Machines and Algorithms
http://www.knovell.org/mna

Research Article,

Deep Learning based Brain Tumor Identification in MRI Images: A Comparative Analysis of Advanced CNN Architecture

Nimra Adrees[1, *]

[1]Department of Computer Science, The Women University, Multan, 60000, Pakistan
[*]Corresponding Author: Nimra Adrees. Email: nimraadrees759@gmail.com
Received: 01 Oct 2025; Revised: 06 Nov 2025; Accepted: 02 Dec 2025; Published: 21 Dec 2025
AID: 004-03-000057

Abstract: The classification of brain tumor is a very important section of medical diagnosis because correct and early diagnosis can greatly enhance the outcome of the patient. To aid in automated and dependable diagnostic methods, this paper examines the efficacy of already trained deep learning model to classify brain magnetic resonance imaging (MRI) scans into four categories, namely, pituitary tumor, meningioma, glioma, and no tumor. This was done using a publicly available set of 7,023 brain MRI images. DenseNet121, ResNet50, Xception and MobileNet four advanced convolutional neural network (CNN) models were also fine-tuned with transfer learning, as well as image preprocessing and data augmentation methods. Transfer learning was used to facilitate effective model adaptation with a low computation complexity and improved classification. According to the experimental findings, DenseNet121 performed better as it obtained the highest classification accuracy of 98.47% and the highest F1 score of 98.47%. The assessed models presented a high level of generalization and steady performance under a variety of evaluation criteria, which also indicates their possible application in the clinical decision support systems. These encouraging findings, however, need additional performance to improve recall on individual tumor classes, as well as the interpretability of deep-learning predictions in clinical settings. On the whole, the results highlight the great potential of the deep learning and transfer learning methods in medical image analysis, which is an opportunity to establish a more reliable, scalable and efficient solution to diagnosis. The next phase of work will involve diversifying the datasets, enhancing the model explainability, and assessing the performance of AI-oriented healthcare systems in practice to make them more widely adopted.

Keywords: Convolutional Neural Network; Brain Tumor Diagnosis; Transfer Learning; Xception; ResNet50; DenseNet101

1. Introduction

Brain tumors may cause severe cognitive and physiological issues that may significantly harm the quality of life of a patient, as well as brain tumors are one of the most harmful neurological conditions. They can be in the brain (primary tumors) or be as a result of spread of other organs to the brain (secondary tumors). The most common primary malignancies that present a challenge in diagnosis include gliomas, meningiomas and pituitary tumors [1], [2]. Meningiomas and pituitary tumors are usually harmless, although they may become fatal in case they are not detected and managed in time [2], [3]. Gliomas, however, are generally aggressive and require to be treated in a very aggressive manner.

The main technique of imaging that is still used to diagnose brain tumors is magnetic resonance imaging (MRI) because it has a great soft-tissue contrast capability, and it is non-invasive [4]. Nevertheless, interpretation of MRI scans manually remains tedious, highly radiologically skilled, and remains susceptible to diagnostic delays and inter-observer variability [5]. These limitations indicate the urgency of reliable and computer-aided diagnostic tools that are automated.

Deep learning, more specifically Convolutional Neural Networks (CNNs) has proved to be an impressive and powerful tool when it comes to analysis of medical images since it extracts discriminative features in visual images automatically. Transfer learning enhances accuracy and reduces computing costs and the need to large labelled medical datasets by using CNN models that have been trained on large publicly available datasets (such as ImageNet) [6], [7]. Nevertheless, the heterogeneity of tumors, lack of inter-class separability and variation caused by MRI acquisition methods render fully automated multi-class brain tumor classification challenging [8].

Many existing studies do not offer a study that is clinically useful due to the fact that they are limited to investigating a specific deep learning architecture or perform binary classification (tumor vs. no-tumor). It is still quite evident that the rigorous benchmarking of a few of the high-performing CNN architectures with the same dataset of multi-classes brain tumor MRI at the same conditions of experiment has not been done yet [9].

To seal this divide, this research paper assesses in detail four common CNN models to classify brain tumors with four classes based on MRI images using DenseNet121, ResNet50, Xception, and MobileNet. These models were selected because they had successful architecture, demonstrated performance in medical image tasks and ability to be utilized in a clinical environment with limited resources. The models that we choose give a fair combination of accuracy of prediction, cost of computation, and speed of inference and are therefore more viable in clinical use, although there exist more recent architectures (including EfficientNet and InceptionV3).

More so, MRI images were down-sampled to 128 x 128 pixels to lower training time and computational complexity with no significant effects on the structural integrity of tumor features. This enabled faster inferences and easier implementation on traditional clinical hardware. To train and test, we used a publicly available four-class brain MRI scan-based dataset also containing images of three types i.e., 1) gliomas, 2) meningioma, 3) pituitary tumors as well as no tumors.

The key findings of this work are:

- Transfer learning was applied to 4 states of the art CNN networks to create an automatic multi-class brain tumor classification model.
- A critical performance comparison (accuracy, precise, recall, and F1-score) of the models under the same experimental conditions is used to identify the most appropriate model in clinical decision assistance.
- The clinical feasibility of an automated brain tumor categorization system is shown and this could assist radiologists in coming up with earlier and more accurate diagnosis.

2. Related Work

The deep learning and the hybrid machine learning approaches have contributed significantly to the rapid development of automated brain tumor identification and classification with MRI. In recent studies, much effort has been put in enhancing classification performance, tolerance to noisy, heterogeneous data, processing speed and applicability in the clinic by using developed neural network architectures, transfer learning, and complex preparation methods.

One of the major trends in the literature is the establishment of decentralized learning systems that support privacy. As an illustration, FL-SiCNN model incorporates a Siamese Convolutional Neural Network into a peer-to-peer federated learning environment, eliminating the central server and attaining 98.78% multi-class brain tumor classification with maintaining privacy of data and resistance to data-poisoning attacks [10]. Equally, mixed methods that integrate deep feature data collection with

ResNet101 and DenseNet121, dimensionality lost with principal component analysis (PCA), and Random Forest classification have shown outstanding performance and flexibility with various and noisy MRI data, to a highest accuracy of 99.7% [11]. Collectively, these studies suggest that privacy-conscious and hybrid feature-learning systems can be used to obtain highly accurate and robust tumor classification in realistic and heterogeneous settings.

More complex deep learning models have also been tested to become more indicative of spatial and sequential characteristics of brain MRI data. A four-stage pipeline of adaptive filtering in combination with enhanced k-means to employ clustering, GLCM for texture features extraction, and Recurrent Convolutional Neural Network (RCNN), with 95.17% accuracy, 98.42% sensitivity, and 89.28% specificity, was found to perform better than other baseline models, including back-propagation (BP) networks and U-Net [12]. PDCNN has been suggested to handle MRI data using two convolutional pipelines, which produce more detailed spatial feature learning, less overfitting, and accuracy of 96.29% has been observed [13]. All these findings indicate that multi-stage, multi-branch CNN models can achieve substantial gains when it comes to representation of the complex tumor morphology and boundary information.

The use of pre-trained CNNs to perform transfer learning is still an essential approach of high-performing systems. Finetuned ResNet50 and EfficientNet networks have also achieved high performance with ResNet50 pipelines commonly getting a accuracy of about 98.5%, and they have also done better than earlier networks like VGG16 and InceptionV3 when carrying out brain tumor MRI classification tasks [14], [15]. EfficientNet, with effective preprocessing, including contrast enhancement, saliency-based feature emphasis, and Extreme Learning Machines (ELM) has also exhibited strong generalization and competitive accuracy on several public MRI data sets. Simultaneously, recent lightweight neural networks such as MobileNet and RetinaNet have made it possible to run tumor detection on medical edge devices near real-time, a balance between computational and high detection accuracy, which is crucial in medical low-resource systems [16].

The overall research in the area continually demonstrates that deep learning models (in particular, 3D CNN and attention-based) outplay conventional machine learning algorithms in both segmentation and classification [17]. It has been further demonstrated by the addition of advanced preprocessing (e.g., histogram equalization, homomorphic filtering), attention modules i.e., with Convolutional Block Attention Module (CBAM) and hybrid pipelines (e.g., EfficientNetB2 with equalization and enhanced feature extraction), that reported classification accuracy can now exceed 99% in a few controlled experiments [18]. MRI systems based on deep learning and classical machine learning classifiers such as bespoke CNNs to perform segmentation and then use MobileNetV2 feature extractors then trained on a multi-class SVM have reported 97.47% segmentation accuracy and 98.92% classification accuracy [19]. Similarly, frameworks based on dual modules, which combine adaptive Wiener-filter with neural network-based enhancement with SVM-based segmentation have achieved 98.9% accuracy in with very fast processing times that can be used in clinical workflows [20]. Together, these papers demonstrate that attention-enhanced pipelines and hybrid deep learning can achieve near-perfect accuracy when appropriately selected data are used, particularly with extensive preparation.

In spite of the recent progress, there are a number of critical issues which are yet to be addressed. The application of heterogeneous data, lack of consistent definition of classes and different evaluation schemes in the high performing studies makes direct comparison of existing methodologies and identification of an optimal model to be used in routine clinical deployment difficult. Additionally, most of the previous studies have either determined the performance of a single deep learning architecture alone or analyzed binary classification problems (tumor versus non-tumor), thus restricting to the overall comparison. Thus, a systematic comparison of various cutting edge pre-trained convolutional neural network (CNN) models in a similar set of experimental conditions on a multi-class brain tumor dataset is underexplored.

In order to fill the research gap, the current work is a systematic assessment and comparison of DenseNet121, ResNet50, Xception and MobileNet on four-class brain tumor MRI publicly available data.

The aim is to come up with CNN architectures with an effective balance between accuracy of classification, computational efficiency, and practicality to be applied to clinical systems.

Table 1: A Comparative Evaluation of Advanced Deep Learning Models for Multi-Class Brain Tumor Classification in MRI Images

Study / Method	Key Idea	Dataset / Task	Reported Accuracy
FL-SiCNN	Federated Siamese CNN for privacy-preserving multi-class classification	Multi-class brain tumor MRI	98.78%
Amin et al. hybrid model	ResNet101/DenseNet121 features + PCA + Random Forest	Multi-class, heterogeneous MRI	99.7%
Deepa & Srinivas RCNN	Adaptive filtering + improved k-means + GLCM + RCNN	Multi-class tumor classification	95.17% (Acc), 98.42% (Sens)
Parallel CNN / advanced deep model	Parallel CNN paths to capture richer spatial features	Multi-class brain tumor MRI	96.29%
ResNet50 / EfficientNet TL	Transfer learning with modern CNN backbones	Multi-class tumor classification	≈98–99%
MobileNet / RetinaNet	Lightweight CNNs for edge / real-time detection	Real-time detection on device	High precision, near real-time
3D/attention-based models	3D CNNs + attention (CBAM, etc.) with heavy preprocessing	Multi-class segmentation + classification	>99% in some studies
Hybrid CNN + SVM frameworks	CNN-based feature extraction + SVM classifier; dual-module enhancement + SVM segmentation	Multi-modal MRI segmentation + classification	≈98.9–98.92%

3. Methodology

The presented research uses advanced deep learning models alongside the notion of transfer learning to obtain the high-quality classification of a brain MRI scan. An experimental analysis is performed using The Brain Tumor MRI dataset that we have exploited for experiments in this paper, is publicly available on Kaggle and contains 7,023 MRI images divided into four categories of pituitary tumor, meningioma, glioma and no tumor [21]. As shown in Figure 1, the overall methodology gives a general view of the system architecture and experimental workflow that will be utilized in the given research to grant the fusion of deep learning and transfer learning methods to make it possible to classify brain tumors with multiple classes accurately.

The suggested workflow includes a few major steps such as the partitioning of the dataset and the exploration analysis, image preprocessing, data augmentation, model choice and adaptation, model training, and overall evaluation of performance. First, the MRI images are processed with other processes like resizing and pixel normalization to bring about uniformity and consistency to the dataset. Techniques of data augmentation are then implemented using rotation, flipping and zooming to augment the training data and enhance model generalization.

ResNet50, DenseNet121, Xception and MobileNet architecture of convolutional neural networks are then adapted to that particular task of brain tumor classification [22]. These models utilize discriminative features in the input MRI scans which are extracted automatically using knowledge acquired on large-scale datasets of images. The features obtained are then utilized to classify each image into one of the four types of tumors. Lastly, the effectiveness and reliability of the trained models is systematically

determined based on standard classification measures, such as accuracy, preciseness, recall, and F1-score. [23, 24]

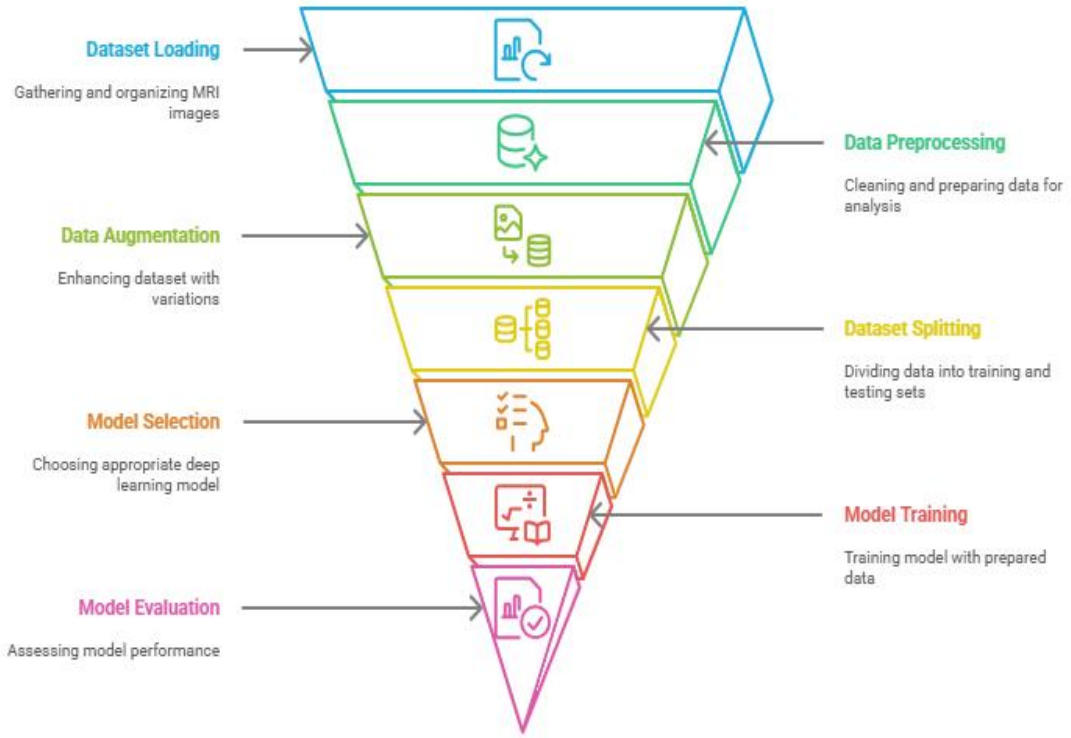

Figure 1: The proposed Brain MRI Classification Methodology's workflow

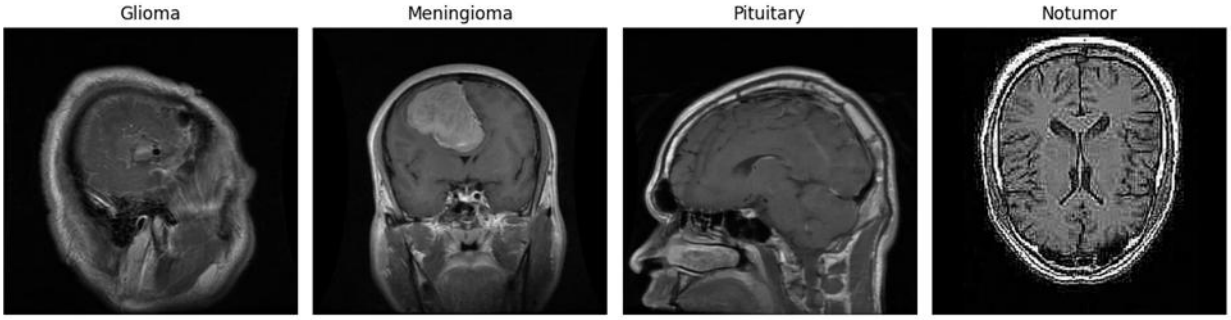

Figure 2: Sample Images from Dataset

3.1. Dataset Overview and Splitting

To conduct this study, a publicly available Brain MRI dataset of Kaggle is utilized under the number of 7,023 MRI scans [21]. Although 1,311 of these images are used as testing, 5,712 as training, it is easier to reproduce the results and compare the classification models across the board. The four categories in which the dataset is classified are pituitary tumors (non-cancerous growths that influence hormone production), meningioma tumors (which is mostly benign but may still cause pressure due to their origin in the protective membranes of the brain), glioma tumors (a more aggressive form of tumor that develops in the glial cells) and normal MRIs with a label no tumor to verify that they do not contain any tumors. The training set has 1,595 no-tumor images, 1,457 pituitary images, 1,339 meningioma images and 1,321

glioma images. This relatively homogeneous distribution across categories allows making comparisons of model performances more accurate and fairer. Moreover, the heterogeneity of tumor types ensured that any classification model was evaluated on a split of 20% allowing to have an 80/20 split of train and validation, such that hyperparameter tuning, early stopping, and model selection were all done on unseen validation data as opposed to on the test set.

Figure 3: Graphical Representation of Class Distribution

3.2. Data Preprocessing

Before training the models, there was a systematic preprocessing pipeline on the MRI images to guarantee the best input of the deep learning models. Each of the images was first down sampled to 128 x 128 pixels to have a standard input size that can fit the chosen convolutional neural network (CNN) architectures. This solution was selected because it minimizes memory and computational requirements without compromising on clinically significant anatomical detailing in regard to brain tumor, thus facilitating possible hardware implementation on common clinical platforms.

Then pixel intensity scores were clustered into the range [0, 1] which facilitates more rapid convergence and enhances numerical stability in training the model. The source color channels of the images were not eliminated to avoid loss of the spatial and intensity related information needed to give proper characterization of tumors and proper extraction of features.

3.3. Data Augmentation

In order to strengthen the model generalization as well as reduce the possibility of overfitting, a number of data augmentation methods were used to augment the training data. These methods comprised random rotations, horizontal and vertical flipping, zooming and spatial shifting. Through this type of introduction of transformations, the models were opened up to more variations of an image, which in essence mimicked the variation that can be seen in the actual clinical acquisition of MRI.

This augmentation approach made the training dataset effective much larger and allowed the models to acquire stronger and more invariant feature representations. This led to the improvement of the classification performance, especially when it comes to the ability to generalize to previously unseen MRI scans.

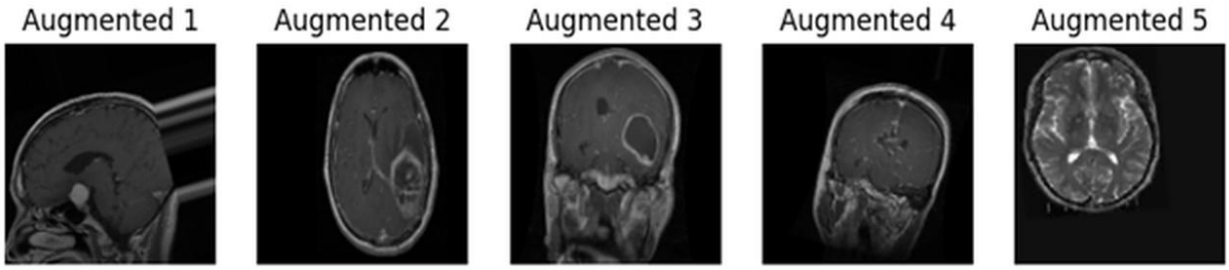

Figure 4: Augmented brain MRI images used for model training

3.4. Implementation of Pre-Trained CNN Models

This work covers brain tumor classification problem with the use of multiple renowned pre-trained convolutional neural network models including Xception model, MobileNet, ResNet50 model, and DenseNet121. These are very known with its unique architectural designs and effective feature extraction features. [25-27]

DenseNet121 has a high connectivity rate as each layer gets feature maps of all other layers. The design enables a good reuse of features, solves the vanishing gradient problem, and is specifically useful when analyzing medical images. ResNet50 employs residual links to facilitate the training of deep networks by enabling gradients to flow directly through shortcut links and it is thus suitable in learning the hierarchical features to differentiate subtle variations in MRI scans.

Xception is also built on depth wise separable convolutions, a subdivision of regular convolution operations into a spatial and a channel-wise one. The design has a significant ability to reduce the computational complexity and still provide a high representational capacity enabling the model to respond to complex spatial patterns. MobileNet is an architecture that is computationally efficient and based on depth wise separable convolutions, and it is therefore fit to be deployed in a resource limited environment whilst still offering an acceptable balance between speed and accuracy.

All models were first pre-trained using the large size ImageNet data, and they thus learnt rich and transferable visual representations. The transfer learning involved substituting the final layers of classification of each of the prepared models with custom dense layers that were specific to the four-class brain tumor classifier problem. The lower-level layers were frozen or fine-tuned during training to fit the learned representations to the attributes of brain MRI data.

Through transfer learning, the models have the advantage of being informed with prior knowledge as a result of varied datasets which facilitates better generalization, quicker convergence, and overall higher performance especially in conditions where there is scarcity of labeled medical information. The architectures that were selected depend on the depth of the architectures, their successful experience in feature learning as well as architecture design innovations like residual connection and dense connection with an aim to attain accurate and reliable brain tumor classification.

3.5. Model Architecture and Model Compilation

The model architecture created in the current paper is aimed at being able to well represent the intricate structure patterns in brain MRI pictures and retain a high generalization capacity. All models, as discussed in Table 1, start with a trained CNN backbone, which is used as a feature extractor and then further classification-specific layers are added.

RGB images of 128 by 128 pixels are fed into the input layer to offer an optimal balance between the efficiency of the calculations and the amount of spatial information. The pre-trained CNN (DenseNet121, ResNet50, Xception or MobileNet) is incorporated as the base model and a Global Average Pooling layer is added to achieve decreasing spatial dimensionality and limit the number of trainable parameters, thus preventing overfitting.

In order to improve the level of generalization further, Dropout layers with a dropout rate of 0.5 are added in front of and after the dense layer. So that the model can learn the non-linear and discriminative representations of features and also regulate the complexity of the model, the fully connected layer contains 64 neurons that use ReLU as the activation function. Additionally, the layer uses a regularization of L2 to ensure that the model is as complex as possible.

The last output layer will be a set of four neurons, which represent the four types of tumors, and utilize a softmax activation function, which is appropriate to multi-class classification. To compile the model, we apply the Adam optimizer. The learning rate was set to 0.001. Moreover, the loss function should be categorical cross-entropy, which is suitable with multi-class labels, one-hot encoded. The number of epochs of training the models is 40, and this is enough time to converge and stabilize the models in good performance.

Table 2: Description of pretrained models

Type of Layer	Description
Input Layer	Accepts input images in RGB format with dimensions 128 × 128 × 3.
Feature Extractor	Employ a pre-trained CNN (e.g., DenseNet121, ResNet50, Xception, MobileNet) for automatic feature extraction.
Global Pooling Layer	Performs global average pooling to reduce the dimensionality and the parameter count.
Regularization (50%)	Implements dropout, randomly deactivating 50% of neurons to help prevent overfitting.
Fully Connected Layer	A dense layer with 64 neurons, employing ReLU activation and L2 regularization to enhance model learning.
Regularization (50%)	Applies a second dropout layer to further regularize the network by deactivating another 50% of neurons.
Output Layer	This layer contains 4 neurons. It uses softmax activation for categorization into one of the four target classes.

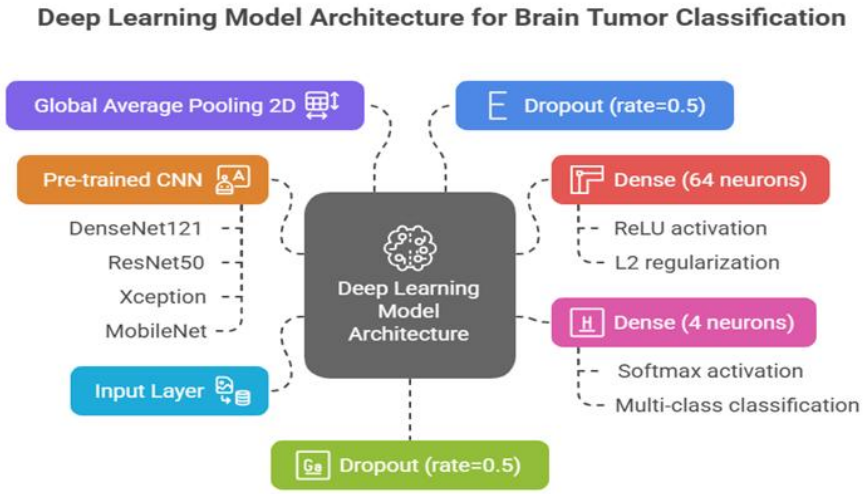

Figure 5: Detailed Deep Learning Model Architecture Representation

The training is typically performed by freezing the lower levels of the pre-trained base model to preserve the general visual features, and then the upper levels of the model and the specific classification

head are learned on the MRI of the brain. This type of transfer learning allows the model to utilize high level features to the targeted task of brain tumor classification. Adaptive and efficient training is done using the Adam optimizer with a learning rate of 0.001. The categorical cross entropy loss is applied because it is highly appropriate when performing a multi-class classification. The accuracy metric is used to measure model performance. The models in this research were trained to a maximum of 40 epochs which was enough to allow the networks to capture the underlying trends in the training data in the selected optimization settings.

The model was trained using mini-batch gradient descent having a batch size of 32. Early stopping and model checkpointing methods were used in order to stop overfitting and enhance training performance. Early stopping is concerned with tracking validation performance and terminating training when no additional improvement is seen, and model checkpointing is saved with the best-performing model throughout training, so nothing will be lost. Weight initialization and shuffling of the data was done with a fixed random seed to enhance reproducibility of the training runs.

3.6. Model Evaluation Metrics

The accuracy, precision, recall and F1-score were also four popular measures that were employed to comprehensively evaluate the performance of the proposed models in the classification of brain tumors. The following measures are essential in the medical image analysis:

- **Accuracy:** This is the percentage of total correct predictions (true positives and true negatives) out of the total predictions.
- **Precision:** This statistic indicates the proportion of cases that were accurately predicted to be positive in comparison to the total number of cases that were anticipated to be positive.
- **Recall (Sensitivity):** Correctly classified actual positives all actual positives.
- The **F1-Score**:is the harmonic average of recall and precision.
- In order to be classified on multiple classes, these metrics were computed on each class and averaged to yield a general performance measure.

4. Results and Discussion

There are four pre-trained deep learning models, MobileNet, ResNet50, Xception and DensNet121 whose results are demonstrated and analyzed in this section. These models have different layer designs and extracted features and show multifarious convolutional neural network (CNN) architecture. The accuracy, precision, recall, F1 score, and loss were important key performance indicators applied in assessing training and testing datasets. The distinct classification features of both of the models are described in the first section, and a performance measure in detail of each model is provided. The relative merits and limitations of both architectures are then discussed as well as their suitability in the classification task of medical images. We also discuss the broader implications of the findings to the improvement of automated image analysis to detect brain tumors.

4.1 Model-wise Performance

Table 3 and 4 as well as the confusion matrix in Section 4.2 confirm that the DenseNet121 model was quite good in classifying brain MRI images. The model was tested on four different tumours measured with the measures of accuracy, precision, recall, and F1 score, as well as support, which gives a complex view of the advantages and disadvantages of the model. The generalizing ability of DenseNet121 on unknown data with the accuracy of 0.97 on the training set and 0.98 on the testing set was good. The questionable difference between the accuracy of training and testing is indicative of little overfitting. The metrics including: 1) precision, 2) recall and 3) F1 score of the test set of the weighted and macro averages all were consistently high macros were 0.98 indicating even performances across the classes.

Table 3: Classification results of DenseNet121

Metrics	GliomaTumor	Meningioma Tumor	No Tumor	Pituitary Tumor	Macro Average	Weighted Average
Precision	98.0%	99.0%	99.0%	97.0%	98.0%	98.0%
Recall	97.0%	96.0%	98.0%	99.0%	97.0%	98.0%
F1-Score	97.0%	97.0%	99.0%	98.0%	98.0%	98.0%

Table 4: Overall DenseNet121 Performance

Metric	Value
Training Accuracy	97.3%
Testing Accuracy	98.1%
Total Training Samples	5,712
Total Test Samples	1,131

On further examination of the per-class results, the model performed extraordinarily on the No Tumor classification with the precision of 0.99, recall of 0.99, and F1 score of 0.99. Pituitary class showed high capabilities of identifying these instances thus it was also very effective with a recall of 1.00 and F1 of 0.99. Also, the Glioma and Meningioma classes were accurately classified with a high level of precision and recall, which demonstrates the effectiveness of DenseNet121 in the classification of tumor types.

Figure 6 demonstrates the training and validation accuracy and loss curves of the DenseNet121 model.

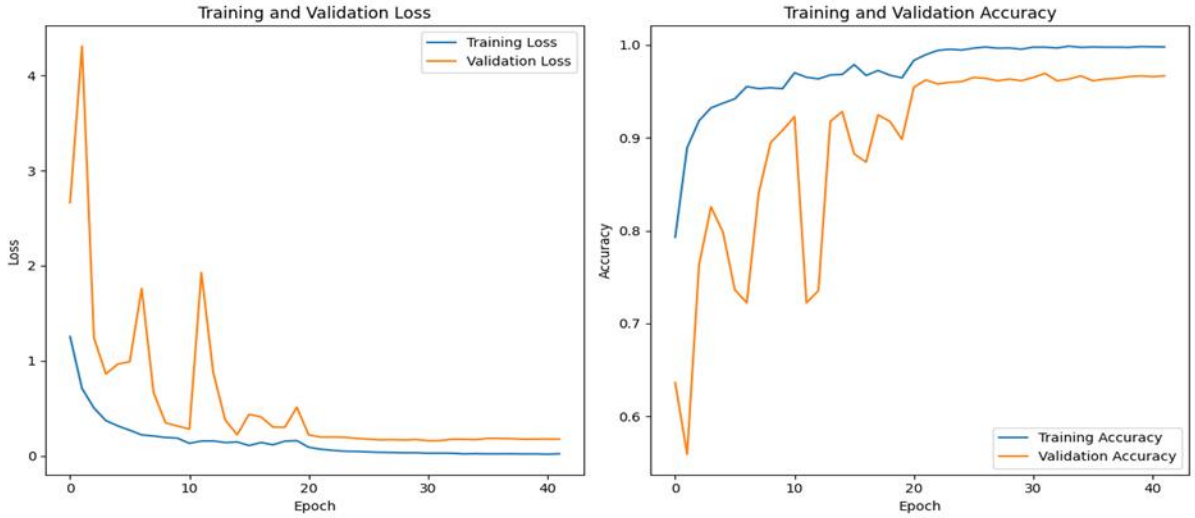

Figure 6: DenseNet121's accuracy and loss curves show rapid convergence, with minimal performance gap, indicating good generalization and limited overfitting.

The plots show that the model converged rather fast, and training and validation accuracy after approximately 20 epochs were above 98 and 95, respectively. The decreasing loss curves and the narrowing gap between training and validation measure indicate effective learning and minimal overfitting. These results demonstrate the extent to which DenseNet121 can be trusted and generalized to classify brain tumors.

ResNet50 also shows great performance (i.e., depicted in table 5 & 6) in brain MRI classification. It achieved good generalization with training accuracy of 94% with a testing accuracy metric of 97% and

macro and weighted averages of precision, recall, and F1score of about 0.97, which means that it has balanced performance across the four classes. According to the per class analysis, the No Tumor category (precision 0.97, recall 0.99, F1score 0.98) and Pituitary class (recall 1.00, F1score 0.97) show great results and indicate that ResNet50 is also reliable in differentiating between the different types of tumors.

Table 5: Classification results of RestNet50.

Metrics	GliomaTumor	Meningioma Tumor	No Tumor	Pituitary Tumor	Macro Average	Weighted Average
Precision	98.0%	97.0%	99.0%	95.0%	97.0%	97.0%
Recall	96.0%	94.0%	97.0%	99.0%	96.0%	97.0%
F1-Score	97.0%	95.0%	98.0%	97.0%	97.0%	97.0%

Table 6: Overall ResNet50 Performance

Metric	Value
Training Accuracy	94.5%
Testing Accuracy	97.2%
Total Training Samples	5,712
Total Test Samples	1,131

Figure 7 shows the training and validation loss and accuracy curve of the model. The metrics of training indicate a continuous and consistent advancement, although the changes in the validation loss and the accuracy indicate certain instability in initial epochs that are corrected with progress of training.

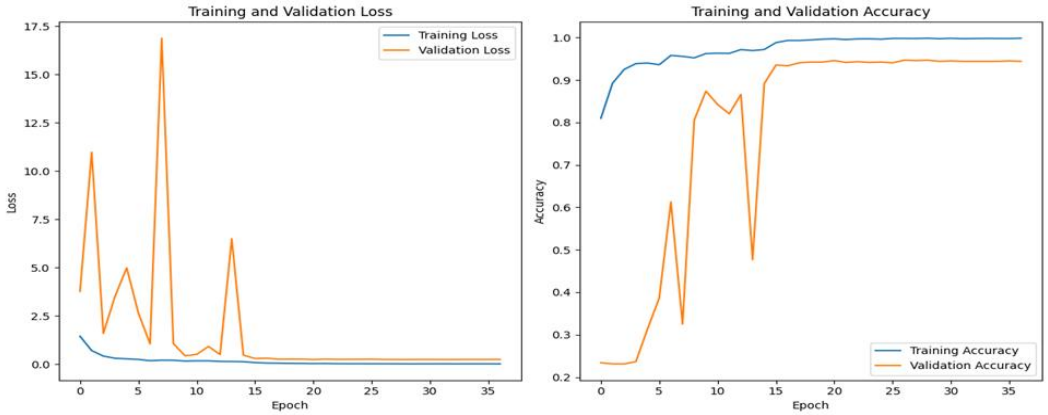

Figure 7: ResNet50's accuracy and loss curves show initial fluctuations in validation loss, which stabilize, indicating eventual convergence with acceptable broadening.

Xception was very effective in brain MRI classification as it is depicted in Table 4. It achieved training accuracy and testing accuracy of 98% and 98% respectively in the test set with macro and weighted precision metric, recall metric, and F1-score, which means that it is highly generalized and well-balanced in all four classes. The No Tumor class attained precision metric, recall metric, and F1 -score of 99% and Pituitary class attained recall of 1.00 and F1 -score of 0.98, where glioma and meningioma were also classified with high precision although it has a lower recall compared to Pituitary. These findings indicate that Xception is powerful and can be used in applications of brain MRI analysis.

Figure 8 shows the training and validation accuracy and loss curve of the Xception model. The convergence in the plots is high, and the performance remains relatively stable, and the validation accuracy is close to the training accuracy which proves that learning is successful and overfitting is not significant.

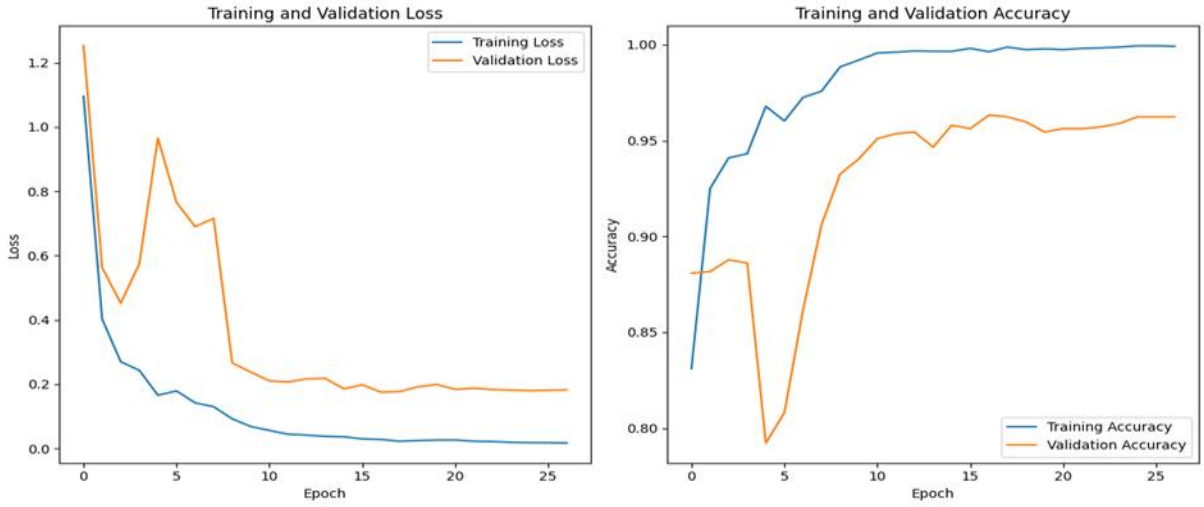

Figure 8: Xception's training and validation accuracy and loss curves show consistent performance, stable learning dynamics, and closely aligned training and validation accuracy.

The MobileNet model was also effective that it was indicated in the confusion matrix in Section 4.2 and summarized in Table 7 and 8. The model was tested on four types of tumors based on accuracy, precision, recall, F1 score, and support measures giving an overall evaluation of the classification capabilities of the model. With the training set MobileNet achieved a training accuracy of 0.96 and a testing accuracy of 0.98 and thus, there is generalization to non-seen data. The macro and weighted averages of the precision, recall, and the F1 score were equal to 0.98 on the test set, which shows the indicators of balanced and strong performance on all classes. The model was also found to be trustworthy in the association against non-tumor cases as shown by the No Tumor model with the highest criteria with a perfection of 98%, recall of 99% and an F1 score of 99%.

Table 7: MobileNet Classification Results

Metrics	GliomaTumor	Meningioma Tumor	No Tumor	Pituitary Tumor	Macro Average	Weighted Average
Precision	97.0%	98.0%	99.0%	96.0%	98.0%	98.0%
Recall	98.0%	96.0%	98.0%	99.0%	97.0%	98.0%
F1-Score	97.0%	97.0%	99.0%	97.0%	98.0%	98.0%

Table 8: Overall MobileNet Performance

Metric	Value
Training Accuracy	95.8%
Testing Accuracy	98.1%
Total Training Samples	5,712
Total Test Samples	1,131

Pituitary class also did remarkably well with a recall of 1.00 and F1 score of 0.98. In addition, the fact that the Glioma and Meningioma classes of tumors were highly perfected and recalled provided further evidence that the model was able to differentiate between the various types of cancers.

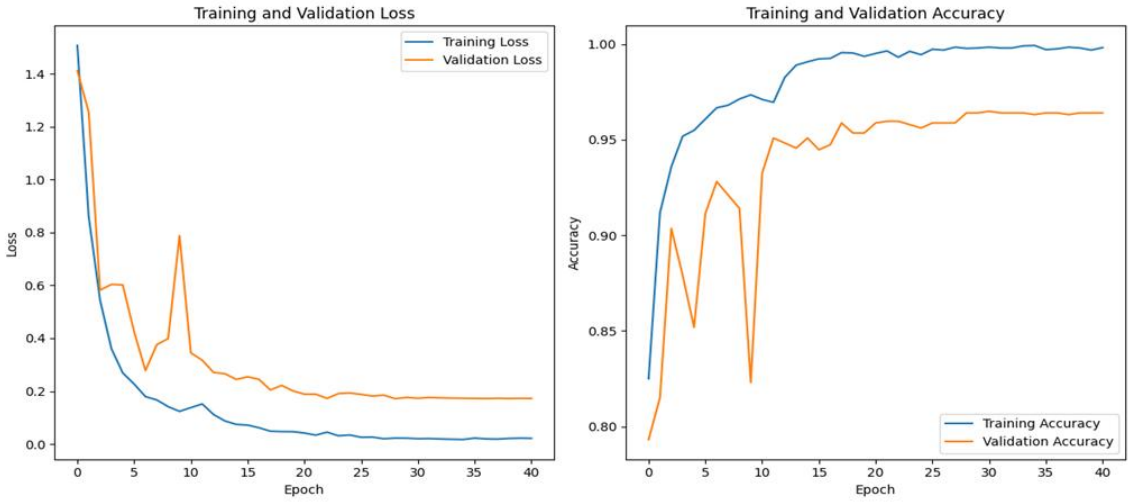

Figure 9: MobileNet's accuracy and loss curves demonstrate its effectiveness as a lightweight model with smooth convergence and consistent performance between training and validation

Figure 9 displays MobileNet's training and confirmation delicacy and loss angles. Effective literacy and little overfitting are indicated by the plots' smooth confluence of the loss and delicacy criteria, with the confirmation delicacy nearly running the training delicacy.

4.2 Overall Model's Performance Comparison

The trained models used in the present study were evaluated in systematic manners by measuring various performance criteria to establish the effectiveness of the models in classification of brain tumors. A table 9 provides a side-by-side comparison of the major metrics-accuracy, precision, recall, and F1 score-of each model, ranked in order of their weighted average of accuracy. This ranking allows one to have a clear view of the extent to which every model can sustain high performance in a wide range of samples sizes and class distribution. DenseNet121 was the best in the weighted average accuracy as indicated in Table 6, with the other models showing Xception (0.9817), MobileNet (0.9786) and ResNet50 (0.9703). The high accuracy, recall and F1 values of all models depict the capability of strong classification, and DenseNet121 is the most reliable model in detecting brain tumor in this research.

Table 9: Comparative Analysis of all Tested Models

Model	Accuracy	Precision	Recall	F1 Score
Xception	0.9817	0.9819	0.9817	0.9817
DenseNet121	0.9847	0.9848	0.9847	0.9847
ResNet50	0.9703	0.9708	0.9703	0.9702
MobileNet	0.9786	0.9788	0.9786	0.9786

Figure 10 presents in the form of a graph the relative performance of all models and focuses specifically on the weighted average of accuracy as a dependent variable of overall performance. This

graphical illustration also indicates the high-quality and the steadiness of the performance of DenseNet121.

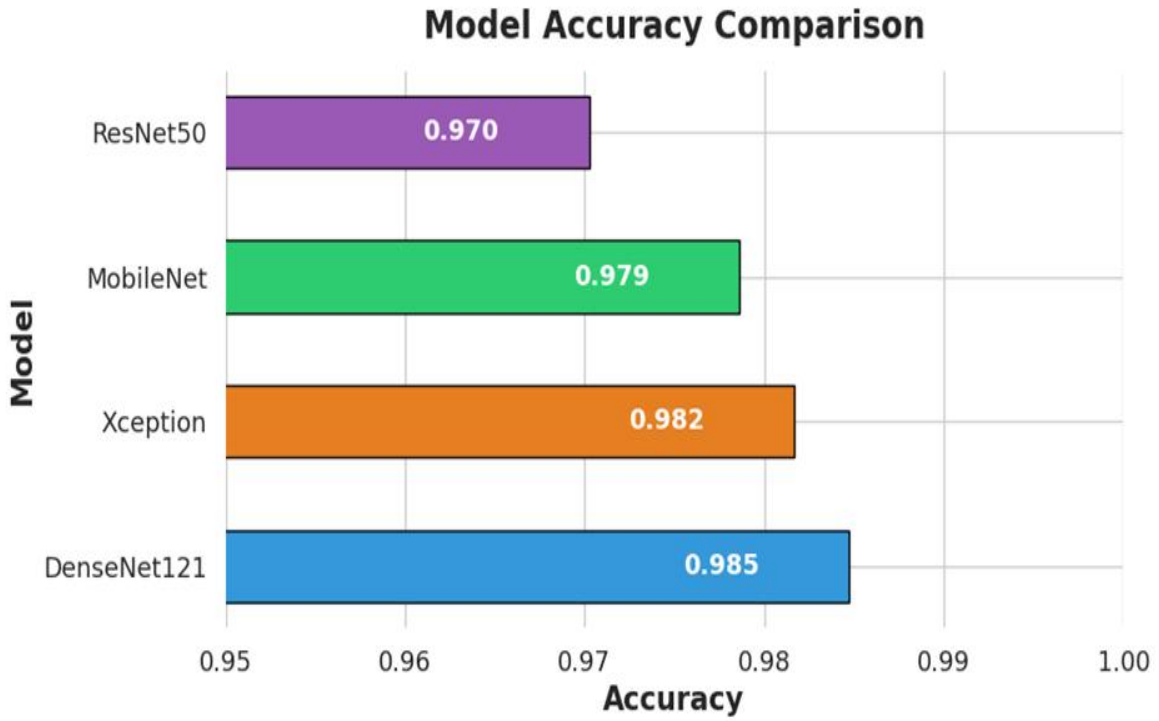

Figure 10: DenseNet121 outperforms Xception, MobileNet, and ResNet50 in weighted accuracy, precision, recall, and F1-score

The high performance of DenseNet121 in this study may be explained by its densely connective structure whereby a layer is provided with feature maps of all the other layers. This connectivity encourages widespread feature reuse, and the gradient flow of the network, and alleviates the vanishing gradient problem, allowing the model to learn more discriminative representations with relatively small MRI data. Conversely, the use of direct inter-layer connections, which are exploited in the ResNet50, Xception, and MobileNet, even though they also attain high results, is probably a contribution to the observed slightly lower accuracy and F1 -scores associated with these two architectures.

This study did not comprehensively apply formal statistical hypothesis tests (e.g. paired t-tests with repeated runs) to all pairwise model comparisons, but DenseNet121 had the highest weighted accuracy and F1-score across repeated training. However, Xception, MobileNet, and ResNet50 yielded a slightly lower but consistently higher result. Though these gaps will be addressed in further research where intensive statistical testing and repeated cross-validation techniques will be used to verify the relevance of these differences, the coherent gaps indicate that the performance gain of DenseNet121 cannot be attributed to randomness only.

Moreover, Figure 11 shows the confusion matrices of each of the models with the four types of tumors, which provide in-depth insight into the classification strengths and possible improvement locations. These matrices give useful information regarding the capacity of each of the models to adequately discriminate between various forms of brain tumors.

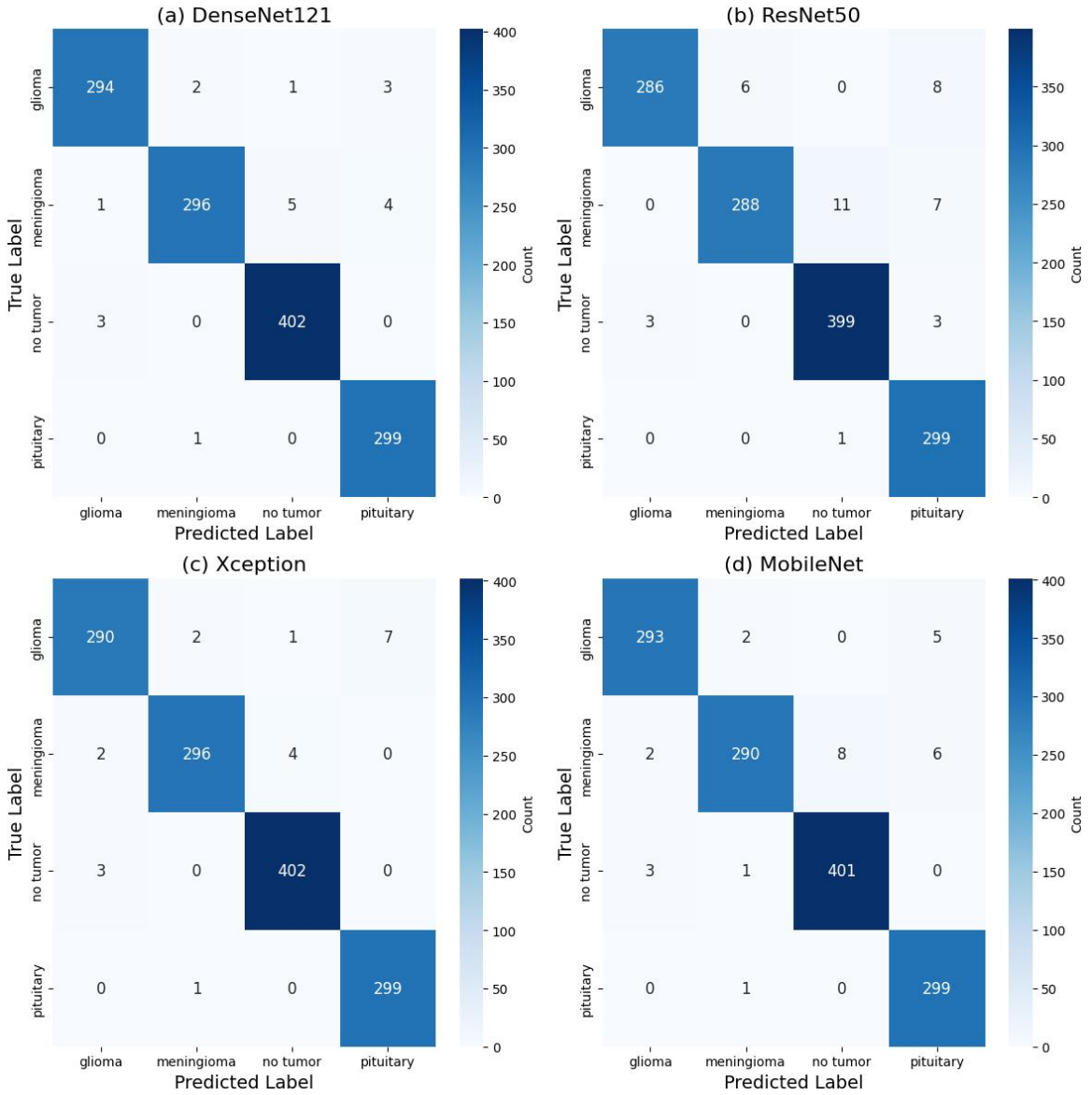

Figure 11: Confusion matrices for (a) DenseNet121, (b) ResNet50, (c) Xception, and (d) MobileNet show most of errors occur between meningioma and glioma, however no-tumor and pituitary cases generally are correctly diagnosed.

5. Conclusion and Future Work

This paper shows that DenseNet121 was able to classify four types of tumors on MRI scans with a transfer learning accuracy of 98.47 with the systematic evaluation and comparison to ResNet50, Xception, and MobileNet in brain tumor classification. The findings prove that transfer learning using these.

architectures can produce unbelievably accurate and evenly distributed categorization and the best result is achieved by DenseNet121 (98.47%), with Xception coming in close second (98.17%), MobileNet (97.86%), and ResNet50 (97.03%).

There were good performance and high generalization of all models on important measures such as precision, recall, and F1 score. Our findings show that, due to their ability to provide fast, reliable and automated tumor diagnosis, deep learning-based systems have significant potential in helping and guiding clinical decision-making processes in neuro-oncology. In practice, these models can be incorporated as

decision support tools to enable radiologists to perform quicker screening of MRI scans, identify high risk cases and lessen the time of reporting, especially on high-volume or resource constrained settings. Nevertheless, they should undergo additional testing in multicenter data and attentive consideration of integration into the current clinical processes to be used safely and in a proper manner.

Even though the results are promising, it is possible to identify several areas that require further research. To increase the generalizability of the models and reduce the influence of the class imbalance, the future studies should focus on the expansion of the dataset containing more diverse and multi-center MRI images. In the case of challenging tumor types, further accuracy of classification might be achieved by integrating ensemble or hybrid modelling approaches and advanced data augmentation techniques. Furthermore, through further elucidating the model judgements, we are going to employ explainability methods such as the Grad-CAM or saliency maps towards the formation of therapeutic trust. Finally, studying the latest topologies, including Vision Transformers (ViT), Swin Transformer, and other new-generation models, can lead to further improvement of performance and allow practical clinical applications. These procedures will be necessary in order to translate deep learning-based brain tumor classification out of research into standard clinical practice.

Ethical and interpretability issues will involve clinical adoption besides the technical developments. To make predictions of the model accepted by clinicians instead of being treated as black box outputs, the future systems must be evaluated in terms of potential biases in different patient subgroups, ensure patient data safety, and provide understandable visualization (e.g., heatmaps to identify suspicious areas).

Funding Statement: No funding received.

Conflicts of Interest: No conflicts of interest exist.

Data Availability: Brain Tumor MRI dataset is available publicly on Kaggle website.

References

[1] Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., Wang, Z., & Feng, Q. (2015). Enhanced performance of brain tumor classification via tumor region augmentation and partition. PLOS ONE, 10(10), e0140381.

[2] Ostrom, Q. T., Price, M., Neff, C., Cioffi, G., Waite, K. A., Kruchko, C., & Barnholtz-Sloan, J. S. (2022). CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2015–2019. Neuro-Oncology, 24(Suppl. 5), v1–v95.

[3] Louis, D. N., Ohgaki, H., Wiestler, O. D., & Cavenee, W. K. (Eds.). (2007). WHO classification of tumours of the central nervous system. World Health Organization.

[4] Martucci, M., Russo, R., Schimperna, F., D'Apolito, G., Panfili, M., Grimaldi, A., Perna, A., et al. (2023). Magnetic resonance imaging of primary adult brain tumors: State of the art and future perspectives. Biomedicines, 11(2), 364.

[5] Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Transactions on Medical Imaging, 35(5), 1240–1251.

[6] 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.

[7] Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging, 35(5), 1285–1298.

[8] Brindha, P. G., Kavinraj, M., Manivasakam, P., & Prasanth, P. (2021). Brain tumor detection from MRI images using deep learning techniques. In IOP Conference Series: Materials Science and Engineering (Vol. 1055, No. 1, Article 012115). IOP Publishing.

[9] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van der Laak, J. A. W. M., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

[10] Onaizah, A. N., Xia, Y., & Hussain, K. (2025). FL-SiCNN: An improved brain tumor diagnosis using siamese convolutional neural network in a peer-to-peer federated learning approach. Alexandria Engineering Journal, 114, 1-11.
[11] Khan, M. A., Khan, A., Alhaisoni, M., Alqahtani, A., Alsubai, S., Alharbi, M., ... & Damaševičius, R. (2023). Multimodal brain tumor detection and classification using deep saliency map and improved dragonfly optimization algorithm. International Journal of Imaging Systems and Technology, 33(2), 572-587.
[12] Vankdothu, R., & Hameed, M. A. (2022). Brain tumor MRI images identification and classification based on the recurrent convolutional neural network. Measurement: Sensors, 24, 100412.
[13] Rahman, T., Islam, M. S., & Uddin, J. (2024). MRI-based brain tumor classification using a dilated parallel deep convolutional neural network. Digital, 4(3), 529-554.
[14] Jain, J., Kubadia, M., Mangla, M., & Tawde, P. (2024). Comparison of transfer learning techniques to classify brain tumours using MRI images. Engineering Proceedings, 59(1), 144.
[15] Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 15(16), 4172.
[16] Uke, S., Sabale, M., Misal, S., Mogal, S., & Bhojane, M. (2025, April). MRI-Based Brain Tumor Classification Using Transfer Learning. In 2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI) (Vol. 3, pp. 142-148). IEEE.
[17] Khan, M. A., Khan, A., Alhaisoni, M., Alqahtani, A., Alsubai, S., Alharbi, M., Malik, N. A., & Damaševičius, R. (2023). Multimodal brain tumor detection and classification using deep saliency map and improved dragonfly optimization algorithm. International Journal of Imaging Systems and Technology, 33(2), 572–587.
[18] Rahman, A. M. J. Z., Gupta, M., Aarathi, S., Mahesh, T. R., Kumar, V. V., Kumaran, S. Y., & Guluwadi, S. (2024). Advanced AI-driven approach for enhanced brain tumor detection from MRI images utilizing EfficientNetB2 with equalization and homomorphic filtering. BMC Medical Informatics and Decision Making, 24(1), 113.
[19] Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 15(16), 4172.
[20] Asiri, A. A., Soomro, T. A., Shah, A. A., Pogrebna, G., Irfan, M., & Alqahtani, S. (2024). Optimized brain tumor detection: A dual-module approach for MRI image enhancement and tumor classification. IEEE Access, 12, 42868–42887.
[21] Nickparvar, M. Brain Tumor MRI Dataset: a dataset for classify brain tumors. 2021.
[22] Asiri, A. A., Soomro, T. A., Shah, A. A., Pogrebna, G., Irfan, M., & Alqahtani, S. (2024). Optimized brain tumor detection: A dual-module approach for MRI image enhancement and tumor classification. IEEE Access, 12, 42868–42887.
[23] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4700–4708).
[24] Powers, D. M. W. (2020). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.
[25] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778).
[26] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1251–1258).
[27] 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.