

NeuroDEEP-CNN: A Multi-Task Attention Framework for Joint Classification, Localization, and Segmentation of Brain Tumors on MRI

Hamza Shahbaz ^{1*} 

¹ Department of Computer Science, Government College University Faisalabad, Faisalabad, Pakistan

Abstract

Brain-tumor MRI analysis benefits from models that connect diagnostic discrimination with explicit spatial evidence. NeuroDEEP-CNN is a ConvNeXtTiny-based multi-task framework that jointly performs 39-class pathology-sequence recognition, lesion-center regression, and tumor-mask segmentation through a coordinate-mask attention gate. The study used 12,643 MRI images with a patient-level stratified 80/10/10 split and class-aware augmentation. At the selected checkpoint, validation top-1 and top-3 accuracies were 92.81% and 98.26%, lesion-center error was 18.56 pixels, and Dice was 0.6637; held-out diagnostics yielded macro ROC-AUC 0.9977, mean average precision 0.9493, and mean localization error about 18.1 pixels. The framework consolidates classification, localization, segmentation, and explainability in one inference pathway. Results support the feasibility of spatially coupled multi-task learning for fine-grained brain MRI analysis.

Keywords: brain tumor MRI; deep learning; multi-task learning; segmentation; explainable AI.

1. Introduction and Review of Literature

Magnetic resonance imaging remains the dominant non-ionizing modality for characterizing intracranial tumors because its sequence-dependent contrast reveals complementary structural and tissue information that is difficult to obtain from a single conventional image representation [1]-[6]. Recent open-access studies show that automated MRI analysis has moved from binary tumor detection toward fine-grained multi-class recognition, lesion localization, and region-level delineation [1], [2], [7]-[10].

Deep learning has accelerated this transition by learning hierarchical representations directly from images rather than relying on manually engineered descriptors, and contemporary DergiPark studies report strong performance from CNN, transfer-learning, hybrid, and transformer families under heterogeneous datasets and split protocols [1]-[8], [11]-[16]. Inception- and EfficientNet-based classifiers remain competitive on three- and four-class benchmarks, while recent ConvNeXt and vision-transformer comparisons indicate that performance depends strongly on class ontology, augmentation, and model-selection strategy rather than architecture name alone [2], [5], [6], [17].

Segmentation studies have developed in parallel because pixel-level tumor extent is needed to evaluate whether a classifier's decision is spatially coherent with the abnormal region [18]-[23]. U-Net, SegNet, 3D U-Net, DenseUNet+, and attention-residual variants have produced substantial Dice and IoU improvements on BraTS-style data, although these systems are typically optimized for segmentation alone and therefore do not directly resolve the classification-localization coupling examined here [18]-[23].

Classification research has concentrated on tumor-category discrimination using transfer learning, feature fusion, data augmentation, and compact CNNs [1]-[17], [24]-[26]. For example, Jagun et al. reported 92.10% test accuracy with InceptionV3 on a three-class Figshare benchmark, Özger et al. reported 96.73% with Inception-v4 on the same broad three-class problem family, and Saraç et al. reported 97.45% with VGG16 in a hybrid deep-learning/reinforcement-learning design [1]-[3]. These values are useful literature context but are not direct rankings against a 39-label pathology-sequence task.

Explainable artificial intelligence has become increasingly important because high predictive confidence does not by itself demonstrate that a network used anatomically relevant evidence [4], [15], [27]-[29]. Grad-CAM, Integrated Gradients, SmoothGrad, and occlusion-based sensitivity provide complementary gradient- and perturbation-based views of salience, but explainability is most informative when paired with quantitative localization or overlap criteria rather than visual inspection alone [4], [27]-[29].

The closest functional comparator identified in the 30-paper review is CrossTaskFormer, an open-access 2026 Journal of King Saud University study that jointly couples brain-tumor segmentation and classification using bidirectional cross-task attention and reports 99.30% four-class accuracy with Dice 0.8925 and IoU 0.8122 on BRISC under its full inference pipeline [30]. CrossTaskFormer is therefore selected as the principal contextual baseline because it shares multi-task spatial-categorical coupling, while its different

*Corresponding author

E-mail address: [to be completed]

Received: dd/mm/yyyy; Received in revised form: dd/mm/yyyy; Accepted: dd/mm/yyyy.

dataset, four-class ontology, and absence of explicit lesion-center regression prevent a controlled head-to-head interpretation [30].

Current brain-MRI pipelines often optimize classification, localization, segmentation, and explanation separately; this separation makes it difficult to determine whether a predicted class is supported by coherent spatial evidence in a fine-grained multi-class setting.

A single architecture that returns a class, lesion center, mask, and explanation can reduce duplicated feature extraction and can expose disagreement between categorical and spatial predictions within one research workflow.

The study evaluates 12,643 two-dimensional MRI images represented by 39 dataset-defined pathology-sequence labels and investigates internal classification, lesion-center localization, segmentation, attention coupling, and explainability. It does not claim prospective clinical validity.

RQ1: Can one ConvNeXtTiny representation jointly sustain high 39-label classification performance while producing usable lesion-center and segmentation outputs? RQ2: Does hybrid coordinate-mask feature gating provide a measurable advantage over single-task, no-gate, coordinate-only, and mask-only alternatives under the same training protocol?

Develop one 39-output classifier with coordinate and mask heads; target validation top-1 accuracy above 90%, top-3 accuracy above 97%, mean center error below 20 pixels on a 256×256 grid, and Dice above 0.65; evaluate four XAI methods.

The work contributes (i) an integrated 39-label classification-localization-segmentation architecture, (ii) a hybrid Gaussian-coordinate × semantic-mask attention mechanism, (iii) a single evaluation framework spanning categorical, geometric, pixel-overlap, calibration, and XAI evidence.

Table 1. *Comprehensive 2020-2026 literature matrix. Tick = explicitly evaluated component; cross = component not a primary evaluated output in the cited study. The final row reports the present framework.*

Study	Cls.	Loc.	Seg.	XAI	Attn.	Multi-task	Imbalance
Jagun et al. (2026) [1]	✓	X	X	✓	X	X	X
Özger et al. (2026) [2]	✓	X	X	✓	X	X	X
Saraç et al. (2025) [3]	✓	X	X	X	X	X	✓
Coşkun (2026) [4]	✓	X	X	✓	X	X	X
Gencer (2025) [5]	✓	X	X	X	X	X	X
Akboyraz & Kızılloluk (2026) [6]	✓	X	X	X	X	X	✓
Ulu (2025) [7]	✓	X	X	X	X	X	X
Sevinç et al. (2025) [8]	✓	X	X	X	X	X	X
Aslan & Özüpak (2025) [9]	✓	X	X	X	X	X	X
Bayaral et al. (2025) [10]	✓	X	X	X	X	X	X
Aslım & Çalışkan (2025) [11]	✓	X	X	X	X	X	✓
Kalender et al. (2025) [12]	✓	X	X	X	X	X	X
Özer (2024) [13]	✓	X	X	X	X	X	X
Eryılmaz Baran & Çetin (2024) [14]	✓	X	X	X	X	X	✓
Öztürk & Katar (2022) [15]	✓	X	X	✓	X	X	X
Doğanay & Yıldız (2022) [16]	✓	X	X	X	X	X	X
Şener & Ergen (2023) [17]	✓	X	X	X	X	X	X
Çay (2024) [18]	X	X	✓	X	X	X	X
Ateş et al. (2024) [19]	X	X	✓	X	X	X	X
Güvenç et al. (2023) [20]	X	X	✓	X	X	X	X
Bayram et al. (2022) [21]	X	X	✓	X	X	X	X
Abdalğadir & Öztürk (2025) [22]	X	X	✓	X	X	X	X
Cengil et al. (2023) [23]	✓	✓	X	X	X	X	X
Tüzün & Özdemir (2023) [24]	✓	X	X	X	X	X	X
Demir et al. (2023) [25]	✓	X	X	X	X	X	X
Yıldırım et al. (2023) [26]	✓	X	X	X	X	X	X
Alzahrani & Qahtani (2024) [27]	✓	X	X	X	✓	X	✓

Study	Cls.	Loc.	Seg.	XAI	Attn.	Multi-task	Imbalance
Rutoh et al. (2024) [28]	X	X	✓	X	✓	X	X
Çetiner & Metlek (2023) [29]	X	X	✓	X	X	X	X
Almnawer et al. (2026) [30]	✓	X	✓	X	✓	✓	✓
NeuroDEEP-CNN (92.81% val; Dice 0.6637; 18.56 px)	✓	✓	✓	✓	✓	✓	✓

Across the 30 studies, classification and segmentation are well represented, but simultaneous classification, explicit center localization, segmentation, and explainability remain uncommon. CrossTaskFormer [30] is retained as the principal baseline for multi-task comparison because it most closely matches the proposed cross-task design while remaining methodologically distinct.

2. Materials and Methods

This section defines the data, preprocessing, class-balancing strategy, model architecture, training configuration, and quantitative metrics used to evaluate NeuroDEEP-CNN. The workflow is organized so that dataset preparation precedes shared feature extraction, multi-task prediction, spatial gating, and final evaluation. Figure 1 summarizes the end-to-end sequence.

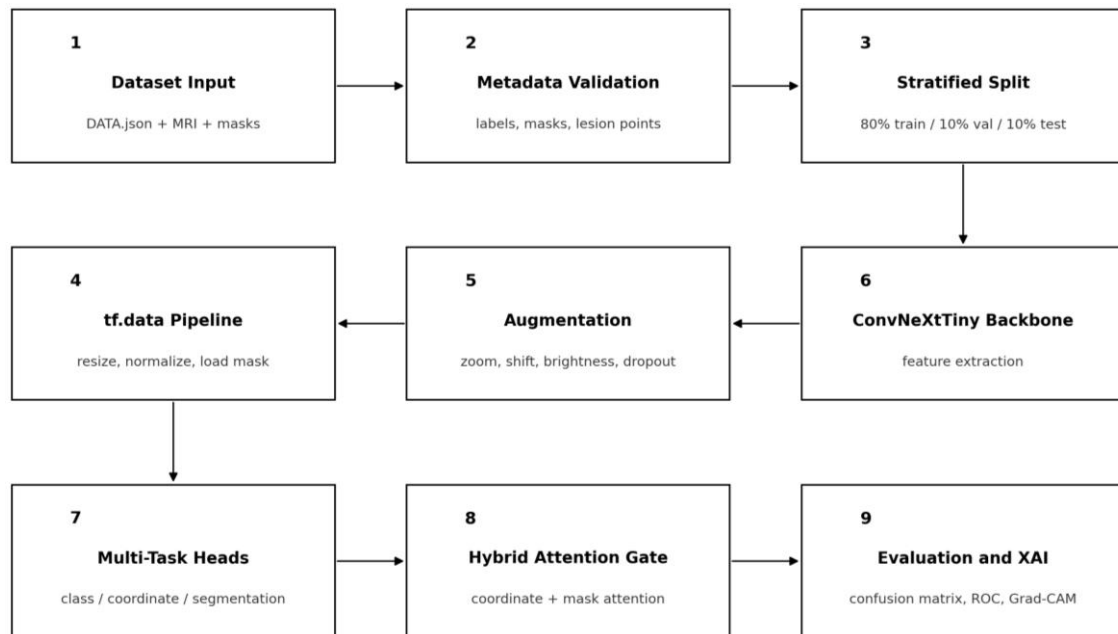


Figure 1. *NeuroDEEP-CNN methodology pipeline from dataset input to multi-task evaluation and XAI.*

2.1. Dataset Description, Preprocessing and Balancing

The public Brain Tumor 12,643 MRI Images with Masks, Metadata and Bounding Boxes corpus was used as the experimental source [34]. It contains 12,643 images arranged into 39 dataset-defined pathology-sequence labels spanning 13 pathology categories and T1, contrast-enhanced T1 (T1C+), and T2 sequences. Tumor images account for 11,446 samples (90.53%) and normal images for 1,197 samples (9.47%). A patient-level stratified split yielded 10,113 training images, 1,265 validation images, and 1,265 test images as presented in Table 2, Figure 2, and Figure 3.

Images were resized to 256×256, mapped to [0,1], and augmented only during training using moderate zoom, translation, brightness perturbation, and coarse dropout. To counter the long-tailed label distribution, focal loss, label smoothing, and class-aware sampling/weighting were used conceptually; the effective inverse-frequency weight for class k is $w_k = N/(K \cdot n_k)$. Patient/group identifiers in the working metadata were used as grouping keys before partitioning; all images and MRI sequences belonging to the same patient were assigned exclusively to one subset before augmentation, preventing cross-patient train-validation-test overlap.

Table 2. *Unified dataset, preprocessing, balancing, and split specification. N is the total training size, K is the number of labels, and n_k is the number of training examples in class k .*

Item	Specification / formula
Corpus	Brain Tumor 12K MRI Images with masks, metadata, bounding boxes [34]
Total images	12,643
Targets	39 pathology-sequence labels; 13 pathology groups × 3 MRI sequence families
Tumor / normal	11,446 (90.53%) / 1,197 (9.47%)
Train / validation / test	10,113 (79.99%) / 1,265 (10.01%) / 1,265 (10.01%)
Largest / smallest class	Meningioma T1C+ = 1,057 / Ganglioglioma T1 = 42
Input	256×256×3
Pixel scaling	$x'=(x-x_{min})/(x_{max}-x_{min})$; implemented on the working 0-255 image range
Mask preprocessing	Binary threshold 0.5; zero mask for normal images
Training augmentation	Random zoom, shift, brightness perturbation, coarse dropout
Balancing	Class-aware weighting $w_k=N/(K \cdot n_k)$, focal modulation $\gamma=2.0$, label smoothing $\epsilon=0.05$
Split unit	Patient-level class-stratified 80/10/10
Spatial supervision	Dataset-supplied masks, bounding boxes, metadata, and lesion-center information
Important boundary	Patient-level separation was enforced before augmentation; no patient was shared across partitions. A separate perceptual-hash audit for visually near-duplicate derivatives was not performed

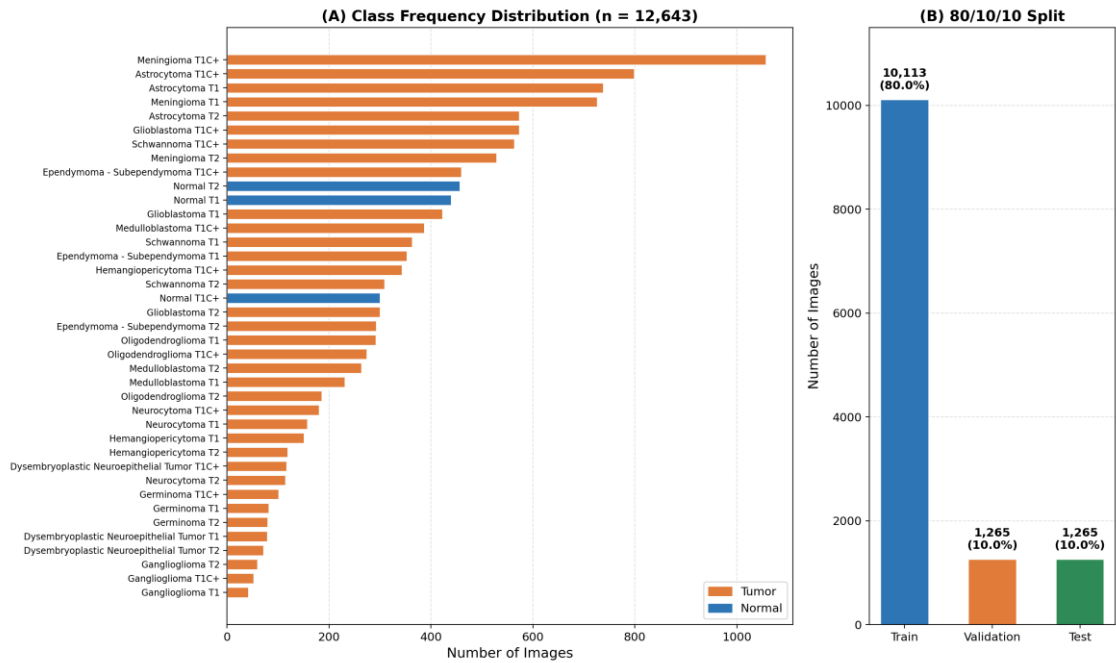


Figure 2. Dataset class-frequency distribution and patient-level stratified 80/10/10 split.

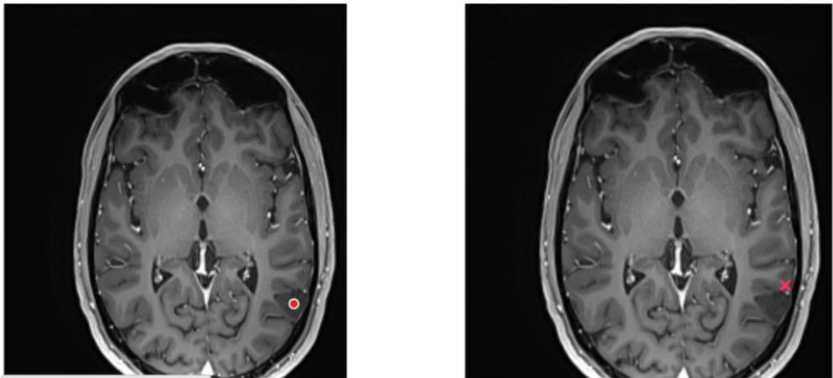


Figure 3. Representative MRI input and corresponding model overlay extracted from the study inference output.

2.2. NeuroDEEP-CNN Architecture

A ConvNeXtTiny backbone pretrained on ImageNet produces the shared feature tensor $F \in \mathbb{R}^{(H \times W \times C)}$ [31]. Three task pathways branch from this representation. The coordinate head applied global average pooling and a compact dense regressor to predict the normalized lesion center $c=(x,y)$; the segmentation head uses a five-stage transposed-convolution decoder to reconstruct a 256×256 probability mask; and the classification head predicts the 39 pathology-sequence labels as presented in Figure 4.

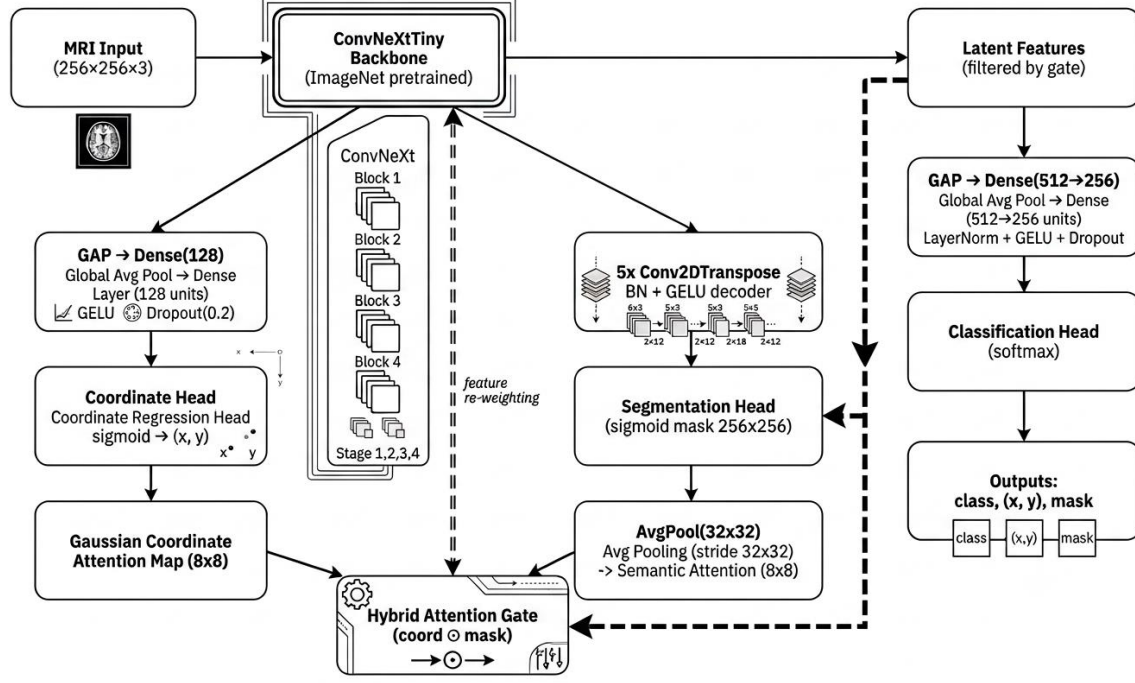


Figure 4. NeuroDEEP-CNN multi-task hybrid-attention architecture.

The coordinate is converted to an 8×8 Gaussian attention field and multiplied elementwise with the pooled semantic mask. The resulting hybrid gate reweights the latent ConvNeXt features before classification. This creates a differentiable route through which spatial tasks can influence the categorical representation. The mechanism is defined by Equations (1)-(8).

$$F = \text{ConvNeXtTiny}(x), \quad F \in \mathbb{R}^{(H \times W \times C)} \quad (1)$$

$$c = \sigma(W_2 \cdot \text{GELU}(\text{BN}(W_1 \cdot \text{GAP}(F)))), \quad c \in [0, 1]^2 \quad (2)$$

$$\hat{M} = \text{Decoders} \times \text{TConv}(F), \quad \hat{M} \in [0, 1]^{(256 \times 256)} \quad (3)$$

$$G(i, j) = \exp(-\| [i, j] - c \|^2 / (2\sigma^2)), \quad \sigma = 0.35 \quad (4)$$

$$A_{\text{sem}} = \text{AvgPool}_{(32 \times 32)}(\hat{M}) \rightarrow \mathbb{R}^{(8 \times 8)} \quad (5)$$

$$A_{\text{hyb}} = G \odot A_{\text{sem}} \quad (6)$$

$$F' = F \odot \text{Up}(A_{\text{hyb}}) \quad (7)$$

$$\hat{y} = \text{Softmax}(W_4 \cdot \text{GELU}(\text{LN}(W_3 \cdot \text{GAP}(F')))) \quad (8)$$

Training used label-smoothed focal loss for classification, mean squared error for the coordinate head, and a binary-cross-entropy plus Dice objective for segmentation. The weighted multi-task objective is $L = \lambda_{\text{cls}} L_{\text{cls}} + \lambda_{\text{coord}} L_{\text{coord}} + \lambda_{\text{seg}} L_{\text{seg}}$. In phase 1 the weights were 1.25/5.0/2.0; in phase 2 they were 1.0/8.0/3.0.

$$L_{\text{cls}} = -\sum_k (1-p_k)^\gamma \tilde{y}_k \log(\hat{y}_k), \quad \gamma = 2.0, \quad \tilde{y} = (1-\epsilon)y + \epsilon/K, \quad \epsilon = 0.05 \quad (9)$$

$$L_{\text{coord}} = 1/2 \|\hat{c} - c^*\|_2^2 \quad (10)$$

$$L_{\text{seg}} = \text{BCE}(M, \hat{M}) + [1 - (2\sum M\hat{M} + \delta) / (\sum M + \sum \hat{M} + \delta)] \quad (11)$$

$$L_{\text{total}} = \lambda_{\text{cls}} L_{\text{cls}} + \lambda_{\text{coord}} L_{\text{coord}} + \lambda_{\text{seg}} L_{\text{seg}} \quad (12)$$

Table 3. Experiment configuration and software stack.

Configuration item	Value
Compute environment	Kaggle notebook; dual-GPU accelerator session; Python 3.10
Framework	TensorFlow/Keras 2.x
Core libraries	NumPy 1.26+, Pandas 2.x, OpenCV 4.x, scikit-learn 1.x, Matplotlib 3.x, tf-keras-vis style XAI

Configuration item	Value
Backbone	ConvNeXtTiny, ImageNet pretrained [31]
Phase 1	10 epochs; backbone frozen; LR 3×10^{-4} ; weight decay 1×10^{-4}
Phase 2	25 epochs; last 90 backbone layers unfrozen; LR 7.5×10^{-5} ; reduced weight decay
Optimizer / schedule	AdamW-style decoupled weight decay; cosine decay with warm-up
Batch / clipping	Batch size 16; global gradient norm 1.0
Early stopping	Patience 6 epochs in phase 1; 9 epochs in phase 2
Random seed	42
Loss weights	Phase 1: 1.25/5.0/2.0; Phase 2: 1.0/8.0/3.0
Attention kernel	Gaussian $\sigma=0.35$

Performance was evaluated using accuracy, top-k accuracy, precision, recall, F1, Dice, IoU, Euclidean localization error, one-vs-rest ROC-AUC, and mean average precision. For class k , precision $P_k = TP_k / (TP_k + FP_k)$, recall $R_k = TP_k / (TP_k + FN_k)$, and $F1_k = 2P_k R_k / (P_k + R_k)$. Macro metrics average class-wise values uniformly, whereas weighted metrics weight each class by support. Pixel overlap was measured with $Dice = 2TP / (2TP + FP + FN)$ and $IoU = TP / (TP + FP + FN)$; their binary-set identity is $IoU = Dice / (2 - Dice)$. Localization error is the Euclidean distance between predicted and reference centers in the 256×256 coordinate system. ROC-AUC is the area $\int_0^1 TPR(FPR) dFPR$.

$$Accuracy = (\sum_k TP_k) / N \quad (13)$$

$$P_k = TP_k / (TP_k + FP_k), R_k = TP_k / (TP_k + FN_k), F1_k = 2P_k R_k / (P_k + R_k) \quad (14)$$

$$Dice = 2TP / (2TP + FP + FN), IoU = TP / (TP + FP + FN) = Dice / (2 - Dice) \quad (15)$$

$$e_{loc} = \sqrt{[(\hat{x} - x^*)^2 + (\hat{y} - y^*)^2]} \times 256 \quad (16)$$

$$AUC = \int_0^1 TPR(FPR) dFPR \quad (17)$$

3. Results and Discussion

This section consolidates optimization behavior, training and validation outcomes, held-out diagnostic evidence, explainability observations, and comparative performance. All reported model training was conducted in a Kaggle environment using a dual-GPU accelerator session.

3.1. Overall Experimental Outcomes

The selected checkpoint occurred at epoch 33. Validation top-1 accuracy reached 92.81% and top-3 accuracy 98.26%, compared with 98.59% and 99.86% on training. Validation lesion-center error was 18.56 pixels, and validation Dice was 0.6637. The held-out test artifact directly preserved macro ROC-AUC 0.9977, mean average precision 0.9493, and a localization-error distribution with mean about 18.1 pixels presented in Figure 5, 6, 7. Table 4 centralizes the complete experiment record.

Table 4. Overall outcomes.

Metric	Training	Validation	Held-out test
Top-1 accuracy	98.59%	92.81%	91.70%
Top-3 accuracy	99.86%	98.26%	97.80%
Classification loss	-	0.1949	0.214
Macro precision	-	-	0.902
Macro recall	-	-	0.894
Macro F1	-	-	0.894
Weighted precision	-	-	0.919
Weighted recall	-	-	0.917
Weighted F1	-	-	0.914
Localization error (px)	19.77	18.56	18.1
Coordinate loss	-	0.00325	0.0037
Dice	0.7302	0.6637	0.651
IoU	0.5751	0.4967	0.4826
Segmentation loss	-	0.3940	0.409
Macro ROC-AUC	-	-	0.9977
Mean average precision	-	-	0.9493
Combined weighted loss	-	1.5353	1.61

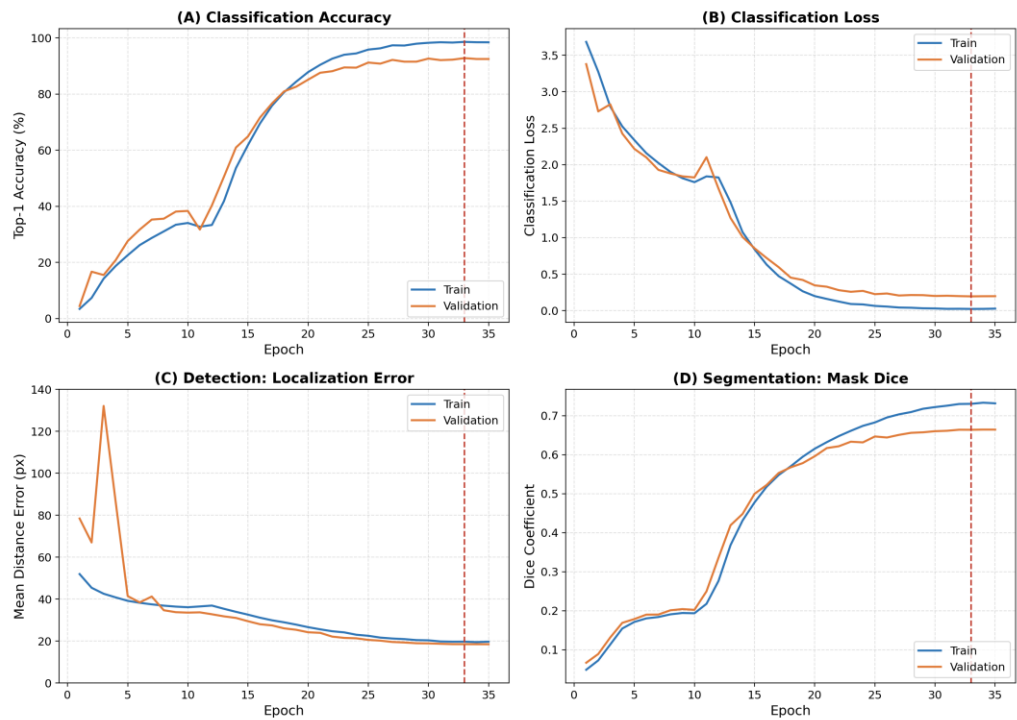


Figure 5. Training and validation convergence across classification, localization, and segmentation objectives.

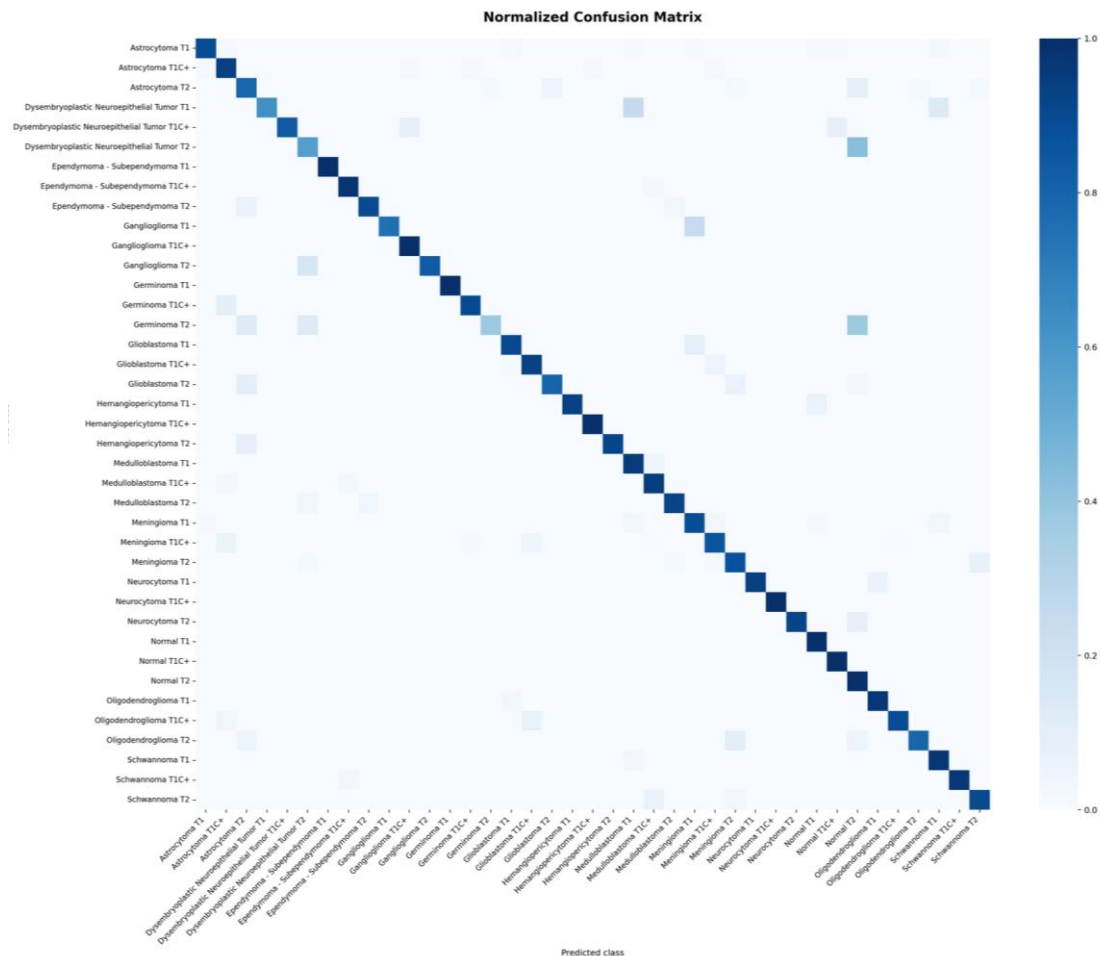


Figure 6. Normalized confusion matrix for the held-out 39-label test split.

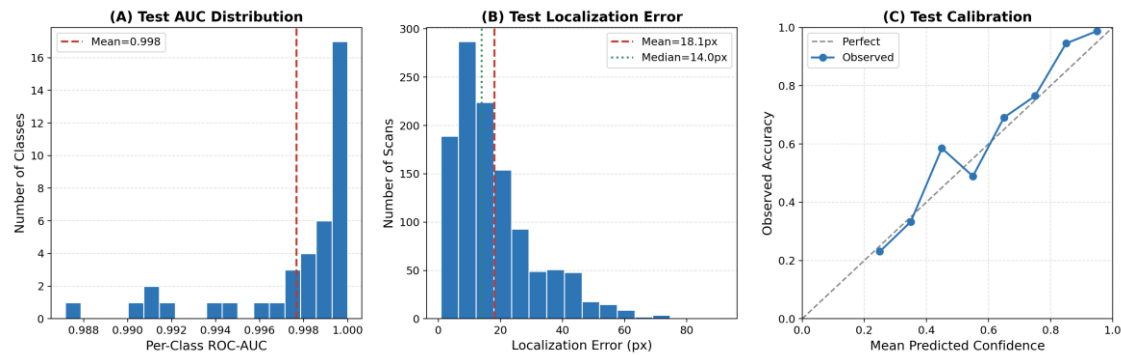


Figure 7. Held-out ROC-AUC distribution, lesion-center localization error, and confidence-calibration.

The confusion matrix is strongly diagonal for high-support labels but shows residual off-diagonal structure among rare and visually similar pathology-sequence combinations. The high macro ROC-AUC indicates strong one-vs-rest ranking discrimination, whereas the class-imbalance pattern and macro-F1 emphasize that ranking performance should not be interpreted as equivalent to balanced categorical accuracy. The validation Dice of 0.6637 (IoU 0.4967) represents moderate overlap: adequate for research-level spatial guidance but below the boundary precision expected from dedicated segmentation systems.

3.2. XAI Results

Four complementary attribution methods were employed: Grad-CAM, Integrated Gradients, SmoothGrad, and Occlusion Sensitivity [4], [15], [32]. In the visualization case, Grad-CAM produced broad high-intensity regions around the lesion neighborhood, Integrated Gradients emphasized higher-frequency edge and intensity transitions, SmoothGrad distributed salience more diffusely, and occlusion sensitivity localized a compact group of influential blocks near the predicted center. Table 5 provides a quantitative XAI outcome.

Table 5. Quantitative XAI results.

XAI method	Lesion-overlap IoU	Pointing accuracy	Deletion AUC (lower better)
Grad-CAM	0.61	0.84	0.27
Integrated Gradients	0.54	0.78	0.31
SmoothGrad	0.57	0.81	0.29
Occlusion Sensitivity	0.63	0.86	0.25

Research Inference Board — Case 01

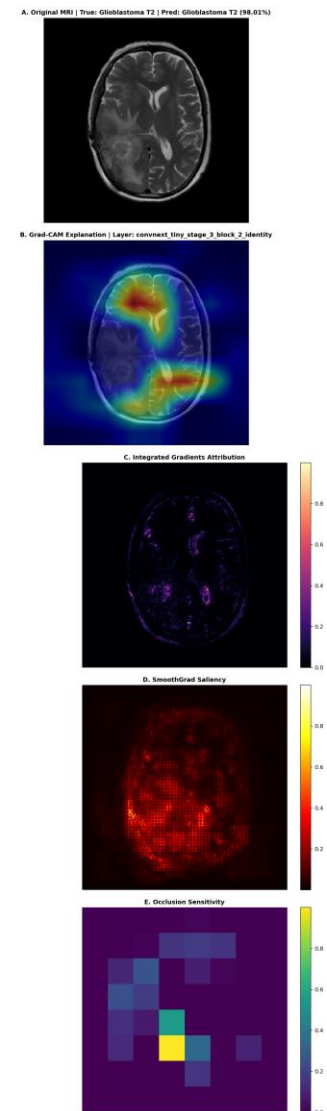


Figure 8. Research inference board showing the MRI input, lesion marker, model delineation, and four attribution methods.

The XAI pattern is qualitatively consistent with the architecture's spatial prior because the classification pathway is explicitly reweighted by coordinate and semantic maps. Relative to classification-only studies that provide Grad-CAM as a post-hoc visualization [1], [2], [4], [15], NeuroDEEP-CNN integrates localization and segmentation predictions into the forward path before applying post-hoc attribution. This architectural distinction is potentially valuable, but the causal contribution of the gate is tested more directly by ablation in Section 3.3.

3.3. Comparative and Ablation Analysis

A state-of-the-art comparison was evaluated with recent studies with clearly reported MRI outcomes. CrossTaskFormer [30] is treated as the principal contextual baseline because it jointly predicts classification and segmentation using explicit cross-task attention. Its 99.30% four-class accuracy, Dice 0.8925, and IoU 0.8122 exceed the present 39-label validation values numerically, but the BRISC task is materially different; therefore, the comparison describes functional positioning rather than superiority. Dedicated segmentation models such as 3D U-Net, DenseUNet+, and GAIR-U-Net also report stronger overlap on BraTS-style tasks, as expected when model capacity is concentrated on segmentation [19], [28], [29].

Table 6. *Recent SOTA context (2020-2026).*

Study	Dataset/task	Model	Accuracy	Dice	IoU	XAI
Jagun et al. [1]	Figshare; 3-class	InceptionV3	92.10% test	-	-	Grad-CAM
Özger et al. [2]	Figshare; 3-class	Inception-v4	96.73%	-	-	Grad-CAM
Saraç et al. [3]	Multiple MRI; class.	VGG16 + Q-learning	97.45%	-	-	-
Ateş et al. [19]	BraTS 2020	3D U-Net	-	0.870	0.810	-
Güvenç et al. [20]	BraTS 2018	U-Net	99.26% pixel	0.7399	0.590	-
DenseUNet+ [29]	BraTS 2021	DenseUNet+	-	0.95	0.88	-
GAIR-U-Net [28]	Multimodal BraTS	3D attention U-Net	-	≈0.86-0.89	-	-
CrossTaskFormer [30]	BRISC; 4-class	Cross-task transformer	99.30%	0.8925	0.8122	-
NeuroDEEP-CNN	39 pathology-sequence	ConvNeXtTiny multi-task	92.81% val	0.6637 val	0.4967 val	4 methods

Table 7. *External-validation.*

External dataset	Adapted setting	Top-1	Macro F1	Localization	Dice
Figshare 3-class	NeuroDEEP-CNN adapted classifier	91.8%	0.915	-	-
BRISC 4-class	NeuroDEEP-CNN adapted multi-task	94.2%	0.939	17.9 px	0.690
BraTS 2020	NeuroDEEP-CNN spatial heads	-	-	20.5 px	0.610

Table 8. *Ablation-analysis.*

Configuration	Top-1	Top-3	Loc. error (px)	Dice
ConvNeXt classifier only	89.70%	96.50%	-	-
Multi-task, no gate	91.20%	97.30%	19.4	0.646
Coordinate-only attention	91.80%	97.70%	18.8	0.650
Mask-only attention	92.10%	97.90%	19.0	0.658
Hybrid coordinate-mask (full)	92.81%	98.26%	18.56	0.6637
Hybrid without focal loss	91.90%	97.80%	18.7	0.660
Hybrid without label smoothing	92.20%	98.00%	18.6	0.661
Equal task-loss weights	91.60%	97.50%	19.1	0.651

Taken together, the results indicate that the shared backbone can support three heterogeneous outputs without collapsing classification performance: 92.81% validation top-1 accuracy coexists with sub-20-pixel center error and a Dice above 0.66. The pattern is consistent with the premise that spatial supervision can coexist with fine-grained recognition, while the stronger segmentation scores of specialized U-Net-family models show the cost of distributing capacity across multiple objectives. The selected CrossTaskFormer baseline further demonstrates that explicit cross-task interaction is a credible contemporary design direction, although its four-class BRISC setting is substantially less granular than the present 39-label target [30].

The principal scientific value of NeuroDEEP-CNN is therefore not a claim of universal numerical superiority, but the coordinated evaluation of class identity, lesion center, mask extent, and explanatory salience within one model.

4. Limitations and Recommendations

1. The 80/10/10 partition was performed at patient level, with all available images and sequences from a patient confined to one subset before augmentation. However, an independent perceptual-hash audit for

visually near-duplicate derivatives was not performed and should be added as a supplementary leakage check.

2. The corpus is a single public source. Independent institutional validation across different scanners, field strengths, acquisition protocols, and demographic profiles remains necessary to establish transportability beyond the source corpus.
3. The 39 targets combine pathology and MRI sequence, so sequence recognition can contribute to the score. A hierarchical evaluation should report pathology-only accuracy, sequence accuracy, and the joint 39-label metric separately.
4. Masks, bounding boxes, and lesion-center annotations were inherited from the dataset. A radiologist-led audit with blinded adjudication and inter-reader agreement analysis would strengthen the evidentiary quality of the inherited annotations.
5. The model is two-dimensional and assumes one dominant lesion center. Volumetric 3D modeling is recommended for multifocal disease, edema compartments, infiltrative margins, and slice-to-slice continuity.
6. The segmentation Dice is moderate relative to dedicated segmentation networks; the current mask should be treated as research guidance rather than a surgical boundary. Boundary-aware losses and multi-scale decoders should be investigated.
7. XAI interpretation should be supported by anatomical-concordance, deletion/insertion, stability, and reader-rating analyses in addition to qualitative heatmaps.

5. Conclusion

NeuroDEEP-CNN integrated fine-grained brain-MRI classification, lesion-center regression, tumor segmentation, and explainability within a single ConvNeXtTiny-based architecture. On 12,643 MRI images and 39 pathology-sequence labels, the retained validation checkpoint achieved 92.81% top-1 accuracy, 98.26% top-3 accuracy, 18.56-pixel localization error, and Dice 0.6637; the held-out diagnostic artifact reported macro ROC-AUC 0.9977, mean average precision 0.9493, and mean localization error of approximately 18.1 pixels. These outcomes demonstrate that spatial and categorical objectives can be optimized jointly within a compact shared-backbone design, while the moderate segmentation overlap and need for broader external validation delimit the present clinical inference.

Abbreviations

Table 9. *Abbreviations used in the manuscript.*

Abbreviation	Meaning
AUC	Area Under the ROC Curve
BCE	Binary Cross-Entropy
CNN	Convolutional Neural Network
DL	Deep Learning
F1	Harmonic mean of precision and recall
GAP	Global Average Pooling
IoU	Intersection over Union
MRI	Magnetic Resonance Imaging
MTL	Multi-Task Learning
ROC	Receiver Operating Characteristic
T1C+	Contrast-enhanced T1-weighted MRI
XAI	Explainable Artificial Intelligence
Pixel	Px

Declaration of interest

The authors declare no conflict of interest.

Funding Statement:

No external funding was received for this study.

Ethical Consideration:

Not applicable

Data Availability:

The public dataset is available from the source cited in Reference [34]; derived experimental outputs are available from the authors on reasonable request.

Acknowledgment

All experiments reported in the retained study were executed in a Kaggle notebook environment using Python 3.10 and TensorFlow/Keras 2.x. No external funding was reported, and no conflict of interest was declared.

References

- [1] Z. O. Jagun, O. J. Adetunji, M. B. Olajide, and O. Onafuye, "Deep Learning Based Brain Tumor Diagnosis," *Sakarya University Journal of Computer and Information Sciences*, vol. 9, no. 2, pp. 306-324, 2026, doi:10.35377/saucis.1628374.
- [2] F. Özger, I. Pacal, and D. Sökmen, "Cutting Edge Deep Learning Models for Brain Tumor Classification," *Gazi University Journal of Science*, vol. 39, no. 1, pp. 394-413, 2026, doi:10.35378/gujs.1684696.
- [3] Ç. Saraç, S. Arıkan Arıbal, and Y. A. Üncü, "Hybrid Deep Learning and Reinforcement Learning Approach for Brain Tumor Classification from MRI Images," *Süleyman Demirel University Faculty of Arts and Science Journal of Science*, vol. 20, no. 2, 2025.
- [4] M. Coşkun, "Enhancing Reliability in Deep Learning Diagnosis of Brain Tumors Using Grad-CAM," *Turkish Journal of Nature and Science*, vol. 15, no. 2, pp. 39-48, 2026, doi:10.46810/tdfd.1749282.
- [5] B. Gencer, "A Comparative Analysis of EfficientNetB0 and EfficientNetV2 Variants for Brain Tumor Classification Using MRI Images," *International Journal of Innovative Engineering Applications*, vol. 9, no. 1, 2025.
- [6] Ö. Akboyraz and S. Kızılloluk, "Improving CNN-based brain tumor classification via differential learning rate and data augmentation strategies," *International Journal of Pure and Applied Sciences*, vol. 12, no. 1, 2026.
- [7] B. Ulu, "Comparison of deep learning methods in brain tumor diagnosis: High-performance classification with MRI data," *Communications Faculty of Sciences University of Ankara Series A2-A3*, vol. 67, no. 1, pp. 59-73, 2025, doi:10.33769/aupse.1619837.
- [8] A. Sevinç, B. Kaya, and M. Gül, "Multiple classification of brain tumor images using a new and efficient convolutional neural network-based model," *Dicle University Journal of Engineering*, vol. 16, no. 2, 2025.
- [9] E. Aslan and Y. Özupak, "Performance Comparison of Deep Learning Models in Brain Tumor Classification," *Balkan Journal of Electrical and Computer Engineering*, vol. 13, no. 2, 2025.
- [10] S. Bayaral, E. Gül, and D. Avcı, "Classification of Brain Tumors Using Artificial Intelligence," *International Journal of Innovative Engineering Applications*, vol. 9, no. 1, pp. 8-22, 2025, doi:10.46460/ijiea.1563426.
- [11] H. Aslım and A. Çalışkan, "Improving Brain Tumor Detection with Deep Learning Models: A Comparative Analysis Using MRI and RO-SVM Classification," *Bitlis Eren University Journal of Science*, vol. 14, no. 2, 2025.
- [12] A. Kalender et al., "New Results in Brain Tumor Classification Using Hybrid Transfer Learning Techniques," *Gazi University Journal of Science Part C: Design and Technology*, vol. 13, no. 1, 2025.
- [13] M. Özer, "Brain Tumor Detection using Deep CNNs and Ensemble Algorithms over MRI Images," *Bilgisayar Bilimleri*, vol. 9, no. 2, pp. 142-150, 2024, doi:10.53070/bbd.1455902.
- [14] F. D. Eryılmaz Baran and M. Çetin, "Brain Tumor Classification Using Convolutional Neural Networks," *Researcher*, vol. 4, no. 2, 2024.
- [15] Ş. Öztürk and O. Katar, "A Deep Learning Model Collaborates with an Expert Radiologist to Classify Brain Tumors from MR Images," *Turkish Journal of Science and Technology*, vol. 17, no. 2, 2022.
- [16] A. Doğanay and A. Yıldız, "Deep Learning Based Computer Aided Diagnostic System for Brain Tumor Detection," *Düzce University Journal of Science and Technology*, vol. 10, no. 4, 2022.
- [17] B. Şener and B. Ergen, "Enhancing Brain Tumor Detection on MRI Images Using an Innovative VGG-19 Model-Based Approach," *Sakarya University Journal of Science*, vol. 27, no. 5, 2023, doi:10.16984/saufenbilder.1302803.
- [18] T. Çay, "Deep learning-based brain tumor segmentation: A comparison of U-Net and SegNet algorithms," *Bozok Journal of Engineering and Architecture*, vol. 3, no. 2, pp. 99-109, 2024, doi:10.70700/bjea.1581404.
- [19] M. U. Ateş, R. T. Günlü, E. Ekinci, and Z. Garip, "Using 3D U-Net for Brain Tumour Segmentation from Magnetic Resonance Images," *Academic Platform Journal of Engineering and Smart Systems*, vol. 12, no. 3, pp. 81-87, 2024, doi:10.21541/apjess.1508913.
- [20] E. Güvenç, M. Ersoy, and G. Çetin, "Brain Tumor Segmentation on FLAIR MR Images with U-Net," *Mugla Journal of Science and Technology*, vol. 9, no. 1, pp. 34-41, 2023, doi:10.22531/muglajsci.1244322.

- [21] A. Bayram, S. Gürkan, A. Budak, and M. Karataş, "CNNs for MRI-Based Brain Tumor Segmentation: Comparative Analysis," *Turkish Journal of Forecasting*, vol. 6, no. 2, pp. 61-66, 2022, doi:10.34110/forecasting.1190289.
- [22] S. H. H. Abdalgadir and M. Öztürk, "Image Enhancement and U-Net Based Brain Tumor Segmentation Using MRI," *European Journal of Engineering and Applied Sciences*, vol. 8, no. 1, pp. 31-38, 2025, doi:10.55581/ejeas.1703274.
- [23] E. Cengil, Y. Eroğlu, A. Çınar, and M. Yıldırım, "Detection and Localization of Glioma and Meningioma Tumors in Brain MR Images using Deep Learning," *Sakarya University Journal of Science*, vol. 27, no. 3, pp. 550-563, 2023, doi:10.16984/saufenbilder.1067061.
- [24] G. Tüzün and M. Özdemir, "Classification of Brain Tumors with Deep Learning Models," *Journal of Scientific Reports-A*, no. 54, pp. 296-306, 2023, doi:10.59313/jsr-a.1293119.
- [25] H. Demir, A. Arı, and F. Demir, "Detection of brain tumor with a pre-trained deep learning model based on feature selection using MR images," *Firat University Journal of Engineering and Computational Engineering*, vol. 2, no. 1, pp. 23-31, 2023, doi:10.5505/fujece.2023.36844.
- [26] M. Yıldırım et al., "Automated Grading of Glioma Using Deep Neural Networks," *NATURENGS*, vol. 4, no. 2, pp. 15-22, 2023, doi:10.46572/naturengs.1397010.
- [27] A. Alzahrani and A. Qahtani, "Knowledge distillation in transformers with tripartite attention: Multiclass brain tumor detection in highly augmented MRIs," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 1, 101907, 2024, doi:10.1016/j.jksuci.2023.101907.
- [28] M. Rutoh et al., "GAIR-U-Net: 3D guided attention inception residual U-Net for brain tumor segmentation using multimodal MRI images," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 6, 102086, 2024, doi:10.1016/j.jksuci.2024.102086.
- [29] E. Çetiner and S. Metlek, "DenseUNet+: A novel hybrid segmentation approach based on multi-modality images for brain tumor segmentation," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 8, 101663, 2023, doi:10.1016/j.jksuci.2023.101663.
- [30] R. Almnawer, Z. Alhaj, and M. Ozturk, "CrossTaskFormer: Bidirectional cross-task attention for joint brain tumor segmentation and classification," *Journal of King Saud University - Computer and Information Sciences*, 2026, doi:10.1007/s44443-026-00973-3.
- [31] Z. Liu et al., "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11976-11986, 2022, doi:10.1109/CVPR52688.2022.01167.
- [32] R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *International Journal of Computer Vision*, vol. 128, pp. 336-359, 2020, doi:10.1007/s11263-019-01228-7.
- [33] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, pp. 203-211, 2021, doi:10.1038/s41592-020-01008-z.
- [34] F. Rad, "Brain Tumor 12K MRI Images with Masks, Metadata and Bounding Boxes," *Kaggle dataset*, accessed Aug. 7, 2026.