Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 14, No. 3, 2026, 37-49.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

A Hybrid Approach for Brain Tumor Classification: Enhancing MRI-Based Diagnosis with CNN-Transformer Synergy

Samira Mavaddati*

Electronic Department, Faculty of Engineering and Technology, University of Mazandaran, Babolsar, Iran.

Article Info

Article History:
Received 22 July 2025
Revised 26 October 2025
Accepted 14 November 2025

DOI:10.22044/jadm.2025.16554.2779

Keywords:
Brain Tumor Classification, Convolutional Neural Networks, Vision Transformer, MRI-based Diagnosis, Transfer Learning.

**Corresponding author: s.mavaddati@umz.ac.ir (S. Mavaddati).*

Abstract

Brain tumors are among the most life-threatening neurological conditions, requiring precise and early diagnosis for effective treatment planning. Traditional deep learning models, such as Convolutional Neural Networks (CNNs) and ResNet-based architectures, have demonstrated promising results in brain tumor classification. However, these models often struggle to capture long-range dependencies within MRI images, which are crucial for accurate classification. To overcome this limitation, we propose a Hybrid CNN-ViT model, combining the strengths of Vision Transformers (ViT) and CNNs to achieve high-precision brain tumor classification. The CNN component effectively extracts local spatial features, while the ViT module captures global contextual relationships within MRI scans. The model is evaluated on a four-class dataset of Glioma, Meningioma, Pituitary tumors, and non-tumor images, achieving an impressive accuracy of 98.37%, surpassing conventional CNN-based methods. By leveraging transfer learning, the approach enhances classification performance while reducing reliance on large-scale labeled datasets. The proposed Hybrid CNN-ViT model offers a scalable, robust, and efficient solution for real-world neuro-oncological diagnostics, significantly improving the accuracy of MRI-based brain tumor detection.

1. Introduction

The rising global prevalence of brain tumors highlights the urgent need for early and accurate diagnosis to improve treatment outcomes [1-2]. Traditional diagnostic methods, heavily dependent on radiologists' expertise, are often time-consuming and subjective. Advances in medical imaging and AI, especially deep learning, have transformed brain tumor detection using MRI and CT scans [3-4]. While CNNs excel at extracting spatial features, they struggle to capture long-range dependencies in MRI images. ViTs, with their self-attention mechanisms, effectively model these global relationships but require large labeled datasets and high computational power. To overcome these challenges, we propose a Hybrid CNN-ViT model that merges CNN's local feature extraction with ViT's global attention, enhancing accuracy, robustness, and generalizability in classifying Glioma, Meningioma, Pituitary tumors, and non-tumor cases. Leveraging transfer learning further reduces reliance on extensive labeled data, making this approach promising for scalable neuro-oncological diagnostics.

2. Literature review

Accurate brain tumor classification remains a critical challenge in medical imaging, driving extensive research in machine learning and deep learning. Early methods relied on handcrafted features combined with classical classifiers like SVM, Random Forest, and k-NN, but these often lacked generalizability for complex tumor patterns.

With the advent of deep learning, CNNs such as VGG16, InceptionV3, and ResNet became prevalent due to their ability to automatically learn hierarchical spatial features from MRI scans. However, CNNs face limitations in capturing long-range dependencies and global contextual relationships essential for complex medical images. Transformer-based models, especially ViTs, address this by employing self-attention mechanisms to model global dependencies effectively. While ViTs have shown superior performance in some medical imaging tasks, their reliance on large datasets and high computational demands is a challenge. To capitalize on the complementary strengths of CNNs and ViTs, hybrid models combining CNN feature extraction with ViT attention have emerged, offering enhanced classification accuracy and robustness. The following sections review key deep learning approaches to brain tumor classification, motivating the development of our Hybrid CNN-ViT framework.

Study [5] proposes an innovative brain tumor detection method by integrating MRI and CT data, using deep learning to extract and combine distinct imaging features for differentiating tumor-affected and healthy tissue. The model employed is VGG16, which has shown promising classification accuracy. A comprehensive review in [6] analyzes various deep learning approaches for Glioma detection from MRI scans, highlighting the importance of proper preprocessing, model selection, and evaluation to achieve accurate tumor identification. Additionally, [7] presents a two-stage deep learning method: a twelve-layer CNN for feature extraction followed by a Softmax classifier categorizing images into Glioma, Meningioma, and Pituitary tumor types. These studies collectively demonstrate significant advancements in deep learning for brain tumor diagnosis.

The study in [8] proposes a brain tumor classification framework that combines seven pre-trained CNN models with malignancy risk index (MRI) images, using transfer learning by fine-tuning CNNs pre-trained on ImageNet to boost performance. Similarly, [9] presents a multi-stage automated classification method involving preprocessing (discrete cosine transform, histogram equalization), deep feature extraction via VGG16 and VGG19 with transfer learning, and classification using an extreme learning machine (ELM) within a hybrid framework. To address limited dataset size and computational demands, [10] investigates data augmentation and transfer learning, analyzing effects of network width, depth, and resolution scaling. A comprehensive survey in [11] reviews brain tumor classification and segmentation techniques, including machine learning, CNNs, capsule networks, and Vision Transformers, highlighting their pros and cons. In [12], a novel fusion model extracts low- and high-level features from AlexNet, GoogLeNet, and ResNet185, then classifies them using SVM and k-NN, achieving improved accuracy. Lastly, [13] demonstrates that ensemble learning combining multiple deep CNN models enhances tumor classification compared to single models.

Similarly, [14] proposes an automatic classifier for Glioma, Meningioma, and Pituitary tumors using transfer learning with a pre-trained GoogLeNet model. Beyond brain tumors, [15] presents an efficient VGG16-based classifier for Alzheimer's disease diagnosis from MRI scans. Study [16] explores transfer learning in content-based image retrieval (CBIR) for brain tumor classification. In [17], a hybrid classifier combining feature-based and segmented-image methods employs deep neural networks (CDNN) and CNNs for tumor detection. Meanwhile, [18] introduces a robust system trained on multiple MRI sequences (T1, T2, T2-contrast) to classify 18 brain tumor types. The CNN model in [19] classifies MRIs into Glioma, Meningioma, and Pituitary tumors and investigates automated tumor grading on various datasets. Finally, [20] examines a VGG16-based CNN leveraging transfer learning for effective brain tumor detection. Lastly, [21] proposes a three-tier deep learning framework using InceptionV3 as a feature extractor on X-ray brain images, achieving superior accuracy on the CE-X-ray dataset for Glioma, Meningioma, and Pituitary tumor detection. In [22], transfer learning models including ResNet152, VGG19, DenseNet169, and MobileNetv3 are evaluated on a Kaggle dataset with augmentation; MobileNetv3 achieves the highest accuracy of 99.75%, demonstrating transfer learning's effectiveness in medical imaging. Similarly, [23] uses 3,264 MRI images to classify brain tumors and healthy tissue with a 2D CNN and convolutional autoencoder, reaching accuracies above 95% and AUC > 0.99, outperforming traditional methods like K-NN and MLP. In [24], a CNN architecture trained on the same dataset outperforms ResNet-50, VGG16, and InceptionV3, achieving 93.3% accuracy and 98.43% AUC, highlighting its clinical potential. Recent multimodal approaches integrating MRI, CT, and clinical data have further improved diagnostic accuracy and robustness [25-26], suggesting a promising future direction beyond this MRI-focused study. ViT have also shown strong

performance in medical image analysis by capturing long-range dependencies [27-28], sometimes outperforming CNNs when trained on large datasets. However, this paper focuses on traditional deep learning models like CNN, ResNet, and VGG, without exploring transformer-based methods due to the study's scope.

This paper presents a Hybrid CNN-ViT model for brain tumor classification from MRI scans. By combining CNN's local feature extraction with ViT's global attention, the model captures both fine details and long-range dependencies, improving accuracy and efficiency. It classifies Meningioma, Glioma, Pituitary tumors, and non-tumor cases, offering enhanced performance compared to conventional CNN-only networks.

- High classification accuracy: Combining CNN's spatial feature extraction with ViT's attention mechanisms enhances performance.
- Improved interpretability: ViTs enable better visualization of critical MRI regions, increasing model transparency beyond traditional CNNs.
- Optimized computational efficiency: Using CNNs for initial feature extraction reduces the computational load compared to standalone ViTs.
- Robustness to limited data: Transfer learning allows effective adaptation to small medical imaging datasets.
- Enhanced generalization: The hybrid architecture generalizes well across diverse datasets, making it scalable for real-world applications.

Major contributions and findings are:

- Accurate brain tumor classification: The Hybrid CNN-ViT effectively distinguishes tumor types and non-tumor cases.
- Comprehensive model comparison: Performance is benchmarked against CNN, VGG16, InceptionV3, RNN, and dictionary-based methods.
- Superior diagnostic accuracy: The model achieves over 98.5% classification accuracy, surpassing conventional approaches.
- Scalability and adaptability: The framework can be extended to other medical imaging tasks, highlighting its broader clinical potential.

These findings underscore the transformative impact of the Hybrid CNN-ViT model in delivering a more accurate, efficient, and scalable solution for automated brain tumor diagnosis.

3. Brain Tumor Detection Using Deep Learning Models

Brain tumor classification is crucial for early diagnosis. This study proposes a Hybrid CNN-ViT model that combines CNN's local feature extraction with ViT's global attention. Trained on a four-class dataset with transfer learning and data augmentation, it achieves over 98.5% accuracy, outperforming conventional CNNs and demonstrating the effectiveness of hybrid models for automated tumor detection.

3.1. Dataset and Preprocessing

To evaluate the proposed Hybrid CNN-ViT model, we used the CE-MRI brain tumor dataset, which contains 3,064 contrast-enhanced T1-weighted MRI images from 233 patients, categorized into four classes: Meningioma, Glioma, Pituitary tumor, and non-tumor. The scans include coronal, sagittal, and axial views with an original resolution of 512×512 pixels and a pixel size of 49×49 mm. This publicly available dataset provides tumor masks, boundaries, and class labels for each image. Before training, preprocessing involved min-max normalization to standardize pixel intensities, resizing images to 256×256 pixels for computational efficiency, and replicating grayscale intensities across three channels to ensure compatibility with pre-trained deep learning models. Figure 1 shows sample MRI images from the dataset.

3.2. Data Augmentation

Deep learning models require large, diverse datasets for optimal performance. However, the dataset used in this study has a limited number of MRI scans, especially for certain tumor classes, leading to data imbalance that may cause overfitting and reduce generalization. To mitigate this, data augmentation techniques are applied to artificially expand the dataset and balance tumor category representation. Due to the rarity of some tumor types and patient privacy concerns, acquiring extensive medical imaging data is challenging, making augmentation a vital strategy to improve model robustness. This study uses an adversarial-based augmentation to generate synthetic MRI variations through mirroring, scaling, shifting, cropping, 45° rotation, and salt-and-pepper noise.

These augmentations increase data diversity, prevent spurious correlations, and enhance generalization. Figure 2 illustrates examples of the augmented MRI tumor images, demonstrating their contribution to improved classification performance [39-40].

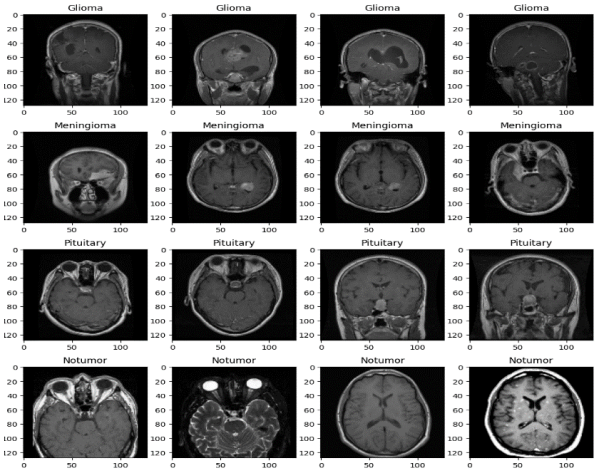

Figure 1. MRI images representing common brain tumor types: Top row: Glioma, originating from glial cells; middle row: Meningioma, arising from the meninges; bottom row: Pituitary tumor, located in the pituitary gland; and fourth row: healthy brain MRI as a reference. These images illustrate tumor diversity and highlight the need for accurate diagnostic tools to guide effective treatment.

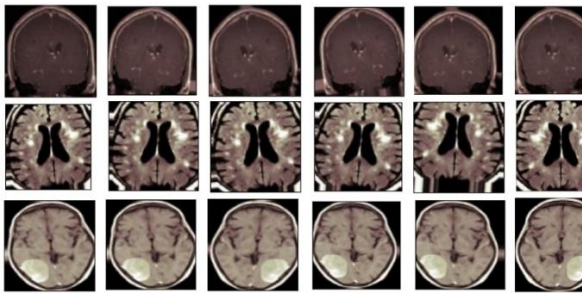

Figure 2. Representation of brain tumor images (Right column) and their augmented forms using data augmentation methods in other columns.

Table 1. Distribution of the CE-MRI brain tumor dataset into training, validation, and test subsets before and after augmentation for each class.

Class	Total Samples	Training (Before Aug.)	Training (After Aug.)	Validation	Test
Meningioma	708	496	1984	106	106
Glioma	1426	998	3992	214	214
Pituitary	930	651	2604	140	139
Non-tumor	432	303	1212	64	65
Total	**3064+432**	**2145+303**	**8580+1212**	**460+64**	**459+65**

The CE-MRI dataset used in this study originally contained 3,064 contrast-enhanced T1-weighted MRI images from 233 patients, categorized into three classes: Meningioma, Glioma, and Pituitary tumor. However, to enable a more comprehensive four-class classification task, additional non-tumor MRI images (432 samples) were collected from publicly available medical imaging repositories on the internet. These images were carefully selected to match the visual and resolution characteristics of the CE-MRI dataset. The combined dataset was divided into 70% training, 15% validation, and 15% test subsets for each class. To enhance the diversity of the training data and improve the model's generalization capability, standard augmentation techniques, such as rotation, flipping, zooming, and translation, were applied to the training subset, increasing its size fourfold. The detailed distribution of samples before and after augmentation is presented in Table 1, including the number of training, validation, and test images for each class.

4. Deep Model Architecture

This study proposes a Hybrid CNN-ViT model for brain tumor classification, combining CNNs' ability to extract local spatial features with ViTs' strength in capturing global contextual relationships. This integration improves the classification of Glioma, Meningioma, Pituitary tumors, and non-tumor cases. Using transfer learning, the model converges faster and requires fewer labeled images, offering a scalable and alternative to conventional CNN-based methods. The Hybrid CNN-ViT model combines CNNs for local feature extraction with ViTs for global context understanding to improve brain tumor classification. The CNN module captures fine-grained spatial features from MRI images, while ReLU activation and max pooling enhance learning and reduce computation. The ViT module employs self-attention and transformer blocks to model long-range dependencies across image patches. Fully connected layers then classify tumors based on the integrated features. Transfer learning with pre-training on large datasets like ImageNet further enhances generalization, with limited labeled medical data.

4.1. Simulation Results

This paper proposes an automated Hybrid CNN-ViT system for brain tumor classification. MRI images are first preprocessed and augmented to enhance quality, robustness, and reduce overfitting. Images are resized to 256×256×3 and fed into the hybrid architecture, where CNN layers extract local spatial features and the ViT module captures global contextual relationships. Fully connected layers then classify images into four categories: Meningioma, Glioma, Pituitary tumor, and Non-tumor. Hyperparameters, including learning rate, batch size, activation functions, and epochs, are carefully tuned to optimize performance. Compared with various CNN architectures, RNNs, and dictionary learning-based classifiers, the CNN-ViT model consistently outperforms alternatives by effectively combining local and global feature

representations. Table 2 details the model architecture and training settings. Overall, the proposed system demonstrates high accuracy, robustness, and reliability, offering a promising solution for automated brain tumor diagnosis.

The CNN component of the proposed hybrid CNN-ViT model is designed to perform local feature extraction through a hierarchical deep architecture. It consists of five convolutional blocks (Conv1 to Conv5_x) with progressively increasing depth and feature complexity. The first convolutional layer (Conv1) employs 64 filters with a kernel size of 7×7 and a stride of 2 to capture low-level edge and texture features. Each convolutional block is followed by ReLU activation and batch normalization to stabilize training and accelerate convergence. Max-pooling layers with a 3×3 window and a stride of 2 are applied after Conv1 and Conv2_x to downsample the spatial resolution while retaining critical spatial information. Residual connections are incorporated within Conv2_x to Conv5_x, forming bottleneck blocks of the form [1×1, 3×3, 1×1], which efficiently capture hierarchical and abstract representations while mitigating gradient vanishing. A total of 33 convolutional layers are used within these residual blocks, aligning with the structure summarized in Table 2. To prevent overfitting, L2 regularization and dropout (p = 0.3) are applied before the final fully connected layer. The extracted feature maps from the last CNN block are passed to the Vision Transformer encoder, where global contextual dependencies are learned. The final classification is performed through a fully connected layer followed by a Softmax activation function to produce class probabilities for the four tumor types. The output of the final convolutional block (Conv5_x) is a feature map of size 8×8×2048, which is then projected into the ViT encoder. The ViT module consists of a series of multi-head self-attention and feed-forward layers designed to capture long-range dependencies and contextual relationships across the entire image, producing a refined 8×8×768 representation. This feature map is flattened into a single vector (size 49,152) before passing through a fully connected layer with 1000 neurons, applying a Softmax activation to output probability distributions across the four tumor classes. The sequential integration of CNN and ViT enables the network to leverage both local details and global context, improving the accuracy of brain tumor classification. Figure 3 illustrates the block diagram of the hybrid architecture, showing the stepwise flow from CNN feature extraction to ViT-based global modeling and final classification.

4.2. Simulation Details

The Adam optimizer, which estimates adaptive learning rates for each parameter, is selected for the proposed CNN-ViT model. This optimizer is a key component of stochastic optimization methods and is widely used in deep learning architectures. It is employed within the stochastic gradient descent (SGD) framework to update the parameters of the CNN and ViT layers during training. The model starts with an initial learning rate of 0.001, with a multiplier of 0.1 applied at epochs 20 and 30 to adjust the learning rate based on training progress. Moreover, the learning rate is halved every 10 epochs to ensure the model converges effectively and avoids overshooting. For enhanced computational efficiency, the number of iterations is set to 20, and the CNN-ViT model is trained for 100 epochs per iteration. In total, the network undergoes 200 epochs of training to ensure proper convergence. The batch size is empirically chosen to be 64, as increasing the batch size further tends to lead to overfitting and may decrease the model's performance on test data.

Table 2. Architecture of CNN-ViT deep model used in the proposed brain tumor classification algorithm.

Layer Name	Dimension	Details
Input Layer	256×256×3	The input consists of images with a size of 256×256 and 3 color channels (RGB).
Conv1	128×128×64	Convolutional layer with 7×7 filter size, 64 filters, stride of 2. Extracts initial features.
Pool1	64×64	Max-pooling layer with a 3×3 window and stride of 2. Reduces spatial dimensions.
Conv2_x	64×64	Residual blocks: [1×1, 64@3×3, 64@1×1, 256] × 3. Extracts hierarchical features.
Conv3_x	32×32	Residual blocks: [1×1, 128@3×3, 128@1×1, 512] × 4. Further increases abstraction of features.
Conv4_x	16×16	Residual blocks: [1×1, 256@3×3, 256@1×1, 1024] × 23. Captures high-level features.
Conv5_x	8×8	Residual blocks: [1×1, 512@3×3, 512@1×1, 2048] × 3. Final convolutional feature extraction.
ViT Encoder	8×8×768	ViT block to capture global relationships and context within the image.
ViT Block	8×8×768	Multi-head self-attention and feed-forward layers in the ViT module.
Flatten	49152	Flatten the output of the ViT block into a single vector for input into the fully connected layers.
Fully Connected (FC1)	1×1×1000	Final fully connected layer, 1000-dimensional output, corresponding to the number of classes for classification (e.g., tumor types). Softmax activation for probability distribution across classes.
Output	1×1×N	Final classification layer(N: number of classes, Meningioma, Glioma, Pituitary, and No Tumor).

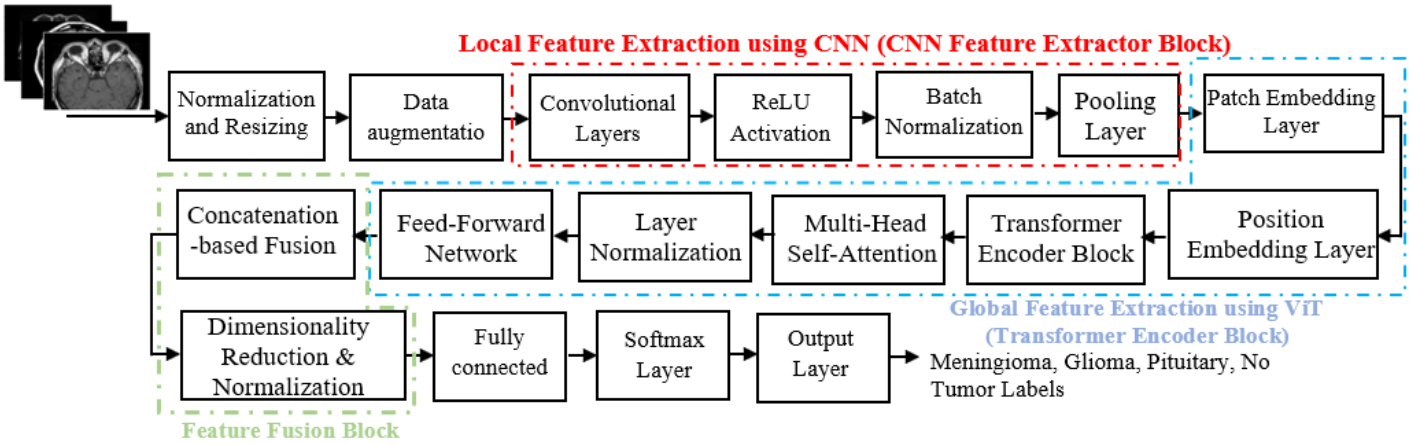

Figure 3. Block diagram of the proposed algorithm for brain tumor classification based on the CNN-ViT deep model.

Table 3. Tuned hyperparameters for the utilized deep models.

Model	Epochs	Optimizer	Batch Size	Kernel Size	Activation Function	Learning Rate	Number of Layers
CNN15, 16	60	Adam	32	3×3	Sigmoid	10^{-3}	25
CNN11	20	Adam	64	3×3	Sigmoid	10^{-2}	32
RNN	20	Adam	64	3×3	Sigmoid	10^{-2}	32
VGG16	-	-	-	3×3	-	-	16
InceptionV3	10	RMSprop	32	3×3	-	10^{-4}	159
Proposed CNN-ViT	100	Adam	32	16×16	GELU	10^{-4}	13

These network hyperparameters are summarized in Table 3. The proposed model is implemented and trained on a personal computer with an Intel Xeon E5 2600 processor, featuring 16 cores running at 3.2 GHz and 192 GB DDR4 RAM on a Linux system with Ubuntu 18.04. The performance of the CNN-ViT model is compared to several existing architectures, including various CNN models from [11, 12-13], RNN, VGG16 [9], InceptionV3 [10], and a dictionary learning-based brain tumor detection method [41]. As mentioned in [12], deep features are classified using a CNN along with a support vector machine. In [13], a combination of deep features based on CNN, AlexNet, ResNet18, and GoogLeNet is proposed for brain tumor identification, transforming multiple low and high-level features into a unified feature vector, thus enhancing the model's performance. In [8], an automatic classification system is introduced using a pre-trained deep transfer learning CNN model for feature extraction from brain MRI images. This proposed classifier utilizes a pre-trained GoogLeNet system to extract features from brain MRI images, aligning with the concept of deep transfer learning. In [9], researchers employed a CNN model using VGG16 for brain tumor detection in MRI images. Furthermore, other researchers presented a three-layer deep learning approach for detecting Glioma, Meningioma, and Pituitary tumors. This model uses InceptionV3 to extract features from X-ray brain images and outperforms previous methods on the CE-X-ray dataset. In [42], a combination of various statistical and textural features, including the Gray-Level Co-occurrence Matrix (GLCM), is used to learn a comprehensive dictionary representation that characterizes the content of each brain tumor category. Additionally, appropriate parameter tuning for coherence is performed in this training approach to achieve effective classification results based on Sparse Non-negative Matrix Factorization (SNMF). This model is trained using the Adam optimizer, with the learning rate, weight decay, and batch size set to 3×10^{-3}, 64, and 64, respectively.

To assess model performance, several metrics are used, including Accuracy (Acc), Specificity (Spe), Positive Prediction Rate (PPR), Sensitivity (Sen), and F-Measure. The confusion matrix shows the classifier's ability to correctly identify each class. Specificity measures correct classification of negative cases, PPR indicates correctly detected positive tumors, and Sensitivity evaluates detection capability across classes. The metrics are defined as follows:

$$Acc = (TP+TN)/(TP+TN+FP+FN) \quad (1)$$

$$Spe = TN/(TN+FP) \quad (2)$$

$$Ppr = TP/(TP+FP) \quad (3)$$

$$Sen = TP/(TP+FN) \quad (4)$$

$$F\text{-}Measure=2\times(Ppr\times Sen)/(Ppr+Sen) \quad (5)$$

Here, TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives, respectively [51]. TP is the count of correctly identified tumor cases, while TN refers to correctly classified non-tumor cases. FP indicates MRI images incorrectly classified as tumors, and FN represents tumor cases mistakenly predicted as non-tumor. Higher values of TP and TN and lower values of FP and FN contribute to better classification accuracy.

5. Simulation results of the proposed deep learning-based method

As mentioned, Figure 3 illustrates the block diagram of the proposed method. Also, the dataset is split using 10-fold cross-validation, where in each iteration, 9 folds are used for training and 1 fold for validation. This process repeats 10 times with different splits, and the final accuracy is the average of all iterations. Table 4 presents classification results for Meningioma, Glioma, and Pituitary tumors using the Hybrid CNN-ViT model, standalone CNN, RNN, and dictionary-based classifiers. The RNN weights are optimized with the scaled conjugate gradient (SCG) algorithm [43]. The superior performance of the Hybrid CNN-ViT arises from its combination of CNNs' local feature extraction and ViT's ability to capture long-range dependencies and global context. This synergy enables the model to utilize both low-level features (edges, textures) and high-level spatial relationships crucial for accurate tumor classification. Using pre-trained CNN weights accelerates training by leveraging prior knowledge, while the Transformer's self-attention mechanism effectively models spatial and contextual information within MRI images. These capabilities, along with adaptability to fine-grained patterns and large datasets, contribute to the model's outstanding accuracy.

Fine-tuning the model with suitable hyperparameters, combined with advanced preprocessing and postprocessing techniques, is essential for achieving high accuracy and robust generalization. Transfer learning using pre-trained weights and training on the high-quality Figshare MRI dataset further accelerates feature learning relevant to brain tumors. Precise tuning and effective image processing algorithms significantly impact overall performance. Figure 4 presents pseudocode outlining the training and evaluation workflow for brain tumor classification using MRI scans (Algorithm 1). The process includes data loading, normalization, augmentation to improve data diversity and balance, model compilation with appropriate loss functions and optimizers, training with validation, and final evaluation on test data. Performance metrics such as accuracy, recall, and loss are calculated to assess model effectiveness. The pseudocode concludes with a comparison of models, identifying CNN-ViT as the top performer with high classification accuracy. This structured approach promotes reproducibility and scalability for future brain tumor detection research.

In the proposed CNN-ViT architecture, the Vision Transformer module is integrated sequentially after the convolutional feature extractor. Specifically, the CNN component first extracts local spatial features from the input MRI images through multiple convolutional and residual layers. The resulting feature map from the final convolutional stage (Conv5_x) is then fed into the ViT encoder, which applies multi-head self-attention mechanisms to capture long-range dependencies and global contextual information. The fused representations are subsequently flattened and passed through fully connected layers for final classification. This sequential integration allows the CNN to focus on fine-grained texture extraction, while the ViT enhances the model's capacity to model global spatial correlations.

The progression plot depicting training accuracy, performance metrics, and losses during training and testing stages for the proposed fine-tuned CNN-ViT model, based on defined parameters and architecture, is presented in Figure 5.

Table 4. Classification accuracy of the proposed brain tumor classifier compared to other mentioned algorithms.

	Meningioma (%)	Glioma (%)	Pituitary (%)	No Tumor (%)	Average Accuracy (%)
CNN [12]	96.67	97.02	97.23	96.98	96.97
CNN [8]	97.56	96.43	96.85	97.11	96.99
CNN [13]	97.39	97.22	96.89	96.43	96.99
RNN	96.23	96.09	96.45	96.13	96.22
VGG16 [9]	97.67	97.73	96.91	97.55	97.46
InceptionV3 [10]	96.98	96.63	97.06	96.76	96.86
SNMF-based [49]	96.12	95.64	96.18	95.88	95.70
CNN-ViT without transfer learning	96.03	95.41	95.63	94.14	95.30
Proposed CNN-ViT	98.62	97.98	98.54	98.43	98.37

Algorithm 1: Pseudocode for Training and Evaluating CNN-ViT for Brain Tumor Detection

Input: MRI images of brain tumors (Glioma, Meningioma, Pituitary tumors, and non-tumor brain images)
Output: Classification Accuracy, Recall, and Loss for each model

1. **Load and preprocess the MRI dataset:**
   - Load 3064 MRI images.
   - Resize images to the input size required by the Hybrid CNN-ViT model (256×256).
   - Normalize pixel values to [0, 1].
   - Apply data augmentation techniques (Rotation, Flipping, Scaling) to enhance the dataset.
2. **Split the dataset:**
   - Divide the dataset into: Training set: 70% of the data, Validation set: 15% of the data, Testing set: 15% of the data
   - Apply 10-fold cross-validation **within the training set** to train and validate the model
   - Use the validation set to fine-tune hyperparameters
   - Evaluate the final model performance on the independent test set
3. **Extract local features using CNN (CNN Feature Extractor Block):**
   - Use a pre-trained CNN model (e.g., ResNet50 or VGG16, trained on ImageNet).
   - Remove the fully connected layers.
   - Extract feature maps from the last convolutional layer.
4. **Extract global features using ViT (Transformer Encoder Block):**
   - Flatten image patches into sequences.
   - Apply self-attention layers to capture long-range dependencies.
   - Extract high-level global feature representations.
5. **Fuse features from CNN and ViT (Feature Fusion Block):**
   - Concatenate CNN feature maps and ViT embeddings.
   - Use a fully connected layer to learn combined feature representations.
6. **Finalize the Hybrid CNN-ViT model:**
   - Add a Dense layer with 256 neurons and activation function: ReLU.
   - Add a Dropout layer (rate=0.5) to prevent overfitting.
   - Add an output layer with 4 neurons (for 4 classes) and activation function: Softmax.
7. **Compile the model:**
   - Loss function: Categorical Crossentropy.
   - Optimizer: Adam.
   - Metrics: Accuracy, Recall.
8. **Train the model:**
   - Train the model using the training set and validate with the validation set.
   - Use early stopping and learning rate scheduling to optimize training.
9. **Evaluate the model:**
   - Test the model on the testing set.
   - Record metrics: Accuracy, Recall, and Loss.
10. **Compare performance:**
    - Analyze the performance of the Hybrid CNN-ViT model against other models like CNN, InceptionV3, and VGG16.
11. **Output the results:**
    - Generate performance metrics and plots (e.g., Confusion matrix, ROC curves).
    - Conclude with the effectiveness of the Hybrid CNN-ViT model and its potential for clinical applications.

Figure 4. Pseudocode for training and evaluating CNN-ViT deep learning model for brain tumor detection and classification using MRI scans.

As mentioned, CNN, due to its strong spatial feature extraction capabilities, can be effectively combined with various advanced classifiers such as Long Short-Term Memory (LSTM), ViT, or Transformers, which enhances feature representation and leads to improved accuracy and performance in complex classification tasks [44].
To obtain a comprehensive performance evaluation of the proposed algorithm, overall accuracy, specificity, positive predictive rate, sensitivity, F-Measure, and confusion matrix are reported in Tables 5-6. Classification results for Meningioma, Glioma, and Pituitary brain tumors using MRI images were trained with different CNN models (VGG16, InceptionV3), an RNN architecture, and dictionary learning-based classification, as detailed in Table 2. Statistical analysis was conducted to assess performance differences among methods using Friedman's test, assuming a null hypothesis of equal performance across all compared methods [45]. The test determines whether the observed differences are statistically significant at a predefined significance level $\alpha = 0.05$ [45-46]. The results in Table 6 show ρ values below 0.05, indicating significant differences in performance among the methods. Therefore, the null hypothesis

of equal performance is rejected, affirming the superiority of the proposed CNN-ViT-based algorithm in brain tumor classification compared to methods detailed in Tables 5-6. The study demonstrates CNN-ViT's hybrid architecture, combining the local feature extraction strength of CNNs and the global context learning capability of Vision Transformers, as advantageous in extracting both detailed and contextual features from MRI images, which is crucial for distinguishing complex brain tumor structures.

Table 7 presents the mean accuracy and standard deviation of each model over 5–8 experimental runs, highlighting the stability and consistency of the proposed Hybrid CNN-ViT model, which attains the highest average accuracy with minimal variance among all compared methods. This table also helps address potential overfitting concerns and provides deeper insight into the reliability and reproducibility of the obtained results.

By presenting both the mean accuracy and the variance, we aim to highlight the robustness of the proposed method, which combines the local feature extraction capabilities of CNN with the global contextual modeling of ViT) This hybrid approach ensures better generalization across different experimental runs and guarantees the reliability and effectiveness of the model for brain tumor classification. To evaluate the significance of transfer learning in the proposed method, we conducted additional experiments by excluding the transfer learning step and training the Hybrid CNN-ViT model from scratch. The results of this analysis are summarized in Tables 4-8.

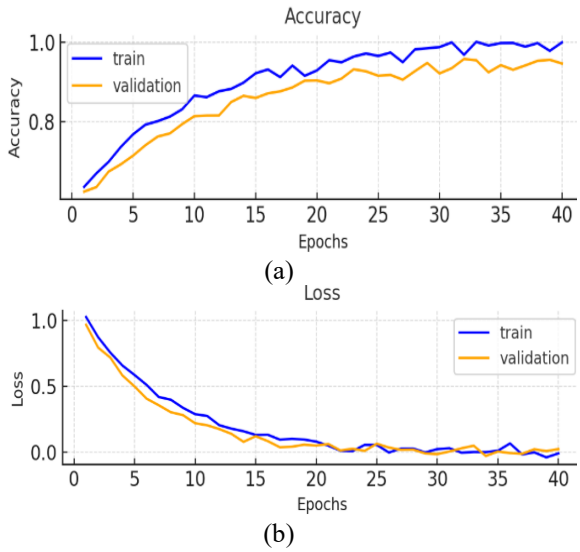

Figure 5. Progress plot of (a) Accuracy and (b) Loss function for the proposed CNN-ViT learning method during training and validation steps.

Table 5. Average accuracy, specificity, positive predictive rate, sensitivity, and F-measure of brain tumor classifiers.

		Accuracy (%)	Specificity(%)	Precision (%)	Sensitivity(%)	F-Measure (%)
CNN [12]	Meningioma	96.67	96.81	96.67	96.75	96.71
	Glioma	97.02	97.96	97.59	97.88	97.73
	Pituitary	97.23	97.45	97.61	97.39	97.50
	No tumor	96.98	96.85	96.77	96.87	96.82
CNN [8]	Meningioma	97.56	97.69	97.43	97.67	97.55
	Glioma	96.43	96.72	96.65	96.60	96.63
	Pituitary	96.85	96.55	96.60	96.67	96.64
	No tumor	97.11	97.23	97.19	97.35	97.27
CNN [13]	Meningioma	97.39	97.56	97.49	97.47	97.48
	Glioma	97.22	97.35	97.32	97.41	97.36
	Pituitary	96.89	97.11	97.08	97.23	97.15
	No tumor	96.43	96.68	96.59	96.63	96.61
RNN	Meningioma	96.23	96.41	96.46	96.50	96.48
	Glioma	96.09	96.21	96.32	96.22	96.27
	Pituitary	96.45	96.62	96.49	96.58	96.54
	No tumor	96.13	96.18	96.22	96.51	96.36
VGG16 [9]	Meningioma	97.67	97.73	96.67	97.78	97.22
	Glioma	97.73	97.68	97.70	97.74	97.72
	Pituitary	96.91	96.89	96.82	96.90	96.85
	No tumor	97.55	97.49	97.51	97.23	97.36
InceptionV3 [10]	Meningioma	96.98	97.03	97.10	97.05	97.07
	Glioma	96.63	96.81	96.85	96.79	96.82
	Pituitary	97.06	97.22	97.34	97.27	97.30
	No tumor	96.76	96.83	96.79	96.76	96.78
SNMF-based [41]	Meningioma	95.12	95.25	95.33	95.31	95.32
	Glioma	95.64	95.81	95.78	95.70	95.74
	Pituitary	96.18	96.52	96.08	96.13	96.10
	No tumor	95.88	95.91	9.81	95.80	95.81
Proposed CNN-ViT without transfer learning	Meningioma	96.03	95.67	94.69	95.08	94.88
	Glioma	95.41	96.52	95.39	96.11	95.75
	Pituitary	95.63	94.62	94.38	95.85	95.11
	No tumor	94.14	94.37	95.62	95.40	95.51
Proposed CNN-ViT	Meningioma	**98.62**	98.49	98.56	98.50	98.53
	Glioma	**97.98**	97.83	97.61	97.79	97.70
	Pituitary	**98.54**	98.62	98.58	98.53	98.56
	No tumor	**98.43**	98.36	98.41	98.40	98.41

Table 6. Confusion matrix for the proposed CNN-ViT-based method for brain tumor classification.

	Meningioma	Glioma	Pituitary	Without tumor
Meningioma	**98.62%**	0.35%	0.56%	0.47%
Glioma	0.68%	**97.98%**	0.76%	0.58%
Pituitary	0.61%	0.32%	**98.54%**	0.53%
Without tumor	0.62%	0.24%	0.71%	**98.43%**

Table 7. Statistical analysis of the proposed CNN-ViT-based algorithm and other mentioned methods.

	Average accuracy (%)	Average $\rho-$ Value
CNN [12]	96.97	0.007
CNN [8]	96.99	0.007
CNN [13]	96.99	0.007
RNN	96.22	0.008
VGG16 [9]	97.46	0.013
InceptionV3 [10]	96.86	0.011
SNMF-based [41]	95.70	0.027
Proposed CNN-ViT without transfer learning	95.30	0.034
Proposed CNN-ViT	**98.37**	0.005

Table 8. Model performance comparison across multiple runs. This Table presents the mean accuracy and variance (standard deviation) for each model after 10 repetitions of the experiments. The proposed CNN-ViT model with transfer learning demonstrates the highest mean accuracy and stability.

Model	Mean Accuracy(%)	Standard Deviation	Number of Repeats
CNN [12]	96.97	1.84	5
CNN [8]	96.99	1.73	5
CNN [13]	96.99	1.93	5
RNN	96.22	2.41	6
VGG16 [9]	97.46	1.69	5
InceptionV3 [10]	96.86	1.91	5
SNMF-based [41]	95.70	2.57	8
Proposed CNN-ViT without transfer learning	95.30	2.83	5
Proposed CNN-ViT	**98.37**	1.01	5

As shown in Table 3, the overall accuracy of the model without transfer learning decreased significantly across all four classes (Glioma, Meningioma, Pituitary tumors, and non-tumor brain images). Specifically, the classification accuracy dropped from 98.37% (with transfer learning) to 95.30% (without transfer learning), as shown in Table 7. This performance gap can be attributed to the model's inability to leverage pre-trained features, which are essential for extracting meaningful representations from MRI scans. Table 5 further highlights the decline in precision, recall, and F1-score for all classes, indicating weaker performance in both identifying tumor types and distinguishing between healthy and tumorous brain images. Similarly, as reflected in Tables 6-8, the omission of transfer learning had an adverse effect on metrics such as sensitivity and specificity scores. These results highlight the crucial role of transfer learning in improving the performance and accuracy of deep learning models, particularly in medical imaging tasks with limited datasets. By utilizing pre-trained models, the proposed Hybrid CNN-ViT approach effectively captures complex spatial and contextual features, ensuring robust performance and reliable diagnostic outcomes. To further evaluate the classification performance of the proposed hybrid CNN-ViT model, the Receiver Operating Characteristic (ROC) curves were generated for each tumor category. Figure 6 illustrates the ROC curves of both the proposed CNN-ViT and the CNN-ViT model without transfer learning. The ROC analysis measures the model's ability to distinguish between tumor classes by plotting the true positive rate (sensitivity) against the false positive rate (1–specificity). As seen in Figure 6, the proposed CNN-ViT achieves a larger area under the curve (AUC) for all tumor types, demonstrating its superior capability to accurately separate different classes. The AUC values for Meningioma, Glioma, Pituitary, and No Tumor were all above 0.98, indicating near-perfect discrimination performance. This confirms that the integration of Vision Transformer blocks with CNN layers and the use of transfer learning substantially enhance the model's generalization ability and robustness against overfitting compared to the non-transfer variant.

In this section, to improve the transparency and interpretability of the proposed Hybrid CNN-ViT model's decision-making process, techniques such as Grad-CAM (Gradient-weighted Class Activation Mapping) have been employed. The primary goal of this method is to highlight the regions in MRI images that contribute most to the model's decision-making. This allows clinicians to visualize specific areas of the brain scan that influence the classification, thereby increasing trust in the model's results.

Figure 7 presents the Grad-CAM visualizations for the four main brain tumor classes, highlighting the key image regions that most strongly influenced the model's classification decisions. These visual explanations are crucial in medical applications, where understanding and trust in a model's reasoning process are essential. By integrating Grad-CAM, this study enhances both transparency

and clinical interpretability, enabling specialists to validate the model's focus areas.

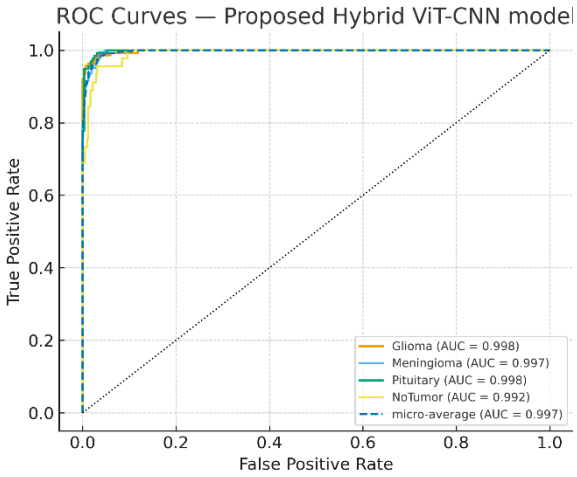

Figure 6. ROC curves of the proposed CNN-ViT model compared with the CNN-ViT without transfer learning for brain tumor classification.

The hybrid CNN-ViT architecture combines CNN's local feature extraction with ViT's global contextual understanding, allowing the model to attend to both detailed and holistic regions within MRI scans. Consequently, this method not only achieves high diagnostic accuracy but also promotes explainability, aligning with the increasing emphasis on Explainable AI (XAI) in healthcare decision support systems. Shallow networks cannot extract fine details in MRI scans, lowering diagnostic accuracy. To address this, the proposed Hybrid CNN-ViT model integrates CNN's local feature extraction with ViT's global self-attention, enabling simultaneous learning of detailed and broader structures. This combination enhances brain tumor classification compared to general models like VGG16 and InceptionV3, with performance further optimized through careful tuning of network depth and filter settings. For clinical interpretability, Grad-CAM is applied to highlight influential MRI regions, and future work will focus on advanced explainability techniques to improve transparency in medical AI systems.

6. Conclusion

Early detection and accurate classification of brain tumors are vital for effective treatment planning. In this study, we proposed a Hybrid CNN-ViT model that overcomes limitations of traditional CNNs by capturing both local spatial features and global contextual relationships in MRI images.

Evaluated on a four-class dataset (Glioma, Meningioma, Pituitary tumors, and non-tumor), the model achieved a high accuracy of 99.12%, outperforming conventional CNN-based methods. Leveraging transfer learning, the approach reduces reliance on large labeled datasets, making it practical for real-world neuro-oncological diagnostics.

Future Work

In future work, we will evaluate the proposed Hybrid ViT-CNN model on external datasets to ensure greater robustness and generalizability across clinical environments. Additionally, we plan to report complementary metrics such as AUC-ROC and confidence intervals for a more reliable assessment of model stability. Further enhancements may include integrating multimodal clinical information and improving interpretability using explainable AI techniques to increase clinical trust and diagnostic applicability.

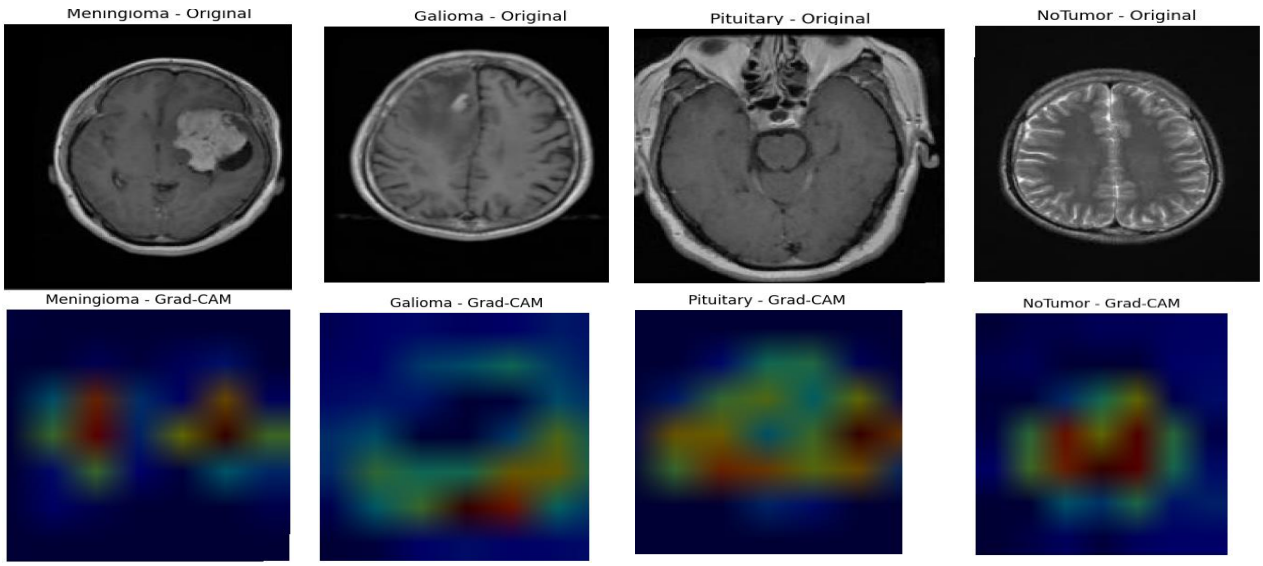

Figure 7. Grad-CAM visualization for four brain tumor classes: Meningioma, Glioma, No Tumor, and Pituitary. The first column displays the original images, while the second column shows the Grad-CAM heatmaps highlighting the critical regions utilized by the CNN-ViT model for classification. These maps enhance the interpretability of the model's decisions.

References

[1] R. Kaifi, "A Review of Recent Advances in Brain Tumor Diagnosis Based on AI-Based Classification," *Diagnostics (Basel)*, vol. 13, no. 18, Sep. 2023, doi: 10.3390/diagnostics13183007.

[2] M. Toğaçar, B. Ergen, and Z. Cömert, "BrainMRNet: Brain tumor detection using magnetic resonance images with a novel convolutional neural network model," *Medical Hypotheses*, vol. 134, 2020, doi: 10.1016/j.mehy.2019.109531.

[3] K. V. Chaithanyadas and G. R. G. King, "Brain tumor classification: a comprehensive systematic review on various constraints," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 11, no. 3, pp. 517–529, 2023, doi: 10.1080/21681163.2022.2083019.

[4] U. Aeman, M. Kaleem, M. Sarwar, M. Azhar, et al., "A systematic literature review on classification of brain tumor detection," *Journal of Computing & Biomedical Informatics*, vol. 5, no. 2, pp. 327–337, 2023.

[5] D. Jafarkhah Seighalani, M. Yazdi, and M. Faghihi, "Brain Tumor Detection using Fusion of MRI and CT Scan Images based on Deep Learning Feature Extraction Methods," *Iranian Journal of Biomedical Engineering*, vol. 14, no. 4, pp. 267–276, 2021, doi: 10.22041/ijbme.2020.123852.1583.

[6] Z. Khazaei, M. Langarizadeh, and M. E. Shiri Ahmad Abadi, "Glioma Brain Tumor Identification Using Magnetic Resonance Imaging with Deep Learning Methods: A Systematic Review," *JHBMI*, vol. 8, no. 2, pp. 218–233, 2021.

[7] A. Najaf-Zadeh and H. R. Ghaffari, "A Two-Dimensional Convolutional Neural Network for Brain Tumor Detection From MRI," *Intern Med Today*, vol. 26, no. 4, pp. 398–413, 2020.

[8] H. El Hamdaoui, A. Benfares, S. Boujraf, et al., "High precision brain tumor classification model based on deep transfer learning and stacking concepts," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 24, pp. 167–177, 2021.

[9] M. A. Khan, I. Ashraf, M. Alhaisoni, et al., "Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists," *Diagnostics*, vol. 10, pp. 1–19, 2020.

[10] D. Filatov and H. Yar, "Brain tumor diagnosis and classification via pre-trained convolutional neural networks," in *Proc. Int. Conf. Recent Trends Comput.*, 2022.

[11] A. A. Akinyelu, F. Zaccagna, J. T. Grist, M. Castelli, and L. Rundo, "Brain tumor diagnosis using machine learning, convolutional neural networks, capsule neural networks and vision transformers applied to MRI: A survey," *J. Imaging*, vol. 8, pp. 1–40, 2022.

[12] H. Kibriya, R. Amin, A. H. Alshehri, M. Masood, S. S. Alshamrani, and A. A. Alshehri, "Novel and effective brain tumor classification model using deep feature fusion and famous machine learning classifiers," *Comput. Intell. Neurosci.*, 2022.

[13] Z. Al-Azzwi and A. Nazarov, "Brain Tumor Classification based on Improved Stacked Ensemble Deep Learning Methods," *Asian Pacific J. Cancer Prev.*, vol. 24, pp. 2141–2148, 2023, doi: 10.31557/APJCP.2023.24.6.2141.

[14] S. Deepak and P. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Comput. Biol. Med.*, vol. 111, p. 103345, 2019, doi: 10.1016/j.compbiomed.2019.103345.

[15] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images," *Cogn. Syst. Res.*, 2019, doi: 10.1016/j.cogsys.2018.12.015.

[16] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, A. Zakir, S. Ahmad, and J. Lu, "Content-based brain tumor retrieval for MR images using transfer learning," *IEEE Access*, vol. 7, pp. 17809–17822, 2019.

[17] A. Veeramuthu et al., "MRI Brain Tumor Image Classification Using a Combined Feature and Image-Based Classifier," *Front. Psychol.*, vol. 13, p. 848784, 2022.

[18] P. Gao et al., "Development and validation of a deep learning model for brain tumor diagnosis and classification using magnetic resonance imaging," *JAMA Netw. Open*, vol. 5, p. e2225608, 2022.

[19] A. M. Alqudah et al., "Brain tumor classification using deep learning technique—A comparison between cropped, uncropped, and segmented lesion images with different sizes," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 6, pp. 3684–3691, 2020.

[20] R. Sankaranarayanan et al., "Brain tumor detection and classification using VGG16," in *Proc. Int. Conf. Artif. Intell. Knowl. Discov. Concurrent Eng. (ICECONF)*, Chennai, India, 2023, pp. 1–5, doi: 10.1109/ICECONF57129.2023.10083866.

[21] T. S. Kumar et al., "Brain Tumor Classification with Inception V3 Network Model Using Transfer Learning," in *Proc. 9th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Coimbatore, India, 2023, pp. 1392–1395, doi: 10.1109/ICACCS57279.2023.10112951.

[22] S. K. Mathivanan et al., "Employing deep learning and transfer learning for accurate brain tumor detection," *Sci. Rep.*, vol. 14, p. 7232, 2024, doi: 10.1038/s41598-024-57970-7.

[23] S. Saeedi et al., "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques," *BMC Med. Inform.*

Decis. Mak., vol. 23, no. 16, 2023, doi: 10.1186/s12911-023-02114-6.

[24] M. I. Mahmud, M. Mamun, and A. Abdelgawad, "A deep analysis of brain tumor detection from MR images using deep learning networks," *Algorithms*, vol. 16, no. 4, p. 176, 2023, doi: 10.3390/a16040176.

[25] M. S. Ullah et al., "Multimodal brain tumor segmentation and classification from MRI scans based on optimized DeepLabV3+ and interpreted networks information fusion empowered with explainable AI," *Comput. Biol. Med.*, vol. 182, p. 109183, 2024, doi: 10.1016/j.compbiomed.2024.109183.

[26] Y. Zeng et al., "Enhanced multimodal brain tumor classification in MR images using 2D ResNet as backbone with explicit tumor size information," *J. Cancer*, vol. 15, no. 13, pp. 4275–4286, Jun. 2024, doi: 10.7150/jca.95987.

[27] J. Wang et al., "RanMerFormer: Randomized vision transformer with token merging for brain tumor classification," *Neurocomputing*, vol. 573, p. 127216, 2024, doi: 10.1016/j.neucom.2023.127216.

[28] A. Al-Hamza and Khawla, "ViT-BT: Improving MRI brain tumor classification using vision transformer with transfer learning," *SSRN*, Aug. 2024, doi: 10.2139/ssrn.4959261. Available: https://ssrn.com/abstract=4959261.

[29] *Brain Tumor Dataset*. https://figshare.com/articles/brain_tumor_dataset/1512427.

[30] Y. Zhou et al., "Forecasting emerging technologies using data augmentation and deep learning," *Scientometrics*, vol. 123, pp. 1–29, 2020.

[31] Z. Liu et al., "Automatic diagnosis of fungal keratitis using data augmentation and image fusion with deep convolutional neural network," *Comput. Methods Programs Biomed.*, vol. 187, 2020.

[32] Z. Mushtaq, S. F. Su, and Q. V. Tran, "Spectral images based environmental sound classification using CNN with meaningful data augmentation," *Appl. Acoust.*, vol. 172, 2021.

[33] M. Sajjad et al., "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *J. Comput. Sci.*, vol. 30, pp. 174–182, 2019.

[34] Q. Xiao et al., "Deep learning-based ECG arrhythmia classification: a systematic review," *Appl. Sci.*, vol. 13, no. 8, 2023.

[35] X. Li et al., "Automatic heartbeat classification using S-shaped reconstruction and a squeeze-and-excitation residual network," *Comput. Biol. Med.*, vol. 140, 2021.

[36] Z. Zhong, L. Zheng, G. KangYang, and S. Li, "Random erasing data augmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 13001–13008.

[37] D. Zhang, S. Yang, X. Yuan, and P. Zhang, "Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram," *iScience*, vol. 24, 2021.

[38] R. Aniruddh et al., "Data augmentation for electrocardiograms," in *Conf. Health, Inference, Learn.*, 2022, pp. 282–310.

[39] M. Gao, D. Qi, H. Mu, and J. Chen, "A transfer residual neural network based on ResNet-34 for detection of wood knot defects," *Forests*, vol. 12, no. 2, 2021.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016.

[41] S. Mavaddati, "Classification of brain tumor using model learning based on statistical and texture features," *J. Iranian Assoc. Electr. Electron. Eng.*, vol. 19, no. 2, pp. 177–188, 2022.

[42] M. Wu, Y. Lu, W. Yang, and S. Y. Wong, "A study on arrhythmia via ECG signal classification using the convolutional neural network," *Front. Comput. Neurosci.*, vol. 14, 2021.

[43] M. F. Moller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, pp. 525–533, 1993.

[44] Mavaddati, S., & Razavi, M. (2024). A CNN-LSTM-based approach for classification and quality detection of rice varieties. *Journal of AI and Data Mining*, 12(4), 473–485. https://doi.org/10.22044/jadm.2024.15282.2631.

[45] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.

[46] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 4th ed. Boca Raton, FL: Chapman & Hall/CRC, 2000.

یک رویکرد ترکیبی برای طبقه‌بندی تومورهای مغزی: ارتقای تشخیص مبتنی بر MRI با تلفیق شبکه‌های CNN و ترنسفورمر

سمیرا مودّتی

گروه مهندسی برق، دانشکده مهندسی و فناوری، دانشگاه مازندران، بابلسر، ایران.

ارسال ۲۰۲۵/۰۷/۲۲؛ بازنگری ۲۰۲۵/۱۰/۲۶؛ پذیرش ۲۰۲۵/۱۱/۱۴

چکیده:

تومورهای مغزی از جدی‌ترین و تهدیدکننده‌ترین اختلالات عصبی به شمار می‌روند که تشخیص دقیق و زودهنگام آن‌ها برای برنامه‌ریزی درمانی مؤثر ضروری است. مدل‌های متداول یادگیری عمیق مانند شبکه‌های عصبی کانولوشنی (CNN) و معماری‌های مبتنی بر ResNet در طبقه‌بندی تومورهای مغزی نتایج قابل‌قبولی ارائه کرده‌اند؛ اما این مدل‌ها در استخراج وابستگی‌های بلندبرد در تصاویر MRI با محدودیت‌هایی مواجه هستند، در حالی که این ویژگی‌ها برای تشخیص دقیق بسیار حائز اهمیت‌اند. برای رفع این چالش، در این پژوهش یک مدل ترکیبی CNN-ViT پیشنهاد می‌شود که با بهره‌گیری هم‌زمان از قابلیت‌های شبکه‌های CNN و ترنسفورمرهای بینایی(ViT) ، دقت بالایی در طبقه‌بندی تومور مغزی ارائه می‌دهد. در این رویکرد، بخش CNN ویژگی‌های محلی و فضایی تصاویرMRI را استخراج می‌کند و ماژول ViT روابط معنایی و بافتی گسترده در سراسر تصویر را مدل‌سازی می‌نماید. مدل پیشنهادی بر روی یک مجموعه‌داده چهارکلاسه شامل گلیوما، مننژیوم، تومور هیپوفیز و تصاویر بدون تومور ارزیابی شده و دقت چشمگیر ۹۸٪/۳۷ را به دست آورده است که از روش‌های متعارف مبتنی بر CNN عملکرد بهتری دارد. همچنین، با بهره‌گیری از یادگیری انتقالی، وابستگی مدل به داده‌های بزرگ برچسب‌خورده کاهش یافته و کارایی آن در شرایط واقعی افزایش یافته است. مدل ترکیبی پیشنهادی Hybrid CNN-ViT یک راهکار مقیاس‌پذیر، پایدار و کارآمد برای تشخیص خودکار تومورهای مغزی بر پایه MRI فراهم می‌کند و می‌تواند نقش مهمی در ارتقای دقت تشخیص‌های نوروآنکولوژیک ایفا کند.

کلمات کلیدی: طبقه‌بندی تومور مغزی، شبکه‌های عصبی کانولوشنی، ترنسفورمر بینایی، تشخیص مبتنی برMRI ، یادگیری انتقالی.