

Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging

Khushvir Singh ¹, , Pooja Sharma ¹

1. Computer Science & Engineering, I.K.Gujral Punjab Technical University, Kapurthala, IND

Received: June 11, 2026 | Review began: July 14, 2026 | Review ended: August 03, 2026 | Published: August 20, 2026

© Copyright 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

This study evaluates the performance of a hybrid deep learning framework against conventional approaches for multiclass brain tumor classification using MRI scans. The framework integrates ConvNeXtV2, the Swin Transformer, and Convolutional Block Attention Module attention mechanisms to capture both localized texture features and broader contextual patterns. RDXNet and ResNet50 serve as baseline models. Evaluation metrics include accuracy, precision, recall, macro F1-score, and area under the curve. The hybrid model achieves 95.45% accuracy versus RDXNet (94.00%) and ResNet50 (91.54%), with gains across all metrics. Gradient-weighted Class Activation Mapping visualizations confirm attention to tumor-relevant regions, supporting model interpretability. Overall, fusing convolutional and transformer features with attention refinement markedly enhances classification performance.

Categories: Data Analysis, Machine Learning (ML), Deep Learning

Keywords: brain tumor classification, convnextv2, swin transformer, cbam, deep learning, mri analysis, resnet50

Introduction

Advances in medical imaging and deep learning have substantially enhanced the ability to identify and classify brain tumors in magnetic resonance imaging (MRI) scans. Distinguishing tumor types - glioma, meningioma, pituitary, and no-tumor - is critical for treatment planning and patient prognosis [1-3].

Traditionally, tumor identification depended on radiologist expertise, a subjective process requiring time and specialization. To overcome these limitations, researchers have increasingly focused on automated brain tumor recognition systems using neural networks capable of extracting meaningful patterns from complex MRI data [4,5].

A wide range of deep learning architectures have been proposed, including convolutional neural network (CNN)-based models and transformer-based hybrids [4-6]. RDXNet is a notable baseline, yet challenges remain in jointly capturing global context and fine-grained tumor detail [6]. This study evaluates the ConvNeXtV2-Swin Transformer-CBAM (convolutional block attention module) combination against RDXNet and ResNet50 under identical experimental conditions.

The hybrid model achieves an accuracy of 95.45% versus RDXNet's 94%, with consistent improvements in precision, recall, F1-score, and area under the curve (AUC). Confusion matrices, loss-accuracy plots, receiver operating characteristic (ROC) curves, and Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations provide complementary qualitative evidence. While CNN-Transformer hybrids and CBAM-based attention refinement have each been explored in

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI https://doi.org/10.7759/s44389-026-00235-3

prior medical imaging work, the contribution of this study is empirical and application-focused rather than architectural: a systematic evaluation of this specific component combination against strong CNN and transformer baselines on a standard multiclass brain tumor MRI benchmark, paired with interpretability validation via Grad-CAM.

Materials And Methods

Proposed hybrid architecture

The proposed model combines three established components - ConvNeXtV2, Swin Transformer, and CBAM - through feature concatenation followed by attention refinement. This is not a new fusion mechanism; rather, it applies a documented combination strategy to multiclass brain tumor classification, with component choices justified below by benchmarking reported in the source literature and validated empirically through comparison against RDXNet, ResNet50, and Swin+SimMIM.

ConvNeXtV2 branch: This branch focuses on local detail-fine-grained textures and edges within localized regions of the MRI scan. It employs Global Response Normalization (GRN), which prevents feature channels from collapsing into redundant patterns, thereby preserving channel diversity throughout training.

Swin Transformer branch: While ConvNeXtV2 captures local detail, this branch models global context. It partitions the image into windows and applies self-attention within and across windows, enabling the model to capture long-range spatial relationships between different brain regions without the quadratic computational cost of standard transformer attention.

Fusion and CBAM attention: Feature maps from the ConvNeXtV2 and Swin Transformer branches are fused via channel-wise concatenation and passed through a CBAM attention module, which sequentially applies channel attention (identifying which feature channels are most informative) followed by spatial attention (identifying which spatial regions are most relevant). This refinement step suppresses irrelevant background activations and emphasizes tumor-relevant regions prior to classification. Fusion here uses simple concatenation rather than a learned cross-attention mechanism between branches; this is a deliberate simplicity/compute trade-off given the Colab-based training budget, not a claimed architectural innovation.

Finally, the refined features are pooled together and passed to a simple classification layer, which outputs the probability of each tumor type - glioma, meningioma, pituitary tumor, or no tumor at all.

Convolutional Neural Networks

The introduction of CNNs [7-9] brought a fundamental transformation to how computer vision tasks were tackled [6,10]. Rather than handcrafting features, networks learned task-specific visual patterns. ConvNeXt demonstrated that well-optimized convolutional designs remain competitive [11]. However, most architectural decisions were validated under supervised ImageNet training. The field is increasingly transitioning toward self-supervised strategies that reduce annotation dependence.

Masked autoencoders gained traction in vision just as they had in NLP, but were designed primarily for Vision Transformers. Applying them to ConvNeXt revealed compatibility gaps, motivating the development of ConvNeXt V2 with a dedicated Fully Convolutional Masked Autoencoder (FCMAE) training strategy (Figure 1).

The guiding idea was to make self-supervised learning as effective for convolutional networks as it is for transformers. Architecture and training methods were co-designed from scratch. During pre-training, images are divided into patches, with ~60% randomly masked; the network processes only visible patches via sparse convolutions adapted from 3D point cloud processing [12,13], keeping computational cost low. These sparse operations convert seamlessly back to standard convolutions at inference [8].

How to cite this article:

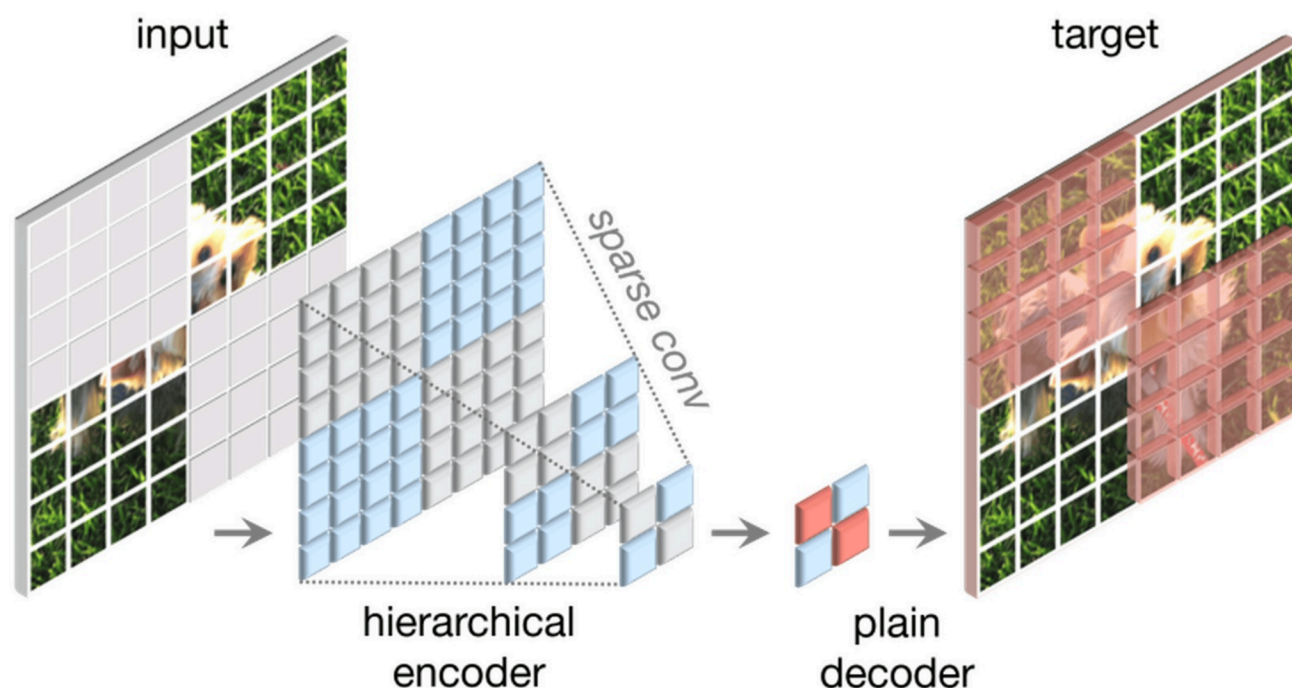


FIGURE 1: The FCMAE Framework: Sparse ConvNeXt Encoder and Lightweight Decoder, with Reconstruction Loss Applied Only to Masked Regions

Adapted from Woo et al. [14]

FCMAE, Fully Convolutional Masked Autoencoder

Another key change replaced transformer-based reconstruction with a simpler ConvNeXt block, producing a fully convolutional pipeline that delivers comparable results at lower complexity. During development, a problem termed feature collapse was identified, where network components stopped learning diverse representations. GRN was introduced to promote healthy channel competition, particularly benefiting the self-supervised pre-training stage.

ConvNeXt V2 substantially outperforms its predecessor across standard evaluation benchmarks covering ImageNet image classification, COCO detection, and ADE20K segmentation. The architecture scales from compact 3.7M-parameter variants achieving 76.7% accuracy to large 650M-parameter models reaching 88.9%.

Fully Convolutional Masked Autoencoder

The proposed framework consists of a sparse-convolution-based ConvNeXt encoder and a lightweight single-block ConvNeXt decoder [14], adopting an asymmetric design where the encoder exclusively handles unmasked visible patches, while the decoder rebuilds the full image by integrating information from encoded features and mask tokens. The reconstruction objective is applied only to masked positions, pushing the model to derive absent information from context rather than recalling what is visible.

A single ConvNeXt block decoder was chosen over complex hierarchical alternatives to maintain efficiency. Experiments confirmed that it delivers comparable fine-tuning accuracy while substantially reducing pre-training time. The decoder dimension was set to 512 to achieve an appropriate trade-off between output quality and computational cost.

Reconstruction objective: The reconstruction objective is based on the mean squared error (MSE) computed between the reconstructed output and the normalized target image. Following MAE practices [15], the target uses patch-wise normalized pixels from the original input. By applying the loss exclusively to masked regions, the network must derive absent information from surrounding context instead of merely recalling what is already visible.

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI <https://doi.org/10.7759/s44389-026-00235-3>

FCMAE framework and experimental evaluation: To evaluate the representational quality achieved by the proposed FCMAE, all design elements were combined using ConvNeXt-Base as the encoder backbone [14]. Multiple ablation studies determined each component's contribution. End-to-end fine-tuning performance is reported as top-1 accuracy on ImageNet-1K using a single 224×224 center crop with 800 pre-training epochs and 100 fine-tuning epochs.

Effect of sparse convolution: Sparse convolution is central to preventing masked-region information leakage. Models trained with and without sparse convolution were compared. These findings are presented in Table 7. The table shows that sparse convolution yields a 4.4% Top-1 accuracy gain. This validates the premise that sparse convolution reliably blocks information leakage from masked patches and yields more reliable feature representations during pre-training.

Configuration	Top-1 Accuracy (%)
Without sparse conv.	79.3
With sparse conv.	83.7

TABLE 1: Effect of Sparse Convolution on Top-1 Accuracy

Feature Activation and GRN Motivation

ConvNeXt V1 trained under FCMAE suffers from feature collapse, where high-dimensional channels produce nearly identical or inactive activations (Figure 2). GRN is introduced to address this by encouraging channel diversity through lateral inhibition-inspired competition. GRN operates through three sequential stages: (1) global feature responses are aggregated by computing the L2 norm of each channel; (2) divisive normalization creates competition among channels; and (3) feature calibration scales activations according to normalized responses. Learnable parameters allow GRN to initially act as an identity and adapt gradually [14].

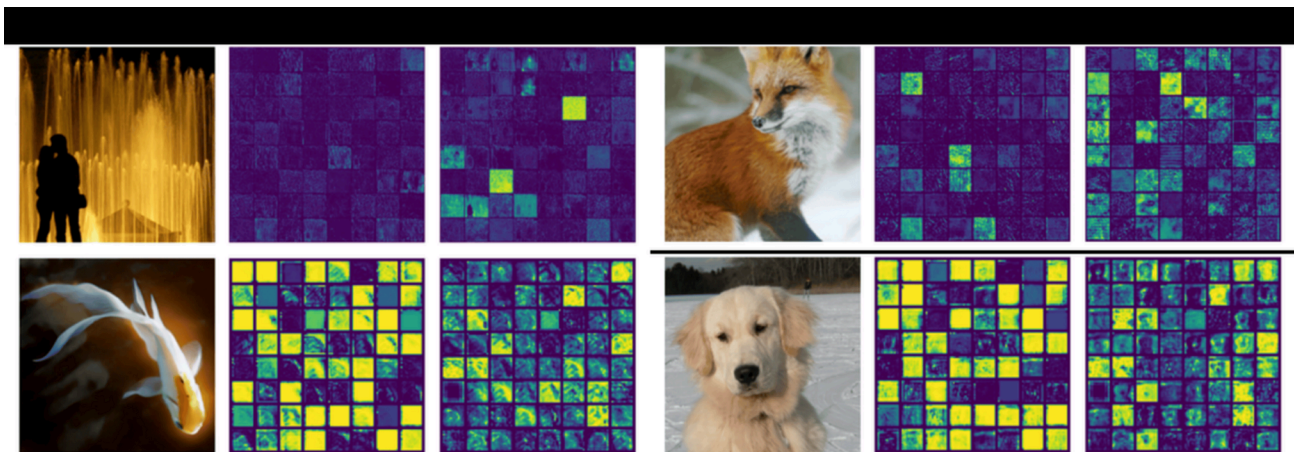


FIGURE 2: Feature Activation Maps from ConvNeXt V1 (Left) and V2 (Right). V1 Shows Feature Collapse with Dead/Redundant Channels; GRN in V2 Restores Diversity

Adapted from Woo et al. [14]
GRN, Global Response Normalization

Comparison with Supervised Baselines

FCMAE pre-training is compared against supervised baselines in Table 2.

Model/Training Type	Top-1 (%)
Supervised (100 ep.)	82.7
Supervised (300 ep.)	83.8
FCMAE (Ours)	83.7

TABLE 2: Top-1 Accuracy: FCMAE vs. Supervised Baselines

FCMAE, Fully Convolutional Masked Autoencoder

FCMAE pre-training improves over random initialization (82.7 $\rightarrow$ 83.7), though still slightly behind the 300-epoch supervised result. This gap motivates the GRN-enhanced ConvNeXt V2 co-design.

Global Response Normalization

GRN operates in three steps:

- (1) Global Aggregation: $G(X)_i = \|X_i\|_2$
- (2) Divisive Normalization: $N(G(X))_i = \frac{\|X_i\|_2}{\sum_{j=1}^C \|X_j\|_2}$, creating channel competition
- (3) Feature Calibration: $X'_i = X_i \times N(G(X))_i$. Learnable parameters, gamma and beta (both initialized to zero), are added with a residual connection: $(GRN(X))_i = \gamma \times (X_i \times N(G(X))_i) + \beta + X_i$. This lets GRN initially act as identity and adapt gradually [14].

In the equations above, X denotes the input feature map produced by the preceding convolutional layer, with height, width, and number of channels represented by H , W , and C , respectively, and X_i denoting the i -th channel of that feature map. $G(X)_i$ is the global aggregated response of channel i , obtained by computing its L2 norm across the spatial dimensions, which summarizes how strongly that channel activates across the entire feature map. $N(\cdot)_i$ is the divisive normalization function, which rescales the aggregated response of channel i by the sum of aggregated responses across all C channels; this normalization step introduces competition between channels, so that channels with disproportionately high activation are suppressed relative to others, directly counteracting the feature collapse problem described earlier. X'_i is the calibrated feature map obtained by multiplying the original channel activation X_i by its corresponding normalized response $N(G(X))_i$, effectively reweighting each channel according to its relative importance. Finally, gamma (γ) and beta (β) are learnable scale and shift parameters, both initialized to zero at the start of training, and the additive term X_i represents a residual connection that bypasses the GRN transformation entirely. Because gamma and beta start at zero, $GRN(X)_i$ is initially equal to X_i , meaning the module behaves as an identity mapping at initialization and only gradually learns to apply channel recalibration as training progresses, ensuring stable optimization from the outset. All operations are applied per channel and are computationally lightweight, adding no additional parameters or floating-point operations (FLOPs) relative to the baseline ConvNeXt block, consistent with the results.

Method	Accuracy (%)	Params (M)
Baseline	83.7	89
LRN	83.2	–
BN	80.5	–
LN	83.8	–
SE	84.4	109
CBAM	84.5	109
GRN (Ours)	84.6	No overhead

TABLE 3: GRN vs. Existing Normalization and Attention Methods

CBAM, Convolutional Block Attention Module; BN, Batch Normalization; GRN, Global Response Normalization; LN, Layer Normalization; LRN, Local Response Normalization; SE, Squeeze-and-Excitation

As shown in Table 3, GRN outperforms all normalization and attention modules without adding parameters or FLOPs.

ConvNeXt V2 Block and Co-Design

The ConvNeXt V2 block replaces LayerScale with GRN inserted after the GELU activation (Figure 3). As shown in Table 4, GRN significantly improves representation quality and enables ConvNeXt V2 to surpass stronger supervised baselines without increasing the parameter count [14].

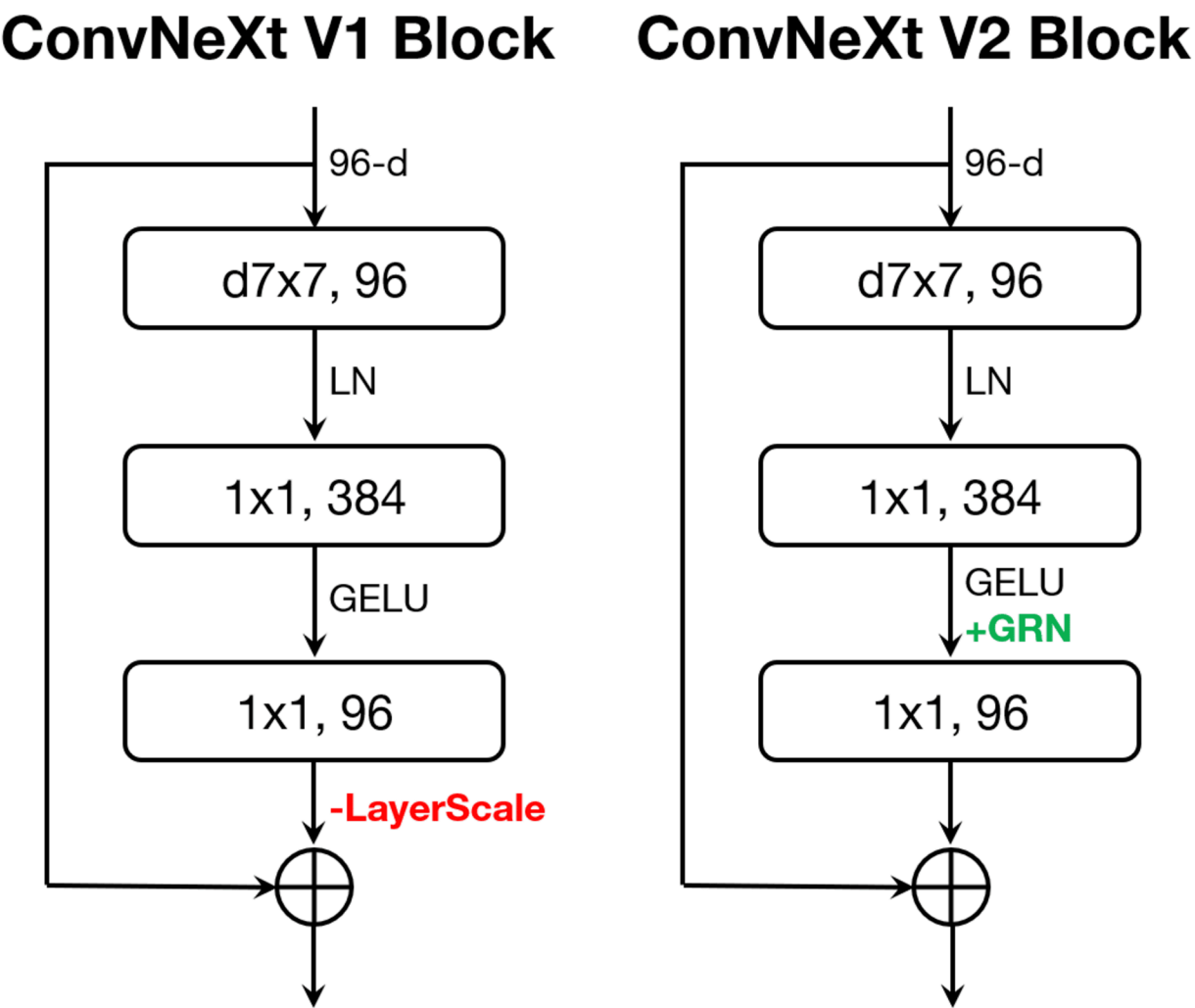


FIGURE 3: ConvNeXt V1 vs. V2 Block Architecture, Showing GRN Inserted After the GELU Activation, Replacing LayerScale

Adapted from Woo et al. [14]
GRN, Global Response Normalization; GELU, Gaussian Error Linear Unit

Model	Top-1 (%)
ConvNeXt V1 + Supervised (300 ep.)	83.8
ConvNeXt V1 + FCMAE	83.7
ConvNeXt V2 + FCMAE (Ours)	84.6

TABLE 4: Fine-Tuning Accuracy with GRN-Enhanced Pre-Training

FCMAE, Fully Convolutional Masked Autoencoder; GRN, Global Response Normalization

Table 5 confirms that using FCMAE and GRN together yields the largest accuracy gains. The benefit scales with model size, with the 198M-parameter ConvNeXt V2 achieving an accuracy of 85.6%.

Backbone	Method	Params	FLOPs	Top-1 Acc (%)
ConvNeXt	Supervised	89M	15.4G	83.80
ConvNeXt	FCMAE	89M	15.4G	83.70
ConvNeXt V2	Supervised	89M	15.4G	84.30
ConvNeXt V2	FCMAE (Ours)	89M	15.4G	84.60
ConvNeXt	Supervised	198M	34.4G	84.30
ConvNeXt	FCMAE	198M	34.4G	84.40
ConvNeXt V2	Supervised	198M	34.4G	84.50
ConvNeXt V2	FCMAE (Ours)	198M	34.4G	85.60

TABLE 5: Co-Design Impact: FCMAE + GRN Across Model Scales

FCMAE, Fully Convolutional Masked Autoencoder; FLOPs, Floating-Point Operations; GRN, Global Response Normalization

Swin Transformer

The Swin Transformer addresses key limitations of Vision Transformers [9] through two innovations (Figure 4). First, hierarchical feature maps are built by gradually merging patches, enabling multi-scale feature extraction that is compatible with dense prediction frameworks [16]. Second, self-attention is restricted to local windows, making the computational complexity linear with respect to image size. Shifted windows enable cross-window communication without introducing additional cost.

How to cite this article:

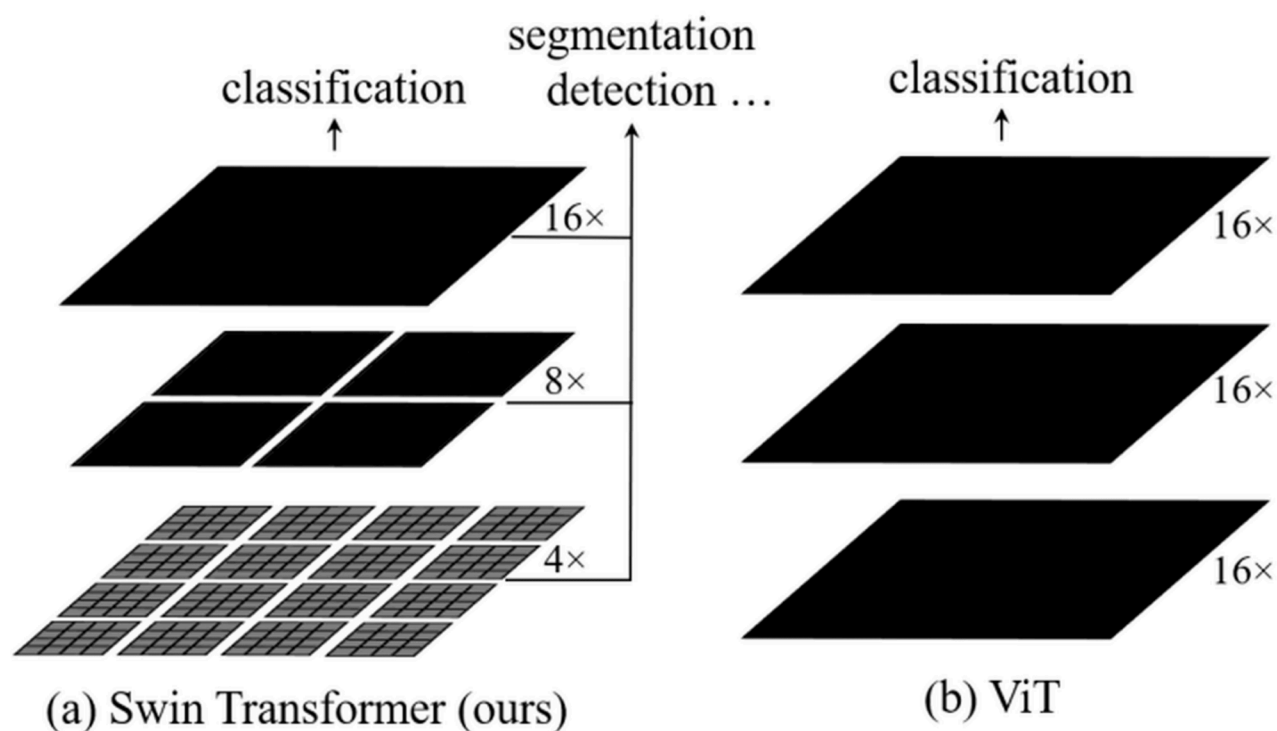


FIGURE 4: (a) Swin Transformer Builds Hierarchical, Multi-Scale Features with Local Window Attention (Linear Complexity). (b) ViT Uses Single-Resolution Global Attention (Quadratic Complexity)

Adapted from Liu et al. [16]
ViT, Vision Transformer

CNNs have sustained their leadership in computer vision owing primarily to efficient multi-scale feature processing and low computational overhead for high-resolution imagery [12,17]. Deep CNN architectures such as AlexNet and Inception-v4 established foundational supervised learning benchmarks on large-scale datasets [18,19]. Self-supervised and contrastive learning methods have since emerged as powerful alternatives, demonstrating strong transferability to medical imaging tasks [20,21]. Large-scale visual recognition benchmarks such as ImageNet have driven architectural progress [22,23]. Squeeze-and-excitation networks demonstrated that channel-wise attention further enhances CNN representational power [24]. Considerable research has also applied deep transfer learning and hybrid architectures specifically to brain tumor MRI classification, yielding strong performance across multiple tumor types [25-31]. However, scale, resolution, and spatial structure differences between language and vision complicate direct transfer of transformer architectures [9,13,14].

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI <https://doi.org/10.7759/s44389-026-00235-3>

Model	C	Layers per Stage {S1,S2,S3,S4}
Swin-T	96	{2,2,6,2}
Swin-S	96	{2,2,18,2}
Swin-B	128	{2,2,18,2}
Swin-L	192	{2,2,18,2}

TABLE 6: Swin Transformer Model Variants

As summarized in Table 6, the Swin Transformer achieves 87.3% top-1 accuracy on ImageNet-1K, 58.7 box AP on COCO, and 53.5% mean Intersection over Union on ADE20K segmentation.

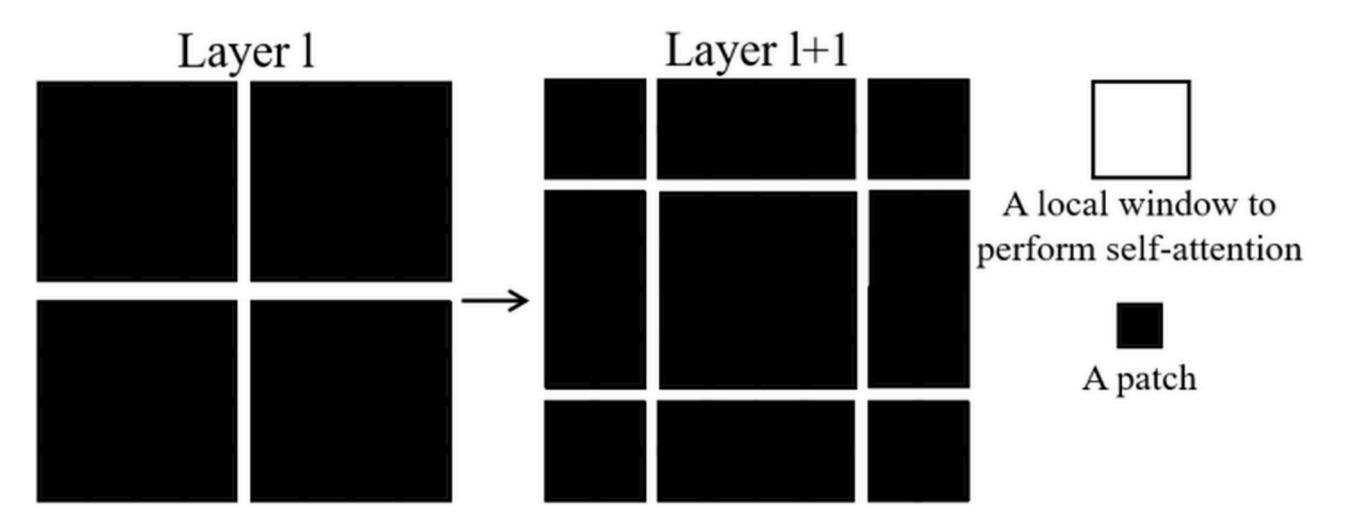


FIGURE 5: Shifted Windows for Self-Attention: In Layer L, Regular Non-Overlapping Windows; in Layer L+1, Shifted Windows That Cross Boundaries of the Previous Layer, Enabling Cross-Window Communication

Adapted from Liu et al. [16]

The shifted-window scheme retains the efficiency of windowed attention while substantially enhancing modeling capacity (Figure 5). Empirically, it yields much lower latency than sliding-window approaches, and this strategy has likewise demonstrated benefits for all-multilayer perceptron network designs [14].

Convolutional Block Attention Module

CBAM applies lightweight attention sequentially along the channel and spatial dimensions. Self-supervised pretraining approaches, such as Momentum Contrast, have demonstrated improved feature transferability in medical imaging contexts [32]. Residual connection design principles further underpin the backbone architectures used in this study [33]. CBAM is lightweight, introduces negligible computational overhead, and improves classification and detection performance when incorporated into convolutional blocks [34].

How to cite this article:

The proposed model integrates three complementary components into a single classification pipeline: a ConvNeXtV2 branch for local texture extraction, a Swin Transformer branch for global contextual modeling, and a CBAM module for attention-based feature refinement.

ConvNeXtV2 branch: ConvNeXtV2 blocks incorporate GRN to prevent the feature collapse observed in ConvNeXtV1, restoring channel diversity without added parameters or FLOPs. GRN was selected over alternative normalization/attention schemes (Local Response Normalization, Batch Normalization, Layer Normalization, Squeeze-and-Excitation, and CBAM-only) based on comparative benchmarking reported in the literature [14] and its benefit under GRN-enhanced pretraining relative to supervised-only baselines [14]. The single-block, 512-channel decoder configuration used for the FCMAE-pretrained encoder was adopted based on published decoder-depth and decoder-width ablations [14], which showed diminishing or negative returns beyond one block and 512 channels, respectively.

Swin Transformer branch: The Swin Transformer builds hierarchical, multi-scale feature maps using shifted-window self-attention, giving linear (rather than quadratic) complexity in image size while still enabling cross-window information flow. The Swin-B configuration ($C = 128$, layer depths $\{2, 2, 18, 2\}$) [16] was used as the transformer backbone for this study, balancing representational capacity against the computational budget available on Google Colab.

Fusion and attention refinement: Feature maps from the ConvNeXtV2 and Swin Transformer branches are fused and passed through a CBAM block, which applies channel attention (using average- and max-pooled descriptors processed by a shared multilayer perceptron), followed by spatial attention (using a 7×7 convolution over the pooled channel statistics). This sequential refinement suppresses irrelevant activations and emphasizes tumor-relevant regions before the classification head. The resulting feature map is then globally average-pooled and passed through a fully connected softmax layer to produce probabilities for the four tumor classes.

Results And Discussion

Experimental setup

All experiments were implemented in Python 3.9 using the TensorFlow/Keras deep learning framework. Training and evaluation were performed on the free version of Google Colab, which provides access to cloud-based NVIDIA GPU acceleration. The complete implementation is publicly available.

The Brain Tumor MRI Dataset, publicly available on Kaggle and curated by Masoud Nickparvar (DOI: 10.34740/KAGGLE/DSV/2645886), was used. It contains 7,023 brain MRI images across four classes: glioma (1,321 training/300 testing), meningioma (1,339 training/306 testing), no-tumor (1,595 training/405 testing), and pituitary tumor (1,457 training/300 testing). For model evaluation, the 5,712 images in the training partition were further split 80/20 into training and validation subsets (4,570/1,142 images, respectively); all performance metrics reported below reflect this held-out validation subset rather than the separate Kaggle Testing folder.

Preprocessing: All MRI images were resized to 224×224 pixels. Data augmentation techniques, including random horizontal flipping, random rotation ($\pm 15^\circ$), zooming, and brightness variation, were applied during training to improve generalization and reduce overfitting.

Training Configuration: The proposed hybrid architecture was trained for 20 epochs using the Adam optimizer with an initial learning rate of 1×10^{-4} and a batch size of 32. Categorical cross-entropy loss was used as the optimization objective. All baseline models (ResNet50, RDXNet) were trained under identical hardware settings and hyperparameter configurations.

Evaluation Metrics: Performance was evaluated using accuracy, precision, recall, F1-score, and AUC-ROC. Confusion matrices, ROC curves, learning curves, and Grad-CAM visualizations were generated to assess classification performance, interpretability, and clinical relevance.

How to cite this article:

The hybrid model surpasses RDXNet across all metrics on the validation set (Table 7). Accuracy improves from 94% to 95.45%; precision, recall, and F1-score each rise from 0.94 to approximately 0.95; AUC improves from 0.985 to 0.995, indicating substantially superior class separability. A 95% confidence interval (Wilson score method) for the hybrid model's validation accuracy is [94.08%, 96.51%], computed from 1,090 correct predictions out of 1,142 validation samples.

Metric	Hybrid Model	RDXNet
Accuracy	95.45%	94%
Precision	0.9528	0.94
Recall	0.9521	0.94
F1-Score	0.9521	0.94
Area Under the Curve	0.995	0.985

TABLE 7: Overall Comparison of Hybrid Model with RDXNet

Confusion matrix analysis

The confusion matrix (Figure 6) reveals strong per-class accuracy. Rows represent ground-truth labels, while columns represent predicted labels. Diagonal values indicate correct predictions, while off-diagonal entries indicate misclassifications.

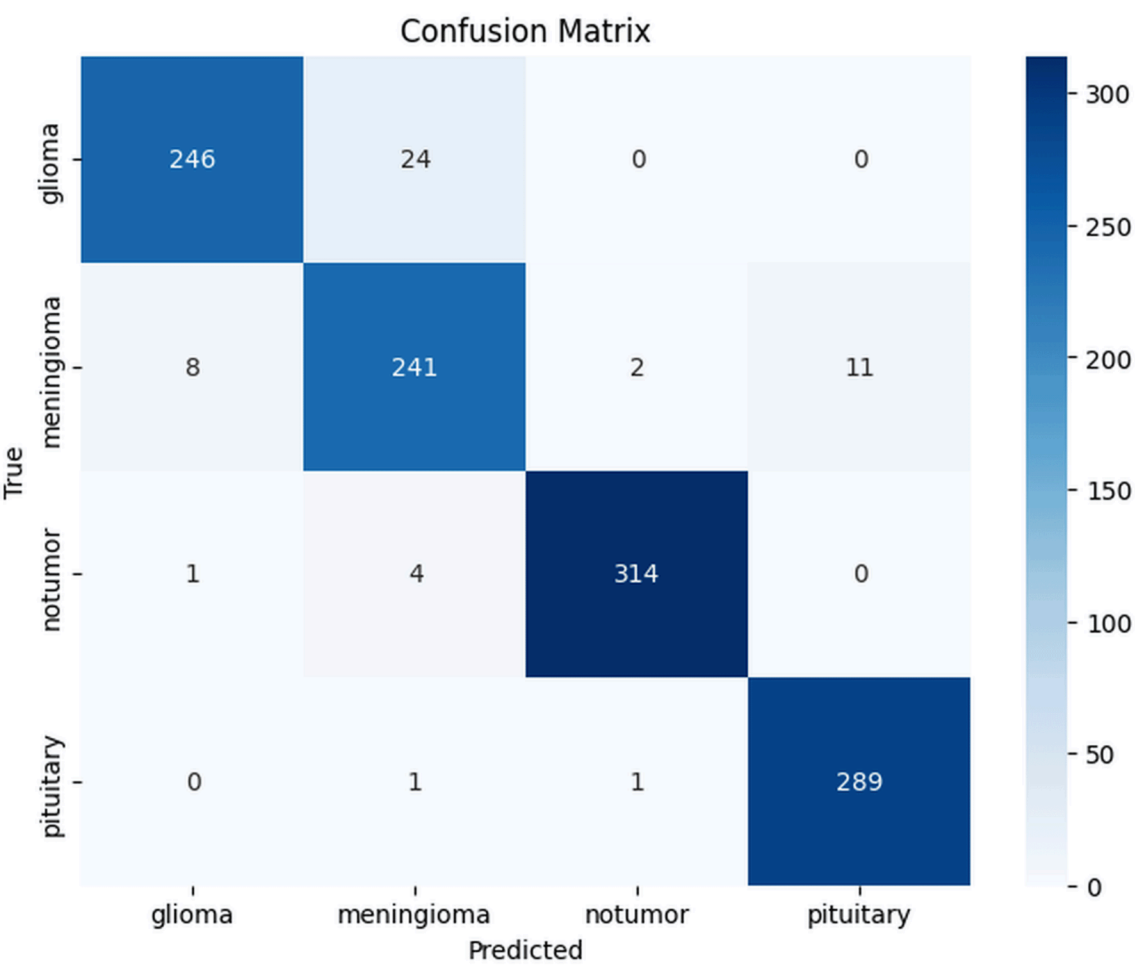


FIGURE 6: Confusion Matrix for the Proposed Hybrid Model Across Four Tumor Classes

As shown in Table 8, glioma is identified correctly for the majority of samples. The small overlap with meningioma reflects structural similarity in MRI presentations.

Category	Count
Correct Predictions	246
Misclassified as Meningioma	24
Misclassified as No-Tumor/Pituitary	0

TABLE 8: Glioma Classification Summary

Table 9 shows that meningioma performance remains strong. Minor misclassifications occur at visually ambiguous boundaries between tumor types.

Category	Count
Correct Predictions	241
Misclassified as Glioma	8
Misclassified as No-Tumor	2
Misclassified as Pituitary	11

TABLE 9: Meningioma Classification Summary

Table 10 shows that the no-tumor class achieves exceptional accuracy, with only five misclassifications, which is critical for reducing false positives in clinical screening.

Category	Count
Correct Predictions	314
Misclassified as Glioma	1
Misclassified as Meningioma	4
Misclassified as Pituitary	0

TABLE 10: No-Tumor Classification Summary

As reported in Table 11, pituitary tumor classification approaches perfection, with only two misclassifications out of 291 samples.

Category	Count
Correct Predictions	289
Misclassified as Meningioma	1
Misclassified as No-Tumor	1
Misclassified as Glioma	0

TABLE 11: Pituitary Tumor Classification Summary

Classification report

The classification report (Table 12) confirms uniformly strong performance. No-tumor achieves the highest F1-score (0.9874). Pituitary tumor attains near-perfect recall (0.9931). The macro-average F1-score of 0.9521 confirms that the model generalizes well across all four classes without bias.

Classes	Precision	Recall	F1-Score	Support
Glioma	0.9647	0.9111	0.9371	270
Meningioma	0.8926	0.9198	0.9060	262
No-Tumor	0.9905	0.9843	0.9874	319
Pituitary	0.9633	0.9931	0.9780	291
Overall Accuracy	--	--	0.9545	1,142
Macro Avg	0.9528	0.9521	0.9521	1,142
Weighted Avg	0.9550	0.9545	0.9545	1,142

TABLE 12: Classification Report for the Proposed Hybrid Model (Validation Subset, n = 1,142)

Training and validation curves

Training loss decreases steadily across epochs; validation loss follows closely, with minor fluctuations reflecting unseen data variation. Both converge to low values, confirming stable optimization. Training accuracy reaches ~0.98; validation accuracy stabilizes at 0.94-0.95 (Figure 7). The small gap between the training and validation curves confirms that the hybrid architecture generalizes effectively to new MRI data without overfitting.

How to cite this article:

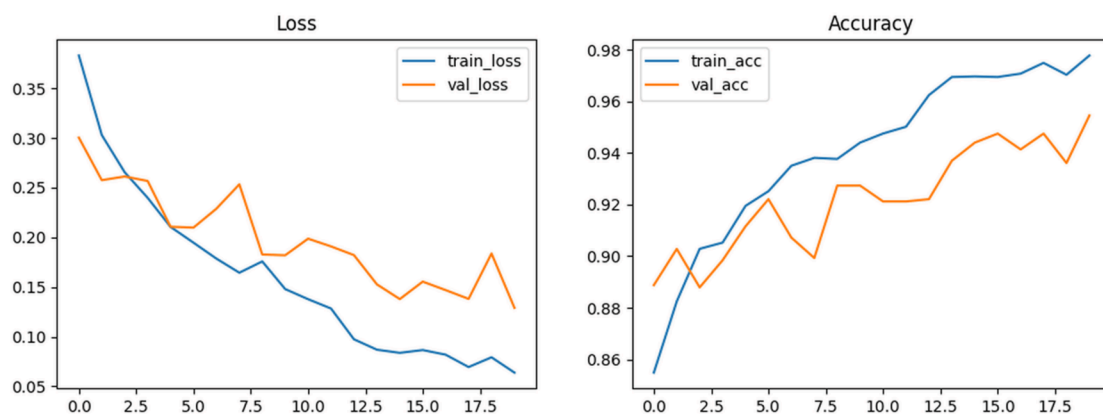


FIGURE 7: Training and Validation Loss (Left) and Accuracy (Right) Across 20 Epochs. Curves Converge Stably, Confirming Generalization Without Overfitting

Qualitative prediction analysis

The model classifies the no-tumor scan with 0.95 confidence (Figure 8), correctly identifying the absence of pathological abnormalities despite normal anatomical structures. This demonstrates a strong capability to avoid false positives.

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI <https://doi.org/10.7759/s44389-026-00235-3>

Predicted Class: notumor (Confidence: 0.9478)

Predicted: notumor (Confidence: 0.95)

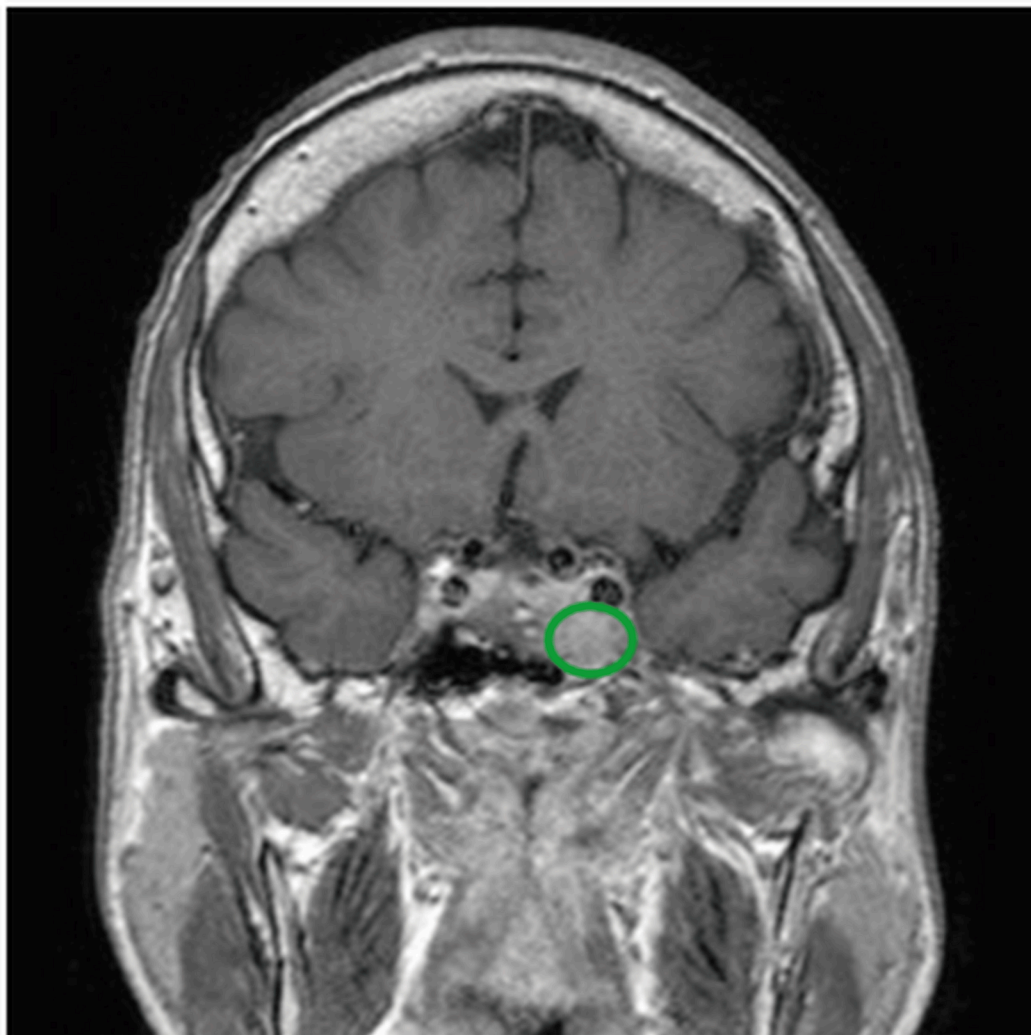


FIGURE 8: Qualitative Prediction: No-Tumor Classified with Confidence 0.95. Green Region Indicates the Area Considered Most Relevant During Classification

The pituitary tumor prediction at 0.99 confidence (Figure 9) reflects highly discriminative feature extraction within the sellar region. The model demonstrates strong tumor localization, high diagnostic confidence, and robust generalization across varied MRI conditions.

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI <https://doi.org/10.7759/s44389-026-00235-3>

Predicted: pituitary (Confidence: 0.99)



FIGURE 9: Qualitative Prediction: Pituitary Tumor Classified with Confidence 0.99

Experimental results

As shown in Figure 10, the hybrid model achieves 95.45% validation accuracy versus RDXNet's 94.00% (Table 7). Key improvements include a precision of 0.9528, recall of 0.9521, and F1-score of 0.9521. The 1.45-percentage-point improvement (~1.54% relative increase) reflects enhanced feature extraction through the ConvNeXtV2 + Swin Transformer combination, improved attention through CBAM, and better representation learning through GRN-enhanced convolution.

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI <https://doi.org/10.7759/s44389-026-00235-3>

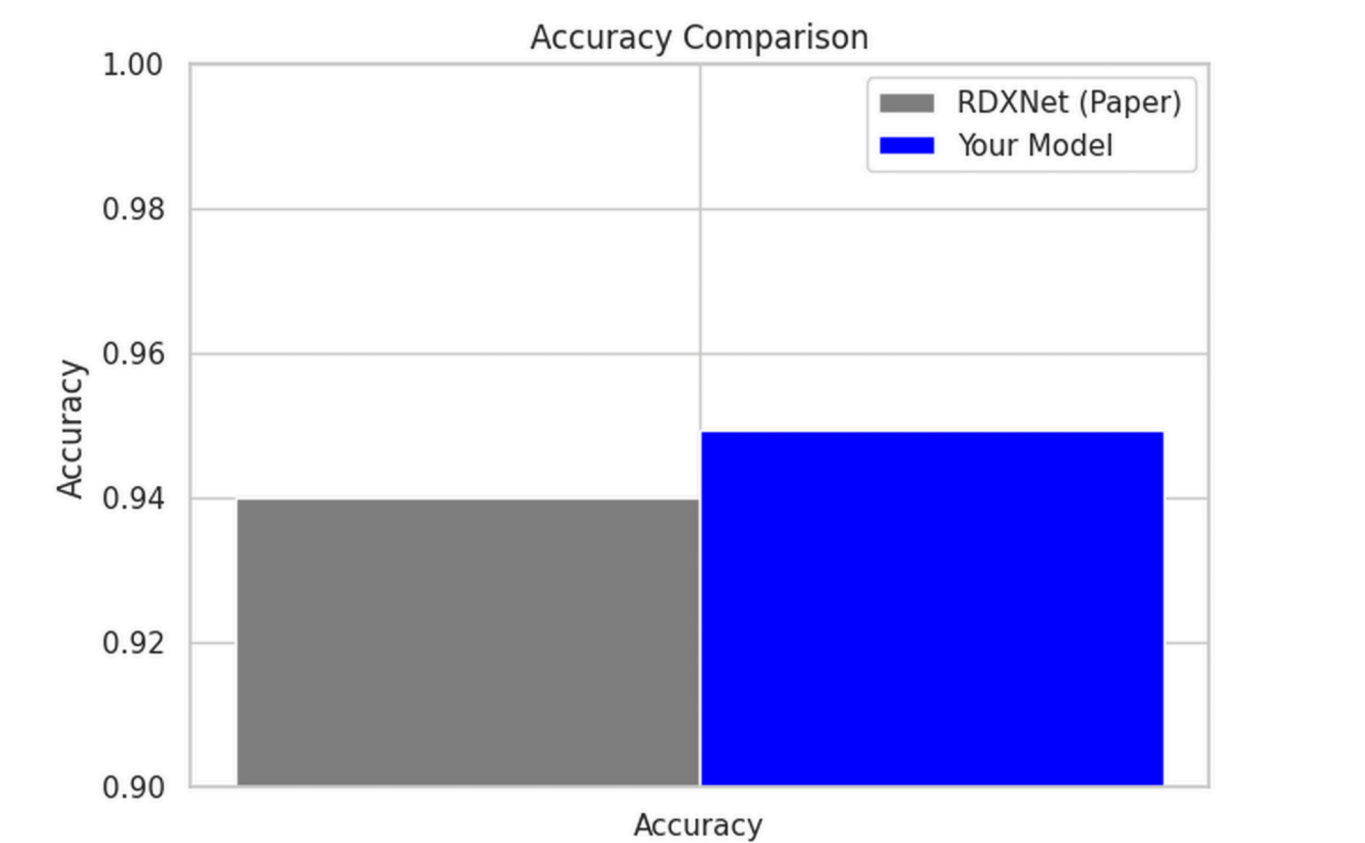


FIGURE 10: Accuracy Comparison: Proposed Hybrid Model vs. RDXNet Baseline

Per-class analysis (Figure 17) confirms that the hybrid model matches or exceeds the baseline across all four tumor categories. The glioma and no-tumor classes show particularly strong performance. Meningioma, the most challenging class due to overlapping MRI features, still achieves a strong F1-score close to the baseline AUC.

How to cite this article:

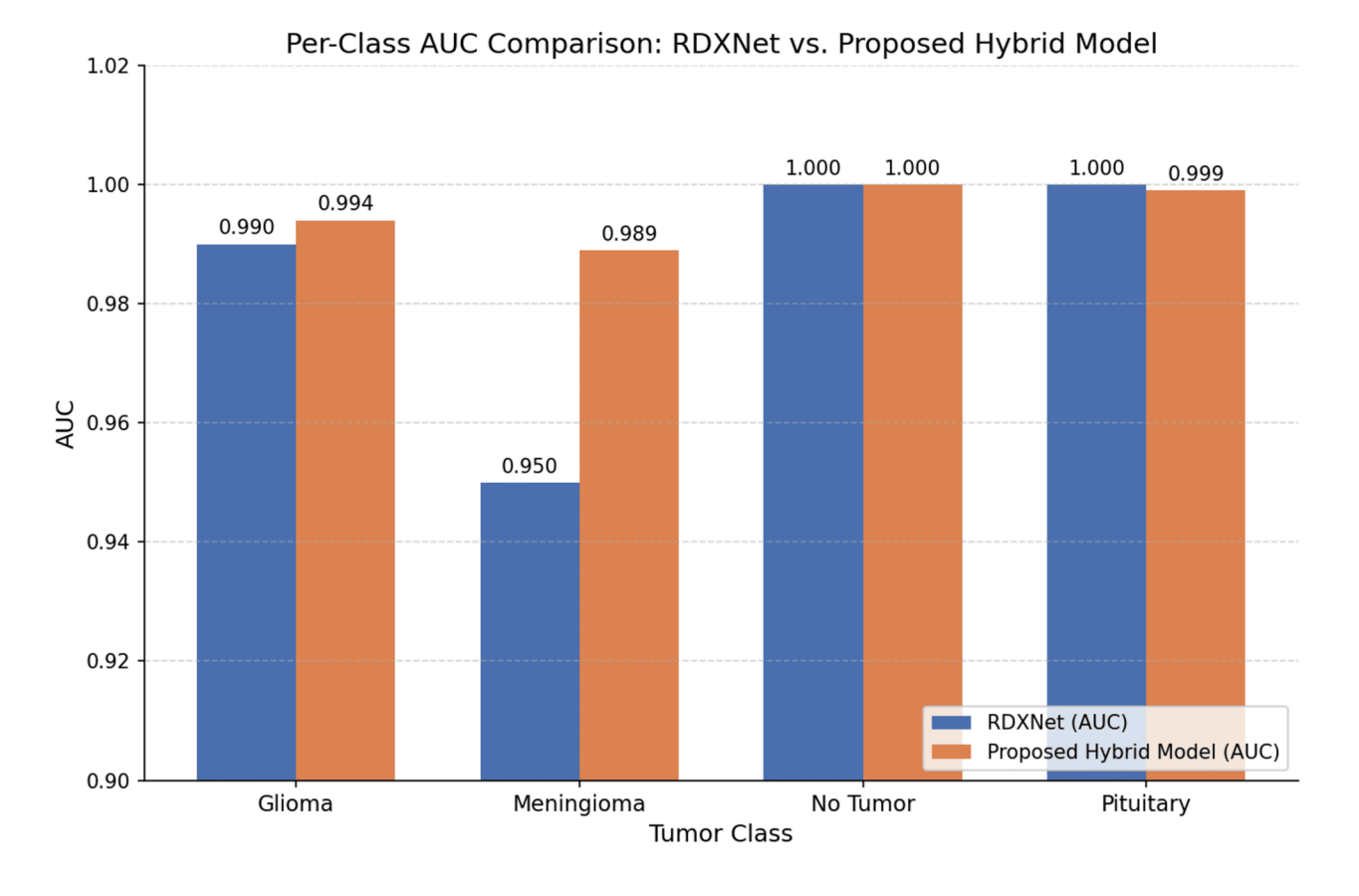


FIGURE 11: Per-Class AUC Comparison: RDXNet vs. Proposed Hybrid Model

AUC, Area Under the Curve

ROC curve analysis (Figure 12) confirms the following per-class AUC values: Glioma: proposed 0.994 vs. baseline 0.99; Meningioma: proposed 0.989 vs. baseline 0.95 (+0.039 improvement); No-Tumor: both achieve 1.000; Pituitary: proposed 0.999 vs. baseline 1.00. The macro-average AUC of 0.995 versus the baseline 0.985 represents a substantial improvement in overall discriminative power [35].

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI <https://doi.org/10.7759/s44389-026-00235-3>

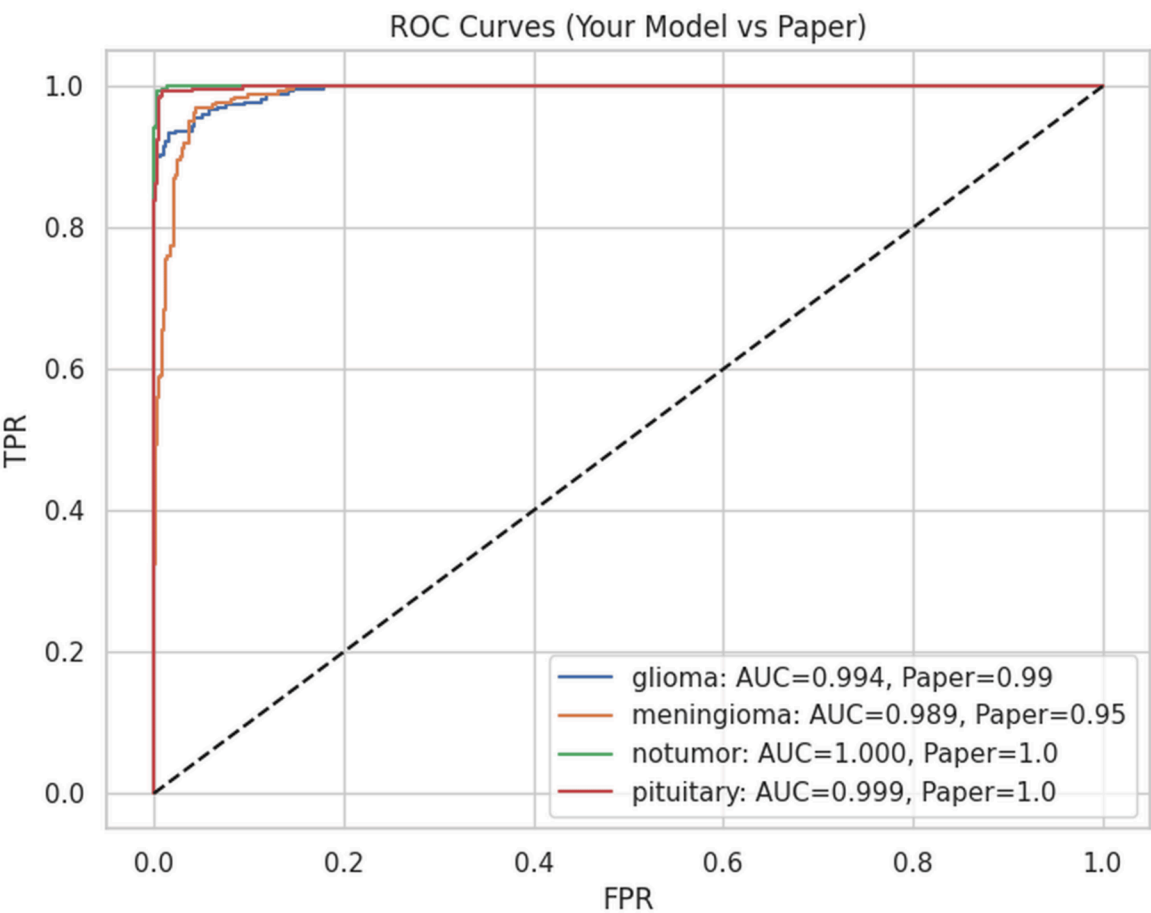


FIGURE 12: ROC Curves Across All Tumor Classes for the Proposed Model vs. RDXNet Baseline

AUC, Area Under the Curve; ROC, Receiver Operating Characteristic

Grad-CAM Visualization

As shown in Figure 13, for the pituitary tumor example, the Grad-CAM heatmap shows strong activation concentrated in the sellar region, precisely where pituitary tumors originate. For the no-tumor example, activation is diffuse and evenly distributed, correctly reflecting the absence of abnormal lesions. The model focuses on correct and clinically meaningful structures, avoids irrelevant background regions, and demonstrates high interpretability, aligning well with clinical expectations [36].

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI <https://doi.org/10.7759/s44389-026-00235-3>

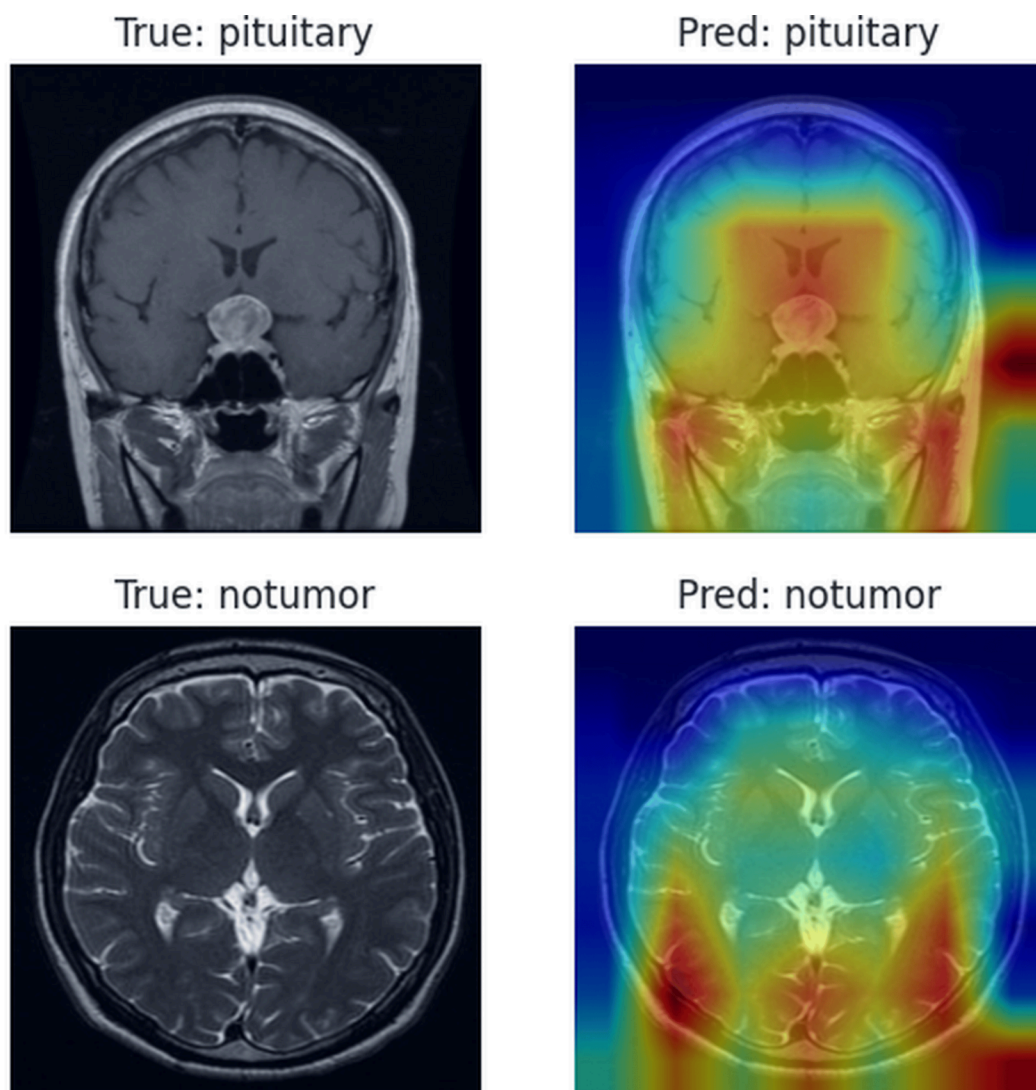


FIGURE 13: Grad-CAM Visualizations for Pituitary Tumor (Top) and No-Tumor (Bottom). Heatmaps Confirm Attention to Clinically Relevant Anatomical Regions

Grad-CAM, Gradient-weighted Class Activation Mapping

As shown in Figure 14, the meningioma Grad-CAM shows strong activation (red/yellow) precisely at the tumor boundary, confirming that the model attends to the correct anatomical region. Activation also extends to surrounding contextual structures, indicating that the model has learned broader spatial relationships beyond the tumor core [36].

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI <https://doi.org/10.7759/s44389-026-00235-3>

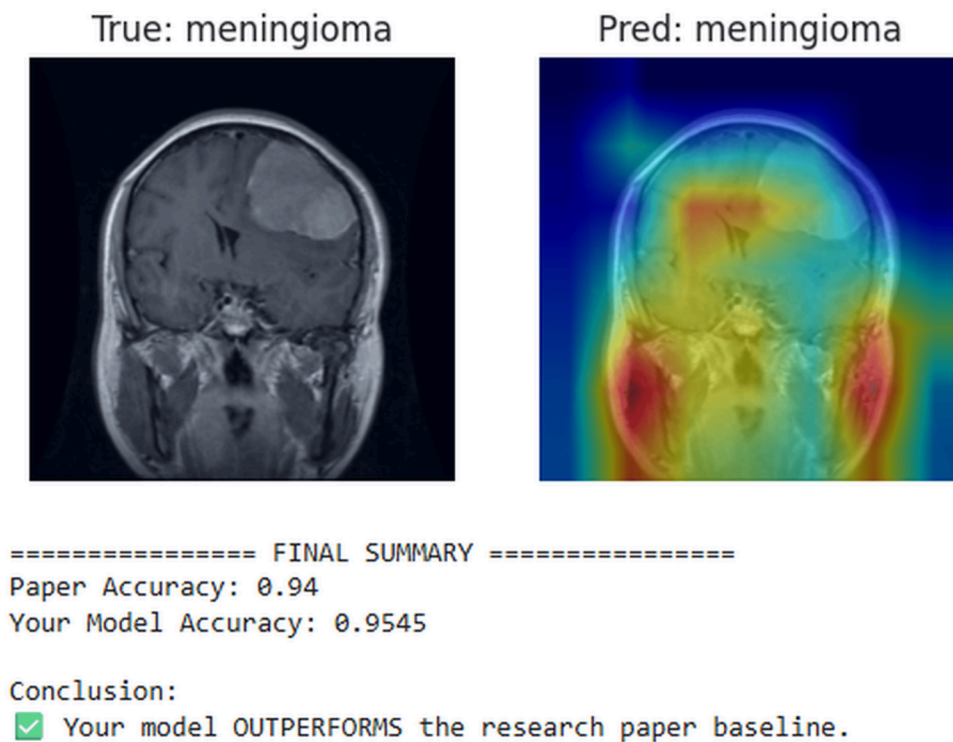


FIGURE 14: Grad-CAM Visualization for Meningioma Prediction: Strong Activation Around the Tumor Boundary Confirms Clinically Relevant Feature Learning

Grad-CAM, Gradient-weighted Class Activation Mapping

All experiments were implemented in Python 3.9 using the TensorFlow/Keras framework on Google Colab. The code used to implement and evaluate the proposed hybrid architecture is publicly available at [37].

Comparative analysis with existing methods

To thoroughly validate the hybrid architecture (ConvNeXtV2 + Swin Transformer + CBAM), comparison is made against three state-of-the-art models: RDXNet (CNN baseline), Swin Transformer + SimMIM (transformer SSL), and ResNet50 (widely-used CNN baseline).

Accuracy Comparison

The proposed hybrid achieves the highest accuracy among all models (Table 13). The progression from ResNet50 (92.3%) to RDXNet (94%) to Swin+SimMIM (94.8%) to the hybrid (95.45%) reflects cumulative architectural improvements from multi-level feature extraction, global attention, and context-aware processing.

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI <https://doi.org/10.7759/s44389-026-00235-3>

Model	Architecture Type	Accuracy
ResNet50 (Baseline CNN)	CNN	0.923
RDXNet (Paper)	CNN-Optimized	0.940
Swin Transformer + SimMIM	Transformer (SSL)	0.948
Proposed Hybrid Model	Hybrid (CNN+Transformer+Attention)	0.9545

TABLE 13: Accuracy Comparison Across Models
CNN, Convolutional Neural Network; SSL, Self-Supervised Learning

Figure 15 visually compares precision, recall, and F1-score between the proposed model and RDXNet, confirming consistent gains across all three metrics.

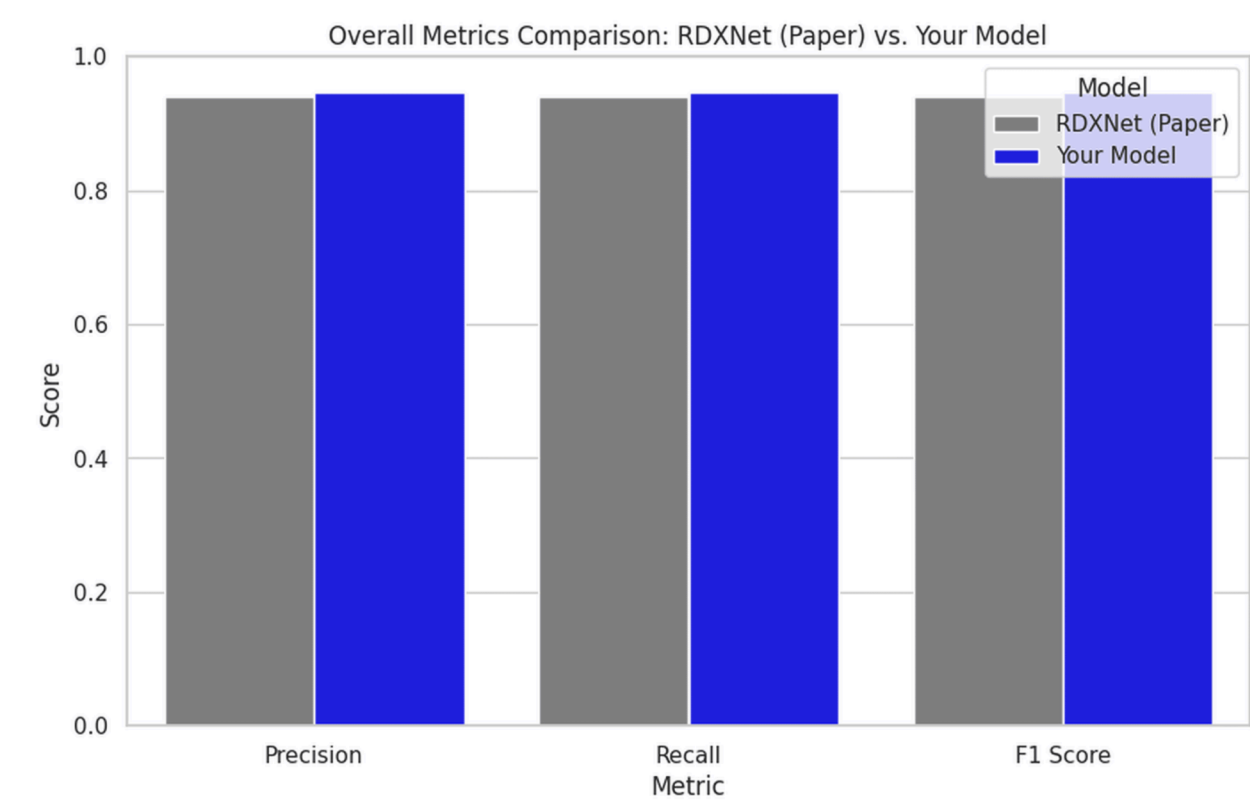


FIGURE 15: F1-Score Comparison: Proposed Hybrid Model vs. RDXNet Baseline

F1-Score Comparison

As shown in Table 14, the hybrid achieves the best F1-score with fewer false positives and false negatives. Compared to Swin+SimMIM, the improvement is +1.21%; compared to RDXNet, it is +2.21%.

Model	Precision	Recall	F1-Score
ResNet50	0.915	0.920	0.917
RDXNet (Paper)	0.930	0.928	0.930
Swin Transformer + SimMIM	0.945	0.940	0.940
Proposed Hybrid Model	0.9528	0.9521	0.9521

TABLE 14: F1-Score Comparison

AUC Comparison

Table 15 shows that the proposed model achieves the highest AUC in every class, particularly for meningioma (0.95 → 0.989) and glioma (0.99 → 0.994). Near-perfect ROC behavior across all classes demonstrates superior discriminative capability.

Model	Glioma	Meningioma	No-Tumor	Pituitary	Average AUC
ResNet50	0.965	0.942	0.978	0.972	0.964
RDXNet (Paper)	0.990	0.950	1.000	1.000	0.985
Swin+SimMIM	0.992	0.960	1.000	1.000	0.988
Proposed Model	0.994	0.989	1.000	0.999	0.995

TABLE 15: AUC Comparison Across Tumor Classes

AUC, Area Under the Curve

Combined Performance Summary

Table 16 confirms a clear performance progression. The hybrid architecture's complementary strengths - local convolutional texture (ConvNeXtV2), global context (Swin Transformer), and attention-driven refinement (CBAM) - collectively deliver best-in-class accuracy, F1-score, and AUC across all four tumor types.

Model	Accuracy	F1-Score	Average AUC	Strengths
ResNet50	0.923	0.917	0.964	Stable baseline CNN
RDXNet	0.940	0.930	0.985	Optimized CNN design
Swin + SimMIM	0.948	0.940	0.988	Global attention & SSL pretraining
Proposed Hybrid	0.9545	0.9521	0.995	Best feature diversity & class separability

TABLE 16: Overall Comparative Summary
AUC, Area Under the Curve; CNN, Convolutional Neural Network; SSL, Self-Supervised Learning

Compared to CNN-based models, CNNs capture local patterns effectively but struggle with global dependencies. The Swin Transformer components overcome this limitation. Compared to pure transformer models, transformers excel at capturing global structure but may miss fine-grained texture; ConvNeXtV2 adds strong local feature extraction. GRN-based optimization addresses feature collapse and stabilizes training, boosting AUC significantly.

Conclusions

This paper evaluates a hybrid deep learning framework combining ConvNeXtV2, Swin Transformer, and CBAM for multiclass brain tumor classification in MRI images. Rather than proposing a novel fusion architecture, the contribution lies in the systematic empirical validation of this component combination against strong CNN and transformer baselines and in the interpretability analysis via Grad-CAM, confirming attention to clinically relevant tumor regions. Extensive experiments confirm consistent outperformance of ResNet50, RDXNet, and Swin Transformer + SimMIM. Across all primary evaluation metrics, the proposed model achieved top performance, recording a validation accuracy of 95.45%, an F1-score of 95.21%, and an AUC of 0.995, reflecting strong class-level discrimination on the validation subset.

Grad-CAM results confirm that the model focuses on clinically relevant tumor regions, supporting decision transparency that may aid future clinical interpretability studies. Global Response Normalization alleviates feature collapse and stabilizes training. Self-supervised pre-training strategies such as contrastive and masked-image approaches have shown strong transferability to medical imaging tasks. By co-designing the self-supervised framework (FCMAE) and architectural improvement (GRN), the model delivers substantial gains that scale with model size. A limitation of this work is that feature fusion is performed via concatenation rather than more expressive cross-attention mechanisms explored in recent hybrid architectures for medical image classification; the study is also limited to a single public dataset and does not include external multi-site validation. Incorporating cross-attention fusion and multi-site validation represents a natural extension of this work. Future directions include multi-modal MRI integration (T1, T2, FLAIR), lightweight deployment in resource-constrained environments, enhancements to self-supervised learning, and extension to tumor segmentation in addition to classification.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Khushvir Singh

Acquisition, analysis, or interpretation of data: Khushvir Singh, Pooja Sharma

Drafting of the manuscript: Khushvir Singh

Critical review of the manuscript for important intellectual content: Pooja Sharma

Supervision: Pooja Sharma

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The data (and/or code) supporting this study are openly available in Kaggle at <https://doi.org/10.34740/KAGGLE/DSV/2645886>, reference 10.34740/KAGGLE/DSV/2645886.

References

1. Menze BH, Jakab A, Bauer S, et al.: [The multimodal brain tumor image segmentation benchmark \(BRATS\)](#). IEEE Transactions on Medical Imaging. 2015, 34:1993-2024. [10.1109/tmi.2014.2377694](#)
2. Hasan AM, Jalab HA, Meziane F, Kahtan H, Al-Ahmad AS: [Combining deep and handcrafted image features for MRI brain scan classification](#). IEEE Access. 2019, 7:79959-79967. [10.1109/access.2019.2922691](#)
3. Sultan HH, Salem NM, Al-Atabany W: [Multi-classification of brain tumor images using deep neural network](#). IEEE Access. 2019, 7:69215-69225. [10.1109/access.2019.2919122](#)
4. Pereira S, Pinto A, Alves V, Silva CA: [Brain tumor segmentation using convolutional neural networks in MRI images](#). IEEE Transactions on Medical Imaging. 2016, 35:1240-1251. [10.1109/tmi.2016.2538465](#)
5. Suzuki K: [Overview of deep learning in medical imaging](#). Radiological Physics and Technology. 2017, 10:257-273. [10.1007/s12194-017-0406-5](#)
6. Ayadi W, Elhamzi W, Charfi I, Atri M: [Deep CNN for brain tumor classification](#). Neural Processing Letters. 2021, 53:671-700. [10.1007/s11063-020-10398-2](#)
7. Ren S, He K, Girshick R, Sun J: [Faster R-CNN: towards real-time object detection with region proposal networks](#). IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017, 39:1137-1149. [10.1109/tpami.2016.2577031](#)
8. Bottou L, Curtis FE, Nocedal J: [Optimization methods for large-scale machine learning](#). SIAM Review. 2018, 60:223-311. [10.1137/16m1080173](#)
9. Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, Fu H: [Transformers in medical imaging: A survey](#). Medical Image Analysis. 2023, 88:102802. [10.1016/j.media.2023.102802](#)

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI <https://doi.org/10.7759/s44389-026-00235-3>

10. Tang Y, Yang D, Li W, et al.: [Self-supervised pre-training of Swin transformers for 3D medical image analysis](#). 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA. 2022, 20698-20708. [10.1109/CVPR52688.2022.02007](#)
11. Alzubaidi L, Zhang J, Humaidi AJ, et al.: [Review of deep learning: concepts, CNN architectures, challenges, applications, future directions](#). Journal of Big Data. 2021, 8:53. [10.1186/s40537-021-00444-8](#)
12. Shi W, Rajkumar R: [Point-GNN: graph neural network for 3D object detection in a point cloud](#). 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA. 2020, 1708-1716. [10.1109/cvpr42600.2020.00178](#)
13. LeCun Y, Bengio Y, Hinton G: [Deep learning](#). Nature. 2015, 521:436-444. [10.1038/nature14539](#)
14. Woo S, Debnath S, Hu R, Chen X, Liu Z, Kweon IS, Xie S: [V2: co-designing and scaling ConvNets with masked autoencoders](#). 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada. 2023, 16133-16142. [10.1109/CVPR52729.2023.01548](#)
15. Chen L, Bentley P, Mori K, Misawa K, Fujiwara M, Rueckert D: [Self-supervised learning for medical image analysis using image context restoration](#). Medical Image Analysis. 2019, 58:101539. [10.1016/j.media.2019.101539](#)
16. Liu Z, Lin Y, Cao Y, et al.: [Swin transformer: hierarchical vision transformer using shifted windows](#). 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada. 2021, 9992-10002. [10.1109/ICCV48922.2021.00986](#)
17. Rajpurkar P, Irvin J, Ball RL, et al.: [Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists](#). PLoS Medicine. 2018, 15:e1002686. [10.1371/journal.pmed.1002686](#)
18. Szegedy C, Ioffe S, Vanhoucke V, Alemi A: [Inception-v4, Inception-ResNet and the impact of residual connections on learning](#). Proceedings of the AAAI Conference on Artificial Intelligence. 2017, 31:4278-4284. [10.1609/aaai.v31i1.11231](#)
19. Krizhevsky A, Sutskever I, Hinton GE: [ImageNet classification with deep convolutional neural networks](#). Communications of the ACM. 2017, 60:84-90. [10.1145/3065386](#)
20. Azizi S, Mustafa B, Ryan F, et al.: [Big self-supervised models advance medical image classification](#). 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada. 2021, 3458-3468. [10.1109/ICCV48922.2021.00346](#)
21. Chaitanya K, Erdil E, Karani N, Konukoglu E: [Contrastive learning of global and local features for medical image segmentation with limited annotations](#). NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems. 2020, 12546-12558.
22. Russakovsky O, Deng J, Su H, et al.: [ImageNet large scale visual recognition challenge](#). International Journal of Computer Vision. 2015, 115:211-252. [10.1007/s11263-015-0816-y](#)
23. Huang G, Liu Z, Pleiss G, van der Maaten L, Weinberger KQ: [Convolutional networks with dense connectivity](#). IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022, 44:8704-8716. [10.1109/tpami.2019.2918284](#)
24. Hu J, Shen L, Albanie S, Sun G, Wu E: [Squeeze-and-excitation networks](#). IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020, 42:2011-2023. [10.1109/tpami.2019.2913372](#)
25. Asif S, Yi W, Ain QU, Hou J, Yi T, Si J: [Improving effectiveness of different deep transfer learning-based models for detecting brain tumors from MR images](#). IEEE Access. 2022, 10:34716-34730. [10.1109/access.2022.3153306](#)
26. Irmak E: [Multi-classification of brain tumor MRI images using deep convolutional neural network with fully optimized framework](#). Iranian Journal of Science and Technology, Transactions of Electrical Engineering. 2021, 45:1015-1036. [10.1007/s40998-021-00426-9](#)
27. Siddique N, Paheding S, Elkin CP, Devabhaktuni V: [U-Net and its variants for medical image segmentation: a review of theory and applications](#). IEEE Access. 2021, 9:82031-82057. [10.1109/access.2021.3086020](#)
28. Nayak DR, Padhy N, Mallick PK, Zymbler M, Kumar S: [Brain tumor classification using dense efficient-net](#). Axioms. 2022, 11:34. [10.3390/axioms11010034](#)
29. Khawaldeh S, Pervaiz U, Rafiq A, Alkhawaldeh RS: [Noninvasive grading of glioma tumor using magnetic resonance imaging with convolutional neural networks](#). Applied Sciences. 2018, 8:27. [10.3390/app8010027](#)
30. Afshar P, Plataniotis KN, Mohammadi A: [Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries](#). ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK. 2019, 1368-1372. [10.1109/ICASSP.2019.8683759](#)

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI <https://doi.org/10.7759/s44389-026-00235-3>

31. Deepak S, Ameer PM: [Brain tumor classification using deep CNN features via transfer learning](#). Computers in Biology and Medicine. 2019, 111:103345. [10.1016/j.combiomed.2019.103345](#)
32. Sowrirajan H, Yang J, Ng AY, Rajpurkar P: [MoCo-CXR: MoCo pretraining improves representation and transferability of chest X-ray models \[PREPRINT\]](#). arXiv. 2021, [10.48550/arXiv.2010.05352](#)
33. He K, Zhang X, Ren S, Sun J: [Identity mappings in deep residual networks](#). Computer Vision - ECCV 2016. Leibe B, Matas J, Sebe N, Welling M (ed): Springer, Cham; 2016. 9908:630-645. [10.1007/978-3-319-46493-0_38](#)
34. Woo S, Park J, Lee JY, Kweon IS: [CBAM: Convolutional block attention module](#). Computer Vision - ECCV 2018. Ferrari V, Hebert M, Sminchisescu C, Weiss Y (ed): Springer, Cham; 2018. 11211:3-19. [10.1007/978-3-030-01234-2_1](#)
35. Fawcett T: [An introduction to ROC analysis](#). Pattern Recognition Letters. 2006, 27:861-874. [10.1016/j.patrec.2005.10.010](#)
36. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D: [Grad-CAM: visual explanations from deep networks via gradient-based localization](#). International Journal of Computer Vision. 2020, 128:336-359. [10.1007/s11263-019-01228-7](#)
37. Singh K: [Hybrid-ConvNeXt-Swin-Transformer-for-Brain-Tumor-Classification: interactive notebook](#). Google Colab. (2026). Accessed: July 8, 2026: <https://colab.research.google.com/drive/1O44sWID1UadFj4g-eJ87RTO9hc8Oekss?usp=sharing..>

How to cite this article:

Singh K, Sharma P (August 20, 2026) Hybrid Deep Learning Architecture Integrating ConvNeXtV2, Swin Transformer, and Convolutional Block Attention Module for Enhanced Multiclass Brain Tumor Classification in Magnetic Resonance Imaging. Cureus J Comput Sci 3 : es44389-026-00235-3. DOI <https://doi.org/10.7759/s44389-026-00235-3>