SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Deep Transferable Representation Learning For Accurate Multi-Class Brain Tumor Classification From MRI Images

Surekha Amol Patil[1] , Dr. Uma Goradiya[2]

[1,2]Department of Computer Science Shree L. R. Tiwari College Of Engineering
[1]surekhashedage22@gmail.com
[2]uma.goradiya@slrtce.in

Abstract- Despite the huge amount of medical imaging data available and the fact that magnetic resonance imaging (MRI) shows nearly unlimited potential for tumor imaging, accurate multi-class classification of brain tumors from MRI is still very difficult due to the heterogeneity of the tumor types, differences in appearance of the images and the limited amount of medical imaging data available to learn sufficiently discriminative features. The goal of this study is to devise a deep representation learning approach that is transferable across brain tumor datasets for automated brain tumor classification. It is an experimental subset of Brain Tumor MRI benchmark dataset containing 3265 images which includes 2609 training, 325 validation and 331 testing images from four classes including glioma, meningioma, pituitary tumor and no-tumor. Three different transfer-learning architectures (ResNet50V2, VGG16, DenseNet121, and EfficientNetB0) are explored in order to extract discriminative representations and classify the data. ResNet50V2 achieves the highest test accuracy of 94.26%, compared with VGG16 (93.66%), DenseNet121 (92.45%), and EfficientNetB0 (28.70%). ResNet50V2 also achieves macro precision of 94.46%, macro recall of 94.29% and macro F1 score of 94.37% with pituitary tumor classification attaining a recall of 100% and 98.38% macro F1 score. The study provides a new framework that integrates the different aspects of transferable representations, class-specific discrimination, and model complexity. The results prove that ResNet50V2 gives the best performance when compared to the other architectures presented for automated multi-class brain tumor MRI classification.

Keywords—Brain Tumor Classification; Transfer Learning; Magnetic Resonance Imaging; Convolutional Neural Networks; ResNet50V2; Multi-Class Medical Image Classification.

I. Introduction

The timely and precise diagnosis of a brain tumour is directly linked to the treatment strategy and survival rate of a patient, and is one of the most serious neurological conditions. The visualisation of brain tissue (anatomical information) is the most important non-invasive imaging modality and manual interpretation of MRI scans is still time consuming, subjective and highly radiologist dependent [1]. While gliomas, meningiomas and pituitary tumors vary in location, shape and texture and intensity characteristics, multi-class discrimination is actually a difficult pattern-recognition problem [2]. Popular conventional machine learning pipelines using hand-crafted features like texture descriptors or shape statistics tend to fail to be generalized across scanners, acquisition protocols and patient cohorts [3]. DCNNs have proven to be powerful tools for automatically learning a hierarchical, discriminative representation directly from raw pixel information, thus avoiding the need for manual feature engineering [4]. Training deep networks on medical imaging datasets, however, is limited by the availability of large-scale annotated MRI scans which would be ideally suitable for this purpose, as it would require the help of expert radiologists and be expensive to obtain [5]. To overcome the lack of data, it extends the ability of convolutional backbones to extract features from a large set of natural images like ImageNet, which are then transferred to and fine-tuned on a relatively small set of medical images. Although the number of studies that employ transfer-learning approaches to classify brain tumors is increasing, these

studies often focus on a single backbone architecture or a limited split of the data set, which raises questions about the behavior of the different kinds of transferable representations under the same experimental conditions for all four types of diagnosis [7]. This encourages the necessity to systematically and comparatively analyze a number of pretrained architectures and assess their superiority in obtaining transferable, class-discriminative representations for multi-class brain tumor classification, under the same experimental protocol. To fill this void, in the present work, a deep transferable representation learning framework is developed and tested to understand the robustness of the four popular convolutional architectures for multi-class brain tumor Magnetic Resonance Imaging (MRI) classification, considering classification accuracy, class-wise discrimination, and model complexity in a uniform experimental setup.

Contributions of this Study

1) We suggest a common deep representation learning framework, based on a transferable approach that can be used to compare the accuracy of four pre-trained convolutional architectures in brain tumor classification.
2) We measure the performance class-wise and overall at the level of precision, recall and F1-score for each of the glioma, meningioma, pituitary and no-tumor category.
3) Given a number of models, parameters and storage, we analyse the accuracy of the models, and select the most efficient transferable architecture.

II. Related Work

Pretrained convolutional networks have been studied for the task of brain tumor classification using numerous studies, which explore transfer learning. Srinivas et al. tested various deep transfer-learning backbones, and they found that fine-tuning models trained on brain MRI data always outperforms networks that were trained from scratch [8]. To enhance the reliability of multi-grade tumor classification, Tandel et al. presented an ensemble of several deep architectures together with explainable-AI techniques by a majority voting method [9]. To fine tune EfficientNet variants for multigrade brain tumor classification, Priyadarshini et al. demonstrated that the light-weight model is sensitive to the hyperparameter and learning-rate configuration [10]. In [11] Shivahare et al. suggested hybrid deep learning system consisting of various feature extractors to attain high precision brain tumor detection and classification in MRI scans. Disci et al. systematically explored a variety of existing CNN backbones and showed that the residual architectures typically perform better on transferring features than the regular CNNs [12]. Gomes and Barbosa studied the interpretability of models along with classification accuracy and stated that transfer-learning models need to be clinically explainable while maintaining the classification accuracy [13]. Abbas et al. used transfer learning and the incorporation of custom convolutional layers to enhance the discriminatory power between similar tumor sub-types with regards to visual similarity [14]. Amarnath et al. suggested an improved pipeline of transfer learning with data augmentation to deal with class imbalance problem in brain tumor MRI data [15]. AlShehri and Busaleh proposed a Multi-scale Representation Learning Approach with morphology information for brain tumor MRI analysis [16] which enhanced the interpretable feature extraction. Hussain and Shouno used gradient-weighted class activation mapping to obtain visual representations of discriminative regions in deep classifiers which enhanced the trust on localization of tumors in automatic systems [17]. Disci et al. also compared the performance of pretrained deep CNNs and found that among the different architectures, the ones incorporating residual connections showed the best generalization from tumor to tumor [18]. Uysal and Erkan compared the performance of a number of deep learning architectures in multiclass brain tumor classification against accuracy in the networks, and they concluded that deeper networks were more accurate, but also required more computational resources [19]. All of these studies validate the effectiveness of transfer learning for brain tumor classification compared to training a CNN from scratch and demonstrate that residual architecture and densely connected architectures are more effective than simple sequential CNNs. In most previous studies, however, only a small set of backbones are evaluated, some studies employ inconsistent splits of their datasets, and some studies fail to provide detailed analysis of the classes and complexities. This inspires the proposed unified comparative framework in this study.

Table 1. Summary of Related Work on Transfer-Learning-Based Brain Tumor Classification

Ref.	Author(s)	Method / Architecture	Key Finding
[8]	Srinivas et al.	Comparative deep transfer learning backbones	Fine-tuned pretrained models outperformed training from scratch

[9]	Tandel et al.	Majority-voting ensemble + Explainable-AI	Ensemble voting improved multi-grade classification reliability
[10]	Priyadarshini et al.	Fine-tuned EfficientNet	Lightweight models are sensitive to learning-rate tuning
[11]	Shivahare et al.	Hybrid deep learning framework	Multi-extractor fusion improved detection precision
[12]	Disci et al.	Multiple pretrained CNN backbones	Residual architectures gave superior transferable features
[13]	Gomes & Barbosa	Interpretability-focused transfer learning	Highlighted accuracy–explainability trade-off
[14]	Abbas et al.	Transfer learning + custom CNN layers	Improved discrimination of visually similar subtypes
[15]	Amarnath et al.	Transfer learning + augmentation	Augmentation mitigated class-imbalance effects
[16]	AlShehri & Busaleh	Morphology-aware multi-scale learning	Improved interpretable multi-scale feature extraction
[17]	Hussain & Shouno	Grad-CAM based explainable CNN	Improved trust via tumor-region visualization

We present the results of ten representative transfer-learning studies in Table 1. Most of these have a single or only two backbones, and only a handful report class-wise metrics and complexity, indicating a need for a single benchmark as proposed here.

III. Brain Tumor MRI Dataset and Data Preparation

This section outlines the dataset of brain tumors used to develop and evaluate the model, a description of the class composition of the dataset, the strategy used to partition it into training, validation and testing subsets, as well as the pre-processing operations applied to ensure the uniformity of image dimensions, pixel intensities, and train it for multi-class classification using transfer learning.

A. Dataset Description

The Brain Tumor MRI Dataset consists of 7,200 MRI images, each labeled according to its tumor type, including glioma, meningioma, pituitary tumor, and no tumor. There were 5,600 images in the dataset with balanced classes distribution to train and 1,600 with unbalanced classes distribution to test. The image sizes of scans differ from each other, so image resizing and normalization is required for each scan. A sample of 2,609 images were selected for training, 325 for validation and 331 for testing, consisting of all four diagnostic classes of this benchmark.

Table 2. Dataset Distribution

Dataset Split	Number of Images
Training	2,609
Validation	325
Testing	331
Total	3,265
Number of Classes	4

The data partitioning applied in the present study is summarized in the Table 2, which shows the training subset is the largest among the four data subsets with the remaining three data subsets (validation and testing) being proportionally balanced.

B. Class Distribution, Training–Validation–Testing Split

The experimental subset has an approximately equal number of cases in each of the four diagnostic classes, and there is no single class dominating the training of the models and class balanced classification metrics are not artificially inflated by class imbalance. A total of 3,265 images were used, divided into 2,609 images (79.9%) for training, to learn discriminative convolutional features by backpropagation; 325 images (10.0%) for validation, to explicitly select the values of the hyperparameters, schedule the learning rate and stop when

overfitting is detected; and 331 images (10.1%) for testing, for an un-biased estimate of the generalization performance. This stratified split maintains the same class proportions between training, validation and testing, thus guaranteeing the proper balance of glioma, meningioma, pituitary tumor and no tumor classes in each set, which is crucial for a fair multi-class evaluation of the classification system, as well as to reduce optimistic bias in the classification results.

C. Preprocessing

The MRI was pre-processed before being fed to the model to ensure uniformity in the different variable-resolution MRI images available in the raw data. All images were resized to a fixed spatial resolution which is suitable for the input layer of the pretrained convolutional backbone, and the intensities of the images were normalized to a standard range to stabilize the optimization process based on the backbones' gradient norm and improve convergence. To ensure good generalization and avoid overfitting due to limited data, the training subset was further expanded by randomly rotating, horizontally flipping and zooming the data. The class labels were one-hot encoded for training and evaluation for the four diagnostic classes: categorical cross-entropy optimization.

$$Inorm(x,y) = \frac{I(x,y) - Imin}{Imax - Imin} \quad (1)$$

To normalize each image to the range [0, 1] the raw pixel values are rescaled by equation (1).

$$Ir(i,j) = I\left(i \cdot \frac{H}{H'}, j \cdot \frac{W}{W'}\right) \quad (2)$$

Equation (2) takes each pixel coordinate in the resized image, and maps it back to the original image resolution.

$$yk = 1\ if\ k = c, else\ 0 \quad (3)$$

One-hot encoding of class label c is defined as equation (3).

$$Iaug = T\theta, f, z(I) \quad (4)$$

To produce the augmented samples, we use a rotation, flipping and a zoom transformation (T) defined in equation (4).

$$LCE = -\sum_{i=1}^{C} yi\ log(\hat{y}i) \quad (5)$$

The categorical cross-entropy loss function to be minimized in the training of networks is given by equation (5) for C classes.

IV. Proposed Deep Transferable Representation Learning Framework

The framework is based on the following ideas: deep image representation learning using convolutional backbones pre-trained on large-scale corpora of natural images, followed by transfer learning for brain tumor classification from MRI images that was restricted to either a shallow representation learning or an architecture based on traditional deep learning classifiers. The preprocessed MRI images are fed into one of four candidate architectures namely, ResNet50V2, VGG16, DenseNet121 or EfficientNetB0, all of which are feature extractors that convert the raw MRI pixel inputs into a concise, high-level representation of the image. The convolutional base of each pretrained network is kept as is, and it is pre-trained with filters that are hierarchical from bottom up: it's filters at low levels capture low level edges and textures, and the deeper levels capture more abstract and tumour-relevant filters. The outputs from the corresponding spatial feature maps are pooled together in a global average pooling layer, before being sent to a fully connected layer that includes dropout regularization and a final softmax layer that generates class probabilities in the four diagnostic categories for a particular task. The training set is used for minimization of categorical cross-entropy loss with an adaptive gradient-based optimizer; and learning-rate adaptation and early-stopping are based on the validation set. By exploiting the visual representations that have been learnt in prior architectures, this transfer-learning approach results in significantly lower amount of labelled MRI data needed for achieving good convergence between the two and a controlled, architecture-level comparison of the quality of the transferable features for multi-class brain tumor classification. The proposed framework consists of MRI preprocessing, partitioning of MRI datasets, transfer-learning models, deep feature extraction and Softmax classification, as shown in Figure 1. ResNet50V2, VGG16, DenseNet121 and EfficientNetB0 models are compared to evaluate performance and ResNet50V2 is found to be the most successful model for four-class tumor classification.

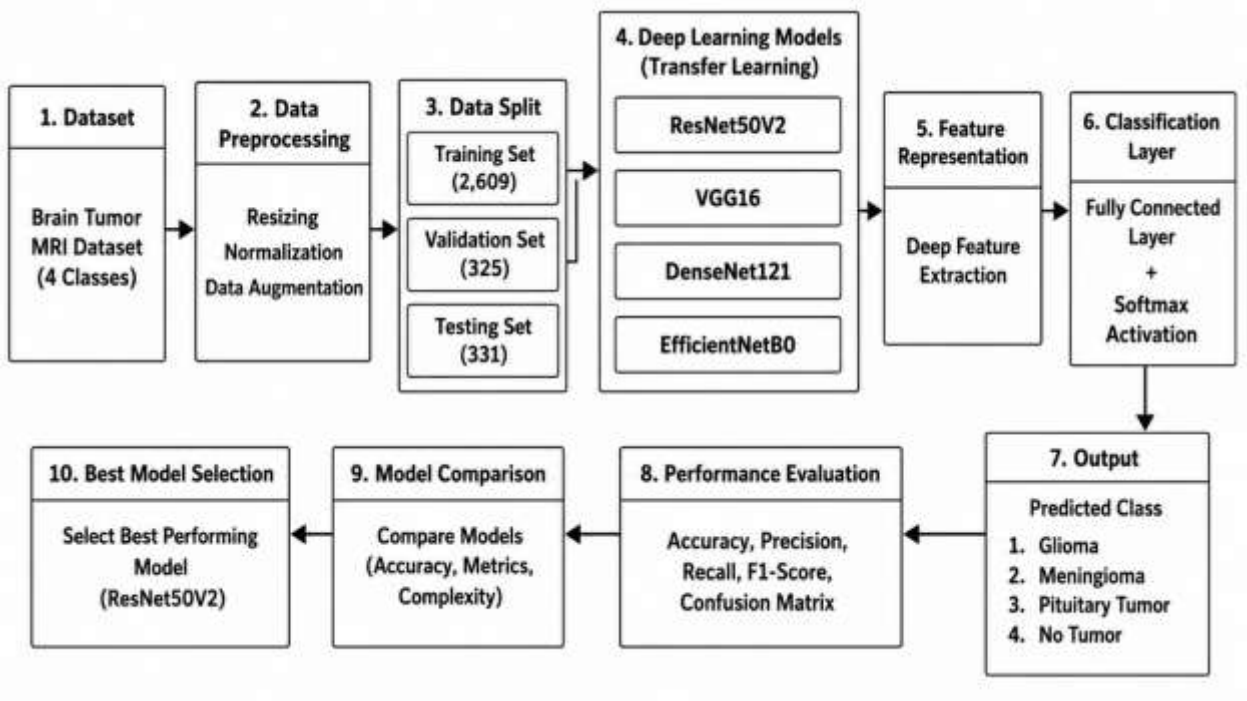

Figure 1: Proposed Deep Transferable Representation Learning Framework for Multi-Class Brain Tumor MRI Classification

A. Deep Feature Extraction Using ResNet50V2, VGG16, DenseNet121, and EfficientNetB0

1) ResNet50V2:

In ResNet50V2, pre-activated residual blocks are used which means that batch norm and activation comes before convolution, and identity shortcut connections are used which enables gradients to pass straight through layers. This residual formulation helps tackling vanishing-gradient degradation in deep networks and helps to retain fine-grained information on the tumor boundaries in the network through 50 layers on the brain tumor MRI dataset while fine-tuning.

$$y = F(x, \{Wi\}) + x \quad (6)$$

$$xl + 1 = xl + F\left(\sigma\left(BN(xl)\right)\right) \quad (7)$$

2) VGG16:

A series of small 3×3 convolutional filters are applied sequentially followed by max pooling layers, where the number of channels is increased and the spatial dimensions are decreased along the way as in VGG16. The uniform architecture results in smooth, hierarchical feature maps however it has the disadvantage of a large number of fully-connected parameters resulting in a large model when compared to more recent architectures.

$$zi,j = \sigma(\Sigma m,n\, wm,n\, xi + m, j + n + b) \quad (8)$$

$$pi,j = max(xi{:}\, i + k, j{:}\, j + k) \quad (9)$$

3) DenseNet121:

The dense block DenseNet121 creates dense connections between all layers in the network, but only by concatenating the features in each layer along the channel dimension, thus reusing features and minimizing the number of parameters needed to obtain good representational power. This tight connectivity enhances the flow of the gradients and facilitates spreading low level tumor texture information to deeper layers of classification.

$$xl = Hl([x0, x1, \ldots, xl - 1]) \quad (10)$$

4) EfficientNetB0:

In addition to mobile inverted bottleneck convolutions and squeeze-and-excitation attention, EfficientNetB0 uses a compound scaling to balance the network depth, width and input resolution simultaneously with a single scaling coefficient. Although the scaling coefficients are optimized for the ImageNet scale data, this design can be a good choice for natural image tasks due to its favorable accuracy-to-parameter ratio, and may need to be carefully re-calibrated when applied to a smaller, domain-specific medical dataset.

$$depth: d = \alpha\varphi,\ width: w = \beta\varphi,\ resolution: r = \gamma\varphi \quad (11)$$

Equations (6)–(11) are the residual mapping, pre-activation residual unit, convolution and max-pooling operations, dense-block concatenation and EfficientNet's compound scaling rule respectively.

B. Feature Representation and Multi-Class Brain Tumor Classification

After convolutional feature extraction, a global average pooling operation was applied to the backbone to project each backbone into a feature vector and then the obtained vectors from each backbone were fed into fully connected layers with dropout regularization to avoid overfitting. The logits obtained are then passed through a final dense layer using a softmax activation function, which outputs the probabilities of the classes for the glioma, meningioma, pituitary tumor and no-tumor classes, and a class is selected as the predicted diagnostic label based on the highest probability produced by the softmax for that class.

$$vc = \left(\frac{1}{H \times W}\right) \cdot \Sigma i, j\ Fc(i, j) \quad (12)$$

Global average pooling is done in each channel and the number of feature maps is reduced to 1 using the following equation as shown below in eq. (12).

$$z = Wfc\ v + bfc \quad (13)$$

In Equation (13), the transformation of pooled features to class logits is fully-connected.

$$\hat{y}i = \frac{ez_i}{\sum_{j=1}^{C} ez_j} \quad (14)$$

In Equation (14), softmax activation function is used to transform logits to class probabilities.

$$\hat{y} = argmaxi\ (\hat{y}i) \quad (15)$$

In equation (15) the class label with the highest probability will be chosen as the predicted class label.

$$d = 1 - \left(\frac{1}{n}\right) \cdot \sum_{i=1}^{n} m_i \quad (16)$$

The dropout regularization (16) is a random deactivation of a portion of the neurons while training.

V. Experimental Setup and Evaluation Metrics

A. Model Configuration, Training Parameters, Implementation Setup

The weights of all four architectures (ResNet50V2, VGG16, DenseNet121 and EfficientNetB0) were initialized using pre-trained convolutional weights from ImageNet and fine-tuned based on the brain tumor MRI training subset. For each backbone, a global average pooling layer, dropout layer and a dense softmax classification head making predictions for the 4 diagnostic classes was used. Adaptive gradient based optimizer with categorical cross-entropy loss was used to train the models and the validation subset was used to track loss converge, reduce learning rate and apply early stopping to avoid overfitting. The held-out testing subset was used for the evaluation of performance with regards to accuracy, precision, recall and F1-score, both at the per class level and macro and weighted averaging, due to the multi-class classification setting.

Table 3. Model Training Hyperparameters

Hyperparameter	Configuration
Input Image Size	224 × 224 × 3
Batch Size	32
Optimizer	Adam
Initial Learning Rate	1×10^{-4}
Loss Function	Categorical Cross-Entropy
Epochs	50 (with early stopping)
Dropout Rate	0.5
Pretrained Weights	ImageNet

Data Augmentation	Rotation, horizontal flip, zoom
Evaluation Metrics	Accuracy, Precision, Recall, F1-score

Table 3 lists the training configuration that is shared by all four architectures. The same hyperparameters were used throughout the models to avoid training-configuration bias, with the sole exception of the number of channels for the 1D and 2D models. All the models had the same hyperparameters except for the number of channels used in the 1D models and 2D models, to avoid training-configuration bias.

VI. Results and Performance Analysis

The comparative classification performance of the four architectures using transfer learning is summarized, and then detailed class-wise and overall analysis is performed for the best performing architecture (ResNet50V2) and a comparison of the classification performance to model complexity is carried out.

A. Comparative Classification Performance of Deep Learning Models

The test accuracy of the four architectures evaluated are compared in Table 4. The high test accuracy of ResNet50V2 (94.26%) and VGG16 (93.66%) was obtained, which is very close to the DenseNet121 (92.45%) accuracy, showing that the residual and densely-connected backbones successfully transferred to the brain tumor MRI domain.

Table 4. Comparative Performance of Deep Learning Models for Brain Tumor Classification

Model	Test Accuracy	Test Accuracy (%)	Rank
ResNet50V2	0.942598	94.26	1
VGG16	0.936556	93.66	2
DenseNet121	0.924471	92.45	3
EfficientNetB0	0.287009	28.70	4

EfficientNetB0, on the other hand, was found to achieve 28.70% accuracy with the same training settings, indicating that their compound-scaling values are not well matched when using this relatively small medical dataset under this relatively small fine-tuning budget. This large performance gap emphasizes the non-uniform transferability across architectures and the ability of residual connections and dense feature reuse over the ability of constrained medical-imaging fine-tuning for compound-scaled mobile-inverted convolutions.

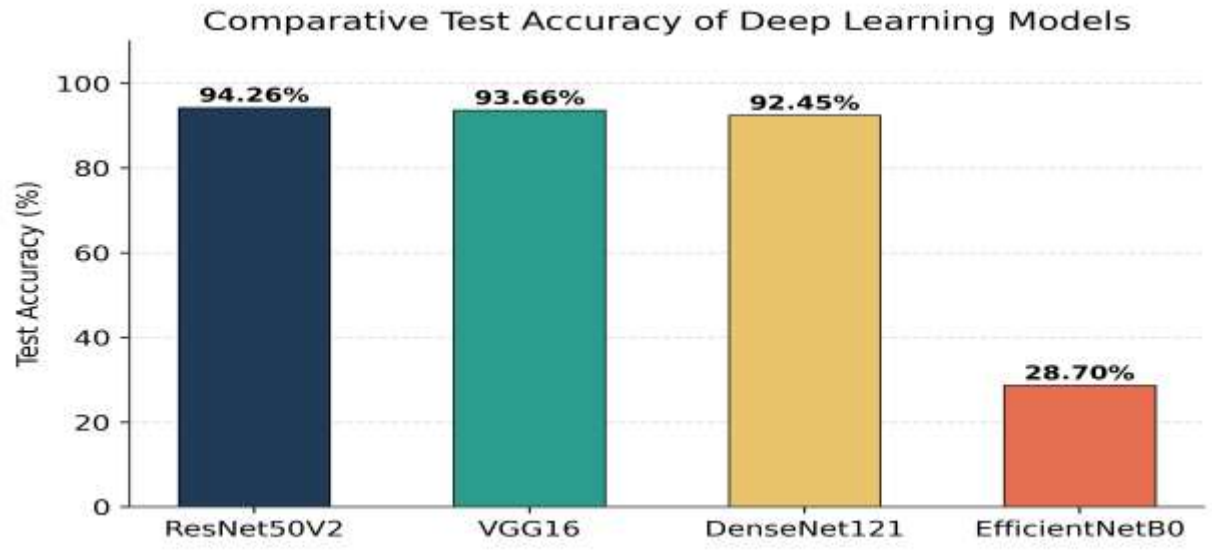

Figure 2. Comparative test accuracy of ResNet50V2, VGG16, DenseNet121, and EfficientNetB0.

As illustrated in Figure 2, there is a significant difference between the accuracy of the sharp architectures and the EfficientNetB0, with ResNet50V2, VGG16, and DenseNet121 above 92%.

B. Class-Wise and Overall Performance of ResNet50V2

The precision, recall and F1-score for class-wise classification results of the ResNet50V2 model is given in Table 5. Results of the classification of pituitary tumors yielded the best precision of 96.81% and a perfect 100% recall, meaning that ResNet50V2 was able to correctly identify all the pituitary tumor cases in the test subset.

Table 5. Class-Wise Classification Performance of the ResNet50V2 Model

Brain Tumor Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Glioma Tumor	90.43	90.43	90.43	94
Meningioma Tumor	94.62	92.63	93.62	95
No Tumor	96.00	94.12	95.05	51
Pituitary Tumor	96.81	100.00	98.38	91
Macro Average	94.46	94.29	94.37	331
Weighted Average	94.24	94.26	94.24	331

The percentage of F1 score was obtained as 95.05% for no-tumor cases as well. Glioma tumor classification had the least accuracy with a precision and Recall of 90.43%, indicating that the cases of glioma have a greater visual similarity to other tumor types and hence are more likely to be misclassified. The overall performance of the meningioma tumor classification was 93.62% in terms of F1-score. Overall, the F1-scores of ResNet50V2 (macro-averaged F1 = 94.37% and weighted-averaged F1 = 94.24%) are very stable across all four classes despite the differing class support.

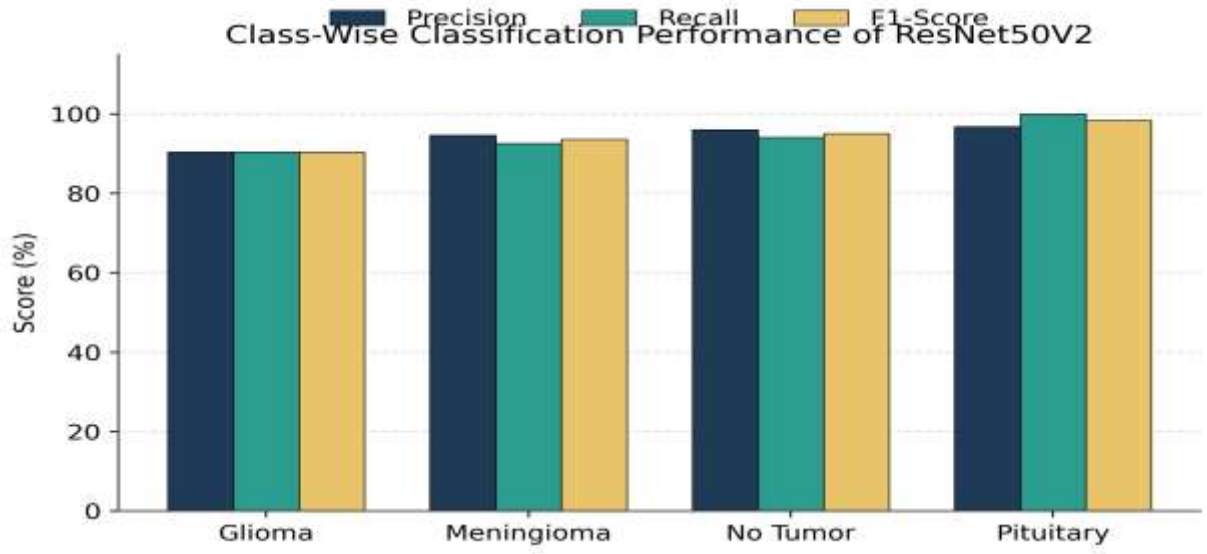

Figure 3. Class-wise precision, recall, and F1-score of the ResNet50V2 model.

Figure 3 demonstrates that the pituitary tumor outscores all three measures, whereas, the glioma tumor consistently has the lowest scores, and has more visual overlap between classes than the other tumors.

Table 6. Overall Classification Performance of the Best-Performing ResNet50V2 Model

Performance Metric	Result (%)
Overall Test Accuracy	94.26
Macro Precision	94.46
Macro Recall	94.29
Macro F1-Score	94.37
Weighted Precision	94.24
Weighted Recall	94.26
Weighted F1-Score	94.24

The overall accuracy and macro/weighted averaged values for ResNet50V2 are summarized in tabble 6. The good agreement of these two averages (both around 94.2–94.4%) suggests that no class-support imbalance is

skewing the performance and that the model does not focus on the majority class type, but rather works well for all four tumor types.

Figure 4. Overall accuracy, macro-averaged, and weighted-average performance metrics of ResNet50V2.

As can be seen in Figure 4, the performance of all seven metrics is within a very small range of 94.2–94.5% when classified as an overall performance metric, thus visually confirming the stability and balance of the classification performance of ResNet50V2.

C. Model Complexity and Classification Performance Analysis

Table 7 shows the relationship between the complexity of the model and accuracy of classification. The most accurate model (ResNet50V2) has a total number of parameters around 23.97 million, 10% of which is trainable (401,412), and requires 91.42 MB of model size, which means that the majority of convolutional layers in the network were kept frozen and only a small classification head was trained.

Table 7. Model Complexity and Classification Performance Analysis

Model	Total Params	Trainable Params	Non-Trainable Params	Model Size (MB)	Test Accuracy (%)
ResNet50V2	23,966,212	401,412	23,564,800	91.42	94.26
VGG16	14,815,044	100,356	14,714,688	56.51	93.66
DenseNet121	7,238,212	200,708	7,037,504	27.61	92.45
EfficientNetB0	4,300,455	250,884	4,049,571	16.40	28.70

With 14.82 million total parameters, VGG16 was able to match the accuracy (93.66%) with a moderate size of 56.51 MB. In terms of accuracy, DenseNet121 had the smallest possible size with 27.61 MB and 7.24 million parameters and provided an attractive trade-off between the accuracy and size. Despite the lowest number of parameters (4.30 million) and smallest architecture size (16.40 MB), EfficientNetB0 only reached the accuracy result of 28.70%, which is a clear proof of the fact that it is possible to be parameter efficient without being transferable as well, and the architecture-architecture compatibility is important to fine-tune the results of this dataset.

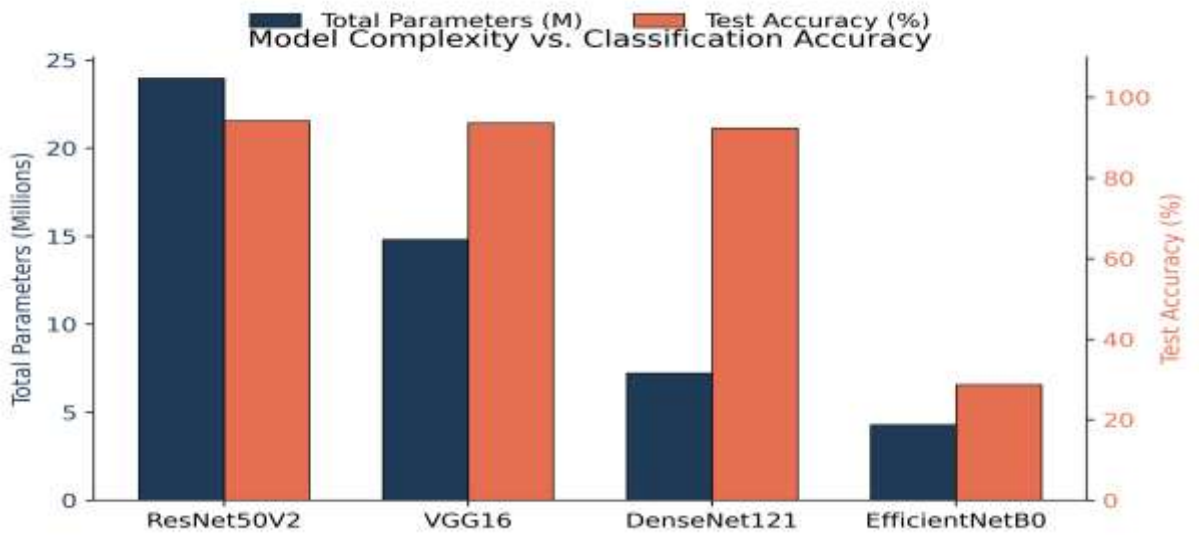

Figure 5. Model complexity (total parameters) versus test accuracy across the four architectures.

However, as seen in Figure5, the number of parameters is not always correlated with accuracy, as in the case of EfficientNetB0 which has a small number of parameters but a low accuracy. This indicates that the transferability is not only dependent on the parameters of the model, but also on other architecture-specific factors.

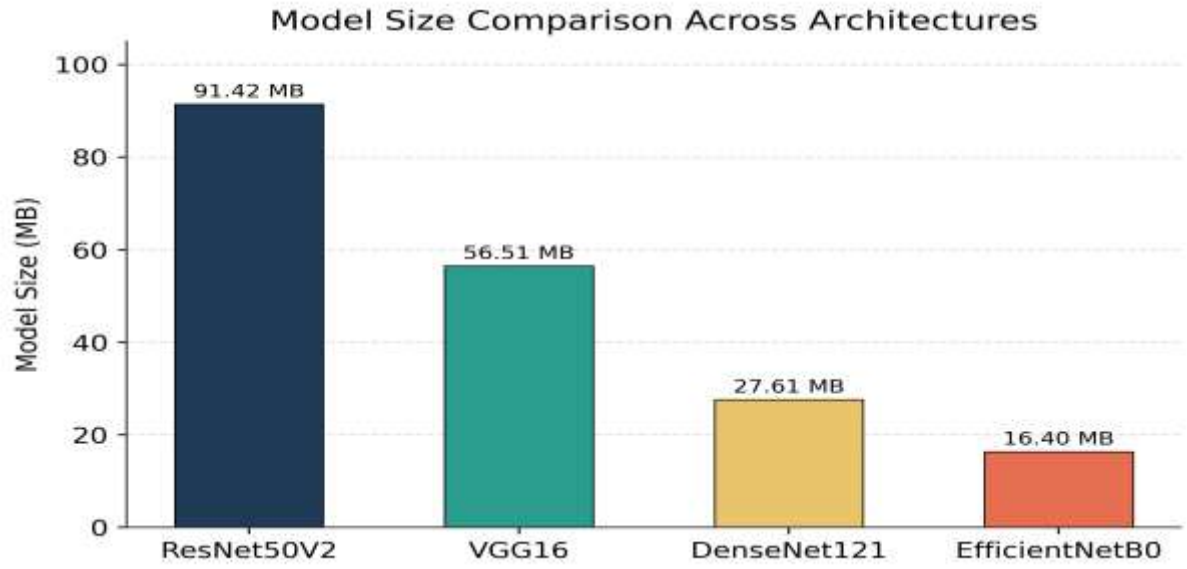

Figure 6. Comparative storage footprint (MB) of the four evaluated architectures.

The significant difference in storage size between ResNet50V2 and the smaller DenseNet121 and EfficientNetB0 models is important for use in deployment in constrained clinical environments, as shown in figure 6.

D. Training Curves and Confusion Matrices from Experimental Runs

The curves and confusion matrices in this subsection show the raw convergence results of model training and validation on the training sets and test sets, respectively, from the actual training runs, as a clear and per-architecture representation of convergence and of the error per class in the training sets.

ResNet50V2: The training accuracy has been approaching almost 1.0 for the first ten epochs, and the validation accuracy has been converging to around 0.88–0.90, while the loss on validation data has been dropping sharply initially followed by a slight increase, showing that the model has been converging quickly with a relatively

bounded validation loss, as seen in figure 7. The confusion matrix in figure 8 shows that the cases of pituitary tumor are correctly classified (91/91), and the same applies for both glioma and meningioma/no-tumor classes, where most of the samples are classified correctly.

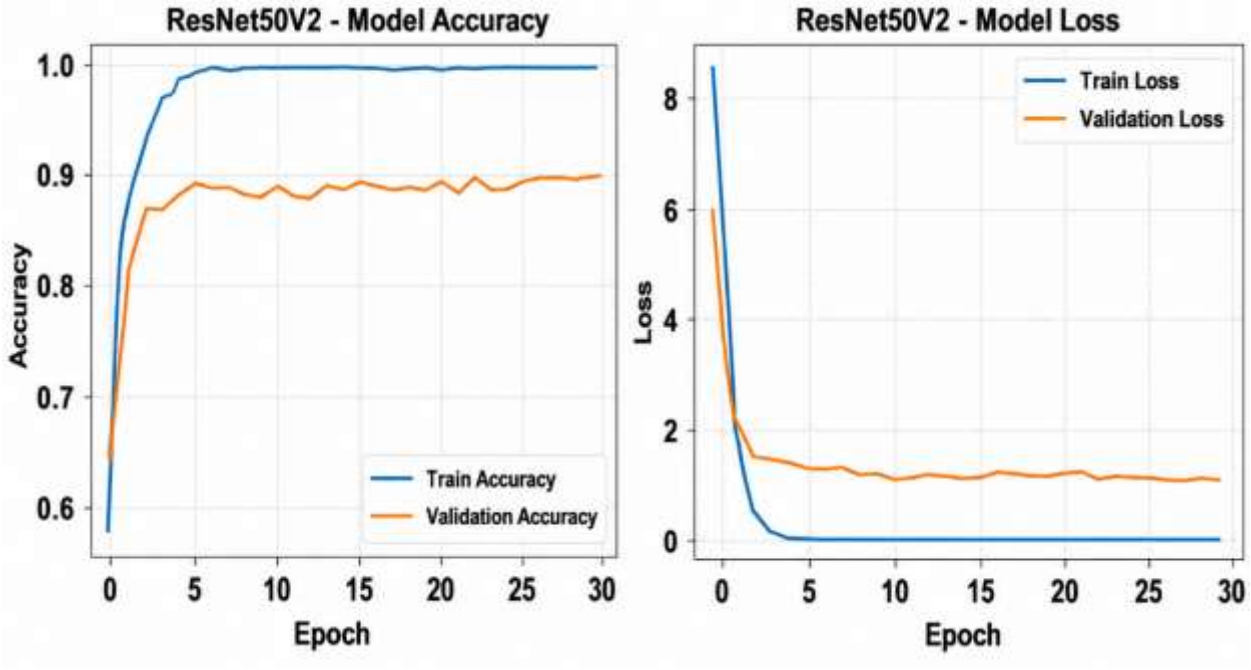

Figure 7. ResNet50V2 training/validation accuracy and loss curves over 30 epochs.

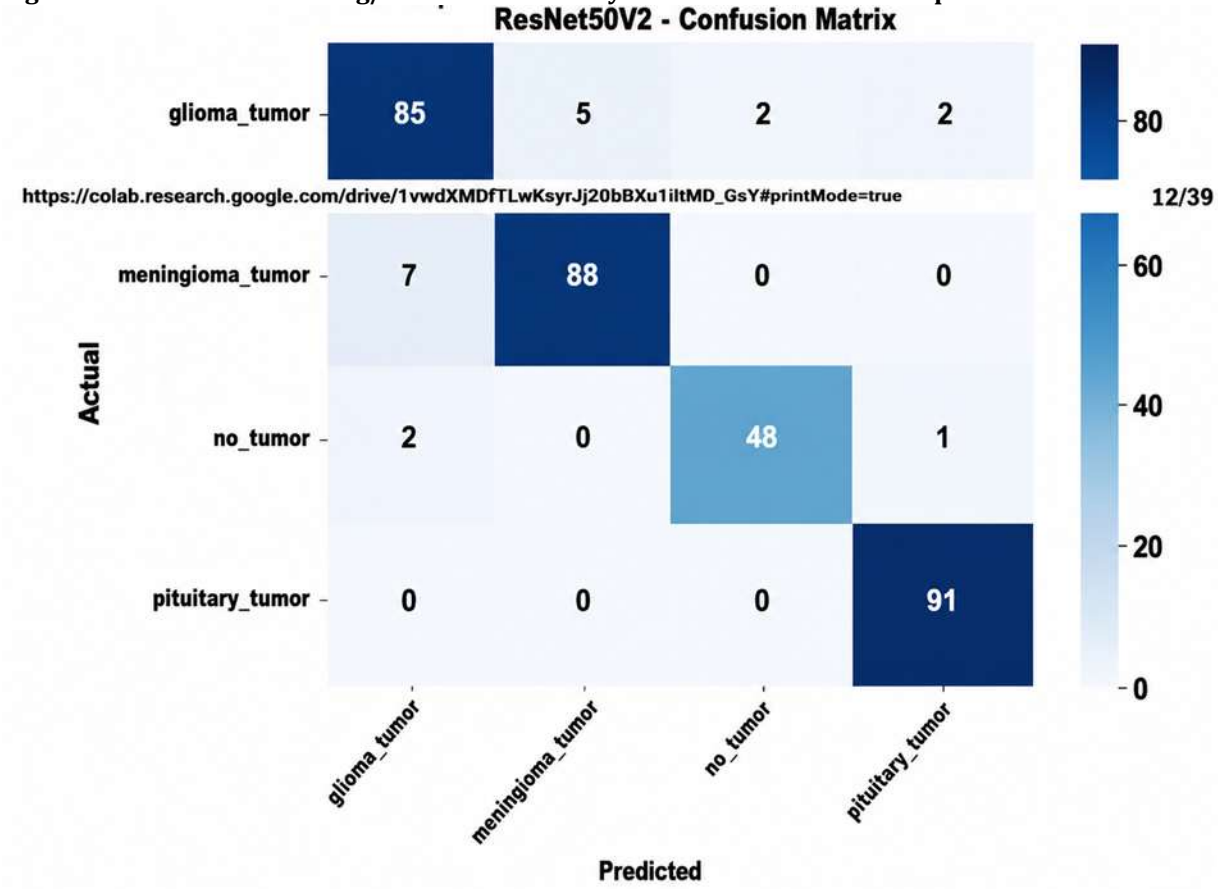

Figure 8. ResNet50V2 test-set confusion matrix (raw prediction counts across the four classes).

DenseNet121: Furthermore, training accuracy was achieved as early as 10 epochs and the validation accuracy increased continuously to 0.92–0.93 with a relatively small gap between train and validation loss as compared to ResNet50V2, which is consistent with DenseNet121's parameter-efficient dense connections as seen in figure 8. The confusion matrix has high diagonal elements in all four classes, the meningioma tumor having the highest number of off-diagonal elements (9 glioma, 3 pituitary) and no-tumor and pituitary tumor are highly separable.

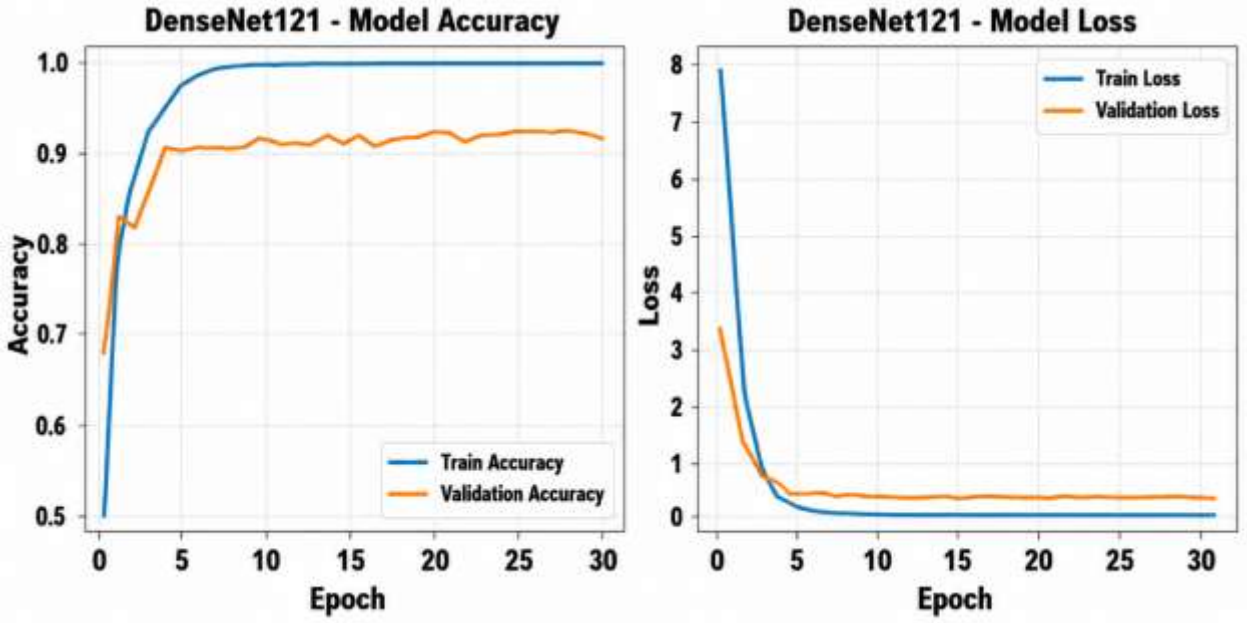

Figure 9. DenseNet121 training/validation accuracy and loss curves over 30 epochs.

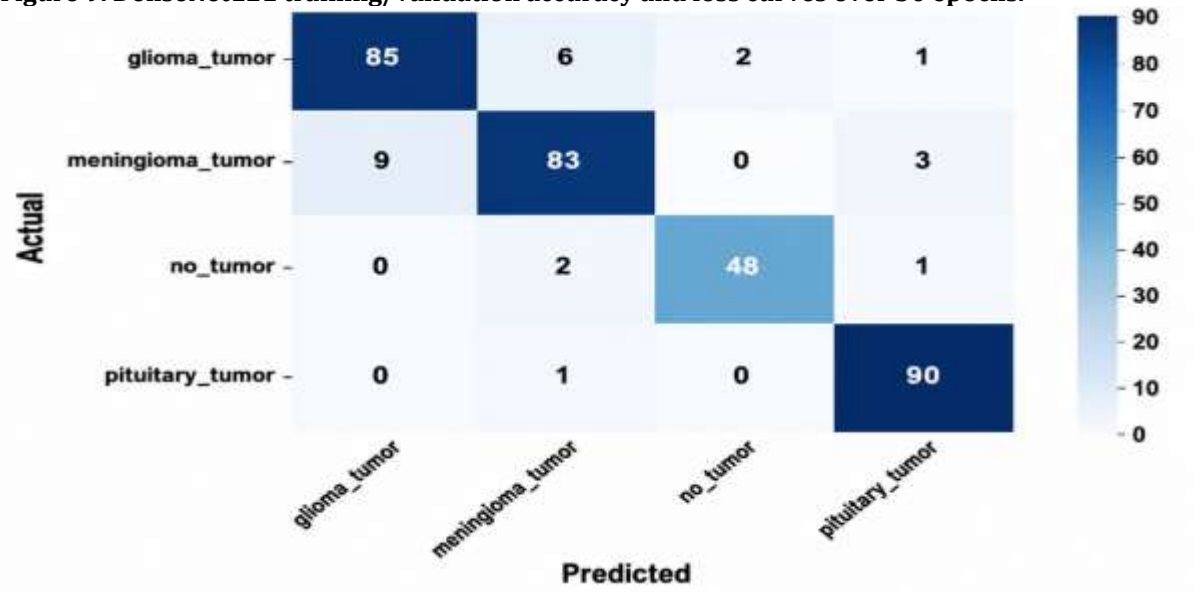

Figure 10. DenseNet121 test-set confusion matrix (raw prediction counts across the four classes)

The confusion matrix for four classes (Brain Tumor) classification is shown in Figure 10. The model accurately classifies 85 glioma cases, 83 meningioma cases, 48 no-tumor cases and 90 pituitary cases, while having relatively few misclassifications, with a few being classified as no-tumor and a few are classified as pituitary.

VGG16: The training curves shown in figure 11 increase at a slow pace compared to the residual/dense architectures, and the validation accuracy still progress steadily to around 0.90–0.92, whereas the validation loss slowed down steadily down to around 0.27–0.28, showing that there is still room to run a few more epochs.

The confusion matrix reveals that the glioma and meningioma tumors are accurately classified with a rate of 85/94 and 85/95 respectively, and again perfectly classified in the case of pituitary tumor (91/91).

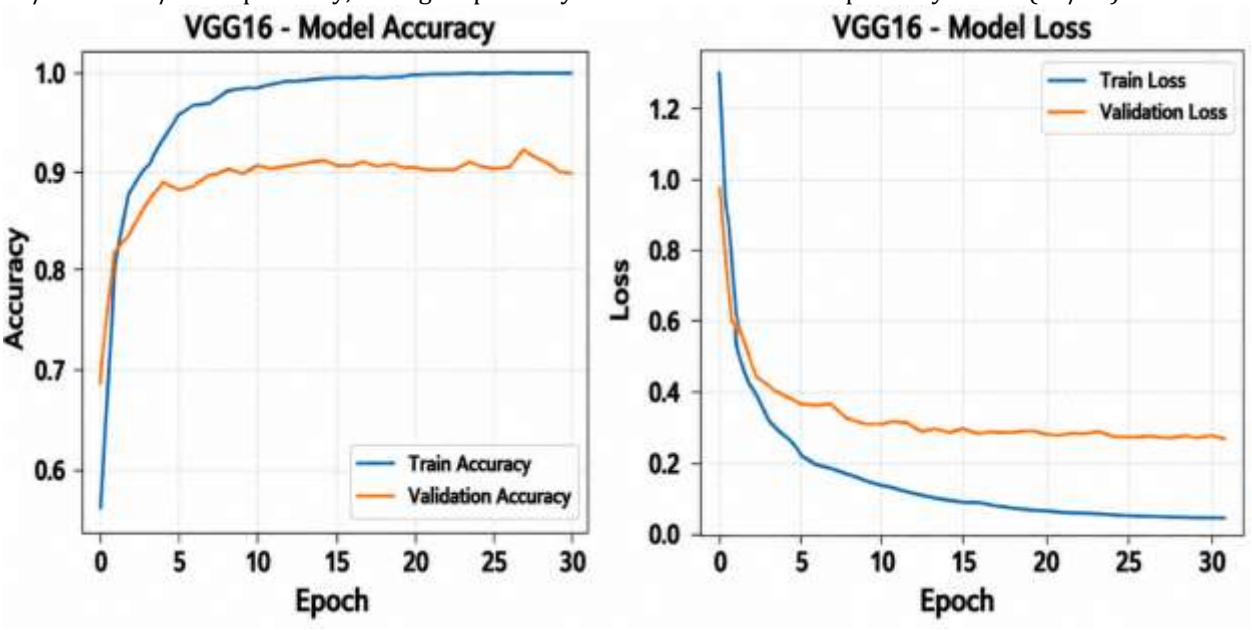

Figure 11. VGG16 training/validation accuracy and loss curves over 30 epochs.

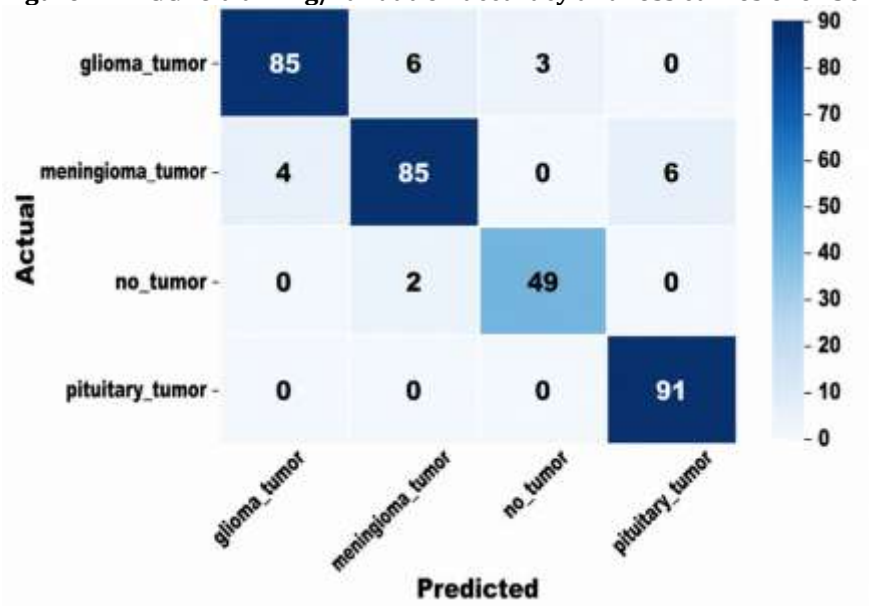

Figure 12. VGG16 test-set confusion matrix (raw prediction counts across the four classes)

The confusion matrix for brain tumor classification into four classes is shown in figure 12 with good discrimination and low inter-class predictions errors for the category glioma, meningioma, no tumor and pituitary (85, 85, 49 and 91 cases respectively correctly classified).

EfficientNetB0: The training curves are quite unstable with both training and validation accuracy oscillating in a small range of 0.15-0.35 across all 30 epochs without touching the convergence as seen from figure 13, and loss doesn't seem to decrease monotonously, but oscillates randomly. The confusion matrix is clear – the model was just not learning a discriminative feature as the confusion is more because of its failure to learn than because of under-performance, as the model has ended up predicting meningioma_tumor for nearly all the test samples irrespective of ground truth class.

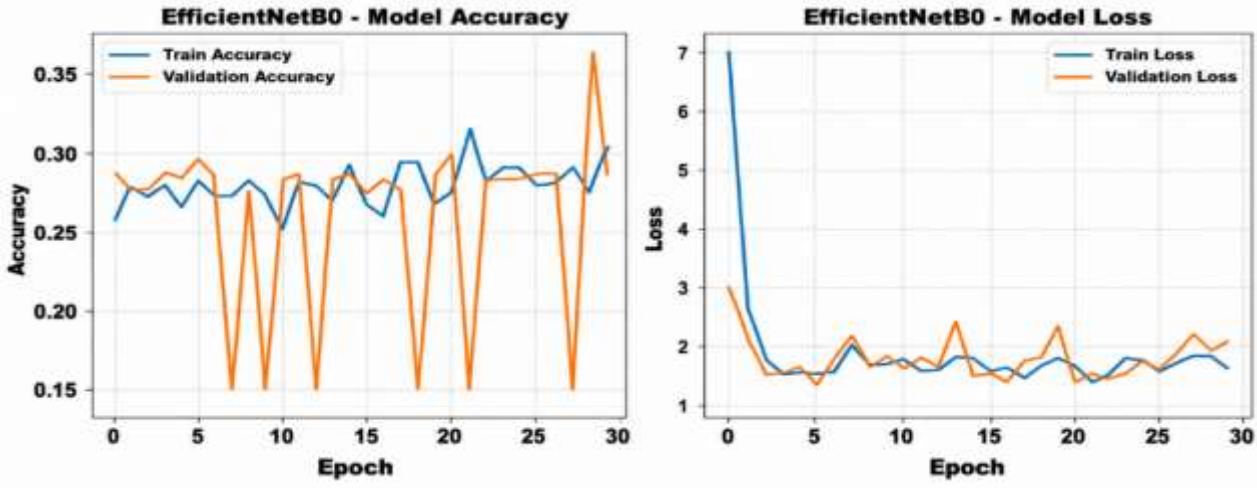

Figure 13. EfficientNetB0 training/validation accuracy and loss curves over 30 epochs

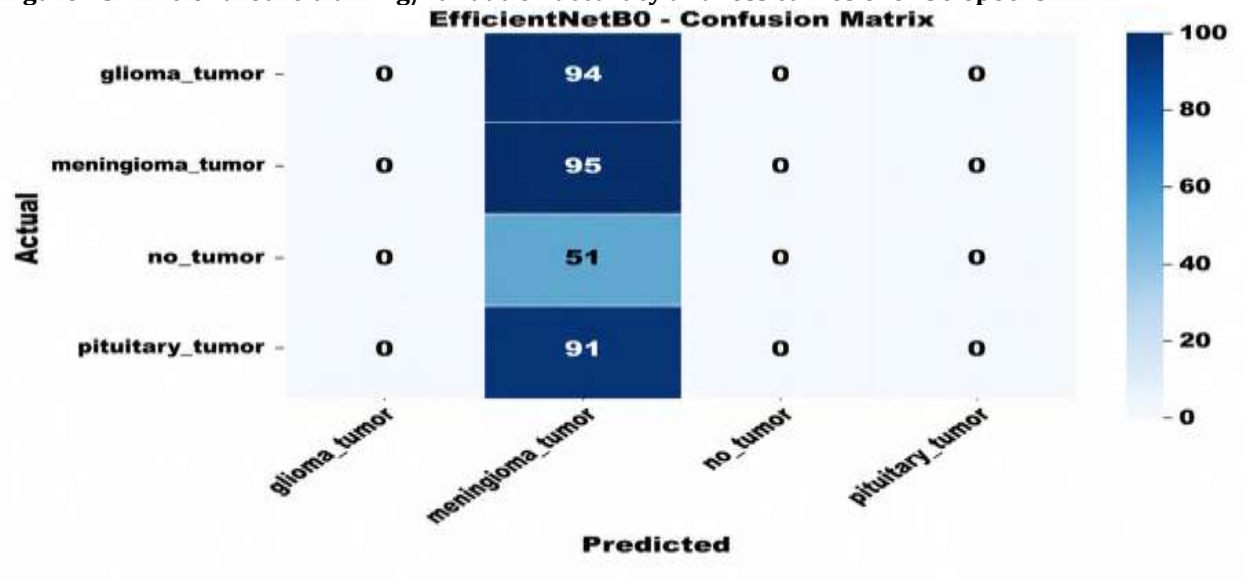

Figure 14. EfficientNetB0 test-set confusion matrix, showing collapse toward a single predicted class

The EfficientNetB0 confusion matrix is shown by figure 14 and there is a high prediction bias in favor of the meningeoma class. The multiclass discrimination and generalization is poor since all glioma, no-tumor and pituitary samples are misclassified as meningioma and 95 meningioma samples are correctly classified.

VII. Discussion

The comparative results show the superiority of the residual/dense networks to the compound-scaled mobile networks in the fine-tuned scenario used in this paper for brain tumor classification on MRI images. The pre-activation residual connections in ResNet50V2 seem to be better at retaining fine-grained information around the tumor boundary and the texture of the tumor, resulting in the best test accuracy, macro F1-score and weighted F1-score of the four models. VGG16 also used a simple, sequential, convolutional stack, with results that had a much lower accuracy, but were reasonably similar to the ones obtained by the best-known 2D CNN. DenseNet121 achieved the best accuracy-to-model-size ratio, thanks to its dense feature-reuse connectivity and is therefore a promising choice for scenarios where storage or inference-time requirements are crucial. The class wise analysis shows that the accuracy and recall of the pituitary tumors were high, probably due to their relatively distinct shape and location in the cranial cavity, while the accuracy and recall of glioma cases are comparatively low, which is in line with the high morphological heterogeneity of gliomas and its infiltrative growth pattern, where gliomas can visually resemble the surrounding tissue or other type of lesions.

The surprisingly poor accuracy of EfficientNetB0 is interesting; the compound-scaling coefficients were optimized on a large collection of natural images, and the small number of fine-tuning epochs and relatively small medical dataset applied here might have been too few to allow the adapted depth-width-resolution to reach accuracy levels in the medical domain. This suggests that not every efficient architecture seems to be equally transferable without further tuning. The results show a clear trade-off from a computational-complexity point of view, with the more accurate model (ResNet50V2) being the larger model, and the smallest model (EfficientNetB0) being the least accurate model, which means that practitioners have to choose between a more accurate model and a smaller model for deployment on the resource-limited clinical hardware.
There are a number of caveats to note, however. Experimental subset is balanced yet relatively small compared to the entire benchmark set which could prevent generalizations to other patient populations or scanners. In addition, cross-institutional external validation, an analysis of explainability of learned features, and segmentation-level localization were not included in the study, which would further enhance the clinical utility. Further work to overcome the limitations outlined above, and more fine-tuning of EfficientNetB0, would provide more evidence of the robustness of the proposed transferable representation learning framework.

VIII. Conclusion and Future Work

The authors designed and tested a deep transferable representation learning approach for multi-class brain tumour MRI classification, basing their experiments on a common protocol and comparing the performance of ResNet50V2, VGG16, DenseNet121 and EfficientNetB0. The overall performance of ResNet50V2 is the best with test accuracy of 94.26%, macro F1 of 94.37% and near perfect classification of the pituitary tumor, whereas the other architectures underperformed significantly, and EfficientNetB0 is significantly worse than the others. A class-wise analysis revealed that cases of glioma were the most difficult to classify and those of pituitary tumor and none were most reliably classified. From the results of the complexity analysis, it can be seen that DenseNet121 has a good accuracy-to-size ratio, which could be considered as a suitable alternative when the storage capacity and calculation rate are limited. All these results demonstrate the usefulness of these general representations for automated multi-class brain tumour diagnosis on MRI, which can be transferred from residual and densely connected convolutional networks. This work will be continued by evaluating on larger, multi-institutional datasets, adding explainable-AI techniques to ensure clinical interpretability of results, introducing tumor segmentation along with classification, and other fine-tuning strategies to make lightweight architectures like EfficientNetB0 more transferable towards real-world clinical deployment.

References

[1] M. A. Khan, M. Z. Hussain, S. Mehmood, M. F. Khan, M. Ahmad, T. Mazhar, T. Shahzad, and M. M. Saeed, "Transfer learning for accurate brain tumor classification in MRI: a step forward in medical diagnostics," Discover Oncology, vol. 16, no. 1, p. 1040, 2025. https://doi.org/10.1007/s12672-025-02671-4

[2] S. K. Mathivanan, S. Sonaimuthu, S. Murugesan, et al., "Employing deep learning and transfer learning for accurate brain tumor detection," Scientific Reports, vol. 14, p. 7232, 2024. https://doi.org/10.1038/s41598-024-57970-7

[3] A. Ghosh, "Attention-Guided Multi-Class Brain Tumor Classification from MRI Scans Using Deep Transfer Learning and Explainable Artificial Intelligence," SSRN, 2026. https://dx.doi.org/10.2139/ssrn.6879940

[4] S. I. Tonni, M. A. Sheakh, M. S. Tahosin, M. Z. Hasan, T. F. Shuva, T. Bhuiyan, M. A. A. Almoyad, N. A. Orka, M. T. Rahman, R. T. Khan, M. S. Kaiser, and M. A. Moni, "A Hybrid Transfer Learning Framework for Brain Tumor Diagnosis," Advanced Intelligent Systems, vol. 7, p. 2400495, 2025. https://doi.org/10.1002/aisy.202400495

[5] A. Ali, X. Li, W. K. Mashwani, M. Abiad, F. K. Karim, and S. M. Mostafa, "Multi-class brain tumor MRI segmentation and classification using deep learning and machine learning approaches," Cancer Imaging, vol. 25, no. 1, p. 131, 2025. https://doi.org/10.1186/s40644-025-00953-2

[6] Gajula, S., Kandula, S.T.R. (2026). Securing Financial Data in Multi-Tenant Clouds Through AI, Blockchain, and Attribute-Based Encryption. In: Nguyen, GN., Swaroop, A., Shukla, P. (eds) Proceedings of Fifth International Conference on Computing and Communication Networks. ICCCN 2025. Lecture Notes in Networks and Systems, vol 1859. Springer, Cham. https://doi.org/10.1007/978-3-032-21499-7_33

[7] E. Zarenia, A. A. Far, and K. Rezaee, "Automated multi-class MRI brain tumor classification and segmentation using deformable attention and saliency mapping," Scientific Reports, vol. 15, p. 8114, 2025. https://doi.org/10.1038/s41598-025-92776-1

[8] M. S. Elhadidy, A. T. Elgohr, M. El-geneedy, S. Akram, and H. M. Kasem, "Comparative analysis for accurate multi-classification of brain tumor based on significant deep learning models," Computers in Biology and Medicine, vol. 188, p. 109872, 2025. https://doi.org/10.1016/j.compbiomed.2025.109872
[9] Sunkara, S. P. (2025). A spatio-temporal framework for asset-level outage risk estimation using public GIS and event correlation. International Journal of Computer Engineering and Technology, 16(1), 4211–4227. https://doi.org/10.34218/IJCET_16_01_286
[10] Kakade, S. V., Dabade, T. D., Patil, V. C., Vidyapeeth, K. V., Ajani, S. N., Bahulekar, A., and Sawant, R., "Examining the Social Determinants of Health in Urban Communities: A Comparative Analysis," South Eastern Eur. J. Public Health, vol. 21, pp. 111–125, 2023.
[11] G. S. Tandel, A. Tiwari, and O. G. Kakde, "Multi-Class Brain Tumor Grades Classification Using a Deep Learning-Based Majority Voting Algorithm and Its Validation Using Explainable-AI," Journal of Imaging Informatics in Medicine, vol. 38, no. 5, pp. 2793–2830, 2025. https://doi.org/10.1007/s10278-024-01368-4
[12] P. Priyadarshini, P. Kanungo, and T. Kar, "Multigrade brain tumor classification in MRI images using Fine tuned efficientnet," e-Prime — Advances in Electrical Engineering, Electronics and Energy, vol. 8, p. 100498, 2024. https://doi.org/10.1016/j.prime.2024.100498
[13] B. D. Shivahare, S. K. Subramaniam, Dafik, et al., "An advanced hybrid deep learning framework for high-precision brain tumor detection and classification in MRI scans," Scientific Reports, vol. 16, p. 19866, 2026. https://doi.org/10.1038/s41598-026-50194-x
[14] R. Disci, F. Gurcan, and A. Soylu, "Advanced Brain Tumor Classification in MR Images Using Transfer Learning and Pre-Trained Deep CNN Models," Cancers, vol. 17, no. 1, p. 121, 2025. https://doi.org/10.3390/cancers17010121
[15] E. F. Gomes and R. S. Barbosa, "Deep Learning Approaches for Brain Tumor Classification in MRI Scans: An Analysis of Model Interpretability," Applied Sciences, vol. 16, no. 2, p. 831, 2026. https://doi.org/10.3390/app16020831
[16] S. K. Ramareddy, "Integrating AI-Driven Data Visualization and Data Warehouse Analytics for Predictive Decision-Making in Financial and Stock Market Systems," 2026 International Symposium of Systems, Advanced Technologies and Knowledge (ISSATK), Hammamet, Tunisia, 2026, pp. 1-6, doi: 10.1109/ISSATK69667.2026.11566934.
[17] M. Abbas, M. Hassan, R.-Z. Wang, and C.-H. Teng, "Brain Tumor Classification in MRI Images Using Combined Transfer Learning and Convolutional Neural Networks," Journal of Imaging, vol. 12, no. 6, p. 233, 2026. https://doi.org/10.3390/jimaging12060233
[18] A. Amarnath, A. Al Bataineh, and J. A. Hansen, "Transfer-Learning Approach for Enhanced Brain Tumor Classification in MRI Imaging," BioMedInformatics, vol. 4, no. 3, pp. 1745–1756, 2024. https://doi.org/10.3390/biomedinformatics4030095
[19] H. AlShehri and M. Busaleh, "Morphology-Aware Multi-Scale Deep Representation Learning for Interpretable Knowledge Extraction in Brain Tumor MRI," Machine Learning and Knowledge Extraction, vol. 8, no. 5, p. 119, 2026. https://doi.org/10.3390/make8050119
[20] Siva Prasad, K. V., & Anand Reddy, N. (2026). A Multi-Stage Hallucination Mitigation Framework for Reliable Retrieval-Augmented Generation Systems. IJISEA Publications. https://doi.org/10.5281/zenodo.20470413
[21] R. Disci, F. Gurcan, and A. Soylu, "Advanced Brain Tumor Classification in MR Images Using Transfer Learning and Pre-Trained Deep CNN Models," Cancers, vol. 17, no. 1, p. 121, 2025. https://doi.org/10.3390/cancers17010121
[22] F. Uysal and M. Erkan, "Multiclass Classification of Brain Tumors with Various Deep Learning Models," Engineering Proceedings, vol. 27, no. 1, p. 30, 2022. https://doi.org/10.3390/ecsa-9-13367