



A Comparative Analysis of Various Transfer Learning-Based Deep Learning Frameworks Using Brain Tumor MRI

Sunil Kumar Agarwal^{1,*} and Yogesh Kumar Gupta¹

¹Department of Computer Science, Banasthali Vidyapith, Rajasthan 304022, India

Abstract

Brain tumors are a critical and challenging health issue worldwide, and they often reduce patient survival if they are not identified at an early stage. This study presents an automatic approach to identify and classify brain tumor magnetic resonance imaging (MRI) scans using several pre-trained deep learning frameworks and compares their relative performance. Four popular frameworks, namely VGG16, Xception, ResNet50, and MobileNetV1, were fine-tuned with a transfer learning technique on 2870 brain tumor MRI training images to predict four classes (glioma, meningioma, pituitary tumor, and no tumor). A quantitative comparison among all applied frameworks was performed using accuracy, precision, recall, and F1-score. On the held-out test set, VGG16 achieved the highest accuracy (70.1%), followed by Xception (63.7%), MobileNetV1 (53.3%), and ResNet50 (50.8%). The results indicate that VGG16 and Xception outperform the other two frameworks under the same transfer

learning setting, while all models show low recall for the glioma class. The study provides a baseline comparison of pre-trained frameworks for brain tumor MRI classification and outlines directions for future research.

Keywords: Brain tumor MRI classification, transfer learning, pre-trained model, VGG16, Xception, ResNet50, MobileNetV1.

1 Introduction

Brain tumors represent a major global health burden, with tens of thousands of new cases diagnosed each year [1]. The brain monitors and controls various body functions, such as imagination, verbal communication, memory, and the movement of body parts. Uncontrolled growth of brain cells can change a person's personality and brain functions, and such unusual symptoms may indicate the presence of a brain tumor [2]. Brain tumors are a significant cause of cancer-related death worldwide, and timely identification is necessary for effective treatment. Brain tumors are generally divided into two categories: benign and malignant [3]. Benign tumors are non-cancerous, whereas malignant tumors are cancerous.

Benign tumors develop slowly. Meningioma is a



Submitted: 29 November 2025

Revised: 29 December 2025

Accepted: 02 January 2026

Published: 19 October 2026

Vol. 2, No. 4, 2026.

10.62762/TMI.2025.919091

*Corresponding author:

✉ Sunil Kumar Agarwal

sun.aipc@gmail.com

Citation

Agarwal, S. K., & Gupta, Y. K. (2026). A Comparative Analysis of Various Transfer Learning-Based Deep Learning Frameworks Using Brain Tumor MRI. *ICCK Transactions on Machine Intelligence*, 2(4), 200–212.

© 2026 ICCK (Institute of Central Computation and Knowledge)

common example of a mostly benign tumor and accounts for roughly one-third of primary brain tumors. In most cases, benign tumors are curable with surgery, but in some cases they can progress to premalignant and then malignant forms. Malignant tumors grow rapidly, are very dangerous, and can be fatal. The main aim of medical professionals is to detect tumors at an early stage, because early detection can improve the survival rate and reduce mortality. Magnetic resonance imaging (MRI) is one of the most popular and reliable techniques used in the diagnosis of brain tumors [4]. MRI provides high-resolution images of human organs that support treatment planning, and various brain abnormalities can be identified accurately using MRI. Owing to advances in medical diagnosis, automatic image analysis has become an essential part of many research studies. The objective of tumor diagnosis is to identify the tumor origin and to estimate its growth rate. Manual analysis of many medical images is time-consuming and can lead to incorrect results. With the progress of research on neural networks, medical professionals have increasingly applied artificial intelligence (AI)-based algorithms in the tumor diagnosis process [5].

However, because of limited data and their dependence on handcrafted features, traditional machine learning techniques can be time-consuming and provide lower accuracy than deep learning methods [6]. In contrast, deep learning algorithms learn features automatically and provide solutions to complex image-related problems [7]. Although deep learning methods are effective, they require large datasets and suffer from problems such as vanishing gradients. Transfer learning-based deep learning models can provide better accuracy even with small datasets [8], and they can perform image identification and classification tasks more reliably. Recently, many transfer learning models that focus on the concept of “knowledge transfer” have been proposed, and transfer learning-based deep convolutional neural network (CNN) models are becoming increasingly popular for problems in the medical sector.

Many researchers have applied AI and deep learning techniques to brain tumor-related problems, but a number of existing methods face common challenges including limited dataset diversity, insufficient evaluation metrics, and suboptimal performance on certain tumor classes [9]. In this work, we review transfer learning-based studies and use four popular frameworks, VGG16, Xception, ResNet50,

and MobileNetV1, to identify and classify brain tumor classes [10]. These deep learning models have already been tested on popular benchmarks and produced excellent results. The main aim of this study is to investigate and compare the efficiency of these frameworks using transfer learning with a fine-tuning process, and to verify the results with several performance metrics. Sample images of the tumor types are shown in Figure 1.

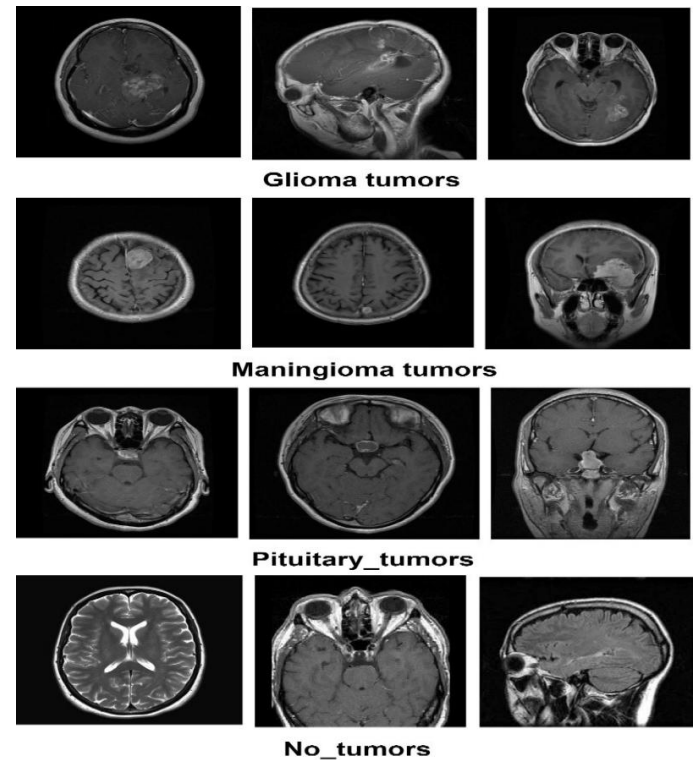


Figure 1. Sample images of the four classes in the dataset.

The main contributions of this work are as follows:

- The performance of four pre-trained deep learning frameworks is evaluated on the same dataset.
- Data augmentation is applied to increase the number of training images and to reduce overfitting.
- Regularization and fine-tuning are applied to improve the accuracy of the frameworks.
- Several metrics are employed to analyze the accuracy and errors of the transfer learning-based models.
- A comparative analysis shows the relative effectiveness of the pre-trained models.

The remainder of this paper is organized as follows. Section 2 reviews previous related studies and their limitations. Section 3 explains the methodology. Section 4 provides the experimental details. Section 5 presents the experimental results and their analysis.

Section 6 concludes the paper with its limitations and future directions.

2 Literature Review

Many researchers have proposed solutions to identify and classify tumors using AI techniques. However, some of these solutions are complex and do not provide high accuracy. In this scenario, a well-trained transfer learning-based deep neural network can play a significant role.

2.1 Classification of Brain Tumor MRI Images Using CNN

Recently, AI frameworks have been widely used to solve MRI classification problems. Among them, the CNN is a very effective and widely used deep learning framework for computer vision problems. However, a CNN does not perform well with limited datasets and requires an extended training period to extract deep features.

Basheera and Satya Sai Ram [11] proposed a brain tumor classification method using a CNN. The method used an improved independent component analysis (ICA) mixture model for tumor segmentation and provided deep feature extraction and performance validation. The study has limitations such as dataset dependency, model complexity, segmentation challenges, and a limited number of tumor types.

Sultan et al. [12] developed a 16-layer CNN framework to categorize images into tumor classes. The framework achieved accuracy rates of 96.13% and 98.7% on two datasets. The method highlights the importance of AI techniques but suffers from limited dataset diversity.

Singh et al. [13] proposed a framework to classify images from the BRATS 2017 dataset, using 320 images for training and 80 images for testing. The framework achieved 89.25% training accuracy and 97.5% validation accuracy. However, the work has limitations, including a simplistic design, a small dataset, a lack of preprocessing details, and insufficient performance metrics.

Abiwinanda et al. [14] proposed a CNN framework to classify three types of brain tumor (glioma, meningioma, and pituitary) from 3064 MRI images. Without any region-based segmentation preprocessing, the model achieved a training accuracy of 98.51% and a validation accuracy of 84.19%. However, the study was limited to three tumor classes

and did not include evaluation on an independent held-out test set.

Minarno et al. [15] conducted an experimental study on brain tumor MRI images with a deep learning framework using hyperparameter tuning. The experiment used 3264 tumor images and achieved a peak accuracy of 96% among three scenarios. However, the study provides limited information on the hyperparameters adjusted, the tuning process, and the computational resources. Mukkapati and Anbarasi [16] introduced an enhanced CNN framework to identify and classify brain abnormalities. The model integrates techniques for segmentation, pattern analysis, and classification, and it achieved 96.85% accuracy on a benchmark dataset. The study shows that the model can effectively segment and classify brain tumors by keeping local and contextual information. However, the paper lacks detailed information on the composition and size of the dataset, the training parameters, and the computational resources.

2.2 Classification of Brain Tumor MRI Images Using Transfer Learning

Deep neural networks, especially CNNs, demand a huge amount of training data, which is not always available. Deep transfer learning largely overcomes this issue: it reduces the computational cost of training and provides better results even on small datasets. Therefore, researchers are now exploring transfer learning techniques to identify and classify brain tumor abnormalities.

Lo et al. [17] developed an automatic glioma grading system based on a pre-trained AlexNet with transfer learning. With data augmentation, the model achieved 97.9% accuracy, compared with 59.85% without augmentation. However, the study is limited by a small dataset, its focus on a specific tumor type, and a lack of external validation.

Rehman et al. [18] proposed a framework using transfer learning to categorize tumor images. The framework explored three pre-trained architectures, VGG16, AlexNet, and GoogLeNet, and the comparative investigation identified the most efficient one. Data augmentation was used to increase the number of images and to reduce overfitting. The fine-tuned VGG16 achieved 98.69% classification accuracy. However, the framework suffers from limitations such as dataset dependency, model complexity, overfitting risks, and a limited number of

tumor types.

Chelghoum et al. [19] proposed an automated system for classifying three types of brain tumor images using contrast-enhanced MRI and transfer learning with CNN architectures. The approach achieved 98.71% classification accuracy and showed that transfer learning is effective with few training samples.

Arbane et al. [20] suggested a deep learning architecture using transfer learning to categorize MRI images. The research investigated pre-trained architectures such as ResNet, Xception, and MobileNetV2, among which MobileNetV2 achieved the best performance. However, the research relies on the quality and diversity of the MRI dataset and is prone to overfitting because of limited data.

Prakash et al. [21] suggested a deep neural architecture based on DenseNet121 with transfer learning to improve classification accuracy. After hyperparameter tuning, the model achieved 97.39% classification accuracy. The main limitation reported was overfitting.

Arya and Agrawal [22] suggested a deep CNN method using transfer learning to categorize tumor images. The work evaluated several frameworks and showed that fine-tuned CNN models achieved high accuracy, with a precision of 96.65%, outperforming traditional classification techniques. However, the work needs to be validated on larger and more diverse datasets.

The above studies indicate that transfer learning is more effective than other techniques, since it can optimize, extract, and learn features with a large number of parameters. Although several studies reported outstanding performance, many of them share common challenges, such as overfitting, time complexity, and low precision for some classes. In addition, some studies were performed on imbalanced data, and a few frameworks only support binary classification. In this work, we therefore carry out a comparative analysis of several popular pre-trained deep learning frameworks under the same transfer learning setting and the same dataset.

3 Materials and Methods

3.1 Dataset

The dataset used in this research was taken from the Kaggle online data repository. It consists of 3264 MRI images classified into four categories: glioma, meningioma, pituitary tumor, and no tumor [23]. Figure 1 shows sample images of all four categories. The data are divided into a training set (2870 images)

and a test set (394 images). Figure 2 shows the percentage of each class in the two sets. During training, 20% of the training images were held out as a validation set to monitor learning progress and detect overfitting.

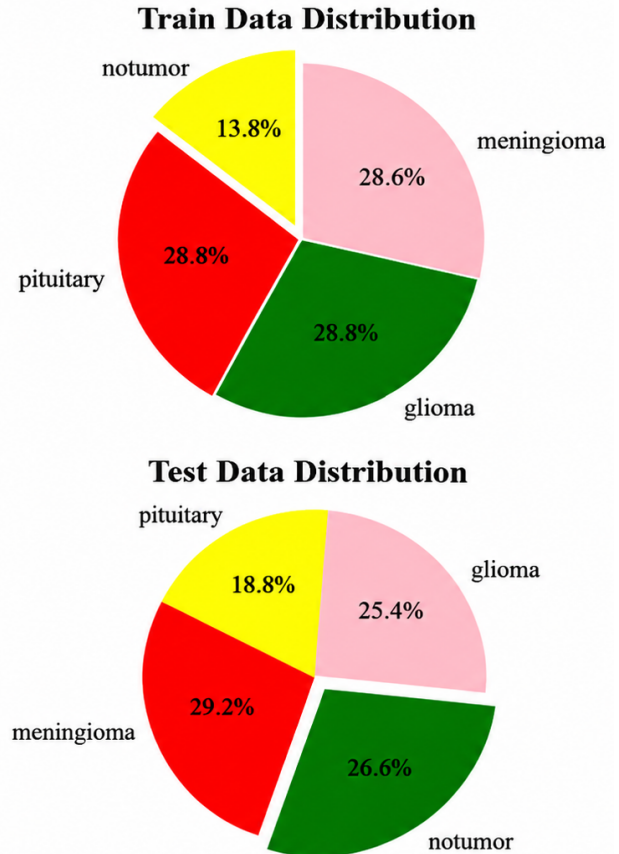


Figure 2. Class distribution of the training and test sets.

3.2 Image Preprocessing

Preprocessing is used to enhance image quality and to standardize the input data. In this experiment, two preprocessing methods were used: image rescaling and data augmentation. Pixel values were rescaled to the range $[0, 1]$ to ensure uniformity, and all images were resized to 128×128 pixels, which reduces the computational cost of training. Data augmentation consists of transformations such as rotation, flipping, zooming, and shifting, which increase the number and diversity of training images [24].

3.3 Transfer Learning

Transfer learning reuses previously acquired knowledge and requires less data for training [25]. In this work, four pre-trained deep learning frameworks were used; they were selected because of their strong performance on the ImageNet benchmark dataset. Figure 3 illustrates the transfer learning technique.

For all four architectures, the convolutional base was initialized with ImageNet weights; the top layers were replaced with a global average pooling layer, a dense layer, and a softmax output layer for four-class classification. The base layers were initially frozen and the newly added layers were trained first; the entire network was then fine-tuned end-to-end [31].

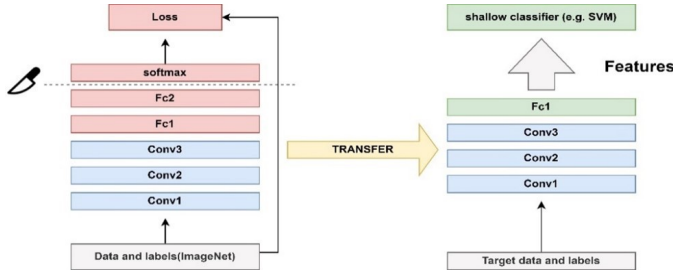


Figure 3. Transfer learning technique.

3.4 VGG16 Pre-Trained Architecture

VGG16 is a deep CNN with 16 weight layers [26]. It is built from multiple convolution layers with 3×3 filters and a stride of 1, followed by max-pooling layers with 2×2 filters and a stride of 2. The architecture repeats this combination of convolution and max-pooling layers, and a softmax function is used for the output. In this work, part of the pre-trained convolutional base was frozen and the remaining layers were fine-tuned, following established practices for adapting pre-trained CNNs to medical image analysis tasks. A short description of VGG16 is presented in Figure 4.

3.5 ResNet50 Pre-Trained Architecture

ResNet50 is a powerful deep CNN framework introduced by He et al. [27] in 2016. The framework consists of 50 layers: 48 convolution layers, one max-pooling layer, and one average-pooling layer. To mitigate the vanishing gradient problem during training, the model uses skip connections. It achieves 92.1% top-5 accuracy on the ImageNet dataset. A short description of the model is depicted in Figure 5.

3.6 Xception Pre-Trained Architecture

The Xception architecture was introduced by Google in 2017 (Figure 6) [28]. It employs depthwise separable convolutions, each of which is decomposed into a depthwise convolution followed by a pointwise (1×1) convolution. This design reduces the number of parameters while maintaining performance. Xception contains 36 convolutional layers organized into 14 modules, and it achieves 94.5% top-5 accuracy on the ImageNet dataset.

3.7 MobileNet Pre-Trained Architecture

MobileNets were introduced by Google in 2017 to meet the increasing demands of smartphones and other handheld devices [29]. They are designed to trade a small loss in accuracy for a large reduction in the number of parameters and in computational cost. They are available in different versions, namely V1, V2, and V3. MobileNetV1 (Figure 7) achieves 89.5% top-5 accuracy on the ImageNet benchmark dataset.

4 Experiments

4.1 Experimental Setup and Implementation

All experiments were implemented in Python using the Keras library with the TensorFlow framework [30], together with the NumPy and Matplotlib libraries. Training and testing were carried out on the Google Colab cloud platform, which was accessed from a local workstation running Windows with an Intel processor, 8 GB of RAM, and a 500 GB hard disk. The experiments were performed with the VGG16, ResNet50, Xception, and MobileNetV1 pre-trained architectures, and each framework was trained on the training set and evaluated on the separate test set.

4.2 Evaluation of Performance Metrics

To evaluate the performance of the models, we used accuracy, precision, recall, and F1-score [32]. These metrics were computed from the confusion matrix, which tabulates the actual classes against the predicted classes. Accuracy is defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}. \quad (1)$$

Precision is defined as

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (2)$$

Recall is defined as

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (3)$$

The F1-score is the harmonic mean of precision and recall:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (4)$$

In Equations (1)–(4), TP , TN , FP , and FN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively. For the four-class problem, these quantities are computed for

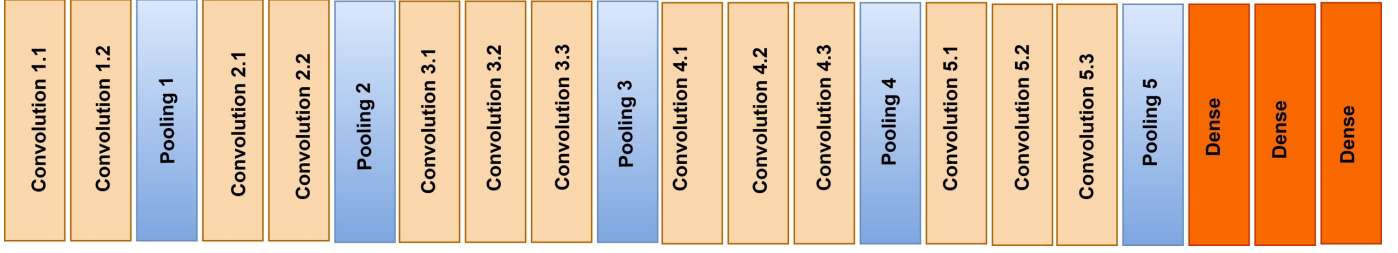


Figure 4. VGG16 architecture.

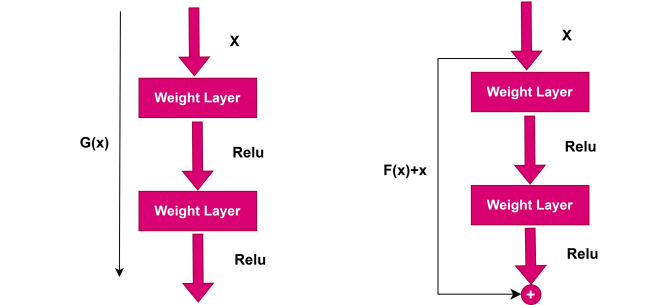


Figure 5. A plain CNN block (left) and a ResNet residual block (right).

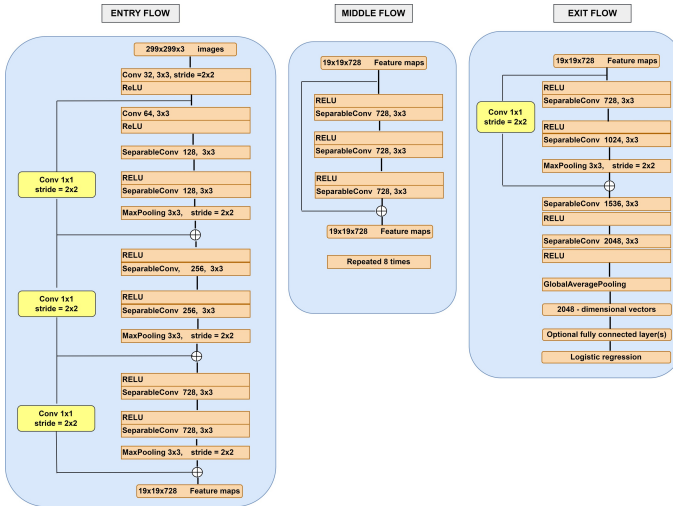


Figure 6. Xception pre-trained architecture.

each class in a one-vs-rest manner; the overall accuracy is the fraction of correctly classified test images, and the overall precision and recall are macro-averaged over the four classes. The lower the numbers of false negatives and false positives, the more reliable the model.

5 Results and Discussion

All pre-trained models were trained with a batch size of 64 and the Adam optimizer [35] with a fixed learning rate of 10^{-3} . The number of epochs differed between the models (30 for VGG16, 10 for Xception, 10 for MobileNetV1, and 50 for ResNet50). The results of each framework are presented as accuracy and loss

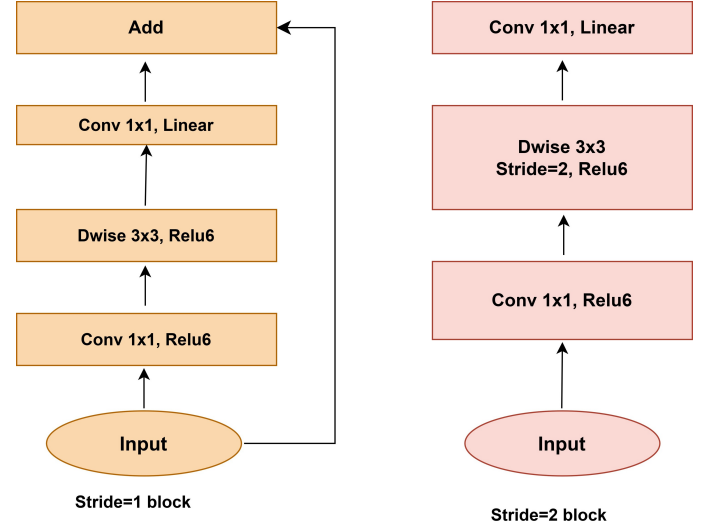


Figure 7. MobileNetV1 architecture.

curves and as a confusion matrix on the test set, and the performance of all frameworks is then compared quantitatively and graphically.

5.1 VGG16

The trainable and non-trainable parameters of the VGG16-based model and the model description are presented in Figure 8. The accuracy and loss curves and the confusion matrix [33] are shown in Figure 9.

The gap between the training and validation loss is small [34], and the model reaches a training accuracy of approximately 91–92% and a validation accuracy of approximately 89–90%. On the held-out test set, however, the accuracy is 70.1% (276 of 394 images), which indicates that the model does not generalize to the test set as well as the validation curves suggest. A detailed presentation of the performance metrics is given in Table 1. The model performs well on the meningioma and no-tumor classes, whereas the recall of the glioma class is only 23.00%: 60 of the 100 glioma test images are misclassified as meningioma.

5.2 Xception

The trainable and non-trainable parameters and the model description are presented in Figure 10. The

Layer (type)	Output Shape	Param #	Remark
vgg16 (Functional)	(None, 4, 4, 512)	14,714,688	Pre-trained VGG16 (feature extractor)
flatten (Flatten)	(None, 8192)	0	-
dropout (Dropout)	(None, 8192)	0	Dropout ($p = 0.5$)
dense (Dense)	(None, 128)	1,048,704	ReLU
dropout_1 (Dropout)	(None, 128)	0	Dropout ($p = 0.5$)
dense_1 (Dense)	(None, 4)	516	Softmax (4 classes)
Total params	15,763,908 (60.13 MB)		
Trainable params	8,128,644 (31.01 MB)		
Non-trainable params	7,635,264 (29.13 MB)		

Functional (VGG16)	
Input shape: (None, 128, 128, 3)	Output shape: (None, 4, 4, 512)
↓	
Flatten	
Input shape: (None, 4, 4, 512)	Output shape: (None, 8192)
↓	
Dropout ($p = 0.5$)	
Input shape: (None, 8192)	Output shape: (None, 8192)
↓	
Dense (128, ReLU)	
Input shape: (None, 8192)	Output shape: (None, 128)
↓	
Dropout ($p = 0.5$)	
Input shape: (None, 128)	Output shape: (None, 128)
↓	
Dense (4, Softmax)	
Input shape: (None, 128)	Output shape: (None, 4)

Figure 8. Trainable and non-trainable parameters and model architecture of the VGG16-based model.

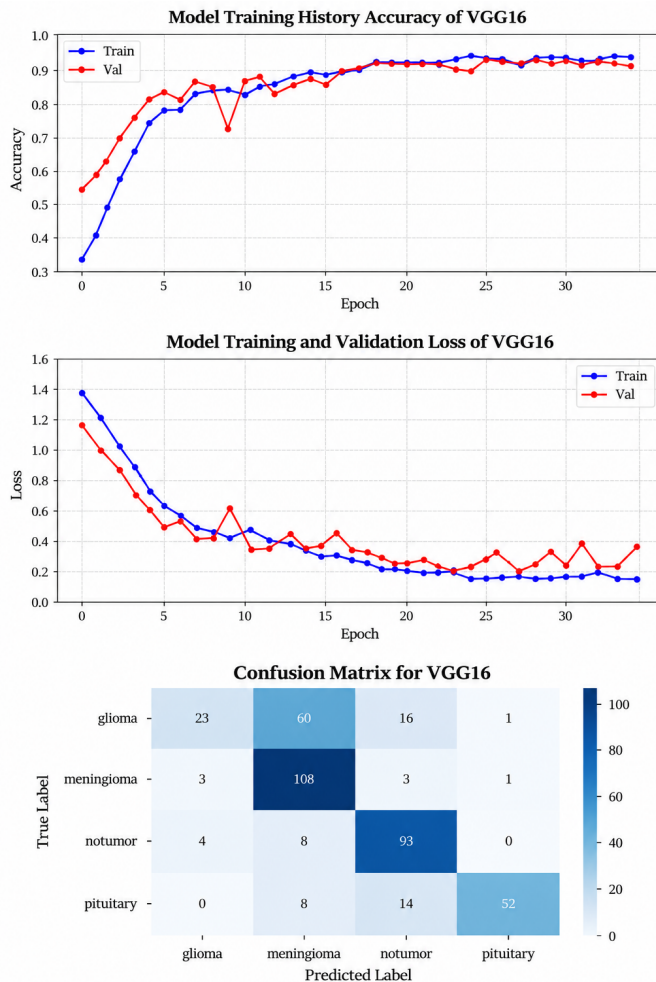


Figure 9. Accuracy and loss curves and confusion matrix of the VGG16-based model.

Table 1. Performance metrics of the VGG16-based model.

Class	Precision	Recall	F1-score
Glioma	76.67%	23.00%	35.38%
Meningioma	58.70%	93.91%	72.24%
No tumor	73.81%	88.57%	80.52%
Pituitary	96.30%	70.27%	81.25%

accuracy and loss curves obtained during training and the confusion matrix of the Xception-based model are shown in Figure 11.

The curves indicate that the Xception model learns effectively, as the accuracy increases and the loss decreases over the epochs. The final training accuracy is approximately 88.61%, and the final validation accuracy is approximately 84.92%. Table 2 gives the performance metrics of the Xception-based model on the test set.

Table 2. Performance metrics of the Xception-based model.

Class	Precision	Recall	F1-score
Glioma	74.29%	26.00%	38.52%
Meningioma	61.31%	73.04%	66.67%
No tumor	63.58%	91.43%	75.00%
Pituitary	63.38%	60.81%	62.07%

5.3 MobileNetV1

The trainable and non-trainable parameters and the model description of the MobileNetV1-based model are presented in Figure 12. The accuracy and

Layer (type)	Output Shape	Param #	Remark
xception (Functional)	(None, 4, 4, 2048)	20,861,480	Pre-trained Xception (feature extractor)
flatten_1 (Flatten)	(None, 32768)	0	-
dropout_2 (Dropout)	(None, 32768)	0	Dropout ($p = 0.5$)
dense_2 (Dense)	(None, 128)	4,194,432	ReLU
dropout_3 (Dropout)	(None, 128)	0	Dropout ($p = 0.5$)
dense_3 (Dense)	(None, 4)	516	Softmax (4 classes)
Total params	25,056,428 (95.58 MB)		
Trainable params	7,358,596 (28.07 MB)		
Non-trainable params	17,697,832 (67.51 MB)		

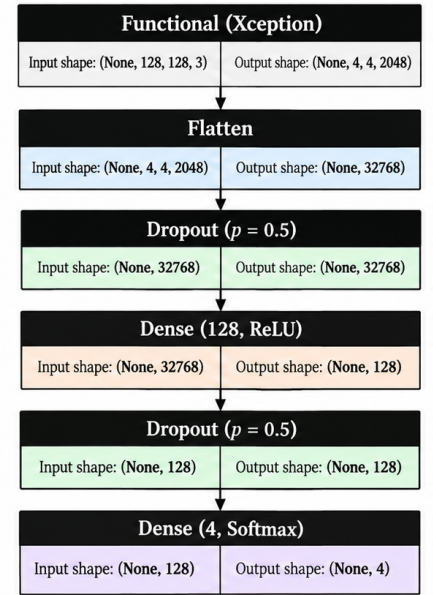


Figure 10. Trainable and non-trainable parameters and model architecture of the Xception-based model.

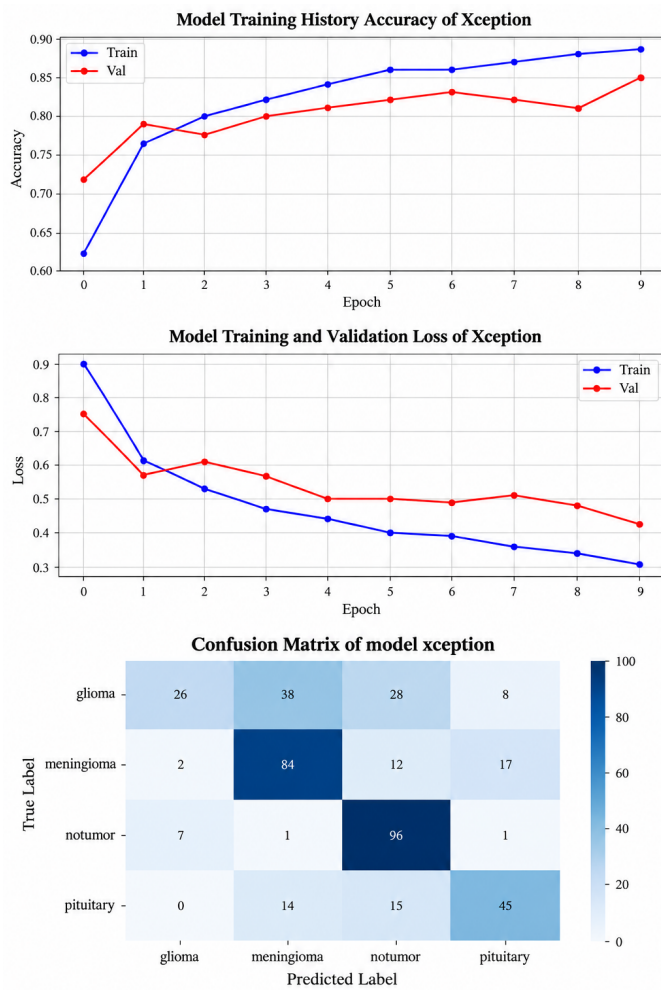


Figure 11. Accuracy and loss curves and confusion matrix of the Xception-based model.

loss curves and the confusion matrix are shown in Figure 13.

The results of the MobileNetV1-based model show the following:

- The training and validation accuracy curves are closely aligned, which suggests that the model does not overfit the training data within 10 epochs.
- The loss curves decrease steadily, indicating effective learning.
- The validation accuracy and loss are slightly better than the training values, which is commonly observed when data augmentation and dropout are active only during training.

The final training accuracy is approximately 76%, and the final validation accuracy is approximately 77%; the test accuracy, however, is only 53.3%. Table 3 summarizes the performance metrics of the MobileNetV1-based model.

Table 3. Performance metrics of the MobileNetV1-based model.

Class	Precision	Recall	F1-score
Glioma	64.52%	20.00%	30.53%
Meningioma	54.95%	43.48%	48.54%
No tumor	55.70%	79.05%	65.35%
Pituitary	46.34%	77.03%	57.87%

Layer (type)	Output Shape	Param #	Remark
mobilenet_1.00_128 (Functional)	(None, 4, 4, 1024)	3,228,864	Pre-trained MobileNetV2 (feature extractor)
flatten_2 (Flatten)	(None, 16384)	0	-
dropout_4 (Dropout)	(None, 16384)	0	Dropout ($p = 0.5$)
dense_4 (Dense)	(None, 128)	2,097,280	ReLU
dropout_5 (Dropout)	(None, 128)	0	Dropout ($p = 0.5$)
dense_5 (Dense)	(None, 4)	516	Softmax (4 classes)
Total params	5,326,660 (20.32 MB)		
Trainable params	3,148,420 (12.01 MB)		
Non-trainable params	2,178,240 (8.31 MB)		

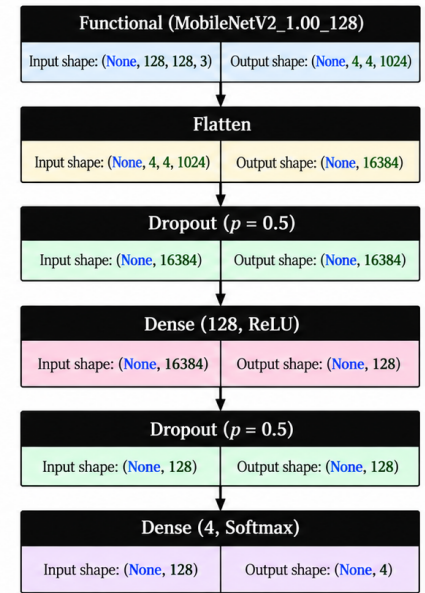


Figure 12. Trainable and non-trainable parameters and model architecture of the MobileNetV1-based model.

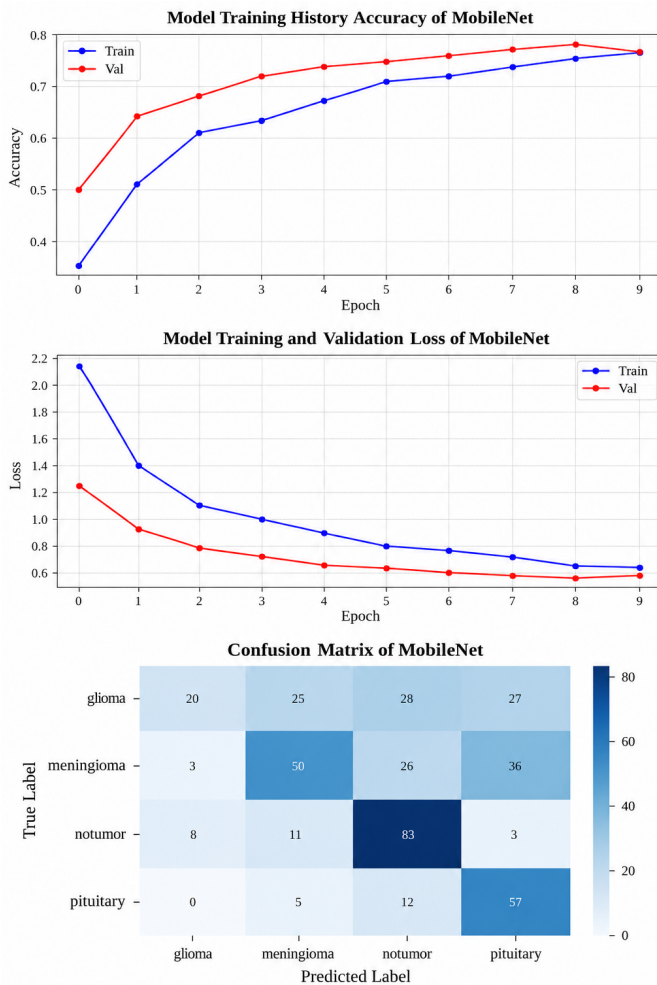


Figure 13. Accuracy and loss curves and confusion matrix of the MobileNetV1-based model.

5.4 ResNet50

The trainable and non-trainable parameters and the model description of the ResNet50-based model are presented in Figure 14. The accuracy and loss curves and the confusion matrix are shown in Figure 15.

The model achieved 66.87% training accuracy and 75.17% validation accuracy after 50 epochs, the lowest values among the four models, and its test accuracy is only 50.8%. Table 4 shows the performance metrics of the ResNet50-based model.

Table 4. Performance metrics of the ResNet50-based model.

Class	Precision	Recall	F1-score
Glioma	48.65%	18.00%	26.28%
Meningioma	59.18%	50.43%	54.46%
No tumor	52.76%	81.90%	64.18%
Pituitary	39.58%	51.35%	44.71%

5.5 Comparative Analysis

Figure 16 presents the confusion matrices of the four pre-trained deep learning models side by side.

According to Figure 16, VGG16 achieves the best accuracy and overall performance, which makes it the most reliable of the four models. Xception ranks second, with a precision and recall about six to eleven percentage points lower than those of VGG16. MobileNetV1 ranks third, and ResNet50 has the lowest accuracy. All four models misclassify most glioma images (recall between 18% and 26%), mainly as

Layer (type)	Input Shape	Output Shape	Param #	Remark
resnet50 (Functional)	(None, 128, 128, 3)	(None, 4, 4, 2048)	23,587,712	Pre-trained ResNet50 (feature extractor)
flatten_3 (Flatten)	(None, 4, 4, 2048)	(None, 32768)	0	-
dropout_6 (Dropout)	(None, 32768)	(None, 32768)	0	Dropout (p = 0.5)
dense_6 (Dense)	(None, 32768)	(None, 128)	4,194,432	ReLU
dropout_7 (Dropout)	(None, 128)	(None, 128)	0	Dropout (p = 0.5)
dense_7 (Dense)	(None, 128)	(None, 4)	516	Softmax (4 classes)
Total params	27,782,660 (105.98 MB)			
Trainable params	5,249,668 (20.03 MB)			
Non-trainable params	22,532,992 (85.96 MB)			

Functional (ResNet50)	
Input shape: (None, 128, 128, 3)	Output shape: (None, 4, 4, 2048)
Flatten	
Input shape: (None, 4, 4, 2048)	Output shape: (None, 32768)
Dropout (p = 0.5)	
Input shape: (None, 32768)	Output shape: (None, 32768)
Dense (128, ReLU)	
Input shape: (None, 32768)	Output shape: (None, 128)
Dropout (p = 0.5)	
Input shape: (None, 128)	Output shape: (None, 128)
Dense (4, Softmax)	
Input shape: (None, 128)	Output shape: (None, 4)

Figure 14. Trainable and non-trainable parameters and model architecture of the ResNet50-based model.

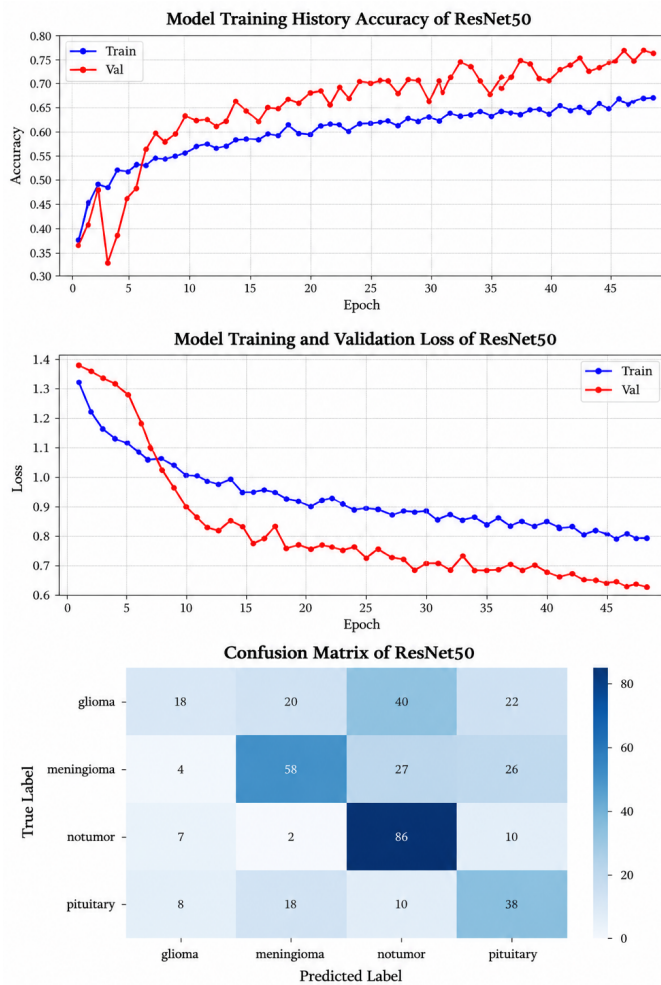


Figure 15. Accuracy and loss curves and confusion matrix of the ResNet50-based model.

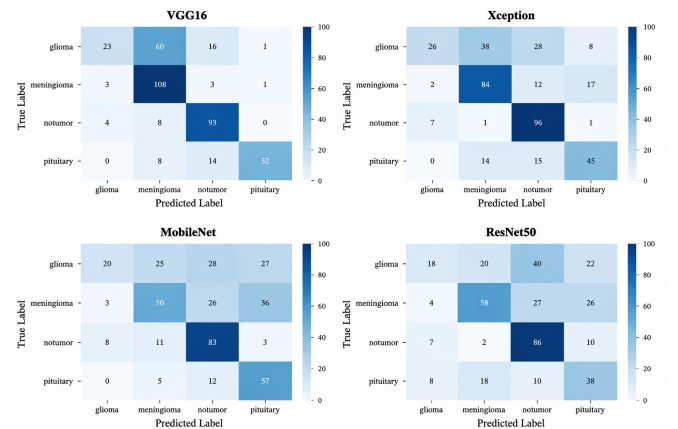


Figure 16. Confusion matrices of the four pre-trained models on the test set.

meningioma or no tumor. Table 5 and Figure 17 show the comparative performance metrics of the models.

Table 5. Comparative analysis of the pre-trained models on the test set (macro-averaged precision and recall).

Model	Accuracy	Precision	Recall
VGG16	0.701	0.764	0.689
Xception	0.637	0.656	0.628
MobileNetV1	0.533	0.554	0.549
ResNet50	0.508	0.500	0.504

Table 6 places the present results in the context of representative studies from the literature. The lower accuracies obtained in this work are attributable to the use of a stricter evaluation protocol—a fully held-out test set that was never seen during training

or validation—whereas many prior studies report cross-validation or validation-set accuracy.

Table 6. Comparison of classification accuracy with selected studies in the literature.

Study	Model	Dataset	Accuracy
Rehman et al. [18]	VGG16	Figshare	98.69%
Chelghoum et al. [19]	CNN (TL)	CE-MRI	98.71%
Prakash et al. [21]	DenseNet121	MRI	97.39%
Minarno et al. [15]	CNN	Kaggle	96.00%
Arbane et al. [20]	MobileNetV2	BRATS	98.24%
This work	VGG16	Kaggle	70.10%
This work	Xception	Kaggle	63.70%
This work	MobileNetV1	Kaggle	53.30%
This work	ResNet50	Kaggle	50.80%

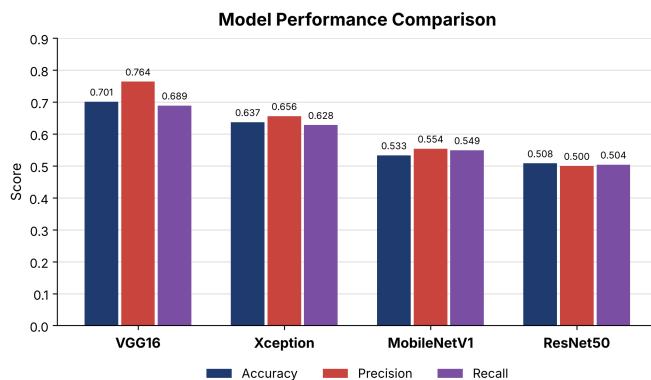


Figure 17. Comparison of the accuracy, precision, and recall of the pre-trained models.

The performance of all pre-trained models is compared in Table 5. It can be observed that the VGG16 and Xception models perform better than the other two models. For VGG16, Xception, and MobileNetV1, the macro-averaged precision exceeds the macro-averaged recall, whereas for ResNet50 the two values are almost equal.

6 Conclusion and Future Scope

This study evaluated four pre-trained frameworks with transfer learning on a public brain tumor MRI dataset using several performance metrics. VGG16 achieved the highest test accuracy (70.1%), followed by Xception (63.7%), MobileNetV1 (53.3%), and ResNet50 (50.8%). These accuracies are lower than those reported in the literature, and there is a clear gap between the validation and test accuracy of every model. Many misclassifications were observed, especially for the glioma class, which are likely caused by the imbalanced dataset, inter-class similarities, intra-class diversity, and differences between the training and test distributions.

In future work, more diverse tumor datasets and additional data augmentation methods should be used to overcome these dataset-related issues. Moreover, the impact of hyperparameter tuning, the number of training epochs, and the choice of loss function was not investigated in this work and should be explored in the future.

Data Availability Statement

The MRI dataset analyzed in this study is publicly available at <https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>. The code and trained models are available from the corresponding author upon reasonable request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable. This study involved secondary analysis of a publicly available, de-identified MRI dataset and did not involve the recruitment of human participants, collection of identifiable personal information, or animal experiments.

References

- [1] Ostrom, Q. T., Cioffi, G., Waite, K., Kruchko, C., & Barnholtz-Sloan, J. S. (2021). CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2014–2018. *Neuro-oncology*, 23(Supplement_3), iii1-iii105. [CrossRef]
- [2] Lapointe, S., Perry, A., & Butowski, N. A. (2018). Primary brain tumours in adults. *The Lancet*, 392(10145), 432-446. [CrossRef]
- [3] Biratu, E. S., Schwenker, F., Ayano, Y. M., & Debelee, T. G. (2021). A survey of brain tumor segmentation and classification algorithms. *Journal of Imaging*, 7(9), 179. [CrossRef]
- [4] Amin, J., Sharif, M., Yasmin, M., & Fernandes, S. L. (2020). A distinctive approach in brain tumor

- detection and classification using MRI. *Pattern Recognition Letters*, 139, 118-127. [CrossRef]
- [5] Iqbal, M. J., Javed, Z., Sadia, H., Qureshi, I. A., Irshad, A., Ahmed, R., ... & Sharifi-Rad, J. (2021). Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. *Cancer Cell International*, 21(1), 270. [CrossRef]
- [6] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1), 53. [CrossRef]
- [7] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60-88. [CrossRef]
- [8] Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., & Azim, M. A. (2022). Transfer learning: a friendly introduction. *Journal of big data*, 9(1), 102. [CrossRef]
- [9] Xie, Y., Zaccagna, F., Rundo, L., Testa, C., Agati, R., Lodi, R., ... & Tonon, C. (2022). Convolutional neural network techniques for brain tumor classification (from 2015 to 2022): Review, challenges, and future perspectives. *Diagnostics*, 12(8), 1850. [CrossRef]
- [10] Babu Vimala, B., Srinivasan, S., Mathivanan, S. K., Mahalakshmi, Jayagopal, P., & Dalu, G. T. (2023). Detection and classification of brain tumor using hybrid deep learning models. *Scientific reports*, 13(1), 23029. [CrossRef]
- [11] Basheera, S., & Satya Sai Ram, M. (2019, March). Classification of brain tumors using deep features extracted using CNN. In *Journal of Physics: Conference Series* (Vol. 1172, No. 1, p. 012016). IOP Publishing. [CrossRef]
- [12] Sultan, H. H., Salem, N. M., & Al-Atabany, W. (2019). Multi-classification of brain tumor images using deep neural network. *IEEE Access*, 7, 69215-69225. [CrossRef]
- [13] Singh, A., Deshmukh, R., Jha, R., Shahare, N., & Verma, S. (2020). Brain tumor classification using convolutional neural network. *International Journal of Advance Research and Innovative Ideas in Education*, 6(2), 1353-1357. <https://scispace.com/papers/brain-tumor-classification-using-convolutional-neural-efzz4asxdy>
- [14] Abiwinanda, N., Hanif, M., Hesaputra, S. T., Handayani, A., & Mengko, T. R. (2018, May). Brain tumor classification using convolutional neural network. In *World Congress on Medical Physics and Biomedical Engineering 2018: June 3–8, 2018, Prague, Czech Republic* (Vol. 1) (pp. 183-189). Singapore: Springer Nature Singapore. [CrossRef]
- [15] Minarno, A. E., Cokro Mandiri, M. H., Munarko, Y., & Hariyady, H. (2021). Convolutional neural network with hyperparameter tuning for brain tumor classification. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 6(2). [CrossRef]
- [16] Mukkapati, N., & Anbarasi, M. S. (2022). Brain tumor classification based on enhanced CNN model. *Revue d'Intelligence Artificielle*, 36(1), 125-130. [CrossRef]
- [17] Lo, C. M., Chen, Y. C., Weng, R. C., & Hsieh, K. L. C. (2019). Intelligent glioma grading based on deep transfer learning of MRI radiomic features. *Applied Sciences*, 9(22), 4926. [CrossRef]
- [18] Rehman, A., Naz, S., Razzak, M. I., Akram, F., & Imran, M. (2020). A deep learning-based framework for automatic brain tumors classification using transfer learning. *Circuits, Systems, and Signal Processing*, 39(2), 757-775. [CrossRef]
- [19] Chelghoum, R., Ikhlef, A., Hameurlaine, A., & Jacquir, S. (2020). Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images. In *Artificial Intelligence Applications and Innovations (AIAI 2020)*, IFIP Advances in Information and Communication Technology (Vol. 583, pp. 189-200). Springer, Cham. [CrossRef]
- [20] Arbane, M., Benlamri, R., Brik, Y., & Djerioui, M. (2021). Transfer learning for automatic brain tumor classification using MRI images. In *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)* (pp. 210-214). IEEE. [CrossRef]
- [21] Prakash, R. M., Kumari, R. S. S., Valarmathi, K., & Ramalakshmi, K. (2023). Classification of brain tumours from MR images with an enhanced deep learning approach using densely connected convolutional network. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11(2), 266-277. [CrossRef]
- [22] Arya, M. K., & Agrawal, R. (2023). Brain tumor classification using deep CNN-based transfer learning approach. *International Journal of Biology and Biomedical Engineering*, 17, 1-7. [CrossRef]
- [23] Bhuvaji, S., Kadam, A., Bhumkar, P., Dedge, S., & Kanchan, S. (2020). Brain tumor classification (MRI) [Data set]. *Kaggle*. Retrieved from <https://doi.org/10.34740/KAGGLE/DSV/1183165>
- [24] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60. [CrossRef]
- [25] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., ... & He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1), 43-76. [CrossRef]
- [26] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. [CrossRef]
- [27] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). IEEE. [CrossRef]

- [28] Chollet, F. (2017, July). Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 1800-1807). IEEE. [CrossRef]
- [29] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. [CrossRef]
- [30] Chicho, B. T., & Sallow, A. B. (2021). A comprehensive survey of deep learning models based on Keras framework. *Journal of Soft Computing and Data Mining*, 2(2), 49-62. [CrossRef]
- [31] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning?. *IEEE transactions on medical imaging*, 35(5), 1299-1312. [CrossRef]
- [32] Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*. [CrossRef]
- [33] Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168-192. [CrossRef]
- [34] Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, 1168(2), 022022. [CrossRef]
- [35] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. [CrossRef]



Sunil Kumar Agarwal is a research scholar pursuing a Ph.D. degree at Banasthali Vidyapith, Rajasthan, India. His research interests include medical image processing, cloud computing, and machine learning. (Email: sun.aipc@gmail.com)



Yogesh Kumar Gupta received the Ph.D. degree in computer science and engineering from Banasthali Vidyapith, Rajasthan, India, in 2018. His research interests include big data analytics, medical image processing, cloud computing, and machine learning. (Email: gyogesh@banasthali.in)