Imaging and Radiation Research 2024, 7(2), 8514.
https://doi.org/10.24294/irr8514

Article

Comparative analysis of vision transformers and fine-tuned transfer learning models for brain tumor classification

Shaimaa E. Nassar[1,2,*], Ibrahim Yasser[1], Hanan M. Amer[1], Mohamed A. Mohamed[1]

[1] Electronics and Communication Engineering Department, Mansoura University, Mansoura 35516, Egypt
[2] Nile Higher Institute of Engineering and Technology, Mansoura 35511, Egypt
*** Corresponding author:** Shaimaa E. Nassar, shaimaaelsabahy@std.mans.edu.eg

CITATION

Nassar SE, Yasser I, Amer HM, Mohamed MA. (2024). Comparative analysis of vision transformers and fine-tuned transfer learning models for brain tumor classification. Imaging and Radiation Research. 7(1): 8514. https://doi.org/10.24294/irr8514

ARTICLE INFO

Received: 10 August 2024
Accepted: 21 September 2024
Available online: 1 November 2024

Abstract: Brain tumors are a primary factor causing cancer-related deaths globally, and their classification remains a significant research challenge due to the variability in tumor intensity, size, and shape, as well as the similar appearances of different tumor types. Accurate differentiation is further complicated by these factors, making diagnosis difficult even with advanced imaging techniques such as magnetic resonance imaging (MRI). Recent techniques in artificial intelligence (AI), in particular deep learning (DL), have improved the speed and accuracy of medical image analysis, but they still face challenges like overfitting and the need for large annotated datasets. This study addresses these challenges by presenting two approaches for brain tumor classification using MRI images. The first approach involves fine-tuning transfer learning cutting-edge models, including SEResNet, ConvNeXtBase, and ResNet101V2, with global average pooling 2D and dropout layers to minimize overfitting and reduce the need for extensive preprocessing. The second approach leverages the Vision Transformer (ViT), optimized with the AdamW optimizer and extensive data augmentation. Experiments on the BT-Large-4C dataset demonstrate that SEResNet achieves the highest accuracy of 97.96%, surpassing ViT's 95.4%. These results suggest that fine-tuning and transfer learning models are more effective at addressing the challenges of overfitting and dataset limitations, ultimately outperforming the Vision Transformer and existing state-of-the-art techniques in brain tumor classification.

Keywords: brain tumors; vision transformer (ViT); artificial intelligence (AI); deep learning (DL); magnetic resonance imaging (MRI)

1. Introduction

An abnormal and uncontrolled growth of cancer cells within the brain or its surrounding tissues is known as a brain tumor. This growth can be either benign or malignant and may originate directly from brain cells, which is classified as a primary brain tumor. Alternatively, it may result from metastasized cells and tissues from other parts of the body, leading to what is called a secondary brain tumor. As reported by CBTRUS, more than 700,000 people, approximately half the population of Hawaii, are currently living with a primary brain tumor. Each year, nearly 84,000 new cases are identified and diagnosed. While brain tumors can occur at any age, they are most frequently diagnosed in older adults and children [1,2]. The most common symptoms of brain tumors include headaches, cognitive changes, and seizures. Additional symptoms may include vision or hearing impairment, limb weakness, and speech or language difficulty [3]. Physicians rely on various factors to detect and categorize brain tumors, including their location, size, and specific imaging characteristics. Meningiomas are noncancerous tumors that develop from

the membranous tissues of the brain, while gliomas and glioblastomas are malignant tumors originating from glial cells and the brain, respectively. Another type of tumor, the pituitary tumor, forms in the pituitary gland, which plays a crucial role in regulating other glands in the body. By considering these distinct characteristics, physicians can precisely diagnose and deal with brain tumors [4]. Brain tumors have a significant negative impact on both patients and their families, making early detection vital for a better prognosis. Various imaging techniques are employed to detect brain tumors, with the most prevalent being magnetic resonance imaging (MRI), that uses magnetic fields and radio waves to identify tumors [5]. Another method used is computed tomography (CT) scan, which employs X-rays to identify brain tumors. Additionally, positron emission tomography (PET) is utilized, where a radioactive substance is injected into the bloodstream to detect tumors. Among surgical options, a biopsy is considered the most accurate for brain tumor detection, involving the examination of a small tumor sample under a microscope to determine its type. These imaging techniques are efficient in detecting brain tumors. However, these conventional methods have significant drawbacks. Imaging techniques can be costly and time-consuming, posing challenges for patients who necessitate frequent follow-up examinations. Additionally, the accuracy of these methods may be affected by the tumor's size, location, and the surrounding tissues, leading to potential inaccuracies. This can result in false outcomes, as indicated by the confusion matrix, which may lead to incorrect diagnoses and delays in treatment [6].

In recent years, advancements in artificial intelligence (AI), machine learning (ML), and deep learning (DL) have transformed the healthcare sector by offering real-time solutions. However, the complexity involved in traditional machine learning operations—such as pre-processing, segmentation, and feature extraction—can diminish the efficiency and accuracy of these models [7]. To overcome the limitations of traditional machine learning methods, deep learning techniques have been introduced to extract and utilize valuable features from input images for more effective diagnosis and classification. Deep learning can enhance the accuracy of tumor identification and classification for doctors. Convolutional neural networks (CNNs) are among the most frequently employed deep learning techniques and are widely applied across various domains [8]. Effective CNN-based classification systems usually need large volumes of visual data for training. To enhance the performance of individual CNN architectures by leveraging pre-existing knowledge, transfer learning can be employed to achieve improved classification accuracy. Transfer learning involves adapting a CNN model that has been trained on a broad dataset like ImageNet to work with domain-specific and smaller datasets. This approach allows the model to leverage previously learned features, and the network parameters are then fine-tuned to enhance performance. The advantage of transfer learning is that it not only boosts classification accuracy but also accelerates the training process [7–9]. Recently, transformers have become prominent models in natural language processing. An adaptation of this model for image analysis, known as the Vision Transformer (ViT), was introduced in Dosovitskiy et al.' research [10]. This study employs two approaches: the first uses transfer learning models (SEResNet, ConvNeXtBase, and ResNet101V2) fine-tuned with global average pooling (GAP) 2D, a dropout layer at 0.2, and a dense layer for classifying four

categories; the second employs the Vision Transformer (ViT) with various data augmentation techniques and hyperparameters, including the AdamW optimizer. These adjustments improve model performance by reducing overfitting, enhancing generalization, and lowering computational complexity [11]. Although the ViT generally performs less effectively compared to the fine-tuned transfer learning models, it achieves higher accuracy than previously reported ViT models in the current state of the art [12].

The primary objectives of the proposed framework are outlined below.

- The study introduces two approaches for classifying brain tumors. The first approach utilizes three fine-tuned transfer learning (TL) cutting-edge models—SEResNet, ConvNeXtBase, and ResNet101V2. These models are enhanced with additional layers, including Global Average Pooling (GAP) 2D, a dropout layer with a rate of 0.2, and a dense layer, to improve the accuracy of brain tumor classification.
- The second approach employs the Vision Transformer (ViT) model, optimized with the AdamW optimizer and extensive data augmentation techniques, achieving a higher accuracy of 95.4% compared to other state-of-the-art ViT models on the BT-Large-4C dataset;
- SEResNet achieves the highest classification accuracy of 97.96% on the BT-Large-4C dataset, which includes MRI images of meningiomas, gliomas, pituitary tumors and healthy brains. This result highlights its superior performance compared to the other investigated models and state-of-the-art methods;
- The study compares fine-tuned transfer learning models with the Vision Transformer (ViT) for classifying brain tumors, finding that the fine-tuned models outperform the ViT, thus demonstrating their superior effectiveness.

The paper is structured in the following way: Section 2 provides a literature review, Section 3 details the methodologies employed, Section 4 outlines the materials used along with the results and discussion, section 5 utilizes limitations and Section 6 provides the conclusion of the study.

2. Literature review

Several methods for classifying brain tumors using MRI scans have been suggested by researchers worldwide. These approaches encompass both conventional machine learning algorithms and sophisticated deep learning models. This section reviews the different findings from these studies on diagnosing brain tumors through MRI images. Ghassemi et al. [13] presented a deep learning approach for classifying brain tumors. They utilized pre-trained networks as GAN discriminators to capture strong features and comprehend the structural details of MRI images. By substituting the fully connected layers with new components and employing techniques such as dropout and data augmentation, their method achieved an accuracy of 95.6% with fivefold cross-validation.

Shaik et al. [14] tackled the complex issue of categorizing brain tumors for medical imaging by introducing MANet, a multi-level attention mechanism. This approach integrates spatial and cross-channel attention mechanisms to highlight

tumors and maintain dependencies across different channels. The method achieved an accuracy of 96.51% for classifying primary brain tumors. Ahmad et al. [15] introduced a deep generative neural network for classifying brain tumors, which merged variational autoencoders with generative adversarial networks to generate realistic MRI images of brain tumors. This approach achieved an accuracy of 96.25%. Munira et al. [16] employed preprocessing methods such as thresholding, cropping, resizing, and rescaling to create a cusomized 23-layer CNN. The extracted features were evaluated using support vector machine (SVM) and random forest (RF) classifiers. The research tested various models, including CNN, CNN-RF, CNN-SVM, and fine-tuned Inception V3, on multi-class brain MRI datasets. Among the models assessed on two publicly available datasets, the CNN-RF model attained 96.52% accuracy on the Figshare dataset, whereas the CNN-SVM model achieved 95.41% accuracy on the BT-large-4c dataset.

Vankdothu et al. [17] developed a model named CNN-LSTM, combining a convolutional neural network with a long short-term memory component. This model demonstrated accuracy rates of 80% and 92% on two separate dataset splits: one with 80% allocated for training and 20% for testing, and the other with 90% for training and 10% for testing, using the BT-large-4c dataset. Hossain et al. [12] employed several pre-trained models, VGG19, VGG16, and InceptionV3, and developed a custom model, IVX16, by combining these top-performing networks. They enhanced their dataset through data augmentation techniques and achieved the highest accuracy of 96.94% with IVX16. Other models reached accuracies between 93.58% and 95.11%. The evaluation was conducted on the BT-large-4c dataset, where 80% of the data was used for training, and the remaining 20% was split equally between validation and testing. Additionally, they tested various Vision Transformer (ViT) models, including SWIN, CCT, and EANet, which attained accuracies of 80%, 74%, and 56%, respectively, on the same dataset. Yurdakul et al. [18] evaluated various Vision Transformer (ViT) models for brain tumor classification using the BT Large 4C dataset. They found that ViT-L/32 achieved the highest accuracy at 92.89%, followed by ViT-L/16 at 92.64%. ViT-B/32 showed the lowest performance with an accuracy of 88.83%. Overall, the top-performing models were ViT-L/32, ViT-L/16, and MobileNet, with accuracies of 92.89%, 92.64%, and 92.89%, respectively.

Divya et al. [19] utilized the ResNet50 algorithm along with data augmentation as preprocessing steps to extract robust features and analyze the structure of MR images. To improve the model's performance, they incorporated three linear modules, two Leaky ReLU modules, two dropout layers, and a soft max classification layer to differentiate tumor types using the Figshare dataset. This approach achieved a maximum accuracy of 98.57%. Pashaei et al. [20] proposed two different approaches for classifying brain tumors. The first approach involved a convolutional neural network (CNN) with 4 convolutional layers followed by 4 pooling layers, a fully connected layer, and additional intermediate layers for data normalization, achieving an accuracy of 81.09%. In their second approach, they used the CNN for feature extraction and employed Kernel Extreme Learning Machine (KELM) for classification, resulting in a higher accuracy of 93.68%.

Table 1. Summarizes the findings of literature survey.

Reference	Dataset	Method	Accuracy
Ghassemi et al. [13]	T1W-CE MRI	pre-trained networks as GAN	96.6%
Shaik et al. [14]	T1W-CE MRI	a multi-level attention mechanism (MANet)	96.51%
Ahmad et al. [15]	Figshare	a deep generative neural network	96.25%
Munira et al. [16]	Figshare	cusomized 23-layer CNN with RF	96.52%
	BT-large-4c	cusomized 23-layer CNN with SVM	95.41%
Vankdothu et al. [17]	BT-large-4c	CNN-LSTM splitting data (80%:20%)	80%
		CNN-LSTM splitting data (90%:10%)	92%
Hossain et al. [12]	BT-large-4c	IVX16	96.94%
		ViT (SWIN)	80%
		ViT (CCT)	74%
		ViT (EANet)	56%
Yurdakul et al. [18]	BT-large-4c	ViT-L/32	92.89%
		ViT-L/16	92.64%
		ViT-B/32	88.83 %
		MobileNet	92.89%
Divya et al. [19]	Figshare	ResNet50 with replacing layers	98.67%
Pashaei et al. [20]	T1W-CE MRI	Custom CNN AND kELM	93.68%
Salih et al. [21]	BT-large-4c	ResNet18 and ResNet50 to extract features	93.74%
Sarada et al. [22]	Figshare, SARTAJ, and Br35H	modified ResNet50V2	96.34%
Suryawanshi et al. [23]	SARTAJ	DL algorithm (CNN-SVM)	95.16%
Jun et al. [24]	Figshare	Dual-attention	98.61%
Kang et al. [25]	BT-large-4c	Feature ensemble SVM	93.72%
Mahmud et al. [26]	BT-large-4c	Developed CNN	93.3%
Nassar et al. [4]	Fishare	Majority voting technique	99.31%

Salih et al. [21] combined feature representations from two distinct deep learning models, ResNet18 and ResNet50, to create more effective feature vectors for classifying different categories. These feature vectors were then fed into a machine learning layer to categorize them into four distinct classes. The preprocessing steps included resizing images to 224 × 224, applying a Gaussian filter, and performing normalization. They achieved an accuracy of 93.74% based on BT-large-4c. Sarada et al. [22] enhanced the ResNet50v2 model by incorporating dropout layers, max pooling, and batch normalization. This improved model achieved an accuracy of 96.34% across three datasets: Figshare, SARTAJ, and Br35H. Sarada et al. [23] employed a convolutional neural network (CNN) and VGG19, combined with the CNN-Support Vector Machines (CNN-SVM) algorithm, achieving an accuracy of 95.16%. This approach was applied to the Brats 2018 dataset for HGG and LGG classification, as well as the SARTAJ dataset for classifying glioma, meningioma, and no tumor. Jun et al. [24] implemented an attention mechanism and a multipath network to improve performance. When tested on a dataset of 3,064 MR images, this approach achieved an overall accuracy of 98.61% based on the Figshare dataset. Nassar et al. [4] employed a majority voting

technique using five deep learning models, achieving an accuracy of 99.31% on the Figshare dataset with minimal preprocessing steps. Kang et al. [25] used features from ShuffleNet V2, DenseNet-169, and MnasNet to train a classical classifier with Support Vector Machines (SVM). They applied preprocessing steps such as cropping, resizing, and data augmentation (rotation and horizontal flipping). This approach achieved a testing accuracy of 93.72% for classifying glioma, meningioma, no tumor, and pituitary tumors using the BT-large-4c dataset, which was split into 80% training and 20% testing data. Mahmud et al. [26] developed a CNN model with three convolutional layers, max-pooling, and a dense layer with 4160 dimensions, using softmax and ReLU activations, along with a dropout rate of 0.5. They applied various data augmentation techniques and achieved a 93.3% classification accuracy for glioma, meningioma, no tumor, and pituitary categories using the BT-large-4c dataset. The findings of the literature survey are summarized in **Table 1**.

Current methods often face with problems like overfitting, which means models perform well on training data but poorly on new data. They may not work well in different situations and can be very costly to compute. Training data biases can make the models less reliable and useful in real-world scenarios. Many methods also require complex and time-consuming preprocessing steps. The study addresses these issues by using advanced models like SEResNet, ConvNeXtBase, and ResNet101V2, with resizing as the only preprocessing step. This method achieves better accuracy in brain tumor classification and provides a thorough comparison with Vision Transformer (ViT) models, which use extensive data augmentation. This highlights how these new methods effectively address the limitations of existing techniques.

3. Methods

This section outlines two proposed methodologies for brain tumor classification. The first approach, depicted in **Figure 1**, provides a detailed description of three fine-tuned transfer learning (TL) models. The second approach, illustrated in **Figure 2,** employs Vision Transformers (ViTs), a novel deep learning technique for computer vision tasks. This approach involves data collection, preprocessing with augmentation techniques, optimizing hyperparameters, and finally fine-tuning the ViT to classify four categories in the new task of brain tumor classification. A comparative analysis of these two approaches is also presented, highlighting their contributions to advancements in brain tumor classification. To ensure accurate brain tumor classification, the first proposed approach involved image data collection, preprocessing, and a reconstructed transfer learning architecture. Models such as SEResNet, ConvNeXtBase, and ResNet101V2 were fine-tuned with additional layers, including global average pooling 2D, a dropout layer with a rate of 0.2, and dense layers, for classifying brain tumors on a specific dataset. To further enhance the research, a Vision Transformer was applied, demonstrating that fine-tuned transfer learning algorithms outperformed in brain tumor classification.

Steps of the first proposed approach:

Step 1: Acquire the BT-Large-4C dataset, obtained from Kaggle [27], was used for the experiments. It contains 3264 MRI images divided into four categories: meningioma, glioma, pituitary tumors, and healthy brains.

Step 2: Preprocess the images by resizing them to 224 × 224 pixels. Then, split the dataset into three subsets: 80% for training, 10% for validation, and 10% for testing.

Step 3: Reconstruct the transfer learning architecture by removing the last three layers following the final activation layer.

Step 4: Fine-tune the model by adding a global average pooling 2D layer, a dropout layer with a rate of 0.2, and a dense layer, tailoring the model for brain tumor classification.

Step 5: Implement well-established transfer learning models, including SEResNet, ConvNeXtBase, and ResNet101V2, within the approach.

Step 6: Evaluate the performance of each transfer learning model using metrics such as accuracy, specificity, recall, precision, and F1-score.

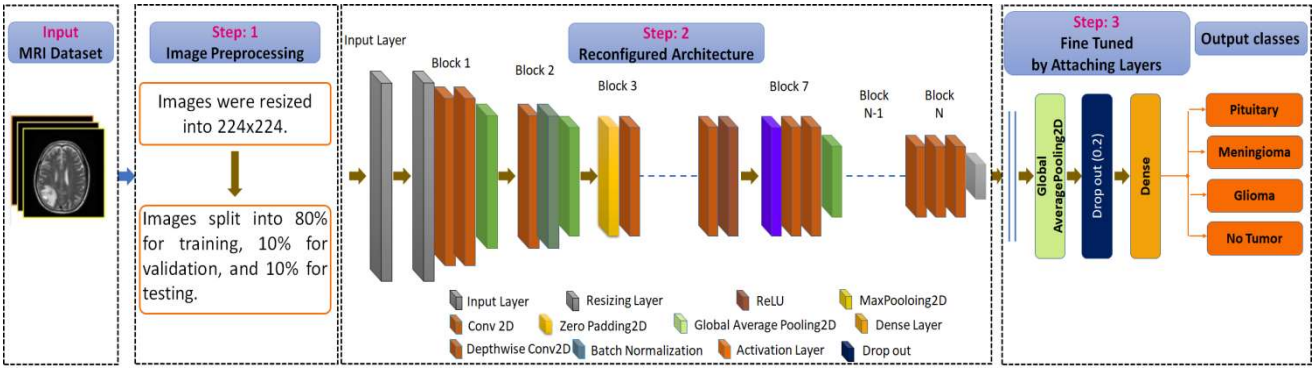

Figure 1. First proposed approach for classifying brain tumors utilizing transfer learning and fine-tuning techniques.

The steps of the second proposed approach:

Step 1: Acquire the BT-Large-4C dataset, obtained from Kaggle [27], was used for the experiments. It contains 3264 MRI images divided into four categories: meningioma, glioma, pituitary tumors, and healthy brains.

Step 2: Conduct preprocessing by resizing all images to 240 × 240 pixels. Subsequently, partition the dataset into training (80%), validation (10%), and testing (10%) subsets. Implement data augmentation techniques to improve model robustness and generalization.

Step 3: Fine-tune the Vision Transformer (ViT) model by meticulously optimizing hyperparameters and employing the AdamW optimizer. Tailor the model to effectively classify the four categories within the dataset.

Step 4: Assess the performance of the ViT model using a comprehensive set of evaluation metrics, including accuracy, specificity, recall, precision, and F1-score, to ensure validation of the model's effectiveness.

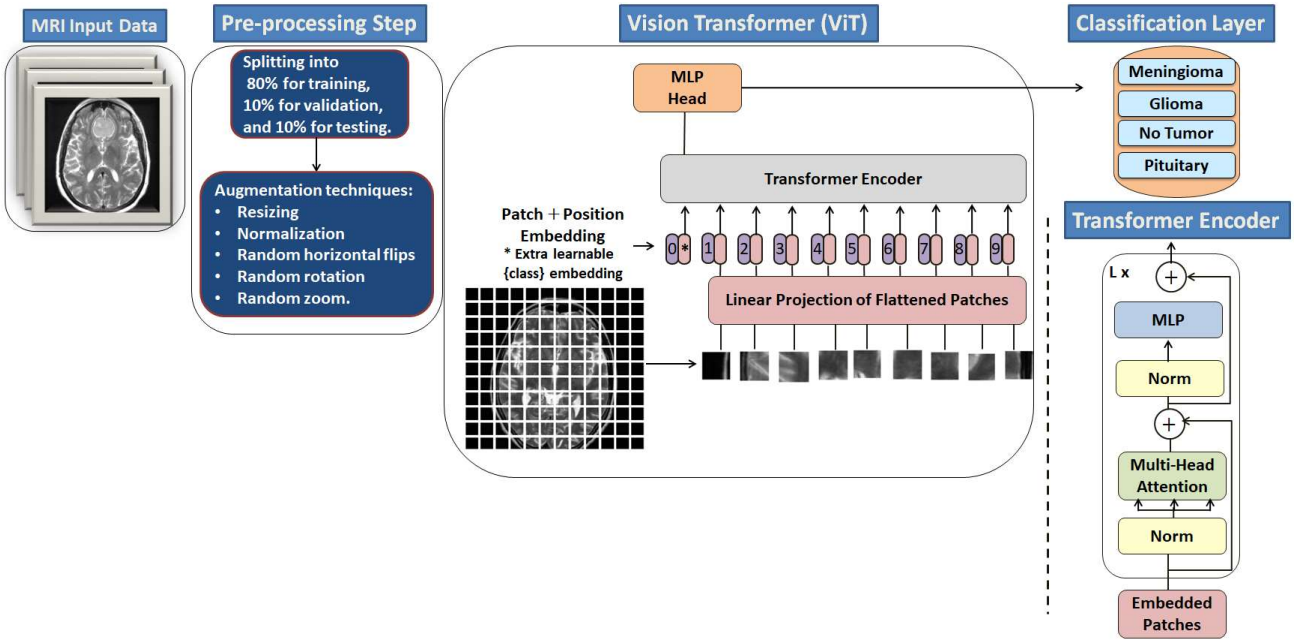

Figure 2. Second proposed approach for brain tumor classification using the Vision Transformer (ViT) model.

After implementing the two proposed approaches on the BT-Large-4C dataset, a comparative analysis was conducted to identify the most effective model, focusing on key performance metrics to improve brain tumor classification.

3.1. Data preprocessing

For the first approach, images are resized to a fixed size of 224 × 224 pixels to meet the input requirements of transfer learning models such as SEResNet, ConvNeXtBase, and ResNet101V2. In contrast, the second approach, which focuses on Vision Transformer (ViT) models, involves more extensive preprocessing. This includes resizing images to 240 × 240 pixels, applying normalization, and using data augmentation techniques such as horizontal flipping, random rotation with a 0.02 factor, and random zoom with width and height factors of 0.2. The AdamW optimizer is utilized to enhance the dataset's suitability for vision-based tasks. Minimal preprocessing for the fine-tuned transfer learning models, such as simple resizing, helps reduce computational load, simplifies model training, preserves image quality, and limits data variability. This approach often leads to superior accuracy compared to ViT models, which require more extensive preprocessing steps.

3.2. Transfer learning

Transfer Learning (TL) aims to boost learning by applying knowledge gained from source tasks to enhance performance on target tasks. TL is an effective approach for reducing training time because it allows leveraging previously acquired knowledge instead of starting the learning process from scratch. Typically, TL involves using a pre-trained deep learning model that was trained on a large dataset, which is then adapted or fine-tuned for the specific target task as shown in **Figure 3**. This approach not only accelerates the training process but also enhances the model's performance by utilizing learned features from related tasks.

This work presents a robust deep learning approach utilizing transfer learning

techniques for categorizing brain tumors. The primary objective is to extract crucial features from a standard dataset to enhance categorization accuracy. The study investigates three deep learning models—SEResNet, ConvNeXtBase, and ResNet101V2—testing and refining their architectures and configurations. The proposed method demonstrates improved performance on the BT Large 4C dataset. In this approach, a significant transfer learning strategy is utilized, involving fine-tuning [28].

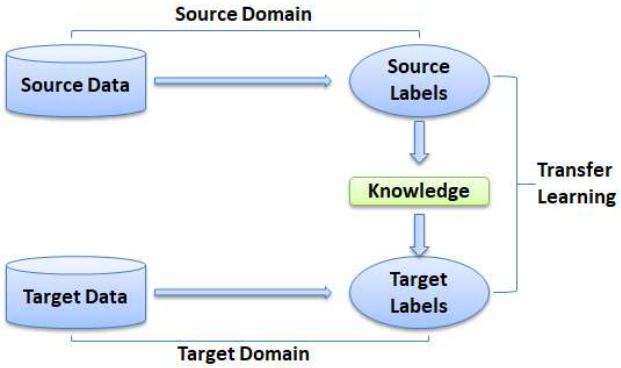

Figure 3. The concept of Transfer Learning using pre-trained models.

3.3. Fine-Tuning

In the fine-tuning process, additional layers are introduced to tailor the architecture for brain MRI classification. Specifically, the last three layers are removed after the final activation layer. A Global Average Pooling 2D layer is then added, followed by a dropout layer with a rate of 0.2 and a dense layer. This configuration aims to enhance the model's ability to effectively classify four categories—meningioma, glioma, pituitary tumor, and no tumor—while also mitigating overfitting and improving generalization.

3.4. Transfer leaning models

SEResNet emphasizes the relationships between convolved feature channels using 1D convolution. The SE block consists of two key operations: the squeeze operation, which aggregates the overall information from each feature map, as well as the excitation operation, that adjusts the significance of each feature map. The squeeze operation captures the critical information from each channel, while the excitation operation calculates the inter-channel dependencies through a fully connected layer with a nonlinear function [29]. The residual block in ResNet effectively utilizes shallow features to extract key feature values, making it a popular choice as the primary structure for feature extraction in image classification and recognition tasks [30]. **Figure 4** illustrates the structural differences between SE-ResNet and traditional ResNet networks. The residual block in ResNet incorporates the SE structure, which not only maximizes the use of shallow features but also reweights each channel to improve key feature extraction. The output of SE-ResNet is obtained, as shown in an Equation (1):

$$y = F(f_{se}(x), (w_i)) + x \quad (1)$$

where, x and y represent the input and output of the SE-ResNet, $f_{se}(.)$ denotes the function of the SE block, and i refers to the weight of the network for the i-th input. However, during the squeeze operation, it is essential to define the scale of the feature image, which significantly influences the reweighting process. Given that the size of each input feature image varies, this paper proposes an adjustable scale based on the size of the feature channel. The output of the j-th SEResNet block, y_j, is defined as shown in Equation (2):

$$y_j = F\left(f_{se}(x_j), (w_{ij})\right) + x_j \tag{2}$$

where, y_j represents the output generated by the j-th SEResNet structure.

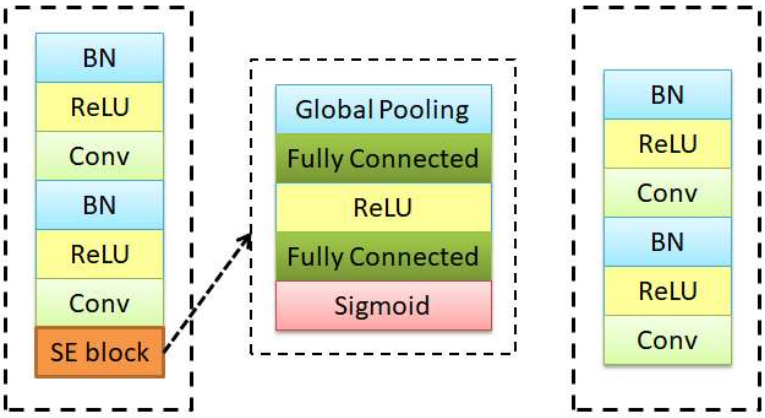

Figure 4. The structural differences between SE-ResNet and traditional ResNet networks.

ConvNeXtBase is an advanced CNN architecture that enhances feature representation and recognition. It aims to boost CNN performance by combining group convolutions with concatenation. Combining group convolutions with concatenation reduces the number of parameters and computational demands needed to train the network, enhancing its efficiency and scalability [31]. ConvNeXts, developed with standard ConvNet modules, perform well against Transformers in terms of accuracy, scalability, and robustness on key benchmarks. Although ConvNeXts are built on traditional ConvNet modules, they are still competitive with Transformers, a different neural network architecture [32]. **Figure 5** depicts the structure of the ConvNeXtBase network.

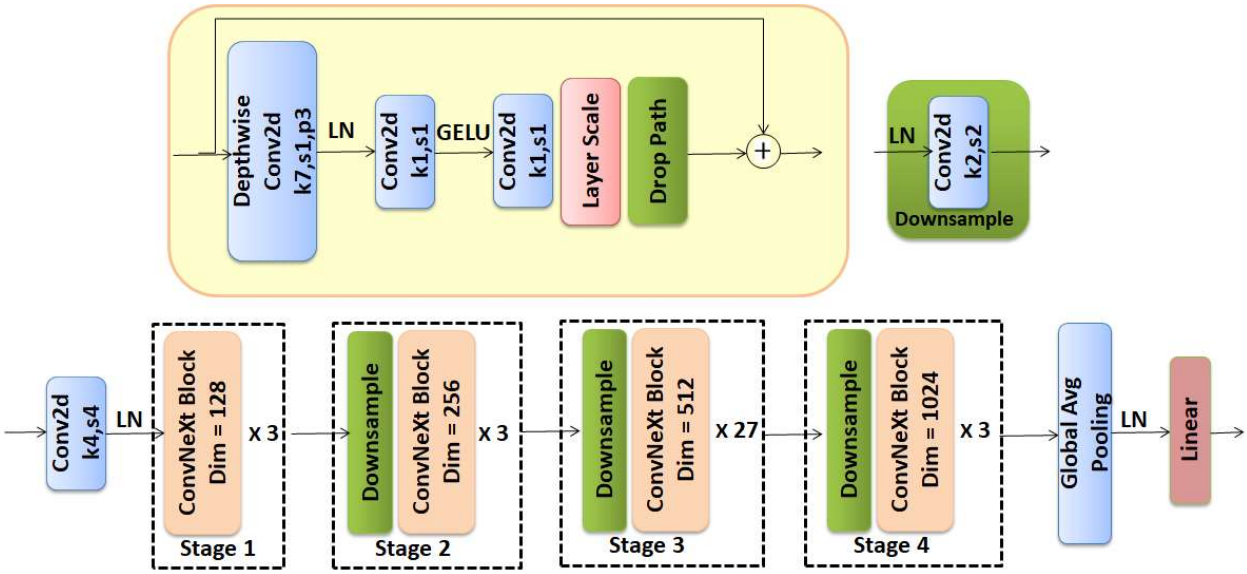

Figure 5. The structure of the ConvNeXtBase network.

ResNet101V2 ResNet101 uses residual connections to maintain gradient flow and prevent vanishing gradients [33]. It features 104 convolutional layers organized into 33 layers and 29 blocks, with residual connections summed at each block. The final blocks use 1 × 1 convolution layers followed by batch normalization to standardize the output. ResNet101V2 networks use pre-activation of weights, which enhances generalization relative to the original ResNet. The key advantage of pre-activation lies in its ability to regularize and normalize the output signal, effectively reducing the likelihood of overfitting [34].

3.5. Vision Transformer (ViT) model

The Transformer architecture is widely utilized in natural language processing (NLP) research [35]. The Vision Transformer (ViT) adapts this architecture for image analysis tasks. Experimental results show that various ViT models have surpassed traditional CNNs in classification tasks on ImageNet, CIFAR, and VTAB datasets. ViT starts by dividing an image into separate patches. These patches, along with their positional information, are fed into the Transformer encoder. The transformer then examines the links between various parts of the image and the overall context. In the final stage, the outputs from the transformer are classified using an MLP head [10]. To ensure compatibility with ViT, the dataset undergoes essential preprocessing steps, including normalization, resizing, random horizontal flips, random rotation (with a factor of 0.02), and random zoom (with height and width factors of 0.2). This preparation ensures that the data meets the necessary criteria for optimal ViT performance [12].

4. Results and discussion

This study evaluates the effectiveness of the first approach, which involves fine-tuning transfer learning models, compared to the second approach based on the ViT model, using the BT-large-4C dataset. This section provides detailed information on the dataset, evaluation metrics, experimental setup, and performance assessment.

4.1. BT-large-4c Dataset

This dataset comprises 3264 JPEG images of MRI scans, depicting three types of brain tumors—meningioma, glioma, and pituitary tumors—alongside images of brains with no tumors. **Figure 6** displays samples of these MRI images in sagittal, coronal, and axial views. It comprises 500 tumor-free scans, 901 pituitary tumor scans, 937 meningioma tumor images, and 926 glioma tumor images [27].

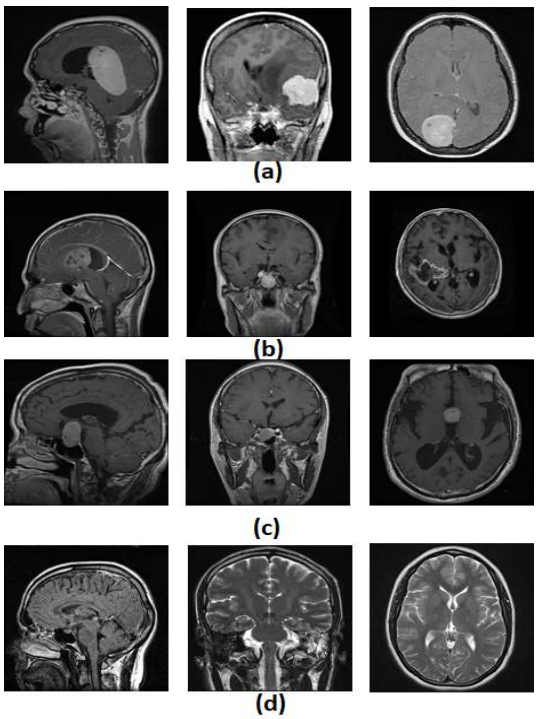

Figure 6. Examples of MRI scans for the four classes. **(a)** Meningioma tumor; **(b)** glioma tumor; **(c)** pituitary tumor; **(d)** healthy brain.
Each class is shown in three different views: axial, coronal, and sagittal (left to right).

4.2. Experimental setup

The experiment was conducted and evaluated using Python within the Kaggle Notebook environment, an open-source platform derived from the Python Notebook project. Python was selected for its object-oriented design, high-level capabilities, and interpretive nature, which were essential for implementing and managing various tasks.

The two proposed approaches were trained using the BT-large-4C dataset, which was partitioned into training, validation, and testing subsets with ratios of 80%, 10%, and 10%, respectively. **Table 2** outlines the hyperparameters used for the different models. The SEResNet, ConvNeXtBase, and ResNet101V2 models were trained with a learning rate of 0.001 over 30 epochs, using a batch size of 32 and the Adam optimizer. In contrast, the Vision Transformer (ViT) model was trained with the same learning rate of 0.001, but over 100 epochs with a batch size of 20, employing the AdamW optimizer—an enhanced version of Adam designed to improve performance [36]. To illustrate the training process, **Figure 7** shows the performance of SEResNet. The model exhibits effective learning, with both training and validation metrics demonstrating significant improvement and eventual

stabilization, indicating successful knowledge acquisition and generalization.

Table 2. Hyper-parameters of utilized models.

Model Name	Learning Rate	No. of Epochs	Batch size	Optimizer
SEResNet	0.001	30	32	Adam
ConvNeXtBase	0.001	30	32	Adam
ResNet101V2	0.001	30	32	Adam
ViT	0.001	100	20	AdamW

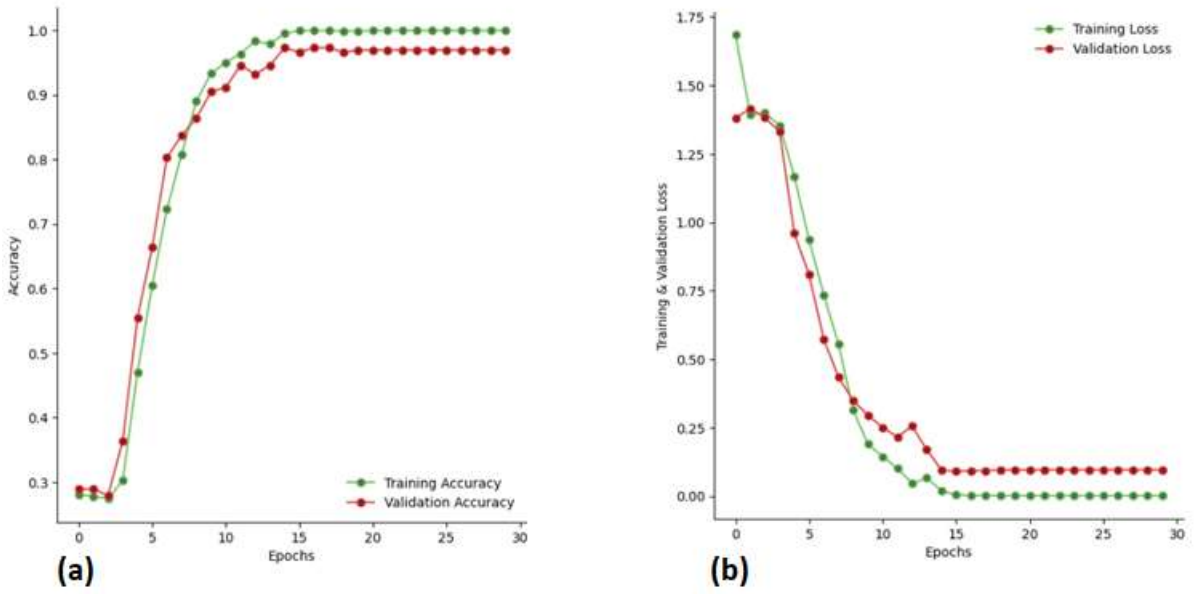

Figure 7. (a) The training and validation accuracy for SEResNet; **(b)** The training and validation accuracy for SEResNet.

4.3. Evaluation metrics

To assess the effectiveness of the proposed technique for brain tumor classification, six key performance metrics were employed: accuracy, sensitivity, precision, specificity, F1-score, and confusion matrices [4]. Accuracy, a basic yet crucial metric, measures the proportion of correctly classified image samples out of the total number of samples, independent of specific class labels. The formula for calculating accuracy is provided in Equation (3).

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (3)$$

Sensitivity, a key performance metric, evaluates how effectively the model identifies brain tumor cases. It is computed using Equation (4):

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (4)$$

Specificity, a vital performance metric, evaluates the model's ability to effectively identify negative samples. It is determined using Equation (5):

$$\text{Specificity} = \text{TN}/(\text{TP} + \text{FP}) \quad (5)$$

Precision, a key performance metric, measures the accuracy of the model's

positive predictions. It is calculated using Equation (6):

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{6}$$

The F1-score, a composite performance measure, provides a balanced assessment of the model's precision and recall. It is calculated using Equation (7):

$$\text{F1} - \text{Score} = 2 \times \text{PrecisionxRecall}/(\text{Precision} + \text{Recall}) \tag{7}$$

where, TP, FP, FN, and TN represent true positives, false positives, false negatives, and true negatives, respectively, in the context of our study.

4.4. Experimental results and discussion

Table 3 presents the performance metrics of the various models utilized in this study for brain tumor classification. SEResNet achieved the highest accuracy at 97.96% along with strong sensitivity, specificity, precision, and F1-score values of 98.10%, 98.05%, 98.14%, and 97.98%, respectively. ConvNeXtBase and ResNet101V2 also performed well, with accuracies of 96.66% and 96.02%, respectively, and strong results across other metrics, though they did not match SEResNet's performance. The Vision Transformer (ViT) achieved a satisfactory accuracy of 95.4%, but it did not surpass the fine-tuned SEResNet model. Overall, SEResNet demonstrated superior performance and shows significant potential for advancing brain tumor classification.

Table 3. Performance metrics of different utilized models.

Model	Accuracy	Sensitivity	Specificity	Precision	F1-score
SEResNet	97.96%	98.10%	98.05%	98.05%	97.98%
ConvNeXtBase	96.66%	96.00%	96.01%	96.80%	96.12%
ResNet101V2	96.02%	96.50%	96.11%	96.32%	96.20%
ViT	95.40%	95.00%	95.20%	95.50%	95.00%

4.5. Comparison of results with related works on the BT-large-4c Dataset

To assess the effectiveness of the proposed approaches, a comparative analysis was conducted using the BT-Large-4C dataset, which is detailed in **Table 4**. The results indicate that the proposed SEResNet model outperforms all other methods, achieving the highest metrics across the board: an accuracy of 97.96%, recall of 98.10%, specificity of 98.05%, precision of 98.14%, and an F1-score of 97.98%. While the second proposed approach, the Vision Transformer (ViT) model, achieved a slightly lower accuracy compared to SEResNet, it still performed well in comparison to other ViT models in the literature. For example, Yurdakul et al. [17] reported an accuracy of 92.89% with their ViT-L/32 model, and even with an ensemble method, their accuracy only slightly improved to 94.92%, which still did not surpass the proposed SEResNet model's performance. Additionally, Hossain et al. [12] demonstrated a relatively high accuracy of 96.94% with their IVX16 model, but it suffered from lower performance in other metrics, such as a recall of 79% and an F1-score of 76%. Other ViT variants reported by Hossain et al., including ViT (SWIN), ViT (CCT), and ViT (EANet), showed even poorer performance, with

accuracies dropping to 80.00%, 74.00%, and 56.00%, respectively. Kang et al. [25] used an ensemble of features with SVM, obtaining an accuracy of 93.72%, while Salih et al. [21] employed ResNet18 and ResNet50 for feature extraction, achieving an accuracy of 92.47% along with a recall of 94.44% and an F1-score of 96.89%. Munira et al. [16] achieved an accuracy of 95.41% using a customized 23-layer CNN combined with SVM. These results underscore the robustness and superior performance of the SEResNet model in brain tumor classification, surpassing existing methods and highlighting its potential for advancing the field of brain tumor classification by addressing the limitations of current techniques.

Table 4. Comparison of results with related works.

Reference	Year	Method	Accuracy (%)	Recall (%)	Specificity (%)	Precision (%)	F1-score (%)
Hossain et al. [12]	2023	IVX16	96.94	79	-	78	76
		ViT (SWIN)	80	-	-	-	-
		ViT (CCT)	74	-	-	-	-
		ViT (EANet)	56	-	-	-	-
Munira et al. [16]	2022	Customized 23-layer CNN with SVM	95.41%	-	-	-	-
Kang et al. [25]	2021	Feature ensemble SVM	93.72	-	-	-	-
Yurdakul et al. [18]	2023	ViT-L/32	92.89	93.53	-	92.89	92.85
		ViT-L/16	92.64	-	93.29	92.58	-
		Ensemble	94.92	95.17	94.92	94.88	-
Salih et al. [21]	2024	ResNet18 and ResNet50 to extract features	92.47	94.44	-	94.37	96.89
Proposed	2024	SEResNet	97.96	98.10	98.05	98.14	97.98
		ViT	95.40	95.00	95.20	95.50	95.00

5. Limitations

The advanced models used, such as SEResNet, ConvNeXtBase, ResNet101V2, and Vision Transformer (ViT), require substantial computational resources, which may not be accessible in all settings. Additionally, the extensive data augmentation techniques used in the ViT approach may not fully reflect real-world conditions, potentially affecting the models' practical application.

6. Conclusions and future work

In future work, additional AI techniques will be explored to further This study presented two distinct approaches for brain tumor classification. The first approach utilized three cutting-edge transfer learning (TL) models—SEResNet, ConvNeXtBase, and ResNet101V2—while the second approach employed the Vision Transformer (ViT) with fine-tuned parameters and the AdamW optimizer. The performance was evaluated using metrics including accuracy, precision, recall, and F1-score. On the BT-Large-4C Brain Tumor Image dataset, SEResNet achieved an accuracy of 97.96%, outperforming other models, while ViT reached 95.40%.

The implications of these models extend across several key applications. They offer significant benefits for clinical diagnosis by providing radiologists with robust tools for accurate tumor identification. Furthermore, they support personalized treatment planning through precise tumor classification and enhance automated medical image analysis, streamlining and expediting the diagnostic process. The integration of these models into clinical workflows has the potential to improve diagnostic efficiency and accuracy, highlighting their practical value and transformative impact on medical imaging.

Future work will focus on enhancing classification performance, and further improvements will be made to the Vision Transformer (ViT) to boost its results for brain tumor classification. Other datasets will be examined to assess the reliability of the proposed system across diverse contexts. Additionally, the system will be applied to other medical classification challenges to evaluate its effectiveness in broader applications.

Author contributions: Conceptualization, SEN and MAM; methodology, SEN; software, SEN; validation, SEN, IY and HMA; formal analysis, SEN; investigation, SEN; resources, SEN; data curation, SEN; writing—original draft preparation, SEN; writing—review and editing, SEN; visualization, SEN; supervision, MAM, IY and HMA; project administration, SEN; funding acquisition, SEN. All authors have read and approved the final version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Pichaivel M, Anbumani G, Theivendren P, et al. An overview of brain tumor. In: Brain Tumors. 2022; pp. 1–10. doi: 10.5772/intechopen.100806.
2. Ostrom QT, Patil N, Cioffi G, et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2013–2017. Neuro-oncology. 2020;22. doi: 10.1093/neuonc/noaa200.
3. Park J, Park YG. Brain tumor rehabilitation: symptoms, complications, and treatment strategy. Brain & Neurorehabilitation. 2022;15(3). doi: 10.12786/bn.2022.15. e25.
4. Nassar SE, Yasser I, Amer HM, et al. A robust MRI-based brain tumor classification via a hybrid deep learning technique. The Journal of Supercomputing. 2023;80(2):2403-2427. https://doi.org/10.1007/s11227-023-05549-w.
5. Abd-Ellah MK, Awad AI, Khalaf AA, et al. A review on brain tumor diagnosis from MRI images: Practical implications, key achievements, and lessons learned. Magnetic Resonance Imaging. 2019; 61:300-318. https://doi.org/10.1016/j.mri.2019.05.028.
6. Asiri AA, Shaf A, Ali T, et al. Exploring the power of deep learning: fine-tuned vision transformer for accurate and efficient brain tumor detection in MRI scans. Diagnostics. 2023;13(12):2094. https://doi.org/10.3390/diagnostics13122094.
7. Rajput IS, Gupta A, Jain V, et al. A transfer learning-based brain tumor classification using magnetic resonance images. Multimedia Tools and Applications. 2023;83(7):20487-20506.
8. Tulbure A-A, Tulbure A-A, Dulf E-H. A review on modern defect detection models using DCNNs—Deep convolutional neural networks. Journal of Advanced Research. 2022; 35:33-48. https://doi.org/10.1016/j.jare.2021.03.015
9. Morid MA, Borjali A, Del Fiol G, et al. A scoping review of transfer learning research on medical image analysis using ImageNet. Computers in Biology and Medicine. 2021; 128:104115.
10. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. 2020.
11. Malla PP, Sahu S, Alutaibi AI. Classification of tumor in brain MR images using deep convolutional neural network and global average pooling. Processes. 2023;11(3):679. https://doi.org/10.3390/pr11030679

12. Hossain S, Chakrabarty A, Gadekallu TR, et al. Vision transformers, ensemble model, and transfer learning leveraging explainable AI for brain tumor detection and classification. IEEE Journal of Biomedical and Health Informatics. 2023;28(3):1261–1272. DOI: 10.1109/JBHI.2023.3266614
13. Ghassemi N, Shoeibi A, Rouhani M. Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images. Biomedical Signal Processing and Control. 2020; 57:101678. http://doi.org/10.1016/j.bspc.2019.101678.
14. Shaik NS, Cherukuri TK. Multi-level attention network: application to brain tumor classification. Signal, Image and Video Processing. 2022;16(3):817–824. https://doi.org/10.1007/s11760-021-02022-0.
15. Ahmad B, Sun J, You Q, et al. Brain tumor classification using a combination of variational autoencoders and generative adversarial networks. Biomedicines. 2022;10(2):223. https://doi.org/10.3390/biomedicines10020223
16. Munira HA, Islam MS. Hybrid deep learning models for multi-classification of tumour from brain MRI. J Inf Syst Eng Bus Intell. 2022; 8:162–174. Available at: http://e-journal.unair.ac.id/index.php/JISEBI
17. Vankdothu R, Hameed MA, Fatima H. A brain tumor identification and classification using deep learning based on CNN-LSTM method. Computers and Electrical Engineering. 2022; 101:107960. Available at: https://doi.org/10.1016/j.compeleceng.2022.107960
18. Yurdakul M, TAŞDEMİR Ş. Brain Tumor Detection with Ensemble of Convolutional Neural Networks and Vision Transformer. In: Proceedings of the 2023 2nd International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI); 2023. IEEE; p. 1–6. DOI: 10.1109/EICEEAI60672.2023.10590129.
19. Divya S, Suresh LP, John A. A deep transfer learning framework for multi-class brain tumor classification using MRI. In: Proceedings of the 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN); 2020. IEEE; p. 283–290. DOI: 10.1109/ICACCCN51052.2020.9362908
20. Pashaei A, Sajedi H, Jazayeri N. Brain tumor classification via convolutional neural network and extreme learning machines. In: 2018 8th International Conference on Computer and Knowledge Engineering (ICCKE); 2018. p. 314–319. IEEE.
21. Salih MS, Abdulazeez AM. A fusion-based deep approach for enhanced brain tumor classification. Journal of Soft Computing and Data Mining. 2024;5(1):183–193. DOI: 10.30880/jscdm.2024.05.01.015
22. Sarada B, Narasimha Reddy K, Babu R, Ramesh Babu BSSV, et al. Brain tumor classification using modified ResNet50V2 deep learning model. International Journal of Computing and Digital Systems. 2023;16(1):1–10.
23. Suryawanshi S, Patil SB. Efficient brain tumor classification with a hybrid CNN-SVM approach in MRI. Journal of Advances in Information Technology. 2024;15(3):340–354. https://doi.org/10.12720/jait.15.3.340-354.
24. Jun W, Liyuan Z. Brain tumor classification based on attention guided deep learning model. International Journal of Computational Intelligence Systems. 2022;15(1):35. https://doi.org/10.1007/s44196-022-00090-9.
25. Kang J, Ullah Z, Gwak J. MRI-based brain tumor classification using an ensemble of deep features and machine learning classifiers. Sensors. 2021;21(6):2222. https://doi.org/10.3390/s21062222.
26. Mahmud MI, Mamun M, Abdelgawad A. A deep analysis of brain tumor detection from MR images using deep learning networks. Algorithms. 2023;16(4):176. https://doi.org/10.3390/a16040176.
27. Bhuvaji S, Kadam A, Bhumkar P, Dedge S, Kanchan S. Brain tumor classification (MRI). Kaggle; 2020. Available at: https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri (Accessed: 2024-07-02).
28. Karim PJ, Mahmood SR, Sah M. Brain tumor classification using fine-tuning based deep transfer learning and support vector machine. International Journal of Computing and Digital Systems. 2023;13(1):83–96. http://dx.doi.org/10.12785/ijcds/130108
29. Chen T, Qin H, Li X, Wan W, Yan W. A non-intrusive load monitoring method based on feature fusion and SE-ResNet. Electronics. 2023;12(8):1909. https://doi.org/10.3390/electronics12081909.
30. He J, Jiang D. A fully automatic model based on SE-ResNet for bone age assessment. IEEE Access. 2021; 9:62460-62466.
31. Vasant Bidwe R, Mishra S, Kamini Bajaj S, Kotecha K. Attention-focused eye gaze analysis to predict autistic traits using transfer learning. International Journal of Computational Intelligence Systems. 2024;17(1):1–33. https://doi.org/10.1007/s44196-024-00491-y
32. Fagbola T, Igwebuike S. Leveraging pretrained models for multimodal medical image interpretation: An exhaustive experimental analysis. In: medRxiv; 2024. https://doi.org/10.1101/2024.08.09.24311762
33. Dhibar S. ResNet101 and DAE for Enhance Quality and Classification Accuracy in Skin Cancer Imaging. arXiv preprint arXiv:2403.14248. 2024.

34. Rahimzadeh M, Attar A. A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. Informatics in Medicine Unlocked. 2020; 19:100360. https://doi.org/10.1016/j.imu.2020.100360
35. Wu H, Xiao B, Codella N, et al. CVT: Introducing convolutions to vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. p. 22–31.
36. Chen X, Liang C, Huang D, et al. Evolved optimizer for vision. In: First Conference on Automated Machine Learning (Late-Breaking Workshop); 2022.