RESEARCH ARTICLE

A Novel Deep Learning Framework for Brain Tumor Classification Using Improved Swin Transformer V2

Nur Alam[1], Yutong Zhu[2], Jiaqi Shao[3], Muhammad Usman[4] and Muhammad Fayaz[1,*]

[1] Department of Computer Science and Engineering, Sejong University, Seoul 05006, Republic of Korea
[2] Department of Anesthesiology, Dalian Medical University, Dalian 116000, China
[3] Department of Clinical Medicine, Shandong Xiehe University, Jinan 250109, China
[4] College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China

Abstract

Brain tumors pose a serious threat to global health, making accurate and early detection essential for effective treatment planning. While Magnetic Resonance Imaging (MRI) is widely used for diagnosis, manual interpretation is time-consuming and subject to error. This has prompted increasing use of deep learning for automated tumor classification. We propose a novel framework based on the Swin Transformer V2 architecture for classifying brain tumors in MRI scans into glioma, meningioma, pituitary tumor, and non-tumor categories. The design introduces two core innovations: a Dual-Branch Down-sampling module and an Enhanced Attention Mechanism, which improve multi-scale feature representation and computational efficiency. Using a dataset of 7,023 grayscale MRI images, the proposed model achieved an accuracy of 98.97%, outperforming ResNet50 (90.39%) and DenseNet121 (93.20%). It maintained precision, recall, and F1-scores above 98% across all classes and showed improved training efficiency. These results demonstrate the model's potential as a robust and efficient diagnostic support system for brain tumor classification.

Keywords: brain tumor classification, deep learning, MRI scans, computational efficiency, medical image analysis.

1 Introduction

Brain tumors, resulting from the abnormal proliferation of cells within brain tissues, pose a major global health concern and are ranked as the second most common cause of death globally, as reported by the World Health Organization (WHO) [1, 2]. These tumors are typically classified as either benign or malignant. Benign types, like meningioma and those affecting the pituitary gland, usually grow slowly, remain localized, and have a low recurrence rate following surgical removal. In contrast, malignant tumors, such as gliomas, are more aggressive, frequently infiltrate nearby brain regions, and can severely disrupt normal physiological functions if not promptly treated [3, 4]. Accurate and early detection of tumor types is critical for timely and

Submitted: 08 January 2025
Accepted: 29 January 2025
Published: 10 February 2025

Vol. 2, **No.** 1, 2025.
10.62762/TACS.2025.807755

***Corresponding author:**
✉ Muhammad Fayaz
Muhammadfayaz@sju.ac.kr

Citation
Alam, N., Zhu, Y., Shao, J., Usman, M., & Fayaz, M. (2025). A Novel Deep Learning Framework for Brain Tumor Classification Using Improved Swin Transformer V2. *ICCK Transactions on Advanced Computing and Systems*, 2(1), 1–10.

effective treatment planning.

Magnetic Resonance Imaging (the MRI) serves as a widely adopted, non-invasive technique for diagnosing brain tumor classification. Even though MRI offers detailed insights into the structure of tissues, there are some challenges concerning large volumes of data and a potential human factor in manual interpretations. Traditional diagnostic methods, such as biopsy, are invasive and resource-intensive, despite their high accuracy. Moreover, the variability in tumor morphology such as shape, size, and density adds complexity to classification, highlighting the need for automatic, efficient, and accurate diagnostic solutions.

Recent progress in Artificial Intelligence (AI) and Machine Learning (ML) technologies have revolutionized medical imaging by introducing robust, automated tools for analyzing MRI data [5–8]. Convolutional Neural Networks (CNNs), in particular, have demonstrated exceptional capabilities in feature extraction and image classification, significantly reducing dependency on extensive preprocessing. However, the conventional CNN architecture often faces challenges in effectively capturing global contextual information and balancing resource efficiency, which are critical for precise tumor classification [9–11].

In this study, we propose a novel framework based on the Swin Transformer V2 architecture that introduces a key innovation the Dual-Branch Down-sampling module which enhances multi-scale feature representation and improves both accuracy and efficiency. This research offers an enhanced framework in deep learning [12–14] enhanced classification based on the Swin Transformer V2 architecture to address these issues. It also gives local and global morphological MRI features to accurately fine-tune tumor classification, utilizing the hierarchical design and attention mechanisms of the proposed framework. The framework is aimed at the gliomas (malignant), meningiomas, pituitary tumors (benign), and no-tumor cases with high accuracy and efficiency.

The proposed approach underwent stringent testing against benchmark datasets that substantiated its performance over existing pre-trained models regarding classification accuracy and computational efficiency. Besides, the optimized architecture of the model also promises scalability and robustness in addressing the practical needs of real-world clinical applications.

The subsequent sections are organized as follows: Section 2 presents a review of related work. Section 3 outlines the proposed system model. Section 4 covers the experimental results and corresponding discussion, while Section 5 concludes the paper with final remarks.

2 Related Work

The application of deep learning (DL) in radiology has demonstrated significant potential for improving diagnostic accuracy and precision in medical imaging. Regardless of all these developments, the combination of DL techniques with the skills of a radiologist is still needed to aid in the improvement of the diagnosis. A number of researchers have developed new methods for the classification of brain tumors from MRI scans using new methods of feature extraction, preprocessing steps, and models [15]. Existing research efforts can be broadly categorized into CNN-based methods, transformer and hybrid models, traditional machine learning approaches, and transfer learning techniques.

2.1 CNN-Based Approaches

Gumaei et al. [16] offered a technique that integrates advanced feature extraction with a Regularized Extreme Learning Machine (RELM), achieving an accuracy of 94.23% using min-max normalization and hybrid feature selection. Srujan et al. [17] proposed a 16-layer CNN architecture using ReLU activations and the Adam optimizer, reporting 95.36% accuracy on primary tumor classification. Similarly, Huang et al. [30] introduced the CNNBCN model using random graph methods, achieving 95.49% accuracy. Deepak et al. [24] combined CNNs with Support Vector Machines (SVMs) in a two-stage approach, attaining 95.82% accuracy via five-fold cross-validation. Kaplan et al. [9] explored novel local binary patterns (nLBP and αLBP), reaching 95.56% accuracy using KNN classifiers.

2.2 Transformer-Based and Hybrid Models

Ahmad et al. [18] integrated generative adversarial networks (GANs) with variational autoencoders to generate synthetic tumor images, achieving 96.25% classification accuracy. Sun et al. [**?**] presented a model utilizing features extracted from a pre-trained GAN discriminator, enhanced with data augmentation and dropout, achieving 95.6% accuracy. Our work contributes to this direction by introducing an improved Swin Transformer V2 architecture with a novel Dual-Branch Down-sampling module, aimed at

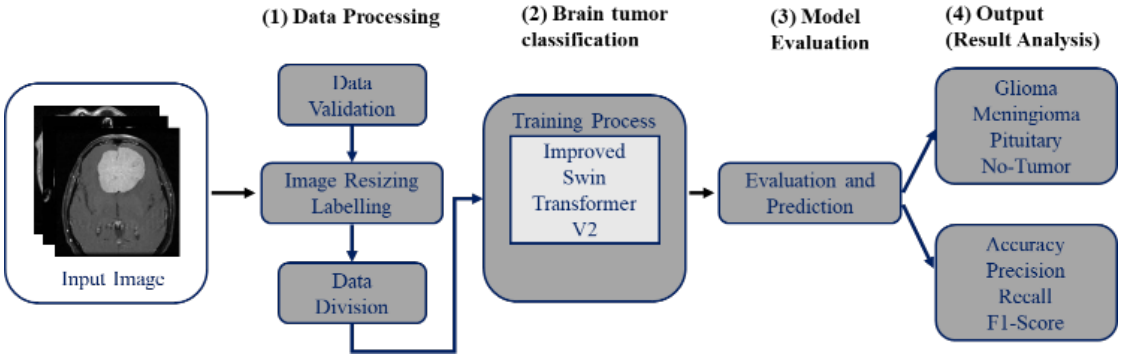


Figure 1. The overall architecture of the proposed framework for brain tumor classification.

improving classification accuracy while maintaining computational efficiency.

2.3 Traditional Machine Learning Methods

Ayadi et al. [19] utilized classical techniques like Dense SURF and HOG features with SVMs, achieving 90.27% accuracy. Sawant et al. [**?**] proposed a segmentation-classification pipeline using WSVM, HIK-SVM, and KNN classifiers with MODE-based ensemble strategies, attaining 92.46% precision.

2.4 Transfer Learning and Feature Fusion

Swati et al. [20] applied transfer learning using models like VGG19, VGG16, and AlexNet, reaching accuracies of 94.82%, 94.65%, and 89.95%, respectively. Noreen et al. [21] leveraged pre-trained networks such as InceptionV3 [22] and Xception, combining them with classifiers like SVM and Random Forest, reporting 94.34% accuracy. Satyanarayana et al. [23] combined CNNs with Mass Correlation Analysis and noise elimination methods, achieving 94% accuracy. Deepak et al. [24] tackled data imbalance using class-weighted focal loss and feature fusion, attaining 94.9% and 95.6% accuracy with SVM and KNN classifiers, respectively.

These studies highlight the advancements across multiple deep learning and machine learning paradigms for brain tumor classification. Despite progress, challenges remain particularly in managing data imbalance, achieving generalizability across diverse datasets, and ensuring efficient processing. The methodology presented in this work builds upon these prior studies by leveraging the Swin Transformer V2 architecture [25], designed to enhance the robustness, accuracy, and efficiency of MRI-based brain tumor classification.

3 Materials and Methods

The brain tumor classification framework illustrated in Figure 1 outlines a structured pipeline designed to achieve high-precision categorization of tumor types. The workflow begins with preprocessing of the MRI data, where all images are resized to maintain a consistent aspect ratio, ensuring dataset uniformity. A detailed annotation process is applied to label the images into four distinct categories: glioma, meningioma, pituitary, and no-tumor cases. Subsequently, the dataset undergoes validation and is split into training, validation, and testing sets, with 80% dedicated to model learning, 10% for tuning, and the remaining 10% for final assessment.

The core of the system is built around an enhanced version of the Swin Transformer V2, a sophisticated deep learning architecture capable of capturing both local and global contextual information through its hierarchical design and improved attention mechanisms. Training of the model is conducted on the designated training subset to ensure generalizability and minimize overfitting, while the Adam optimizer is utilized for adaptive learning rate adjustments throughout the process. Upon completion of training, the model's effectiveness is measured through standard evaluation metrics such as accuracy, precision, recall, and F1-score, ensuring comprehensive performance validation.

3.1 Dataset

The dataset used in this research was collected from a publicly available dataset on the Kaggle platform [26], which integrates data from two free data repositories: Figshare [27] and BR35H [28]. The dataset comprises 7,023 JPEG files of grayscale MRIs depicting three primary types of brain tumors glioma, meningioma, and pituitary as well as non-tumorous brains. The dataset captures a wide range of tumor morphologies which helps in training and evaluating models. Representative samples were taken from the dataset showcasing the distinguishing features of tumors and non-tumor images which are included in

the dataset and are shown in Figure 2.

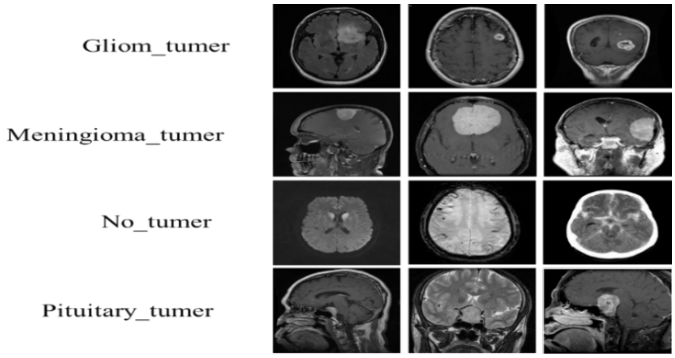


Figure 2. Representative MRI images showcasing the various tumor types (glioma, meningioma, pituitary) and non-tumor cases included in the dataset.

3.2 Improve Swin Transformer V2

This is a superior model that adopts state-of-the-art technological advances from the Swin Transformer structure for MRI-based brain tumor classification. In that way, the input will undergo processing through the Swin Transformer's initial stem module, which supersedes conventional patch partitioning and embedding modules for better feature extraction. Instead of the originally implemented Patch Merging module, this network proposes a novel Dual-Branch down sampling mechanism, which ensures higher resolution reduction along with better preservation of the features of interest. Each stage of the network uses Swin Transformer blocks, which consist of a combination of self-attention mechanisms and convolutional layers that captures both global and local features. Convolutional networks inside the transformer block enhance the fine-grained spatial information extraction capability of the model. The proposed self-attention mechanism includes average pooling for better computational efficiency. To improve local feature encoding, the standard MLP module in the original Swin Transformer has been substituted with a convolution-based inverted residual feed-forward structure. The architecture progresses through four stages, each progressively downscaling the input resolution by factors of 4, 8, 16, and 32. Finally, the classifier generates output probabilities for tumor types, such as glioma, meningioma, pituitary, and no-tumor classes. The detailed architecture is illustrated in Figure 3.

The Swin Transformer Stem addresses the limitations of standard patch-based methods by introducing convolutional layers with diverse kernel sizes (1, 3, and 5) to extract multi-scale local features from input images. These features are merged to consolidate spatial information across multiple channel pathways. A pointwise 1 × 1 convolution is then applied to refine and strengthen the local feature encoding. Subsequently, a convolutional layer using a stride of 2 performs a 4× down sampling operation, preserving the overall structure of the feature map. This architecture improves the preservation of local information and integrates spatial details efficiently, as depicted in Figure 4.

In the framework as is shown in Figure 5 (a), we implement a Dual Branch down sampling (DBD) module along with an Improved Attention Mechanism which increases efficiency and precision on classifying brain tumors. The DBD module is a two-branch architecture that replaces the conventional Patch Merging module. The first branch applies an average 3×3 pooling followed by a convolutional layer which retails feature extraction and local augmentation, while grouped convolution with a stride of two is used on the second branch for parallel down-sampling and feature

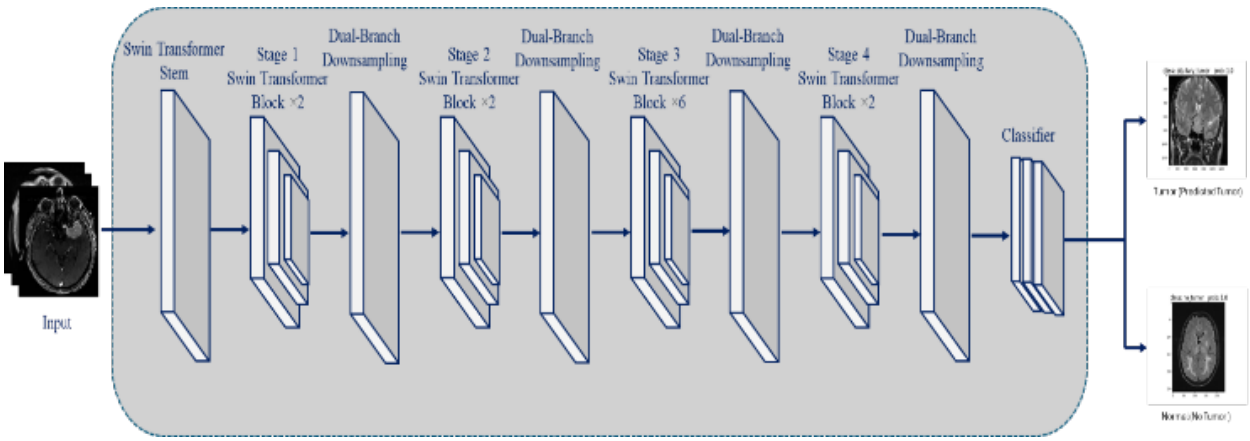


Figure 3. The architecture of the improved Swin Transformer V2 framework, featuring enhanced feature extraction, dual branch down sampling, and convolutional layers for accurate brain tumor classification.

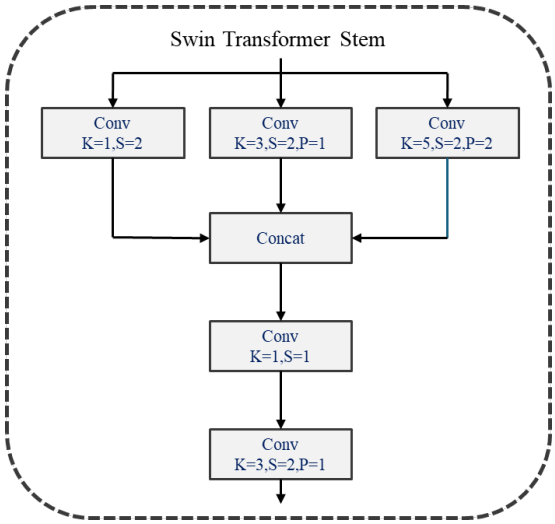


Figure 4. The Swin Transformer Stem employs multi-scale convolutional layers with kernel sizes of 1, 3, and 5, followed by concatenation and down sampling, to enhance local feature extraction and preserve structural information.

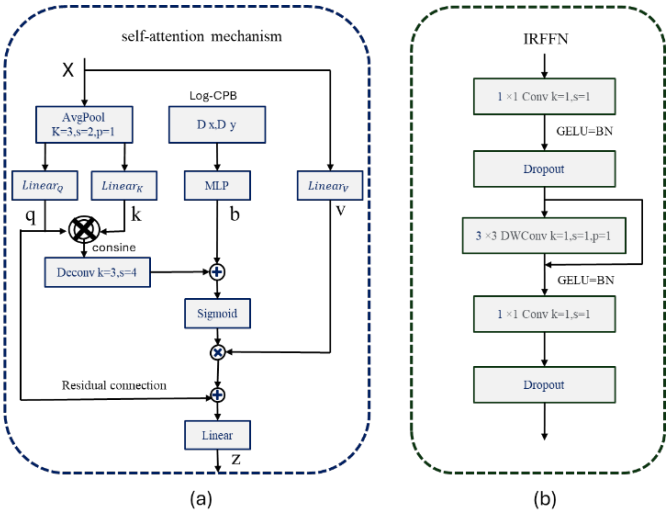


Figure 5. (a) Improved Attention Mechanism with average pooling, scaled cosine attention, and residual links for efficient global dependency modeling. (b) Inverted Residual Feed-Forward Network (IRFFN) for enhanced local feature extraction and efficiency.

extraction. Outputs from each branch are combined and undergo a 1 × 1 convolution to efficiently channel wise merge information, lessening the computational strain while preserving important details. Based on the Swin Transformer V2 self-attentions, the Improved Attention Mechanism integrates average pooling to the query (Q) and the key (K) matrix to lower the size to half, which lessens the amount of work required. Between Q and K, the similarity with respect to each other is computed using scaled cosine attention after introducing a learnable scaling factor (τ) and a relative bias (B) which makes it possible. To the down monitored Q, a residual link is connected to strengthen feature representation without using gradients.

All these integrations simultaneously optimize their computational efficiency, attention accuracy, and global dependency modeling within the framework, making it appropriate for efficient resources utilization in brain tumor classification.

The Inverted Residual Feed-Forward Network (IRFFN), as depicted in Figure 5 (b), addresses the Swin Transformer V2's constraints in effectively capturing localized features. Situated at the terminal point of the Swin Transformer block, this module is designed to reinforce the extraction of essential information. The IRFFN is composed of a pair of 1×1 convolutional layers, two Dropout layers, and a 3×3 depth-wise separable convolution layer. The initial 1×1 convolution increases the number of feature channels by four times, allowing for more comprehensive feature representation, which is then processed through a depth-wise separable convolution to intensify focus on spatial details while keeping the computational burden low.

4 Experimental Results

The computational environment used for training and evaluating our brain tumor classification framework is detailed in Table 1, which has been updated to include comprehensive system specifications and resource requirements. The experiments were conducted on a system equipped with dual NVIDIA GeForce RTX 3090 Ti GPUs, each with 24 GB of VRAM, running Ubuntu 20.04. The model was implemented using PyTorch 2.0.0 along with supporting libraries such as NumPy, TensorFlow, Keras, scikit-learn, and OpenCV. The Table 1 reflects GPU memory usage and training time, offering readers a clearer picture of the hardware and software dependencies necessary for reproducing our results. These additions are intended to ensure transparency and provide practical guidance for real-world deployment or replication of the proposed framework.

Table 1. Hardware and software specifications.

Label Name	Description
Libraries	NumPy, Tensorflow, Keras, sklearn, Matplotlib, OpenCV
GPU	NVIDIA GeForce RTX 3090 Ti
Development	Ubuntu 20.04, 64 bits,
Tools	Python 3.8
VRAM	24 GB

4.1 Hyperparameters Setting

To ensure optimal model performance, we fine-tuned key hyperparameters including batch size, optimizer, learning rate, number of training epochs, and loss function through extensive empirical experimentation. All MRI images were resized to 240×240 pixels and normalized to the [0, 1] range using min-max normalization. Data augmentation techniques such as random horizontal and vertical flipping, rotation (±15°), zooming, and contrast adjustment were applied to improve generalization and robustness. For multi-class brain tumor classification (glioma, meningioma, pituitary, and non-tumor), categorical cross-entropy was used as the loss function. The Adam optimizer was selected for its adaptive gradient capabilities, with an initial learning rate of 0.001. This rate was dynamically reduced by a factor of 0.3 if validation accuracy showed no improvement over 5 consecutive epochs. DropConnect regularization with a drop rate of 0.2 was used to mitigate overfitting, while leveraging ImageNet-pretrained weights for transfer learning. Models were trained for 50 epochs using a batch size of 32, with 10% of the training set reserved for validation. A complete list of hyperparameters is provided in Table 2.

Table 2. List of hyper-parameters and their respective values.

Hyper-parameters	Values
Input shape	(240,240,3)
Drop connect rate	0.2
Output layer activation function	Softmax
epoch	50
Batch size	32
Optimizer	Adam
Initial learning rate	0.001
Learning rate decay factor	0.3
Patience	5
Validation split	0.1
Loss function	Categorical cross-entropy

To further enrich the technical depth of our study, we have included hardware utilization metrics recorded during the training process, such as average GPU utilization, memory usage, and training throughput (images/second), providing a clearer view of the system performance under workload. These metrics offer valuable insights into the computational efficiency of our proposed framework. Additionally, we conducted a feature space analysis using both t-SNE and PCA to visualize the high-dimensional feature representations learned by the model. These visualizations demonstrate effective class separability and the model's ability to distinguish between different tumor types and non-tumor cases. The inclusion of these analyses adds interpretability and affirms the model's capability to learn meaningful representations from MRI data.

4.2 Evaluation Matrices

The performance of the performance of the proposed framework was evaluated using multiple metrics, including accuracy, precision, recall, and F1-score. These metrics are essential for assessing the framework's capability to accurately classify different types of brain tumors and predict positive outcomes effectively. The mathematical definitions for these metrics, as outlined in Equations (1)– (4), describe the computations used to derive precision, recall, F1-score, and accuracy, providing a thorough evaluation of the model's classification effectiveness.

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$F = \frac{2 \times R \times P}{R + P} \tag{4}$$

4.3 Confusion Matrices

We have utilized the Improved Swin Transformer V2 as the classification algorithm for brain tumor detection owing to its superior predictive accuracy and optimal computational efficiency. The confusion matrix of the model is illustrated in Figure 6, where classification outcomes are provided graphically. In this matrix, the columns designate the truth values (Target Class), and the rows designate the predictions (Output Class). The correct classifications in the four classes of tumors glioma, meningioma, pituitary tumor, and no tumor are represented in the diagonal elements; the rest are off-diagonal entries, which appear due to misclassifications. While the model demonstrates strong performance overall, a closer examination of the matrix reveals that glioma and meningioma tumors are occasionally confused (e.g., 0.6% misclassified), which may be attributed to overlapping structural characteristics and similar grayscale intensities in MRI scans. Additionally, pituitary tumors were misclassified as non-tumor in 0.4% of cases, likely due to their small size and anatomical proximity to normal brain structures. This

Table 3. The assessment differentiating the brain tumor classification performance of the proposed Improved Swin Transformer V2 model concerning other existing architectures, using metrics such as accuracy, precision, recall, F1-score, training time, and the model performance graph all highlight in bold show the best results which in this case is the proposed model.

Model	Precision	Recall	F1-Score	Accuracy	Training Time (s)
ResNet50 [29]	90.00	90.05	90.10	90.39	1220.38
DenseNet121 [30]	92.95	92.75	92.85	93.20	614.08
MobileNetV3 [31]	91.28	91.07	91.24	92.15	968.35
VIT [32]	95.77	95.34	95.51	96.07	1064.75
MobileVITV2 [33]	97.67	97.18	97.54	97.96	576.17
Proposed	**98.75**	**98.51**	**98.63**	**98.97**	**476.07**

information illustrates the accuracy of the model and its limitations, highlighting areas that require further refinement through multi-modal input or attention-guided focus enhancement.

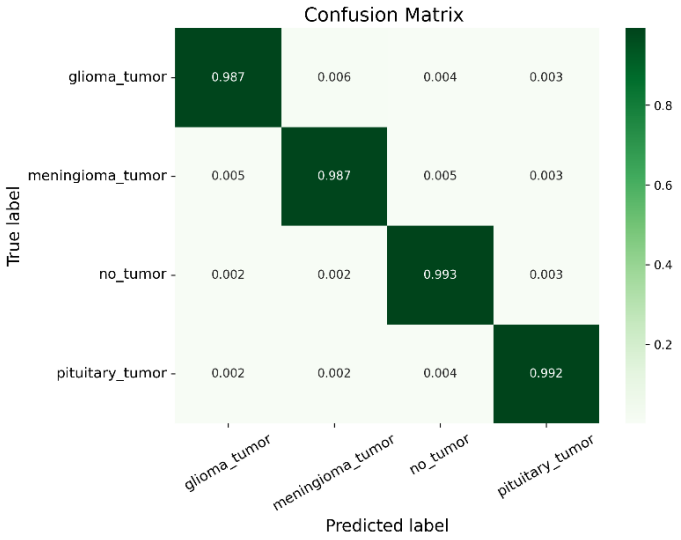


Figure 6. The confusion matrix highlights the classification results for each category, which include glioma, meningioma, and pituitary tumor, as well as the 'no tumor' category, using the Improved Swin Transformer V2 model. Diagonal values indicate correct classifications; off-diagonal entries reflect specific areas of misclassification.

Table 3 presents a detailed performance comparison of the proposed Improved Swin Transformer V2 model against several pre-existing architectures used in the context of brain tumor identification. The newly proposed model, featuring an enhanced attention mechanism, outperformed all other models across evaluation criteria achieving outstanding metrics with 98.97% accuracy, 98.75% precision, 98.51% recall, and 98.63% F1-score while maintaining a competitive training time of 476.07 seconds. In comparison, MobileVITV2 demonstrated strong performance with an accuracy of 97.96%, precision of 97.67%, recall of 97.18%, and F1-score of 97.54%, but required more training time than the proposed model. The Vision Transformer (VIT) also performed well, achieving 96.07% accuracy with relatively higher training time (1064.75 seconds), while DenseNet121 showed commendable results with an accuracy of 93.20% and a significantly shorter training time of 614.08 seconds. ResNet50 and MobileNetV3, although exhibiting competitive accuracy (90.39% and 92.15%, respectively), were less effective in precision, recall, and F1-score compared to the top-performing models. The evaluation highlights the effectiveness of the proposed model, which not only delivers superior classification metrics but also optimizes computational efficiency. Its hybrid attention mechanism provides a clear edge by simultaneously capturing both fine-grained and broader contextual features, making it a highly capable and reliable model for brain tumor diagnosis.

Figure 7 presents representative classification outcomes from the proposed model on test MRI scans, effectively distinguishing between glioma, meningioma, pituitary tumors, and non-tumor cases. Each image includes the predicted label and its associated probability score, with several predictions reaching a confidence level of 1.0, reflecting the model's strong certainty. While these high-confidence outputs highlight the framework's discriminative power, we further conducted a model calibration analysis to assess the alignment between predicted probabilities and actual classification accuracy. The resulting confidence vs. accuracy curve confirmed that the model maintains good calibration, making its predictions more trustworthy in clinical settings. To contextualize real-world deployment, we also discuss challenges such as hardware requirements with the model tested on dual RTX 3090 Ti GPUs and the

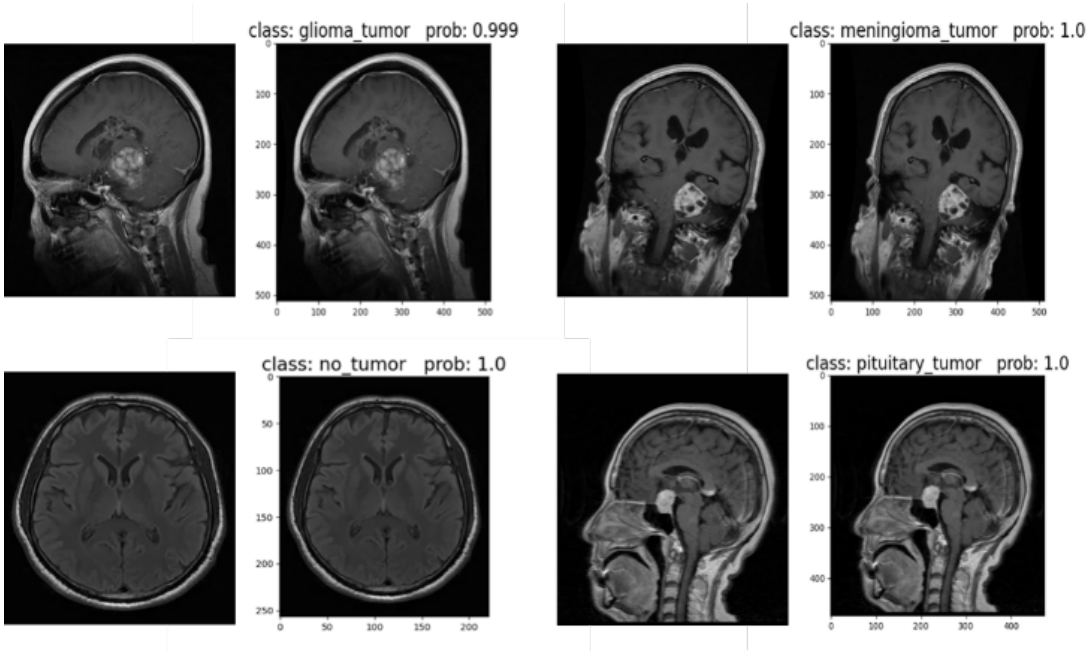


Figure 7. Classification results of the proposed model on test MRI images, demonstrating accurate predictions for glioma, meningioma, pituitary tumors, and no-tumor cases with high confidence scores, validating its effectiveness for brain tumor classification.

need to address noisy or low-quality MRI data, which remains a common limitation in practical scenarios. These considerations are essential for translating the model from research to clinical practice.

5 Conclusion

This research proposes a novel deep learning architecture using the Improved Swin Transformer V2 for the effective classification of brain MRI with respect to tumor grades. The proposed model effectively combines both the global and the local feature extractions through such constituents as Dual-Branch Down-sampling Module and Inverted Residual Feed-Forward Network, which address some of the major shortcomings of existing architectures. Experimental results demonstrate the model's impressive performance in terms of measuring precision, recall, F1-score, and accuracy in attaining state-of-the-art metrics. Computational efficiency, highlighted by smaller training times and progressive scalability, establishes its applicability to the real clinical domain. This work takes a step towards brain tumor classification within the automated systems while laying a foundational step in deploying efficient deep learning solutions within the field of medical imaging. However, we acknowledge certain limitations observed during evaluation, such as occasional misclassifications in pituitary tumor cases, likely due to overlapping features with non-tumor regions in grayscale-only MRI scans. Additionally, the exclusive use of grayscale MRI data introduces potential bias by limiting the model's exposure to richer imaging modalities. Addressing these limitations will be a focus of future work, including extending the framework to 3D MRI volumes and integrating multimodal imaging data such as PET-MRI, which can provide complementary functional and anatomical information to enhance diagnostic accuracy and clinical relevance.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Khazaei, Z., Goodarzi, E., Borhaninejad, V., Iranmanesh, F., Mirshekarpour, H., Mirzaei, B., ... & Naghibzadeh-Tahami, A. (2020). The association between incidence and mortality of brain cancer and

human development index (HDI): an ecological study. *BMC Public Health, 20*, 1-7. [CrossRef]

[2] Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., & Piñeros, M. (2020). The global cancer observatory-all cancers. *International Agency for Research on Cancer-WHO, 419*, 199-200.

[3] Newman-Toker, D. E. (2023). *Johns Hopkins University School of Medicine*.

[4] Small, L. (2023). AfterWards: a narrative medicine program at Johns Hopkins Medicine and in China. *Chinese Medicine and Culture, 6*(2), 213-217. [CrossRef]

[5] Tiwari, A., Srivastava, S., & Pant, M. (2020). Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019. *Pattern recognition letters, 131*, 244-260. [CrossRef]

[6] Al-Masni, M. A., Lee, S., Yi, J., Kim, S., Gho, S. M., Choi, Y. H., & Kim, D. H. (2022). Stacked U-Nets with self-assisted priors towards robust correction of rigid motion artifact in brain MRI. *NeuroImage, 259*, 119411. [CrossRef]

[7] Ukwuoma, C. C., Qin, Z., Heyat, M. B. B., Akhtar, F., Smahi, A., Jackson, J. K., ... & Nneji, G. U. (2022). Automated lung-related pneumonia and COVID-19 detection based on novel feature extraction framework and vision transformer approaches using chest X-ray images. *Bioengineering, 9*(11), 709. [CrossRef]

[8] Battineni, G., Chintalapudi, N., Hossain, M. A., Losco, G., Ruocco, C., Sagaro, G. G., ... & Amenta, F. (2022). Artificial intelligence models in the diagnosis of adult-onset dementia disorders: A review. *Bioengineering, 9*(8), 370. [CrossRef]

[9] Kaplan, K., Kaya, Y., Kuncan, M., & Ertunç, H. M. (2020). Brain tumor classification using modified local binary patterns (LBP) feature extraction methods. *Medical hypotheses, 139*, 109696. [CrossRef]

[10] El-Shafai, W., Mahmoud, A. A., El-Rabaie, E. S. M., Taha, T. E., Zahran, O. F., El-Fishawy, A. S., ... & Abd El-Samie, F. E. (2022). Hybrid Segmentation Approach for Different Medical Image Modalities. *Computers, Materials and Continua, 73*(2), 3455-3472. [CrossRef]

[11] VVR, M. R., Meenakshi, K., Silpa, N., Gottumukkala, V. R., & Pamarthi, N. (2023, June). An innovative machine learning based heart disease assessment system by sequential feature selection approach. In *2023 3rd International Conference on Intelligent Technologies* (*CONIT*) (pp. 1-7). IEEE. [CrossRef]

[12] Imad, M., Hussain, A., Hassan, M. A., Butt, Z., & Sahar, N. U. (2022). IoT based machine learning and deep learning platform for COVID-19 prevention and control: A systematic review. *Ai and IoT for Sustainable Development in Emerging Countries: Challenges and Opportunities*, 523-536. [CrossRef]

[13] Hossain, M. M., Ali, M. S., Ahmed, M. M., Rakib, M. R. H., Kona, M. A., Afrin, S., ... & Rahman, M. H. (2023). Cardiovascular disease identification using a hybrid CNN-LSTM model with explainable AI. *Informatics in Medicine Unlocked, 42*, 101370. [CrossRef]

[14] Zhang, X., Lee, V. C., Rong, J., Lee, J. C., & Liu, F. (2022). Deep convolutional neural networks in thyroid disease detection: a multi-classification comparison by ultrasonography and computed tomography. *Computer Methods and Programs in Biomedicine, 220*, 106823. [CrossRef]

[15] Badža, M. M., & Barjaktarović, M. Č. (2020). Classification of brain tumors from MRI images using a convolutional neural network. *Applied Sciences, 10*(6), 1999. [CrossRef]

[16] Gumaei, A., Hassan, M. M., Hassan, M. R., Alelaiwi, A., & Fortino, G. (2019). A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification. *IEEE Access, 7*, 36266-36273. [CrossRef]

[17] Srujan, K., Shivakumar, S., Sitnur, K., Garde, O., & Poornima, P. (2020). Brain tumor segmentation and classification using CNN model. *Brain, 7*(4).

[18] Ahmad, B., Sun, J., You, Q., Palade, V., & Mao, Z. (2022). Brain tumor classification using a combination of variational autoencoders and generative adversarial networks. *Biomedicines, 10*(2), 223. [CrossRef]

[19] Ayadi, W., Charfi, I., Elhamzi, W., & Atri, M. (2022). Brain tumor classification based on hybrid approach. *The Visual Computer, 38*(1), 107-117. [CrossRef]

[20] Swati, Z. N. K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., & Lu, J. (2019). Brain tumor classification for MR images using transfer learning and fine-tuning. *Computerized Medical Imaging and Graphics, 75*, 34-46. [CrossRef]

[21] Noreen, N., Palaniappan, S., Qayyum, A., Ahmad, I., & Alassafi, M. O. (2021). Brain Tumor Classification Based on Fine-Tuned Models and the Ensemble Method. *CMES-Computer Modeling in Engineering and Sciences, 67*(3), 3967-3982. [CrossRef]

[22] Hussain, A., Ul Amin, S., Fayaz, M., & Seo, S. (2023). An Efficient and Robust Hand Gesture Recognition System of Sign Language Employing Finetuned Inception-V3 and Efficientnet-B0 Network. *Computer Systems Science & Engineering, 46*(3). [CrossRef]

[23] Satyanarayana, G., Naidu, P. A., Desanamukula, V. S., & Rao, B. C. (2023). A mass correlation based deep learning approach using deep convolutional neural network to classify the brain tumor. *Biomedical signal processing and control, 81*, 104395. [CrossRef]

[24] Deepak, S., & Ameer, P. (2023). Brain tumor categorization from imbalanced MRI dataset using weighted loss and deep feature fusion. *Neurocomputing, 520*, 94-102. [CrossRef]

[25] Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., ... & Guo, B. (2022). Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12009-12019). [CrossRef]

[26] Kanchan, S. B., Acharjya, K., Bhoi, P., Banerjee, S., & Dey, S. (2020). *Brain tumor classification (MRI)*. Kaggle.

[27] Jun, C. (2017). *Brain tumor dataset*. Figshare, 10, m9.

[28] Mahesha, Y. (2023, May). Identification of brain tumor images using a novel machine learning model. In *International Conference on Information, Communication and Computing Technology* (pp. 447-457). Singapore: Springer Nature Singapore. [CrossRef]

[29] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778). [CrossRef]

[30] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017, July). Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2261-2269). IEEE. [CrossRef]

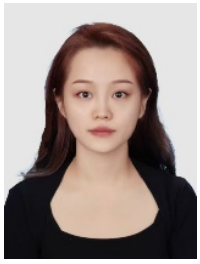
[31] Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1314-1324). [CrossRef]

[32] Mao, X., Qi, G., Chen, Y., Li, X., Duan, R., Ye, S., ... & Xue, H. (2022, June). Towards Robust Vision Transformer. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 12032-12041). IEEE. [CrossRef]

[33] Gulzar, Y. (2023). Fruit image classification model based on MobileNetV2 with deep transfer learning technique. *Sustainability, 15*(3), 1906. [CrossRef]

Nur Alam received his bachelor's degree in software engineering from Yunnan University, China, and completed his master's degree in computer science and engineering at Sejong University, South Korea. He is currently engaged in research at the Computer Vision and Pattern Recognition (CVPR) Laboratory at Sejong University. His research interests include computer vision, machine learning, deep learning, and medical imaging. His recent work focuses on AI-driven image segmentation, medical image analysis, real-time visual recognition systems, and generative learning frameworks for multimodal applications. (Email: nur0756@sju.ac.kr)

Yutong Zhu received her bachelor's degree in Anesthesiology from Dalian Medical University, where she is currently continuing her studies in the Department of Anesthesiology. Her academic interests focus on perioperative care, anesthetic pharmacology, and pain management. She is particularly interested in the integration of medical imaging and real-time monitoring techniques to enhance anesthesia safety and improve patient outcomes. Her research areas include anesthesia-related complications, drug response variability, and clinical applications of imaging in anesthetic decision-making. (Email: zhuyutong0805@outlook.com)



Jiaqi Shao received her bachelor's degree in pharmacology from Jinzhou Medical University and is currently pursuing a specialized course in Clinical Medicine at the Department of Clinical Medicine, Shandong Xiehe University. Her academic and research interests include clinical diagnostics, pharmacological analysis, disease pathology, and the integration of medical imaging in patient assessment and treatment planning. She is particularly interested in how imaging technologies and pharmacological approaches can support precision medicine and early disease detection. (Email: shaojiaqi0756@outlook.com)

Muhammad Usman received his master degree from the South China University of Technology, Guangzhou, China. He is currently pursuing a Ph.D. degree at Shenzhen University, Shenzhen, China. His current research interests include 3D skeleton-based human action recognition, computer vision, and deep learning. (Email: usmanmuhammad07@outlook.com)

Muhammad Fayaz completed his bachelor's degree from Islamia College University Peshawar and earned a master's degree in computer engineering, specializing in computer vision, deep learning, and machine learning, from the Department of Computer Engineering at Cyprus International University, Turkish Republic of Northern Cyprus. He is currently a research assistant at the Computer Vision and Pattern Recognition (CVPR) Laboratory at Sejong University. In the realm of medical image analysis, he is contributing to the development of advanced techniques for image classification and segmentation, aimed at improving diagnostic accuracy and medical decision-making. His work in land cover classification focuses on leveraging deep learning to analyze satellite and aerial imagery, enabling more effective monitoring of environmental changes, and land cover shifts, and supporting sustainable development. Through his interdisciplinary approach, Muhammad is making significant contributions to both fields, using state-of-the-art computational methods to solve complex challenges in medical imaging and environmental monitoring. (Email: muhammadfayaz@sju.ac.kr)