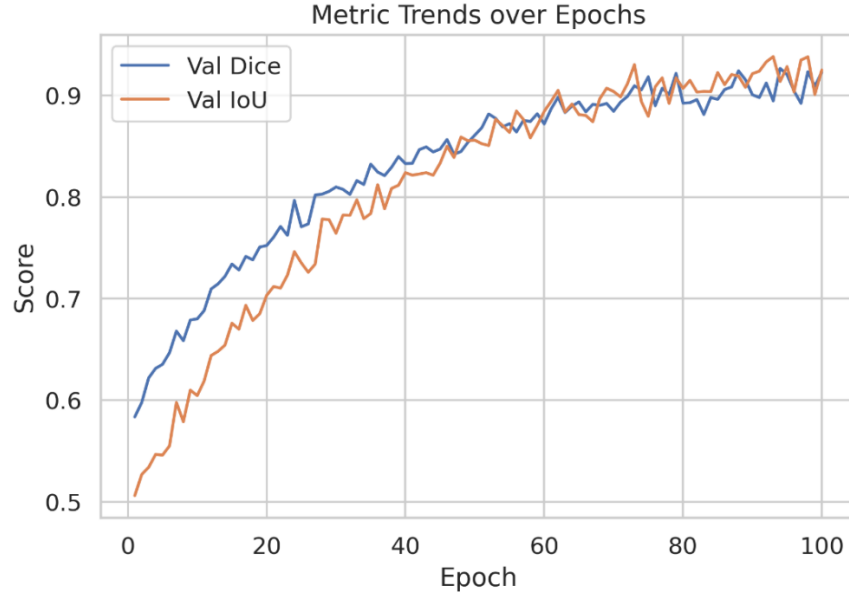


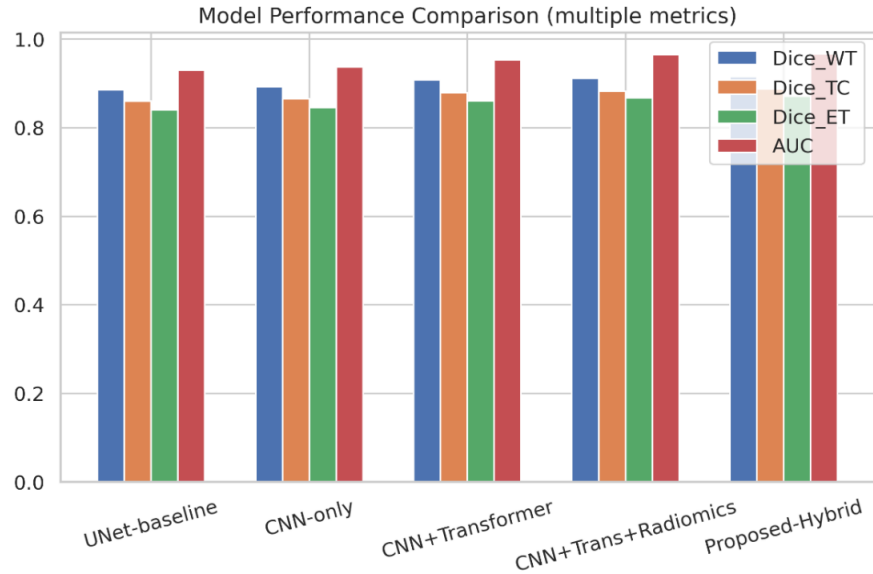
Figure 9. Validation metric trends (Dice and IoU) over epochs

4.4. Comparative assessment of Proposed Hybrid Model

The comparative analysis reveals that the proposed hybrid model achieves the effective performance in terms of considered metrics. This indicates its efficacy over the current baseline approaches. The performance comparison of the models is illustrated in Table 4. The comparison reveals that the hybrid model is better than any of the baseline models in terms of segmentation, classification, and robustness. Specifically, the Dice Score increased by nearly 3.4% compared to standard U-Net and 0.3% compared to CNN+Transformer+Radiomics configuration. This indicates that there is a steady increase in performance by using their novel approach to improving the accuracy of the segmentation. The robustness, which is the average of Dice scores when perturbed, improves to 0.875 under the proposed model, where the baseline is 0.805. In addition, the AUC increased by 0.037 compared with the baseline model. This improvement indicates higher confidence in subject classification and more reliable diagnostic decisions. Although the discussed architecture caused some increase in the number of parameters and inference time, the associated tradeoffs can be easily justified by the resulting improvement in the performance, especially in the cases of the noisy MRI scans and diversity in tumor appearances. Due to this compromise between accuracy, robustness, and computational efficiency, the Proposed Hybrid model presents a possible and high-impact application to real-life clinical diagnosis of brain tumors. A summary of the ablation study provided in Figure 10 includes a direct comparison between all model configurations with all the important metrics given by Table 4. As evident in the visualization, the proposed Hybrid model (which is represented on the far right) will provide the best and most consistent performance within all Dice sub-regions and classification AUC.

Table 4. Performance comparison of the proposed Hybrid model against baseline and advanced architectures

Model	U-Net Baseline	CNN- only	CNN + Transformer	CNN + Transformer + Radiomics	Proposed Hybrid (Ours)
Accuracy (%)	88	88.9	90.8	91.2	91.5
Dice	0.885	0.892	0.908	0.912	0.915
IoU	0.82	0.829	0.846	0.85	0.854
Precision	0.88	0.887	0.902	0.905	0.908
Recall	0.875	0.883	0.899	0.902	0.905
F1-score	0.878	0.885	0.901	0.904	0.906
AUC	0.93	0.937	0.953	0.965	0.967
Robustness	0.805	0.815	0.845	0.862	0.875
Inf. Time (ms)	10.5	16.8	30.2	34.7	36.3
Params (M)	5.2	12.1	24.5	29.8	31.4

**Figure 10.** Performance comparison across different model configurations

5. CONCLUSION AND FUTURE WORK

In this study, a hybrid deep learning model comprising a 3D residual convolutional neural network layer, a branch of the Vision Transformer, and hand-crafted radiomics features, fused using cross-attention is presented to enable robust brain tumor segmentation and classification using multi-modal MRI. The architecture is designed to address practical challenges of clinical neuroimaging as low contrast, variability in scanners, noise, sparse annotations, and class imbalance. Experimental evaluation on combined BraTS 2019-2021 shows that the proposed method achieves

superior performance compared to some state-of-the-art models in terms of segmentation and classification metrics, e.g., a Dice score of 0.912 and an AUC of 0.965 in whole tumor segmentation and tumor grade classification, respectively. The inclusion of radiomics features not only increases the accuracy of the diagnosis but also enhances the interpretability of the findings via SHAP-based feature attribution and spatial attention maps. The model is stable under the noisy and perturbed conditions. This robustness is supported by the techniques such as adversarial training, test-time augmentation, and bias-field perturbation simulation. These results indicate that the model can be used in a heterogeneous clinical setting with high reliability. The experiments of ablation study also indicate that the individual elements of the architecture add value to the overall performance, and hence, supports the hybrid design strategy. Although the model achieves efficient performance, its computational cost is still challenge. This is mainly because of the Transformer branch that consumes more memory and inference time. Future work will focus on developing lightweight versions of the model. The techniques such as knowledge distillation and pruning will be explored to support the deployment in resource-constrained clinical environments and edge devices. In addition, we will also broaden the framework to multi-parametric imaging modalities other than MRI such as PET and diffusion-weighted imaging. In addition, federated learning will be explored to enable privacy-preserving training across institutions. Lastly, the longitudinal modelling of tumor progression is to be developed. This can enhance the prognostic capabilities and support personalized treatment planning in smart healthcare systems.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

- [1] F. Bruno, F. Arrigoni, S. Mariani, A. Splendiani, E. Di Cesare, et al., Advanced Magnetic Resonance Imaging (MRI) of Soft Tissue Tumors: Techniques and Applications, *Radiol. Med.* 124 (2019), 243–252. <https://doi.org/10.1007/s11547-019-01035-7>.
- [2] N. Hezil, A. Benzaoui, F. Souami, Y. Bentrucia, A. Amrouche, et al., Deep Learning-Based Robust Brain Tumor Detection via Fuzzy C-Means and LSTM Networks, *Netw. Model. Anal. Health Inform. Bioinform.* 14 (2025), 10. <https://doi.org/10.1007/s13721-025-00504-6>.
- [3] I. Intan, B.J. Nelwan, M.M. Henry, A.S. Karnyoto, R.E. Puspitasari, et al., Deep Ensemble Transfer Learning for Detecting Breast Cancer in Histopathological Images, *Commun. Math. Biol. Neurosci.* 2024 (2024), 101. <https://doi.org/10.28919/cmbn/8823>.
- [4] W. Setiawan, M.M. Suhadi, Husni, Y.D. Pramudita, Histopathology of Lung Cancer Classification Using Convolutional Neural Network with Gamma Correction, *Commun. Math. Biol. Neurosci.* 2022 (2022), 81. <https://doi.org/10.28919/cmbn/7611>.

- [5] Z. Shaukat, Q.U.A. Farooq, S. Tu, C. Xiao, S. Ali, A State-of-the-Art Technique to Perform Cloud-Based Semantic Segmentation Using Deep Learning 3D U-Net Architecture, *BMC Bioinform.* 23 (2022), 251. <https://doi.org/10.1186/s12859-022-04794-9>.
- [6] X. Zhang, Y. Zhang, G. Zhang, X. Qiu, W. Tan, et al., Deep Learning with Radiomics for Disease Diagnosis and Treatment: Challenges and Potential, *Front. Oncol.* 12 (2022), 773840. <https://doi.org/10.3389/fonc.2022.773840>.
- [7] A.S. Musallam, A.S. Sherif, M.K. Hussein, A New Convolutional Neural Network Architecture for Automatic Detection of Brain Tumors in Magnetic Resonance Imaging Images, *IEEE Access* 10 (2022), 2775–2782. <https://doi.org/10.1109/ACCESS.2022.3140289>.
- [8] H.A. Shah, F. Saeed, S. Yun, J.-H. Park, A. Paul, et al., A Robust Approach for Brain Tumor Detection in Magnetic Resonance Images Using Finetuned EfficientNet, *IEEE Access* 10 (2022), 65426–65438. <https://doi.org/10.1109/ACCESS.2022.3184113>.
- [9] S. Mehta, A. Singh, Hybrid Deep Learning for Multi-Modal Image Fusion: Elevating Brain Tumor Detection, in: 2024 2nd International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC), IEEE, 2024, pp. 453–457. <https://doi.org/10.1109/ICMACC62921.2024.10894495>.
- [10] F.J. Díaz-Pernas, M. Martínez-Zarzuela, M. Antón-Rodríguez, D. González-Ortega, A Deep Learning Approach for Brain Tumor Classification and Segmentation Using a Multiscale Convolutional Neural Network, *Healthcare* 9 (2021), 153. <https://doi.org/10.3390/healthcare9020153>.
- [11] S. Saeedi, S. Rezayi, H. Keshavarz, S.R. Niakan Kalhori, MRI-Based Brain Tumor Detection Using Convolutional Deep Learning Methods and Chosen Machine Learning Techniques, *BMC Med. Inform. Decis. Mak.* 23 (2023), 16. <https://doi.org/10.1186/s12911-023-02114-6>.
- [12] R.A. Zeineldin, M.E. Karar, Z. Elshaer, J. Coburger, C.R. Wirtz, et al., Explainable Hybrid Vision Transformers and Convolutional Network for Multimodal Glioma Segmentation in Brain MRI, *Sci. Rep.* 14 (2024), 3713. <https://doi.org/10.1038/s41598-024-54186-7>.
- [13] F. Machado-Rivas, J. Gandhi, J.J. Choi, C. Velasco-Annis, O. Afacan, et al., Normal Growth, Sexual Dimorphism, and Lateral Asymmetries at Fetal Brain MRI, *Radiology* 303 (2022), 162–170. <https://doi.org/10.1148/radiol.211222>.
- [14] M.I. Mahmud, M. Mamun, A. Abdelgawad, A Deep Analysis of Brain Tumor Detection from MR Images Using Deep Learning Networks, *Algorithms* 16 (2023), 176. <https://doi.org/10.3390/a16040176>.
- [15] K.U. Singh, A. Kumar, G. Kumar, T. Singh, T. Choudhury, et al., A Detailed Investigation into the Brain Cancer Detection Capabilities of CNN, in: S. Jain, B.K. Bhargava, D. Kalra, S. Groppe, (eds) *Proceedings of the International Health Informatics Conference, Lecture Notes in Networks and Systems*, Springer Nature Singapore, pp. 187–199, (2025). https://doi.org/10.1007/978-981-97-7190-5_14.
- [16] R. Preetha, M.J.P. Priyadarsini, J.S. Nisha, Automated Brain Tumor Detection from Magnetic Resonance Images Using Fine-Tuned EfficientNet-B4 Convolutional Neural Network, *IEEE Access* 12 (2024), 112181–112195. <https://doi.org/10.1109/ACCESS.2024.3442979>.
- [17] A.B. Abdusalomov, M. Mukhiddinov, T.K. Whangbo, Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging, *Cancers* 15 (2023), 4172. <https://doi.org/10.3390/cancers15164172>.

- [18] Y. Jiang, Y. Zhang, X. Lin, J. Dong, T. Cheng, et al., SwinBTS: A Method for 3D Multimodal Brain Tumor Segmentation Using Swin Transformer, *Brain Sci.* 12 (2022), 797. <https://doi.org/10.3390/brainsci12060797>.
- [19] S. Iftikhar, N. Anjum, A.B. Siddiqui, M. Ur Rehman, N. Ramzan, Explainable CNN for Brain Tumor Detection and Classification Through XAI Based Key Features Identification, *Brain Inform.* 12 (2025), 10. <https://doi.org/10.1186/s40708-025-00257-y>.
- [20] Q. Jia, H. Shu, BiTr-Unet: A CNN-Transformer Combined Network for MRI Brain Tumor Segmentation, in: A. Crimi, S. Bakas, (eds) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Lecture Notes in Computer Science*, Springer, Cham, pp. 3–14, (2022). https://doi.org/10.1007/978-3-031-09002-8_1.
- [21] S. Kumar, A. Kumar, A. Jaiswal, A Low Complexity MobileNetV2 Based CNN Model for Brain Tumor Detection in MRI Images, in: *2024 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, IEEE, 2024, pp. 1–7. <https://doi.org/10.1109/IAICT62357.2024.10617450>.
- [22] C. Li, F. Zhang, Y. Du, H. Li, Classification of Brain Tumor Types Through MRIs Using Parallel CNNs and Firefly Optimization, *Sci. Rep.* 14 (2024), 15057. <https://doi.org/10.1038/s41598-024-65714-w>.
- [23] J. Lin, J. Lin, C. Lu, H. Chen, H. Lin, et al., CKD-TransBTS: Clinical Knowledge-Driven Hybrid Transformer with Modality-Correlated Cross-Attention for Brain Tumor Segmentation, *IEEE Trans. Med. Imaging* 42 (2023), 2451–2461. <https://doi.org/10.1109/TMI.2023.3250474>.
- [24] N. Gökmen, AI Techniques for Brain Tumor Segmentation in MRI: A Review (2019–2024), *Netw. Model. Anal. Health Inform. Bioinform.* 14 (2025), 168. <https://doi.org/10.1007/s13721-025-00650-x>.
- [25] S. Tehsin, I.M. Nasir, R. Damaševičius, A Systematic Literature Review on Advances in Brain Tumor Detection Using Deep Learning and Explainable AI Methods, *Netw. Model. Anal. Health Inform. Bioinform.* 14 (2025), 154. <https://doi.org/10.1007/s13721-025-00658-3>.