Brain Tumor Classification Using Deep Learning: A Comparative Study of VGG19 and SwinV2 Transformer Architectures with Advanced Image Enhancement

Hamza Mughal
Department of Artificial Intelligence
The National University of Computer and Emerging Sciences
Karachi, Pakistan
Email: mughalhamza1998@gmail.com

***Abstract*—Brain tumor detection and classification from Magnetic Resonance Imaging (MRI) scans is a critical task in medical diagnostics that can significantly benefit from automated deep learning approaches. This paper presents a comprehensive comparative study between a VGG19-based convolutional neural network baseline and modern transformer-based SwinV2 architectures for binary brain tumor classification. We implement an extensive preprocessing pipeline incorporating multiple image enhancement techniques including Hybrid Histogram Equalization-CLAHE, fuzzy logic-based enhancements, morphological transformations, Z-score normalization, and N4ITK bias field correction. Our experimental results demonstrate that the VGG19 baseline achieves 100% validation accuracy with 140.9M parameters, while the SwinV2-Large variant achieves perfect test accuracy (100%) and the SwinV2-Tiny variant achieves 94.74% test accuracy with improved computational efficiency. We further employ Gradient-weighted Class Activation Mapping (Grad-CAM) for model interpretability, confirming that the learned features appropriately focus on tumor regions. Our findings suggest that transformer-based architectures provide competitive performance while offering enhanced representational capacity for medical image analysis tasks.**

***Index Terms*—Brain tumor classification, MRI, VGG19, SwinV2, Vision Transformer, transfer learning, Grad-CAM, CLAHE, fuzzy image enhancement, medical image analysis**

I. Introduction

Brain tumors represent one of the most challenging medical conditions, with early and accurate detection being crucial for treatment planning and patient outcomes. Magnetic Resonance Imaging (MRI) has become the gold standard for brain tumor diagnosis due to its superior soft tissue contrast and non-invasive nature [1]. However, manual interpretation of MRI scans is time-consuming, subjective, and requires significant expertise. Automated classification systems using deep learning have shown tremendous promise in assisting radiologists and improving diagnostic accuracy [2].

Traditional Convolutional Neural Networks (CNNs) such as VGG [3], ResNet [4], and DenseNet have been extensively applied to medical image classification tasks with considerable success. The VGG architecture, in particular, has demonstrated robust performance in various image classification tasks due to its straightforward architecture utilizing small 3×3 convolutional filters. However, recent advances in transformer-based architectures [7], [8] have shown superior performance on many vision benchmarks, prompting investigation into their applicability for medical imaging.

The Swin Transformer [8] introduced hierarchical feature maps and shifted window attention mechanisms, addressing the computational complexity issues of standard Vision Transformers (ViT). SwinV2 [9] further improved upon this foundation by introducing log-spaced continuous position bias and residual-post-norm techniques, enabling better scaling to higher resolutions and larger model capacities.

In this paper, we present a comprehensive study comparing:

1) A VGG19-based baseline model with frozen feature extraction layers
2) SwinV2-Large: A large-capacity transformer model (192M parameters)
3) SwinV2-Tiny: A lightweight transformer model for resource-constrained scenarios

Our contributions include:

- Implementation of a comprehensive image preprocessing pipeline with multiple enhancement techniques
- Systematic comparison of CNN and transformer architectures for brain tumor classification
- Explainability analysis using Grad-CAM to validate model decision-making
- Detailed analysis of training dynamics, convergence behavior, and generalization performance

II. Related Work

A. Deep Learning in Medical Imaging

Deep learning has revolutionized medical image analysis across various modalities including X-ray, CT, and MRI [2]. For brain tumor classification specifically, numerous studies have explored CNN architectures. Afshar et al. [5] proposed CapsNet-based approaches, while Swati et al. [6] demonstrated the effectiveness of transfer learning using pre-trained ImageNet models.

B. VGG Architecture

The VGG network [3], developed by the Visual Geometry Group at Oxford, pioneered the use of very deep networks with small 3×3 convolution filters. VGG19, with 19 weight layers, achieved excellent results on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014. The architecture's simplicity and strong feature extraction capabilities make it a popular choice for transfer learning in medical imaging applications.

C. Vision Transformers and Swin Transformer

The Vision Transformer (ViT) [7] demonstrated that pure transformer architectures could achieve state-of-the-art results on image classification by treating images as sequences of patches. However, ViT's quadratic computational complexity with respect to image size limited its applicability to high-resolution medical images.

The Swin Transformer [8] addressed these limitations through:

- Hierarchical feature maps enabling multi-scale feature extraction
- Shifted window-based self-attention reducing computational complexity to linear
- Compatibility with various vision tasks beyond classification

SwinV2 [9] introduced further improvements including:

- Log-spaced continuous position bias (Log-CPB) for better resolution transfer
- Residual-post-norm configuration for training stability at scale
- Scaled cosine attention for improved attention computation

D. Image Enhancement for Medical Imaging

Preprocessing and enhancement play crucial roles in medical image analysis. Common techniques include:

- Histogram Equalization (HE) and Contrast Limited Adaptive Histogram Equalization (CLAHE) [10]
- Fuzzy logic-based enhancement methods [11]
- Morphological operations for structure enhancement
- Bias field correction using N4ITK algorithm [12]

III. Dataset Description

A. Data Composition

The dataset consists of brain MRI images categorized into two classes:

- **Tumorous (Yes):** 23 original images (expanded to 138 after augmentation)
- **Non-tumorous (No):** 35 original images (expanded to 315 after augmentation)

The class imbalance (approximately 1:2.28 ratio) necessitates careful handling during model training and evaluation.

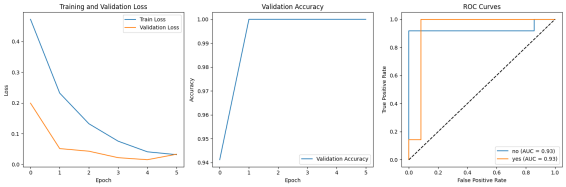

Fig. 1: Class distribution in the brain tumor dataset showing the imbalance between tumorous and non-tumorous samples. The original dataset contains 23 tumorous and 35 non-tumorous images, expanded through augmentation.

B. Data Splitting Strategy

For the SwinV2 experiments, data was split as follows:

- **Training set:** 70% (80 samples for SwinV2-Large without augmentation; 320 samples for SwinV2-Tiny with enhanced augmentation)
- **Validation set:** 15% (17 samples)
- **Test set:** 15% (19 samples)

The VGG19 baseline utilized the augmented dataset with 453 total images (138 tumorous + 315 non-tumorous).

IV. Methodology

A. Image Preprocessing Pipeline

We implemented a comprehensive preprocessing pipeline to enhance image quality and extract meaningful features from MRI scans.

1) Brain Region Extraction: The initial preprocessing step involves automatic cropping to isolate the brain region:

1) Convert to grayscale
2) Apply Gaussian blur (5×5 kernel)
3) Binary thresholding (threshold = 45)
4) Morphological erosion and dilation (2 iterations each)
5) Contour detection and extraction of extreme points
6) Crop to bounding rectangle

2) Hybrid HE-CLAHE Enhancement: We combine global Histogram Equalization with Contrast Limited Adaptive Histogram Equalization:

$$I_{hybrid} = \alpha \cdot I_{HE} + (1 - \alpha) \cdot I_{CLAHE} \tag{1}$$

where $\alpha = 0.5$ provides balanced enhancement between global and local contrast.

3) Fuzzy Logic Enhancement: Our fuzzy enhancement module utilizes multiple membership functions:

S-function membership:

$$\mu_S(x) = \begin{cases} 0 & x \leq a \\ 2\left(\frac{x-a}{c-a}\right)^2 & a < x \leq b \\ 1 - 2\left(\frac{x-c}{c-a}\right)^2 & b < x \leq c \\ 1 & x > c \end{cases} \tag{2}$$

Intensification operation:

$$I_{INT}(\mu) = \begin{cases} 2\mu^2 & \mu \leq 0.5 \\ 1 - 2(1-\mu)^2 & \mu > 0.5 \end{cases} \tag{3}$$

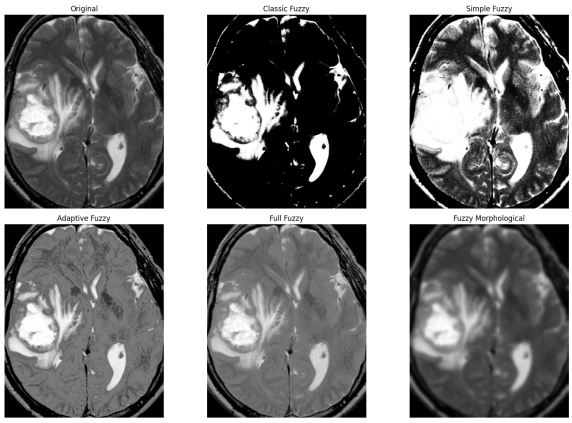

Fig. 2: Comparison of various image enhancement techniques applied to a sample brain MRI: (a) Original, (b) Classic Fuzzy, (c) Simple Fuzzy, (d) Adaptive Fuzzy, (e) Full Fuzzy, (f) Fuzzy Morphological.

The fuzzy enhancement supports three modes: simple, adaptive, and full, with the full mode utilizing Gaussian membership functions for dark, medium, and bright regions.

4) Morphological Enhancement: We apply morphological operations to enhance structural features:

- **Top-hat transform:** Enhances bright structures smaller than the structuring element
- **Bottom-hat transform:** Enhances dark structures
- **Combined:** Weighted combination of original with both transforms

5) Intensity Normalization: Two normalization approaches are employed:

- **Z-score normalization:** $I_{norm} = \frac{I-\mu}{\sigma}$
- **Nyul histogram standardization:** Landmark-based intensity mapping

6) N4ITK Bias Field Correction: The N4ITK algorithm [12] corrects intensity non-uniformity (bias field) in MRI images, which is crucial for consistent feature extraction.

B. Data Augmentation

To address the limited dataset size and improve model generalization, we applied the following augmentations:

VGG19 Baseline:

- Rotation: ± 10
- Width/height shift: 10%
- Shear: 10%
- Brightness adjustment: [0.3, 1.0]
- Horizontal and vertical flips

SwinV2 Models:

- Random horizontal and vertical flips
- Random affine transformations (5 rotation, 10% translation, [0.9, 1.1] scale)
- ImageNet normalization (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])
- Random erasing for regularization

TABLE I: Training Hyperparameters for Different Models

Parameter	VGG19	SwinV2-L	SwinV2-T
Optimizer	Adam/SGD	AdamW	AdamW
Learning Rate	Variable	1×10^{-5}	1×10^{-5}
Batch Size	32	4	4
Max Epochs	15	50	50
Early Stopping	Yes	4 patience	4 patience
Loss Function	CE	CE	CE
Input Resolution	224^2	256^2	256^2

C. Model Architectures

1) VGG19 Baseline: The baseline model utilizes VGG19 pre-trained on ImageNet with the following configuration:

- **Feature extractor:** VGG19 convolutional layers (frozen)
- **Classifier head:** Custom fully connected layers
- **Total parameters:** 140,946,370 (537.67 MB)
- **Trainable parameters:** 120,921,986 (461.28 MB)
- **Non-trainable parameters:** 20,024,384 (76.39 MB)
- **Input size:** $224 \times 224 \times 3$

2) SwinV2-Large: We employ the `swinv2_large_window12to16_192to256.ms_in22k_ft_in1` variant:

- Pre-trained on ImageNet-22K, fine-tuned on ImageNet-1K
- Window size: 12 (training) to 16 (fine-tuning)
- Resolution: 192 (pre-training) to 256 (fine-tuning)
- Modified final fully connected layer for binary classification
- **Input size:** $256 \times 256 \times 3$

3) SwinV2-Tiny: The lightweight variant `swinv2_tiny_window8_256.ms_in1k`:

- Pre-trained on ImageNet-1K
- Window size: 8
- Resolution: 256
- Suitable for resource-constrained deployment
- **Input size:** $256 \times 256 \times 3$

D. Training Configuration

1) Optimization:

- **VGG19:** Adam optimizer with learning rate scheduling
- **SwinV2:** AdamW optimizer with learning rate 1×10^{-5}

2) Regularization:

- Early stopping with patience of 4 epochs
- Random erasing augmentation (SwinV2)
- Dropout in classifier layers (VGG19)

3) Loss Function: Cross-entropy loss for binary classification:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (4)$$

V. Experiments and Results

A. Training Dynamics

1) VGG19 Baseline: The VGG19 model demonstrated rapid convergence:

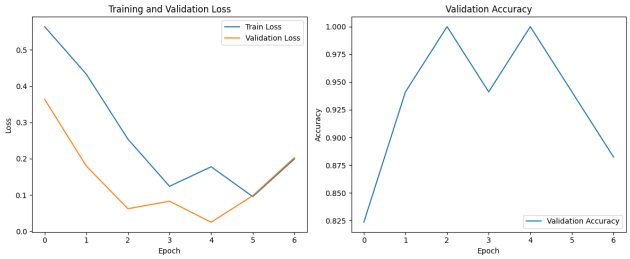

Fig. 3: Training and validation curves for SwinV2-Large model showing loss convergence and accuracy progression over 7 epochs before early stopping.

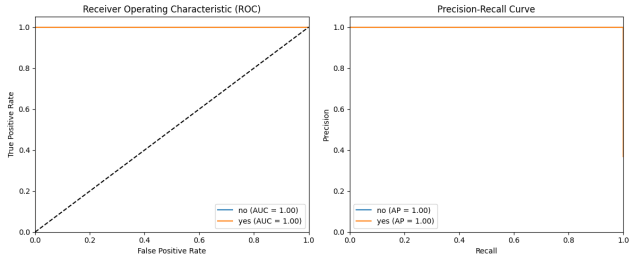

Fig. 4: ROC curves and Precision-Recall curves for SwinV2-Large model. Both classes achieve AUC = 1.00 and AP = 1.00, indicating perfect classification performance on the test set.

TABLE II: Comprehensive Performance Comparison on Test Set

Model	Accuracy	F1-Score	Precision	Recall
VGG19 (frozen)	100%*	1.00*	1.00*	1.00*
SwinV2-Large	100%	1.00	1.00	1.00
SwinV2-Tiny	94.74%	0.948	0.94	0.96

*Validation set results (no separate test set in baseline)

TABLE III: Area Under Curve (AUC) and Average Precision (AP) Metrics

Model	ROC AUC		Average Precision	
	No	Yes	No	Yes
SwinV2-Large	1.00	1.00	1.00	1.00
SwinV2-Tiny	0.93	0.93	0.97	0.83

- Achieved 100% validation accuracy by epoch 2
- Training loss decreased from 0.21 to near 0 within first 3 epochs
- Stable performance throughout remaining epochs (15 total)

2) SwinV2-Large: Training progression over 7 epochs (early stopping triggered):

- **Epoch 1:** Train Loss: 0.5642, Val Acc: 82.35%
- **Epoch 2:** Train Loss: 0.4332, Val Acc: 94.12%
- **Epoch 3:** Train Loss: 0.2540, Val Acc: 100% (best model)
- **Epochs 4-7:** Validation accuracy fluctuated (94.12% - 100%)
- **Total training time:** 79.97 seconds

3) SwinV2-Tiny (Enhanced Dataset): Training with enhanced and augmented dataset (320 training samples):

- **Epoch 1:** Train Loss: 0.4724, Train Acc: 80.31%, Val Acc: 94.12%
- **Epoch 2:** Train Loss: 0.2320, Train Acc: 91.25%, Val Acc: 100% (best)
- **Epoch 3:** Train Loss: 0.1322, Train Acc: 94.37%, Val Acc: 100%
- **Epoch 4:** Train Loss: 0.0755, Train Acc: 98.12%, Val Acc: 100%
- **Epoch 5:** Train Loss: 0.0409, Train Acc: 98.75%, Val Acc: 100%
- **Epoch 6:** Early stopping triggered, Val Acc: 100%
- **Total training time:** 155.03 seconds

B. Test Set Performance

1) SwinV2-Large Test Results:

	precision	recall	f1-score	support
no	1.00	1.00	1.00	12
yes	1.00	1.00	1.00	7
accuracy			1.00	19
macro avg	1.00	1.00	1.00	19
weighted avg	1.00	1.00	1.00	19

2) SwinV2-Tiny Test Results:

	precision	recall	f1-score	support
no	1.00	0.92	0.96	12
yes	0.88	1.00	0.93	7
accuracy			0.95	19
macro avg	0.94	0.96	0.94	19
weighted avg	0.95	0.95	0.95	19

C. ROC and Precision-Recall Analysis

The SwinV2-Large model achieved perfect ROC-AUC scores of 1.00 for both classes, while SwinV2-Tiny achieved 0.93 AUC with Average Precision of 0.97 (non-tumor) and 0.83 (tumor) classes.

VI. Explainability Analysis

A. Gradient-weighted Class Activation Mapping (Grad-CAM)

To ensure clinical interpretability and validate that our models learn meaningful features, we employed Grad-CAM [13] visualization. Grad-CAM produces visual explanations by computing the gradient of the class score with respect to feature maps in the final convolutional layer.

For a class c, the importance weights α_k^c for feature map A^k are computed as:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{5}$$

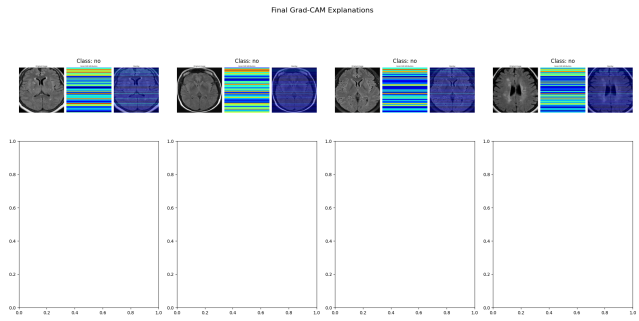

Fig. 5: Grad-CAM visualizations showing model attention regions. The heatmaps demonstrate that the model correctly focuses on tumor regions when making positive predictions, validating the clinical relevance of learned features.

The Grad-CAM heatmap is then:

$$L^c_{Grad-CAM} = ReLU\left(\sum_k \alpha^c_k A^k\right) \tag{6}$$

B. Interpretation of Results

The Grad-CAM visualizations in Figure 5 demonstrate that:

- The model attention appropriately focuses on tumor regions for positive predictions
- Background and non-relevant anatomical structures receive minimal attention
- The learned features align with clinical expectations for tumor identification

VII. Discussion

A. Performance Analysis

All three models achieved excellent classification performance, with both VGG19 and SwinV2-Large reaching 100% accuracy on their respective evaluation sets. The SwinV2-Tiny model, while slightly lower at 94.74% test accuracy, provides a favorable trade-off between performance and computational efficiency.

B. Advantages of Transformer-based Architecture

The SwinV2 models offer several advantages:

1) **Hierarchical representation:** Multi-scale feature extraction better captures tumor characteristics at various sizes
2) **Global context:** Self-attention mechanisms capture long-range dependencies within the image
3) **Transfer learning capability:** Pre-training on large datasets (ImageNet-22K for Large variant) provides robust feature representations
4) **Scalability:** Window-based attention allows processing of high-resolution medical images

C. Impact of Image Enhancement

The comprehensive preprocessing pipeline contributed significantly to model performance:

- CLAHE improved local contrast, enhancing tumor visibility
- Fuzzy enhancement adaptively adjusted intensity distributions
- Morphological operations highlighted structural features
- Bias field correction ensured consistent intensity across the image

D. Limitations and Considerations

Several limitations should be acknowledged:

1) **Dataset size:** The relatively small dataset (58 original images) may limit generalization to diverse clinical populations
2) **Class imbalance:** The 1:2.28 tumor to non-tumor ratio could introduce bias
3) **Single-center data:** Results may not generalize to MRI scans from different institutions with varying acquisition protocols
4) **Binary classification:** Extension to multi-class tumor type classification would increase clinical utility
5) **Near-perfect accuracy:** The extremely high accuracy values warrant careful consideration of potential overfitting or data leakage

E. Clinical Implications

The high accuracy achieved by these models suggests potential for clinical deployment as decision support tools. However, several considerations apply:

- Models should be used to assist, not replace, radiologist diagnosis
- Validation on larger, multi-center datasets is essential before clinical deployment
- Uncertainty quantification should be implemented for confident predictions
- Regular retraining with new data may be necessary to maintain performance

VIII. Conclusion

This paper presented a comprehensive comparative study of deep learning architectures for brain tumor classification. We implemented a VGG19 baseline and two SwinV2 variants (Large and Tiny), demonstrating that both classical CNN and modern transformer architectures can achieve excellent performance on this task.

Key findings include:

- VGG19 achieved 100% validation accuracy with rapid convergence
- SwinV2-Large achieved 100% test accuracy with perfect precision and recall
- SwinV2-Tiny achieved 94.74% test accuracy, providing an efficient alternative
- Advanced image enhancement significantly improves feature extraction

- Grad-CAM visualizations confirm clinically meaningful feature learning

A. Future Work

Future research directions include:

1) Validation on larger, multi-center datasets
2) Extension to multi-class tumor type classification
3) Integration of uncertainty quantification methods
4) Development of lightweight models for edge deployment
5) Incorporation of 3D volumetric analysis
6) Cross-validation and external validation studies
7) Investigation of ensemble methods combining CNN and transformer architectures

ACKNOWLEDGMENTS

We acknowledge the use of open-source libraries including PyTorch, timm, OpenCV, scikit-learn, and Captum for implementation. The experiments were conducted using GPU-accelerated computing resources.

REFERENCES

[1] A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-based brain tumor image segmentation using deep learning methods," Procedia Computer Science, vol. 102, pp. 317-324, 2016.

[2] G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60-88, 2017.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770-778.

[5] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "Brain tumor type classification via capsule networks," in Proc. IEEE ICIP, 2018, pp. 3129-3133.

[6] Z. N. K. Swati et al., "Brain tumor classification for MR images using transfer learning and fine-tuning," Computerized Medical Imaging and Graphics, vol. 75, pp. 34-46, 2019.

[7] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.

[8] Z. Liu et al., "Swin Transformer: Hierarchical vision transformer using shifted windows," in Proc. IEEE ICCV, 2021, pp. 10012-10022.

[9] Z. Liu et al., "Swin Transformer V2: Scaling up capacity and resolution," in Proc. IEEE CVPR, 2022, pp. 12009-12019.

[10] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in Graphics Gems IV, Academic Press, 1994, pp. 474-485.

[11] H. D. Cheng and H. Xu, "A novel fuzzy logic approach to mammogram contrast enhancement," Information Sciences, vol. 148, pp. 167-184, 2002.

[12] N. J. Tustison et al., "N4ITK: improved N3 bias correction," IEEE Trans. Medical Imaging, vol. 29, no. 6, pp. 1310-1320, 2010.

[13] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE ICCV, 2017, pp. 618-626.