International Journal on Robotics, Automation and Sciences

# Brain Tumor Detection Using MRI Images

Lim Xin Tian, and S Prabha Kumaresan*

***Abstract* – Brain tumors represent an important clinical disease of the nervous system that requires early and accurate diagnosis for improved patient prognoses. Magnetic Resonance Imaging (MRI) is still the modality of choice for the evaluation of brain tumor due to its better soft tissue contrast and non-invasive nature. Nevertheless, manual interpretation of MRI examinations is difficult and subject to inter-observer variation. The present study aims at a thorough comparative evaluation of deep learning architectures for fully automated, multi-class brain tumor classification based on MRI data. A unified experimental framework is designed to empirically evaluate a custom convolutional neural network (CNNs) trained with the scratch method and four pre-trained transfer learning methods, including VGG16, ResNet50, DenseNet121, and MobileNetV3. The models are trained and validated using a large, consolidated data set of four classes: glioma, meningioma, pituitary tumor and healthy brain images. In order to make sure the comparability between models is true, same preprocessing, data augmentation and dataset partition strategy and the evaluation measure are followed by all the models. Performance is measured in terms of accuracy, precision, recall, F1 score, confusion matrices and roc-auc analysis. Results show that transfer learning models significantly exceed the custom CNN. Among the architectures evaluated, ResNet50 can achieve the best classification performance, test accuracy is 98.38% and macro-averaged F1-score is 0.9837, which can be regarded as good classification and generalisability. These results highlight the power of deep residual learning for solving some of the difficulties in classifying brain MRI, such as inter-class similarity and feature heterogeneity. The study adds a rigorous benchmarking mechanism and empirical evidence for adopting ResNet50 as a robust model for multi-class brain tumor diagnosis. Future research may extend this work by develop a web-based application that enables users to upload MRI images for automated brain tumor detection and classification.**

***Keywords—Brain Tumor Detection, Brain Tumor Classification, Magnetic Resonance Imaging, Deep Learning, Transfer Learning, ResNet50***

## I. INTRODUCTION

Brain tumors are caused by an undesirable proliferation of neural or supporting tissues and can severely affect physiological function with resultant cognitive deficits, visual disturbances and a reduction in quality of life. Early and accurate diagnosis is imperative for good therapeutic planning and increased survival. MRI is the imaging modality of choice for the diagnosis of brain tumor because of the high spatial resolution and excellent soft tissue contrast it provides compared to CT and PET imaging. Despite these benefits, the MRI interpretation process is still formidable requiring expert radiological interpretation, as well as considerable time, especially in distinguishing between tumor types.

Recent progress in deep learning techniques has made it possible to perform automatic analysis of medical images, in particular convolutional neural networks (CNNs), which are able to learn hierarchical

*Corresponding Author email: prabha.kumaresan@mmu.edu.my, ORCID: 0000-0002-0969-7428
Lim Xin Tian is with Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia. (email: 1211110799@student.mmu.edu.my)
S Prabha Kumaresan is with Faculty of Computing and Informatics, Center of Natural Language Processing (CNLP), CoE for Artificial Intelligence, Multimedia University, Cyberjaya, Malaysia. (prabha.kumaresan@mmu.edu.my)

International Journal on Robotics, Automation and Sciences (2026) 8, 2:106-117
https://doi.org/10.33093/ijoras.2026.8.2.13
Manuscript received: 22 Feb 2026 | Revised: 23 Apr 2026 | Accepted: 29 Apr 2026 | Published: : 31 Jul 2026

Published by MMU PRESS. URL: http://journals.mmupress.com/ijoras

feature representations directly from the input medical image data. Numerous experiments have reported a high classification accuracy in detection of brain tumor by CNN based techniques. Nonetheless, existing studies often lack systematic evaluation protocols, often have limited dataset sizes, focus on binary classification problems or lack a systematic comparison between architectures.

The current paper overcomes these shortcomings by proposing a well-framed and unified assessment of multiple deep learning models for multi-class brain tumor classification. A custom CNN along with four popular transfer learning architectures, namely, VGG16, ResNet50, DenseNet121, and MobileNetV3 architectures are evaluated in the same experimental setup.

The main contributions of this work are:

1. A large-scale, balanced multi-class MRI classification framework.
2. A fair and reproducible benchmarking protocol across diverse architectures.
3. A detailed performance analysis identifying ResNet50 as the most effective model.

## II. Literature Review

### A. Traditional Machine Learning Approaches

Early works that studied classification of brain tumors have relied on handcrafted feature extraction methods and classical classification algorithms. Amin et al. [1] proposed a pipeline in which the MRI enhancement, tumor segmentation, extraction of handcrafted feature descriptors such as Local Binary Patterns (LBP) and Gabor Wavelet Transform (GWT), and classification using support-vector machines and ensemble learners were adopted. Although the authors described high accuracy on BRATS datasets, they emphasised the high dependence of their results on segmentation quality and feature engineering.

A detailed survey by Amin et al. [2] further confirmed the fact that traditional machine learning techniques are able to execute well in controlled circumstances but often will not be robust in a heterogeneous dataset. The reliance on handcrafted features, extensive preprocessing, and dataset-specific tuning appeared to be a main factor that limits and is therefore the motivation to consider moving towards the deep learning-based representations.

### B. Custom CNN Models Trained from Scratch

With the arrival of deep learning, several publications have been undertaken to learn convolutional neural network (CNN) architectures from the raw MRI image without using transfer learning. Hossain et al. [3] have shown that shallow CNN models are indeed effective for screening of binary brain tumor; a five-layered compact model gave 97.87 % accuracy with low cost of calculation. Similar results have been reported by Febrianto et al. [4] who demonstrated that limited depth compact CNNs can achieve competitive accuracy while allowing training efficiency.

The tendency of shallow efficiency was further supported by the work of Martinez-Del-Rio-Ortega et al. [5] who systematically tested small CNNs with a varying number of convolutional layers (two to four). Using grid search-based hyper-parameter optimisation, their optimum architecture achieved 97.5 percent accuracy and 99.2 per cent sensitivity and confirmed that shallow CNNs, when carefully tuned, remain clinically competitive. Likewise, Choudhury et al. [6] have proposed an even more minimalistic three-layer CNN to perform binary tumor detection on Kaggle MRI data with an accuracy of 96.08 and tend to show the importance of regularisation over architecture complexity.

In contrast, deeper CNN architectures that are trained from scratch have been studied to be used in order to increase the representational capacity for more complex diagnostic tasks. Khan et al. [7] proposed a 23-layer CNN for binary and multiclass classification of brain tumors and found high-performance results but poor generalisation and over-fitting when the training data were small. Similar conclusions were made by Lamrani et al. [8], who proved that deeper CNNs can be used to outperform handcrafted pipelines for binary detection but are still a function of dataset size, preprocessing quality and regularisation strategy.

Extending custom CNNs from binary screening, Tiwari et al. [9] found multi-class brain tumor classification with a deep CNN trained using a four-class dataset of brains from the Kaggle repository due to MRI scan, healthy, glioma, meningioma, and pituitary tumors. Their architecture - which included repeated convolutional blocks with batch normalisation, pooling and dropout - managed to achieve an overall accuracy of close to 99%. Nonetheless, the authors highlighted persistent challenges associated with a lack of annotated data and heterogeneity in datasets and a reliance on fully automated end-to-end pipelines.

Overall, these studies show that shallow CNNs are still effective and computationally efficient and can be used for binary tumor screening; the deeper CNNs trained from scratch can achieve strong multiclass performance but are highly sensitive to the availability of the data, the quality of the annotation and the hyperparameter regularisation. These limitations taken together drive the growing popularity of the transfer learning approaches to improve the generalisation and robustness of the brain tumor classification.

### C. Transfer Learning for Brain Tumor Detection

Transfer learning has become a dominate solution to brain tumors MRI classification due to its ability to use pre-trained visual representations. Bairagi et al. [10] used pre-trained AlexNet and VGG16 models and reported accuracies of over 98% and demonstrated the usefulness of transfer learning models for limited medical datasets.

Khaliki and Basarslan [11] performed a multiclass comparison of a number of pre-trained CNN backbones that demonstrated the observation that transfer learning models are better than custom CNNs. Rastogi et al. [12] further supported these

results by fine-tuning the light architecture like MobileNet and Xception and providing good performance with less computational needs.

Despite all these advances, most approaches in transfer learning only assess one or two architectures, hindering the discovery of the most suitable model for balanced multiclass brain tumor classification under stable experimental conditions.

### *D. Hybrid Deep-Classical Models*

To enhance the robustness and readability, several studies were done that combined deep feature extraction and classical classifiers. Khairandish et al. [13] proposed CNN-SVM hybrid framework and showed better classification accuracy as compared with standalone CNN models. Kolla et al. 14 also combined CNN features with handcrafted texture descriptors and reported an improved discriminative capability.

A large-scale evaluation from Rahman et al. [15] has shown that feature fusion strategies based on deep and handcrafted features could be used to increase the accuracy of the classification process. However, these hybrid approaches raise the complexity of the system and they are also sensitive to preprocessing choices, as well as feature engineering choices.

### *E. Segmentation and Object Detection-Based Frameworks*

Beyond image-level classification, there are a number of studies which have paid attention to tumor localisation based on segmentation and object detection methods to enhance reliability of the diagnosis and clinical interpretability. Soomro et al. [16] published a detailed review of the brain tumor MRI segmentation approaches categorized into the traditional approaches including thresholding, regional/ clustering techniques (e.g. FCM, K-means), watershed based methods and the deep learning method including U-Net and its variants. Their survey showed Consistency of CNN-based segmentation methods compared to the classical pipelines on BraTS dataset for Dice score and Intersection over Union but also underlined consistent drawbacks on MRI noise, intensity inhomogeneity, scanner variability, ambiguous tumor boundaries, and the scarcity of high-quality annotated ground truth.

Building up on these results, Hemanth et al. [17] proposed an end-to-end framework that combines CNN-based segmentation with following classifier. Using a LinkNet encoder-decoder architecture for tumor - mask generation performances followed with successive feature extraction and classification, theirs method shows better performance as compared with classical baselines like SVM, CRF, and genetic algorithms. Nevertheless, like Soomro et al. [16], their method remained sensitive to preprocessing quality and poorly defined tumor boundaries.

More recently, the employment of object-detection approaches to provide fast and explicit tumor-localisation have been introduced. Abdusalomov et al. [18], suggested an improved YOLOv7-based differentiable model for multiclass brain tumor detection and localisation using multi-scale feature fusion techniques and attention mechanisms. Their model was found to be highly precise and recall, showing the potential for the use of object detection systems to support real time diagnosis. Nevertheless, such approaches require carefully annotated bounding box labels on the other hand, which are expensive to acquire and restrict scalability especially for small or subtle tumor regions.

Overall, although segmentation and object-detection frameworks help enhance spatial interpretability and localisation capability, the dependence upon detailed annotations and the sensitivity to data quality present a key challenge in the clinical deployment in a large scale.

### *F. Survey and Systematic Review Studies*

Survey and systematic review studies act as an important role in synthesis trends, challenges and research gaps in brain-tumor MRI analysis in deep-learning and machine learning paradigms. Solanki et al. [19] carried out a detailed survey of intelligent brain tumor detection and classification techniques, categorising in a systematic way traditional handcrafted feature-based machine learning pipelines, such as GLCM and DWT coupled with SVM and Random Forest, using CNN deep learning and transfer learning methods. Their analysis indicated that deep-learning approaches usually outperform traditional ML approaches when there are enough data available whereas classical approaches may be competitive in small data. However, specific problems remained as noted in the survey, such as noise of MRI, diffuse boundary between tumor and normal, finding of small tumor, scanner and protocol variations, and absence of uniform benchmarking across tumor classes.

More recently, a systematic review based on the PRISMA protocol was carried out by Rasool and Bhat [20] to analyse ML and DL studies for the detection and segmentation and classification of brain continuum for brain tumors. By organising previous work based on the type of task and the model family, their work has supported the prevalence of deep learning methods, specifically encoder-decoder models such as U-Net, for segmentation and classification problems. Nevertheless, they found common limitations across the literature, such as class imbalance, heterogeneous imaging conditions, lack of annotated datasets and cross scanner generalisation. The authors underlined the importance of larger shared datasets, robust data augmentation strategies, multi-scale and uncertainty-aware deep learning models, and increased model interpretability in order to facilitate clinical use.

Collectively, these survey studies provide a strong argument to conclude that, while deep learning has become the dominant methodology for brain-tumor MRI analysis, there still are unanswered issues surrounding data diversity, annotation quality and the learning generalisation. These results provide stimulus for systematic benchmarking under the use of unified experimental protocols, as used in this study.

### G. Research Gap and Motivation

Although prior studies demonstrate the effectiveness of CNNs and transfer learning for brain tumor detection, several gaps remain. First, many works focus on binary classification rather than balanced multi-class diagnosis. Second, most studies evaluate a limited number of architectures without a unified preprocessing and evaluation protocol. Third, systematic benchmarking of widely used architectures—such as VGG16, ResNet50, DenseNet121, and MobileNetV3—under identical experimental conditions remains limited.

To address these limitations, this study presents a structured, side-by-side comparison of five deep learning architectures using a large balanced MRI dataset and a consistent experimental framework, providing clearer empirical justification for model selection in multi-class brain tumor classification.

Although convolutional neural networks have been extensively used in the classification of brain tumors, the novelty of the presented study is in its organized and reproducible benchmarking framework. In contrast to most previous studies that test small architectures in non-uniform experimental conditions, this experiment guarantees a fair comparison, by using the same preprocessing, training policies, and performance measures on all models. This enables more reliable and objective evaluation of the model performance to classify brain tumor in multi-class.

## III. RESEARCH METHODOLOGY

### A. Overview of the Research Framework

This study uses an organized and reproducible research methodology to develop and evaluate a deep learning–based framework for multi-class brain tumor detection using MRI. The research methodology is designed as shown in Figure 1 to ensure fair comparison across multiple deep learning architectures by enforcing a unified preprocessing pipeline, consistent training strategy, and standardized evaluation metrics.

The proposed framework includes the follow sequential stages: dataset acquisition, data preprocessing and augmentation, model development, training strategy implementation, hyperparameter optimization and performance evaluation. A custom CNN and four transfer learning models, VGG16, ResNet50, DenseNet121, and MobileNetV3 are implemented and compared with the same set of data and experimental condition to remove bias in the comparative analysis of results.

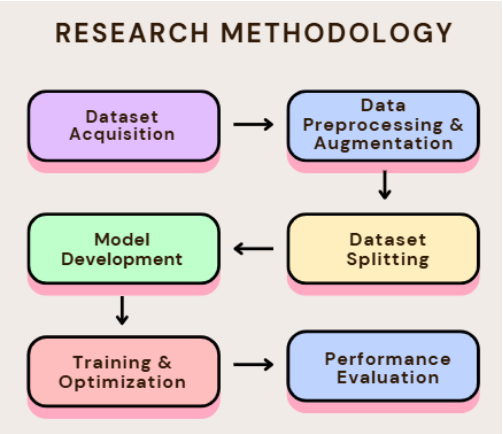


**FIGURE 1. Flowchart of Research Methodology.**

### B. Dataset Description

The experiments are conducted on Brain Tumor MRI Multi-Class Dataset from Kaggle. The dataset is an aggregation derived from a lot of MRI pictures from other unbiased repositories, allowing evaluation of fully automated brain tumor classification systems.

Four classes in the dataset:

- Glioma
- Meningioma
- Pituitary
- Healthy

In total, 16,269 MRI images are available, of which 3,325 glioma, 3,266 meningioma, 2,974 pituitary tumor, and 6,704 healthy images are available. All MRI scans are 2D slices saved as JPG or PNG files with very variable spatial resolutions, from small images to large images.

To support robust learning and evaluation, a dataset is divided into training set with 80% of the data for training set, 10% for validation and 10% for testing set. This split ensures there is sufficient data used for learning, and independent samples used for unbiased evaluation.

Despite the large and balanced dataset used in this research, it is downloaded from Kaggle and is a collection of MRI images from disparate sources. Consequently, the dataset might not represent the variability of clinical practice, including scanner-specific variations, protocols and hospital-specific acquisition parameters. Variations in different MRI scanners and settings can lead to intensity variations. Despite the use of normalization methods to reduce such inconsistencies, the data may not be as consistent as single-institution clinical data. As such, future research needs to include primary clinical data sets for validation to ensure the robustness and applicability of the model in clinical practice.

### C. Data Preprocessing and Augmentation

Given the heterogeneous nature of the dataset, a large amount of preprocessing is exercised, in order to achieve a better quality of the data, while also considering the compatibility with the deep learning architectures. Invalid or unreadable image files are eliminated from others and from images with too small spatial dimensions which less than 50 pixels wide or

height is invalidated because of insufficient anatomical information.

Duplicate images are identified with MD5 hashing of the original image arrays. Images with a hash match are accepted as duplicates and only one instance of those images is kept to avoid leakage of data between the different split datasets.

The images are converted to the RGB format to match the input format of the CNN model. To mitigate variations due to different MRI scanners and intensity mismatches in the aggregated dataset, we introduced a normalization step. This involved first implementing a z-score normalization to standardize the distribution of intensities of each image, followed by a min-max normalization to rescale values to a standard range. This technique helps mitigate the effects of different MRI scanners and imaging protocols, enhancing the transferability of the models to different sources.

Subsequently, all images are resized to 224x224 pixels which is the typical input resolution for VGG, ResNet, DenseNet and MobileNet architectures.

Online data augmentation is only used for the training set with the help of the ImageDataGenerator framework. Data augmentation operations include random rotations (±10°), zooming (up to 15%), spatial shifts (±5%), and brightness variation (0.9 – 1.1). Horizontal flipping is purposely avoided because of possible anatomic asymmetry of brain MRI scans. The validation as well the test dataset is stored without augmentation so that there are no integrity issues regarding the unbiased evaluation.

### *D. Model Development*

All models are designed to accept input images of size 224 × 224 × 3 and output predictions for four tumor classes. A custom CNN and four transfer learning architectures are evaluated.

Figure 2 showing the proposed CNN follows a hierarchical feature extraction strategy consisting of multiple convolutional blocks with increasing filter depths (32 → 64 → 128 → 256). Each block includes convolutional layers, batch normalization, Swish activation, max pooling, and spatial dropout for regularization. Squeeze-and-Excitation (SE) blocks are incorporated to boost channel-wise feature recalibration.

The classification head consists of global average pooling, fully connected layers with dropout, and a Softmax output layer. Focal loss is employed to emphasize hard-to-classify samples, and optimization is performed via learning rate of $3\times10^{-4}$ of Adam optimizer.

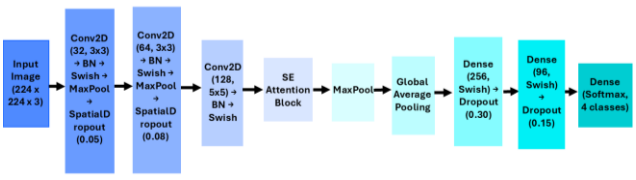


**FIGURE 2. Custom CNN Architecture Diagram.**

Figure 3 shows the architecture of VGG16, which was used for the present study. VGG16 has been chosen because it is a simple and uniform architecture, comprising stacked convolutional layers with small receptive fields (3 x 3). This architecture has shown great performance in the field of visual recognition by learning hierarchical features progressively through depth instead of architectural complexity.

The VGG16 base network has been initialized with ImageNet pretrained weights and the original full connected classification layers were removed. In the first training phase, all the convolution part was fixed, in order to retain the low and mid-level visual feature information that was learned from large-scale natural image data. A customized classification head was added, which is a Global Average Pooling layer, and then some fully connected layers were added with batch normalization and dropout, both to enhance generalization and battle over-fitting. The last output layer had four neurons with softmax activation, which corresponds to each brain tumor classes.

Multiple VGG16 configurations were explored during fine tuning as the depth of the classifiers, the dropout rate, and the depth of fine tuning. The configuration chosen gave the best balance between classification accuracy and training stability. Fine-tuning was then done by unfreezing the top convolutional blocks of the VGG16 backbone so that the network learned to fit the patterns of MRI higher features. The model was optimised using the Adam Optimiser and trained with categorical cross-entropy loss function with the classification accuracy monitored during validation.

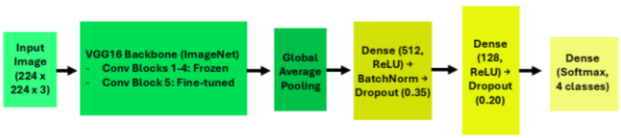


**FIGURE 3. VGG16 Architecture Diagram.**

Figure 4 shows the architectural details of ResNet50 used in this study The reason I chose ResNet50 is because it has a very deep learning paradigm, the model includes short cut connections to mitigate the vanishing gradient issue that is common in very deep neural network models. These residual connections enable stable training of deep architectures while preserving representational richness.

The ResNet50 backbone was trained using ImageNet pre-trained weights and the original fully connected classification layer was removed. During the initial training stage, the convolutional backbone was frozen to maintain the strong feature hierarchies that can be used as strong feature hierarchies are learned from large-scaled data. A custom classification head was then added consisting of Global Average Pooling followed by dense layers with batch-normalisation and dropout regularisation. The terminal classification layer used Softmax activation function for 4 class prediction.

Several Configurations of ResNet50 were scrutinised during hyper - parameter optimisation. The best configuration produced better validation and test

performance than other variants. Partial fine-tuning has been attempted by unfreezing the upper residual blocks, which partly led to the network drawing in domain specific features from the MRI data while preserving stable lower level features. Optimisation has been done using Adam optimiser with a carefully chosen learning rate to ensure stable convergence and avoid overfitting.

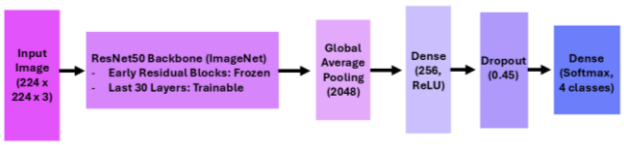


**FIGURE 4. ResNet50 Architecture Diagram.**

The architecture of DenseNet121 adopted in this study is shown in Figure 5. DenseNet121 was selected because of its denser connected model, where each layer of the model takes the feature maps from all previous layers as inputs. This kind of configuration encourages discontinuous reuse of features, is conducive to gradient flow, and reduces the number of the trainable parameters compared with the deep networks.

The DenseNet121 baseline was pre-trained on ImageNet and the original classification layers were removed. First, all the convolutional layers were frozen in order to keep the pretrained feature representations. A task specific classification head was appended which consisted of global average pooling, fully connected layers, dropout regularization and a SoftMax output layer for a four-class classification task.

Comprehensive approaches for hyperparameter optimization and tunings of depths, classifiers and regularization methods were attempted. The configuration selected provided the best possible trade-off between accuracy and stability of the training. Fine-tuning was only applied to the upper dense blocks due to both the need of adapting to the characteristics of MRI images as well as the risk of overfitting. A model optimised using Adam optimiser and accuracy for validation was used as an major metric to track the accuracy of the model.

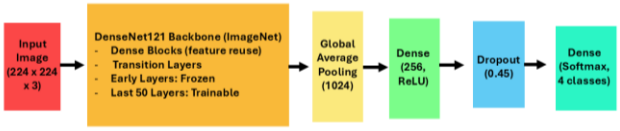


**FIGURE 5. DenseNet121 Architecture Diagram.**

The MobileNetV3 architecture used in this study shown in Figure 6. The inclusion of MobileNetV3 was an attempt to test the effectiveness of lightweight neural network designs that are computationally efficient for rejected brain tumor MRI scans. The model is based on depthwise separable convolution to significantly reduce the computational cost and parameter size while Squeeze-E& excitation modules are used to capture channel-wise attention mechanism to highlight the informative feature maps. Moreover, the architecture uses the h-swish activation function, and it represents a favorable balance between representational expressiveness and computational efficiency.

The MobileNetV3- Large variant pretrained on ImageNet was used for the backbone, with the classification head of the original version removed. In contrast to other transfer learning paradigms all the layers of the MobileNetV3 backbone were made trainable from the beginning, which allows comprehensive fine tuning. This approach allows the lightweight network to adapt both low level and high-level representations to the individual characteristics of the brain MRI.

A Global Average Pooling layer was used to aggregate spatial features, and then batch normalization and a dense layer with 256 units activated with ReLU were used. A dropout layer was used with the probability of 0.5 to alleviate the overfitting risk. The terminal output layer had four nodes using Softmax activation. Training was done using the Adam optimization algorithm with a small learning rate value 1 x 10-5, categorical cross entropy loss, and categorical accuracy was utilized as the performance metric.

Several versions of MobileNetV3 were tested for hyperparameter optimisation. A fully fine-tuned MobileNetV3-Large variant gave the best validation performance and showed the most stable convergence between the evaluated lightweight models.

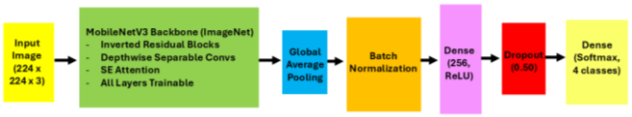


**FIGURE 6. MobileNetV3 Architecture Diagram.**

### *E. Training Strategy and Hyperparameter Optimization*

To make sure that a fair and rigorous comparison across all architectures evaluated has been taken, a unified training strategy and systematic hyperparameter optimization procedure were adopted. Although the models vary in terms of architectural complexity and design philosophy, all the experiments were conducted under the same data preprocessing, data augmentation, dataset partitioning, and evaluation protocols. This way, experimental bias is minimized and comparative analysis is made possible.

Hyperparameter optimization has been done separately for all the model architectures using the validation set. Both architectural and training-related parameters were investigated in order to identify specific parameters to maximize classification performance within a stable convergence and generalization region. Candidate configurations were tested using validation accuracy and macro averaged performances, and the best performing configuration of each model was then tested on the independent testing data.

*Custom CNN Hyperparameter Optimization*

For the custom CNN, a lot of tuning was performed, both in architecture depth and training

parameters. The number of convolution layers varied between four to six layers and several strategies of filter progression were tested such as 32→64→128→192, 32→64→128→256, and 32→64→128→256→256. Convolutional kernel sizes of 3 x 3 and 5 x 5 were considered to compromise the spatial feature resolution with computation cost. Activation functions that were tested were LeakyReLU (α = 0.1) and Swish, with Swish demonstrating superior training stability.

To improve channel wise feature recalibration, Squeeze-and-Excitation(SE) blocks of reduction ratio 16 were added in some configurations. Some of the regularization techniques applied were spatial dropout rates in convolutional blocks between 0.05 and 0.20 and dropout rate in fully connected network layers from 0.15 to 0.45. L2 Weight regularization was used with coefficients of 1 x $10^{-5}$ for convolutional layers and coefficients of 1 x $10^{-4}$ for dense layers. For classifier head designs were evaluated some dense layer combinations of 384→128, 256→64, and 256→96.

Training optimization has been done with Adam and AdamW with learning rates of 5 x $10^{-5}$, 1 x $10^{-4}$ and 3 x $10^{-4}$. Learning rate scheduling strategies were a fixed learning rate, ReduceLROnPlateau, and Cosine Decay Restarts. Loss functions tested were Categorical Cross-Entropy and Focal Loss, focal loss parameters were set to $\gamma$ = 2.0 and $\alpha$ = 0.25 and 0.5. The final configuration of the CNN was chosen according to a performance that was consistent in the validation period of the results and the stability of convergence.

*VGG16 Hyperparameter Optimization*

Hyperparameter tuning appeared in the VGG16 focused mostly on transfer learning depth and classifier head designing. Two training methods were tested, feature extraction (where all convolutional layers are frozen) and partial fine-tuning (where the upper blocks of VGG are unfrozen). The number of unfrozen blocks varied between zero and two, making it possible to control the adaptation of higher-level features to MRI-specific patterns.

Classifier head configurations were experimented with dense layer sizes of 256, 384 and 512 neurons, and one or two dense layers. Activation functions tried were ReLU and Swish and dropout rates were between 0.40 to 0.55 to reduce the over fitting problem. Optimization was done using the learning rates of 1 x $10^{-3}$ of Adam optimizer for the feature extraction stage and the reduced learning rates of 1 x $10^{-4}$ and 1 x $10^{-5}$ for the fine-tuning stage. The selected configuration achieved the best balance between accuracy and training stability.

*ResNet50 Hyperparameter Optimization*

For ResNet50, hyperparameters were adjusted so as to focus on fine-tuning depth and classifier regularization. Two configurations were tested, feature extraction with all convolutional layers frozen and partial fine tuning at upper residual blocks. The number of the unfrozen layers changed between the top 30 and 50 layers to observe progressive adaptation to the MRI features without destabilizing pre-training representations.

Classifier head architectures were dense with 256 and 384 neurons, dropout rates of 0.40 - 0.50, and ReLU activation. Training was optimised with learning rates between 1 x $10^{-4}$ to 1 x $10^{-5}$ of Adam optimizer. The last ResNet50 model showed better performance in validation and test experiments, which suggests successful domain adaptation and very good generalisation.

*DenseNet121 Hyperparameter Optimization*

DenseNet121 was hyperparameter tuned extensively as it consists of dense connectivity pattern, hence it needs careful optimisation to keep the gradient flow stable. The number of the trainable layers was varied from feature extraction only to fine tuning last 50, 60, 80, 120 and 150 layers. Batch normality methods were also tested; with or without frozen BatchNorm layers during fine tuning.

Classifier head configurations of dense layer sizes of 256, 384, 512 neurons, activation functions of ReLU, ELU and Swish and dropout rates of 0.45-0.55 were tested. L2 regularisation with coefficient of 1 x $10^{-4}$ was applied to dense layers. The loss functions used for evaluation were Categorical Cross-Entropy, Cross-Entropy with label-smoothing (0.05 and 0.10), and Focal Loss with $\gamma$ = 2.0 and α = 0.25.

Optimization strategies were Adam, AdamW and SGD with momentum 0.9, with learning rates of 1 x $10^{-4}$, 1 x $10^{-5}$, and 2 x $10^{-5}$. Learning rate schedulers were ReduceLROnPlateau and Cosine Decay. Selected DenseNet121 combination gave stable converging and competitive performance.

*MobileNetV3 Hyperparameter Optimization*

Hyperparametric tuning for the MobileNetV3 focused on computational efficiency and generalisation. Model capacity was experimentally varied on a range of width multipliers between 0.75 and 1.0. Classifier head configurations were tested involving dense layer sizes ranging from 128 to 256 neurons, dropout rates ranging from 0.20 to 0.40, and the activation function, ReLU and Swish.

Optimisation was done with the Adam optimiser using a range of learning rates from 1 x $10^{-4}$ to 1 x $10^{-3}$. The final MobileNetV3 configuration showed good validation performance and stability in training as well as a significantly reduced computational footprint in comparison to deeper architectures.

*Final Hyperparameter Configuration*

Table 1 shows the final hyperparameter configuration of CNN, VGG16, ResNet50, DenseNet121 and MobileNetV3 model. These hyperparameter selection for each model has achieved the highest accuracy for each model.

**TABLE 1. Final Hyperparameter Configuration of selected models.**

| Hyperparameter | CNN | VGG16 | ResNet50 | DenseNet121 | MobileNetV3 |
|---|---|---|---|---|---|
| Base Architecture | Custom CNN | VGG16 (ImageNet) | ResNet50 (ImageNet) | DenseNet121 (ImageNet) | MobileNetV3 (ImageNet, Large by default) |
| Input Size | 224×224×3 | 224×224×3 | 224×224×3 | 224×224×3 | 224×224×3 |
| Trainable Layers | All | Stage 1: base frozen; Stage 2: block5 only | Last 30 layers trainable | Last 50 layers trainable | All base trainable |
| Conv Filters | 32→64→128→256→256 | Pretrained | Pretrained | Pretrained | Pretrained |
| Kernel Sizes | 3×3 and 5×5 | Pretrained 3×3 | Pretrained | Pretrained 3×3 | Pretrained (depthwise separable internally) |
| Activation (Head/custom) | Swish | ReLU | ReLU | ReLU | ReLU |
| Attention | SE block (ratio=16) | - | - | - | - |
| Global Pooling | Global Average Pooling | Global Average Pooling | Global Average Pooling | Global Average Pooling | Global Average Pooling |
| Dense Head | 256 → 96 | 256 → 96 | 256 → 96 | 256 → 96 | 256 → 96 |
| Dropout | Spatial: 0.05/0.08/0.12; Head: 0.30 & 0.15 | Spatial: 0.05/0.08/0.12; Head: 0.30 & 0.15 | Spatial: 0.05/0.08/0.12; Head: 0.30 & 0.15 | Spatial: 0.05/0.08/0.12; Head: 0.30 & 0.15 | Spatial: 0.05/0.08/0.12; Head: 0.30 & 0.15 |
| L2 Regularization | Conv: 1e-5; Dense: 1e-4 | Conv: 1e-5; Dense: 1e-4 | Conv: 1e-5; Dense: 1e-4 | Conv: 1e-5; Dense: 1e-4 | Conv: 1e-5; Dense: 1e-4 |
| Loss | Focal Loss | Categorical Cross-Entropy | Categorical Cross-Entropy | Categorical Cross-Entropy | Categorical Cross-Entropy |
| Focal params | γ=2.0, α=0.5 | - | - | - | - |
| Optimizer | Adam | Adam | Adam | Adam | Adam |
| Learning Rate | 3e-4 | 1e-4 → 1e-5 | 1e-5 | 1e-5 | 1e-5 |
| LR Scheduling | ReduceLROnPlateau | ReduceLROnPlateau | ReduceLROnPlateau | ReduceLROnPlateau | ReduceLROnPlateau |
| Batch Size | 16 | 32 | 32 | 32 | 16 |
| Epochs | 30 | 25 (15 + 10) | 35 | 35 | 30 |

### *F. Evaluation Metrics*

The performance of the model is evaluated using several metrics, such as accuracy, precision, recall, F1 - score, confusion matrices, and ROC - AUC. To ensure a balanced evaluation between the different types of tumors, macro-averaging is used for the precision, recall, and the F1-score metrics.

## IV. Results And Discussions

### *A. Quantitative Performance Comparison*

The quantitative performance of all the evaluated models is tested by their test accuracy and macro-averaged F1-score as shown in Figure 7 and Figure 8, respectively. After running an experiment as shown in Figure 7, the custom CNN baseline test accuracy is 82.13% far less than all transfer learning models. In contrast, all pretrained architectures have an accuracy of more than 94%, which shows the effectiveness of transfer learning in brain tumor MRI classification.

Out of models tested, ResNet50 yields the highest test accuracy of 98.38% followed by VGG16 (97.25%), MobileNetV3 (96.38%) and DenseNet121 (94.25%). This trend can always be observed in a comparison of the macro-averaged F1-scores illustrated in Figure 8. ResNet50 has the highest macro F1-score of 0.9837. The superior macro F1-score follows that ResNet50's performance of classification in all classes of tumors also rather than just dominant ones.

The given comparison results in Figure 7 and Figure 8 collectively indicate deeper transfer learning architectures achieved a significant improvement as compared to the performance of the custom CNN whereby ResNet50 indicated the highest generalization ability and class-wise consistency.

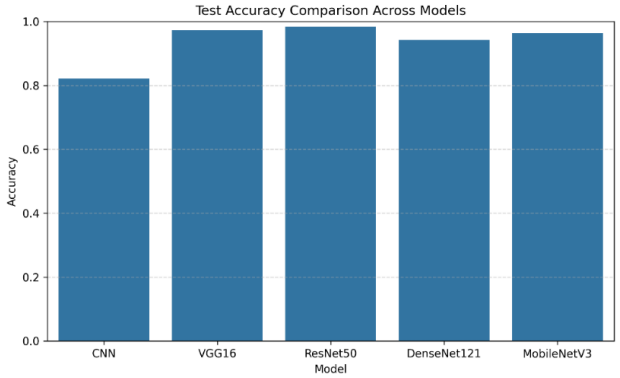


**FIGURE 7. Test Accuracy Comparison Across Models.**

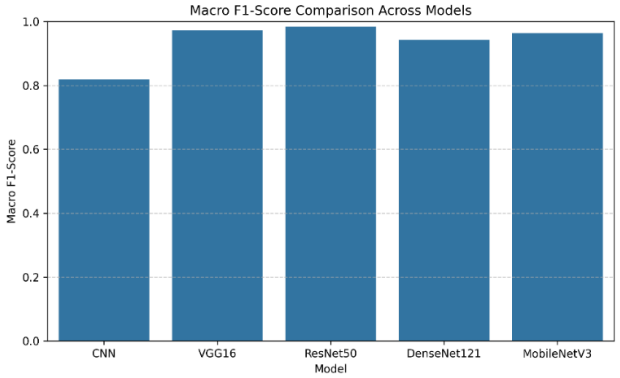


**FIGURE 8. Macro F1-Score Comparison Across Models.**

### *B. ROC-AUC Analysis and Class Separability*

An extra evaluation of class separability was performed using the comparison of the macro-averaged ROC-AUC values of all models as shown in Figure 9. A surprising and highly intriguing result was found to show that all versions of transfer learning achieve macro ROC-AUC scores in excess of 0.99; this indicates a strong discriminative ability. On the contrary, the bespoke CNN exhibits a significantly lower macro ROC - AUC, which is indicative of reduced class - separation performance.

ResNet50 registers the highest macro ROC-AUC of 0.9996 as shown in Figure 9, suggesting an almost perfect discrimination between glioma, meningioma or pituitary tumor classes and the healthy class from MRI scans. To investigate class specific performance, per class ROC Curves are examined for each model.

The per-class ROC curves of the custom CNN, presented in Figure 10, show that the discrimination is lower in comparison, especially for the meningioma class. In contrast, the per class ROC curves for VGG16 are shown in Figure 11 and show significantly increased class separation across all classes. The per-class roc curves of ResNet50, as presented in Figure 12, indicate something very close to perfect discrimination among the four classes, with roc curves approaching the upper-left corner of the roc space.

DenseNet121 per-class ROC curves in Figure 13 and MobileNetV3's curves in Figure 14 also show good performance; however, some loss of separability is apparent between particular tumor classes compared to ResNet-50. Collectively, ROC-AUC proves that ResNet50 provides the highest and most exemplary class separability among all architectures under analysis.

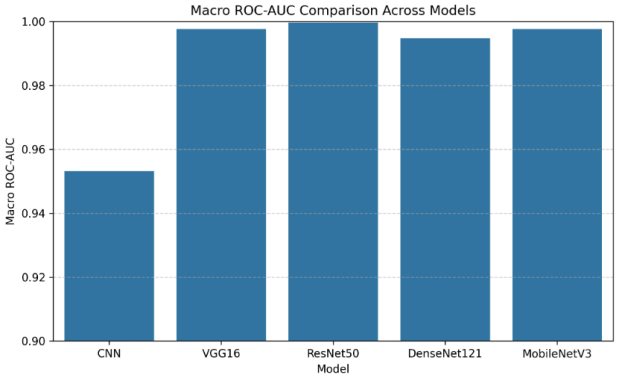


**FIGURE 9. Macro ROC-AUC Comparison Across Models.**

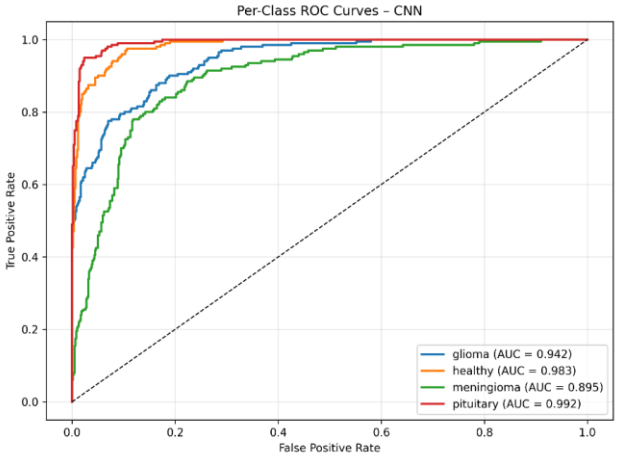


**FIGURE 10. CNN Per-Class ROC Curves.**

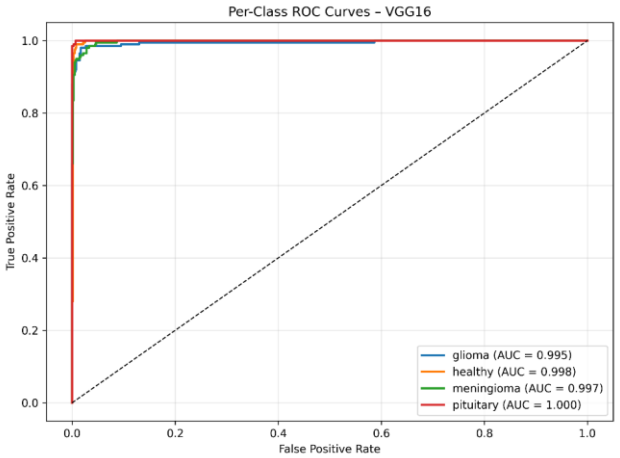


**FIGURE 11. VGG16 Per-Class ROC Curves.**

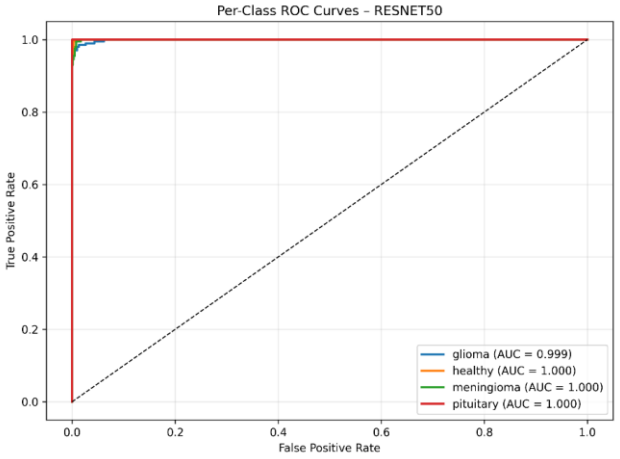


**FIGURE 12. ResNet50 Per-Class ROC Curves.**

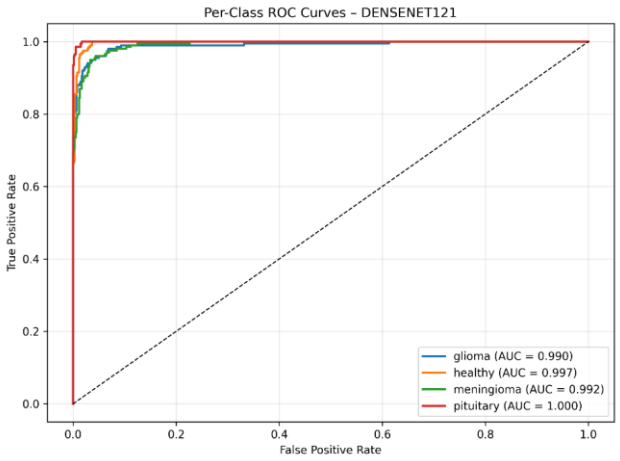


**FIGURE 13. DenseNet121 Per-Class ROC Curves.**

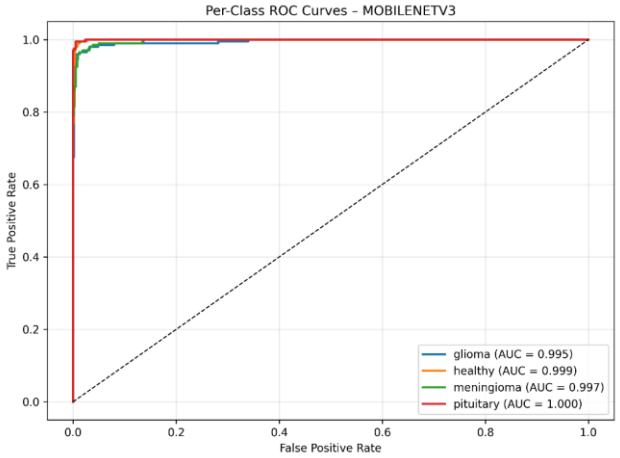


**FIGURE 14. MobileNetV3 Per-Class ROC Curves**

### *C. Qualitative Analysis*

In addition to the quantitative evaluation, a qualitative analysis was performed to get a visual assessment of the classification behaviour of the best-performing model. Figure 15 shows representative examples of MRI images that are correctly classified by the ResNet 50 model for all the four classes. These examples show that ResNet50 can detect different sizes, shapes and locations of tumors precisely and thus it is robust to the anatomical variations.

Misclassified examples produced by ResNet50 are shown in Figure 16. Visual inspection of these cases reveals that most errors occur in MRI scans with ambiguous tumor boundaries or tumor regions that closely resemble healthy tissue. Such misclassifications are consistent with limitations reported in prior studies and reflect inherent challenges in brain tumor MRI interpretation rather than model instability.

Together, the qualitative results in Figure 15 and Figure 16 support the quantitative findings and further confirm that the strong performance of ResNet50 arises from effective feature learning rather than overfitting.

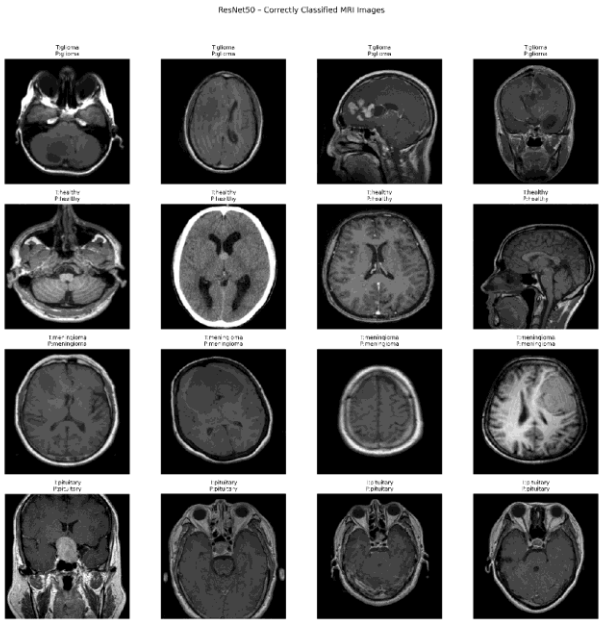


**FIGURE 15. ResNet50 Correctly Classified MRI Images Results.**

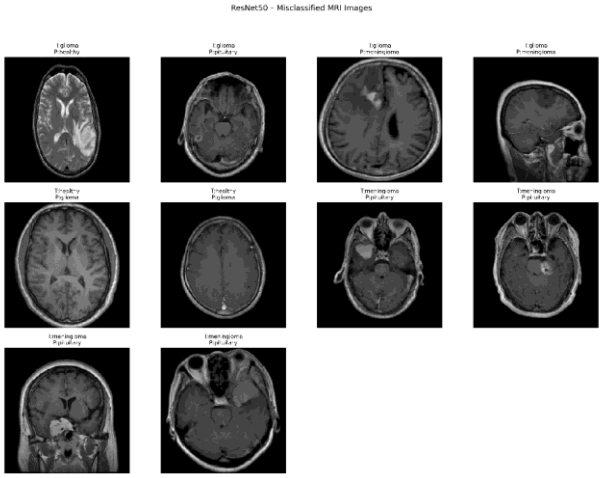


**FIGURE 16. ResNet50 Misclassified MRI Images Results.**

### *D. Discussion Summary*

The experimental results shown in Figure 17 along with the accompanying quantitative performance table are very clear to show that ResNet50 outperforms all other evaluated architectures on the main performance assessment criteria. In particular, ResNet50 achieved the best classification accuracy of 98.38%, macro-precision, macro-recall, macro F1-Score of 0.9837, macro ROC-AUC of 0.9996, which indicated the superior overall classification performance, as well as the balanced class discrimination.

Compared to CNN baseline without transfer learning, each transfer learning model offers significant improvement in the performance, thus validating the effectiveness of pre-trained feature representations for complex medical imaging problems. Among the transfer learning methodologies, ResNet50 is the one that exhibits the most robust and consistent behaviour. Its residual learning method aids stable training for the deep networks, as well as promotes effective gradient propagation, which aids in better feature discrimination across all classes of tumor. This

advantage is especially true in the near perfect macro ROC - AUC score, suggesting excellent class separability.

Although VGG16 and MobileNetV3 also have impressive results, their results are slightly worse in all the macro-averaged metrics than ResNet50. DenseNet121, although benefiting from dense feature re-users, has comparatively lower accuracy and macro F1-score, indicating that dense connectivity does not necessarily provide optimal performance for such a task without wise refinements on the architectural component and training.

Superior performance of ResNet50 can be attributed to a coincidence of three important factors: use of large and balanced multi-class MRI dataset, uniform and equitable experimental protocol and systematic hyperparameter optimisation dedicated for each one. Unlike many previous studies with small dataset sizes, binary classification or the use of a default setup for the training configuration, this work shows that careful optimisation and controlled comparative evaluation is imperative when looking for the most appropriate model for multi-class brain tumor classification.

It is also worth mentioning that the data used in this study comes from different sources, which could affect the scanner types and image acquisition conditions. While normalization techniques were utilised, these may still have impacted model performance. Thus, additional validation using clinical data with a uniform protocol is needed before clinical implementation.

In conclusion, the results validate the effectiveness of the multiple class brain tumor classifier based on the ResNet50 network in terms of robust, reliable, and generalisable model for automatic brain tumor classification from MRI images that succeeded to overcome common limitations found in previous works.

**TABLE 2. Results of CNN, VGG16, ResNet50, DenseNet121 and MobileNetV3.**

| Model | Accuracy | Precision | Recall | F1 | ROC_AUC |
|---|---|---|---|---|---|
| CNN | 0.821 | 0.84 | 0.821 | 0.818 | 0.953 |
| VGG16 | 0.973 | 0.973 | 0.973 | 0.972 | 0.998 |
| ResNet50 | 0.984 | 0.984 | 0.984 | 0.984 | 1.0 |
| DenseNet121 | 0.943 | 0.942 | 0.943 | 0.942 | 0.995 |
| MobileNetV3 | 0.964 | 0.964 | 0.964 | 0.963 | 1.0 |

## V. CONCLUSION

This study introduced an entire framework of the deep learning challenge of classifying brain tumors with multiple classes using magnetic resonance imaging (MRI) data. Systematic and reproducible methodology was implemented that includes dataset acquisition, thorough preprocessing and augmentation, unified training protocols, comprehensive hyperparameter optimisation and comprehensive multi-metrical evaluation of the results. Both a custom convolutional neural network, as well as four state-of-the-art transfer learning architectures, namely: VGG16, ResNet50, DenseNet121, MobileNetV3, were implemented, and evaluated under the same experimental conditions to ensure a fair and unbiased comparison.

Experimental results show that the transfer-learning architectures clearly outperform the custom CNN which shows the effectiveness of pretrained feature representations for complex medical imaging tasks. Among all the models considered, ResNet50 was the model with the best overall performance (highest classification accuracy, macro-averaged score of variances, recall, F1 and the macro value of the ROC). The residual learning mechanism of ResNet50 facilitated better gradient flow and strong feature learning, and therefore, enabled near-perfect class separability of glioma, meningioma, pituitary tumor and healthy MRI images, as verified by ROC curve analysis.

The large-scale hyperparameter optimisation strategy also contributed to performance improvement by ensuring the operating conditions of each architecture were close to optimal conditions instead of default settings. Qualitative analysis of the result of the prediction confirmed that the high quantitative performance of ResNet50 was not the result of overfitting but rather reflected its good ability of generalisation across diverse tumor appearances and imaging variations.

Overall, the proposed framework overcomes some limitations that were identified in previous research studies, such as: reliance on small datasets, binary classification setups, rather constrained architectural comparisons, and a lack of hyperparameter tuning. By combining access to a large, normalised data set, algorithmic examination of multiple architectures and powerful optimisation approaches this work represents a robust, uniform and scalable solution for multi-class brain tumor classification.

Future work can be further developed into a decision-support system, especially for automated radiotherapy treatment planning systems. The model can be coupled with tumor segmentation and localization modules to guide clinicians in treatment strategy decisions. For instance, tumor classification can be integrated with radiotherapy dose prescription and treatment strategies based on tumor types. Integration with hospital imaging systems like PACS systems may also allow real-time integration with clinical workflows. Future research should also aim to validate the system using large multi-center clinical datasets to ensure it is reliable, safe and compliant for deployment in healthcare settings.

## ACKNOWLEDGMENT

The authors wish to acknowledge the anonymous reviewers who made some very important comments and suggestions to help improve the quality and readability of this paper.

## FUNDING STATEMENT

There is no funding agencies supporting the research work.

## AUTHOR CONTRIBUTIONS

Lim Xin Tian: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;

S Prabha Kumaresan: Project Administration, Writing – Review & Editing

## CONFLICT OF INTERESTS

No conflict of interests were disclosed.

## ETHICS STATEMENTS

This study follows the publication ethics guidelines of the Committee on Publication Ethics (COPE). The job required no clinical trials, animal experiments or direct medical intervention. The information that is relevant to human beings in this research was gathered through publicly available datasets and anonymized survey responses. All participants in the survey gave informed consent, and no personal identifiable information was gathered or stored.The data supporting the findings of this study are derived from publicly available datasets and anonymized survey responses. Derived data supporting the conclusions of this work are available from the corresponding author upon reasonable request.

## DECLARATION OF AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

No artificial intelligence tools were used in the research or writing of this article.

## REFERENCES


[1] J. Amin, M. Sharif, M. Raza, T. Saba and M.A. Anjum, "Brain tumor detection using statistical and machine learning method," *Computer Methods and Programs in Biomedicine*, vol. 177, pp. 69-79, 2019. DOI: https://doi.org/10.1016/j.cmpb.2019.05.015

[2] J. Amin, M. Sharif, A. Haldorai, M. Yasmin and R.S. Nayak, "Brain tumor detection and classification using machine learning: a comprehensive survey," *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 3161-3183, 2022. DOI: https://doi.org/10.1007/s40747-021-00563-y

[3] T. Hossain, F.S. Shishir, M. Ashraf, M.A.A. Nasim and F.M. Shah, "Brain Tumor Detection Using Convolutional Neural Network," *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pp. 1-6, 2019. DOI: https://doi.org/10.1109/ICASERT.2019.8934561

[4] D.C. Febrianto, I. Soesanti and H.A. Nugroho, "Convolutional Neural Network for Brain Tumor Detection," *IOP Conference Series: Materials Science and Engineering*, vol. 771, no. 1, pp. 012031, 2020. DOI: https://doi.org/10.1088/1757-899X/771/1/012031

[5] R. Martínez-Del-Río-Ortega, J. Civit-Masot, F. Luna-Perejón and M. Domínguez-Morales, "Brain Tumor Detection Using Magnetic Resonance Imaging and Convolutional Neural Networks," *Big Data and Cognitive Computing*, vol. 8, no. 9, pp. 123, 2024. DOI: https://doi.org/10.3390/bdcc8090123

[6] C.L. Choudhury, C. Mahanty, R. Kumar and B.K. Mishra, "Brain Tumor Detection and Classification Using Convolutional Neural Network and Deep Neural Network," *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, pp. 1-4, 2020. DOI: https://doi.org/10.1109/ICCSEA49143.2020.9132874

[7] M.S.I. Khan, A. Rahman, T. Debnath, M.R. Karim, M.K. Nasir, S.S. Band, A. Mosavi and I. Dehzangi, "Accurate brain tumor detection using deep convolutional neural network," *Computational and Structural Biotechnology Journal*, vol. 20, pp. 4733-4745, 2022. DOI: https://doi.org/10.1016/j.csbj.2022.08.039

[8] D. Lamrani, B. Cherradi, O.E. Gannour, M.A. Bouqentar and L. Bahatti, "Brain Tumor Detection using MRI Images and Convolutional Neural Network," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, 2022. DOI: https://doi.org/10.14569/IJACSA.2022.0130755

[9] P. Tiwari, B. Pant, M.M. Elarabawy, M. Abd-Elnaby, N. Mohd, G. Dhiman and S. Sharma, "CNN Based Multiclass Brain Tumor Detection Using Medical Imaging," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1-8, 2022. DOI: https://doi.org/10.1155/2022/1830010

[10] V.K. Bairagi, P.P. Gumaste, S.H. Rajput and C.K. S., "Automatic brain tumor detection using CNN transfer learning approach," *Medical & Biological Engineering & Computing*, vol. 61, no. 7, pp. 1821-1836, 2023. DOI: https://doi.org/10.1007/s11517-023-02820-3

[11] M.Z. Khaliki and M.S. Başarslan, "Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN," *Scientific Reports*, vol. 14, no. 1, 2024. DOI: https://doi.org/10.1038/s41598-024-52823-9

[12] D. Rastogi, P. Johri, M. Donelli, L. Kumar, S. Bindewari, A. Raghav and S.K. Khatri, "Brain Tumor Detection and Prediction in MRI Images Utilizing a Fine-Tuned Transfer Learning Model Integrated Within Deep Learning Frameworks," *Life*, vol. 15, no. 3, pp. 327, 2025. DOI: https://doi.org/10.3390/life15030327

[13] M. Khairandish, M. Sharma, V. Jain, J. Chatterjee and N. Jhanjhi, "A Hybrid CNN-SVM Threshold Segmentation Approach for Tumor Detection and Classification of MRI Brain Images," *IRBM*, vol. 43, no. 4, pp. 290-299, 2022. DOI: https://doi.org/10.1016/j.irbm.2021.06.003

[14] M. Kolla, R.K. Mishra, S.Z.u. Huq, Y. Vijayalata, M.V. Gopalachari and K. Siddiquee, "CNN-Based Brain Tumor Detection Model Using Local Binary Pattern and Multilayered SVM Classifier," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1-9, 2022. DOI: https://doi.org/10.1155/2022/9015778

[15] A. Rahman, M. Hayat, N. Iqbal, F.K. Alarfaj, S. Alkhalaf and F. Alturise, "Enhanced MRI brain tumor detection using deep learning in conjunction with explainable AI SHAP based diverse and multi feature analysis," *Scientific Reports*, vol. 15, no. 1, 2025. DOI: https://doi.org/10.1038/s41598-025-14901-4

[16] T.A. Soomro, L. Zheng, A.J. Afifi, A. Ali, S. Soomro, M. Yin and J. Gao, "Image Segmentation for MR Brain Tumor Detection Using Machine Learning: A Review," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 70-90, 2023. DOI: https://doi.org/10.1109/RBME.2022.3185292.

[17] G. Hemanth, M. Janardhan and L. Sujihelen, "Design and Implementing Brain Tumor Detection Using Machine Learning Approach," *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1289-1294, 2019. DOI: https://doi.org/10.1109/ICOEI.2019.8862553

[18] A.B. Abdusalomov, M. Mukhiddinov and T.K. Whangbo, "Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging," *Cancers*, vol. 15, no. 16, pp. 4172, 2023. DOI: https://doi.org/10.3390/cancers15164172

[19] S. Solanki, U.P. Singh, S.S. Chouhan and S. Jain, "Brain Tumor Detection and Classification Using Intelligence Techniques: An Overview," *IEEE Access*, vol. 11, pp. 12870-12886, 2023. DOI: https://doi.org/10.1109/ACCESS.2023.3242666

[20] N. Rasool and J.I. Bhat, "Brain tumour detection using machine and deep learning: a systematic review," *Multimedia Tools and Applications*, vol. 84, no. 13, pp. 11551-11604, 2024. DOI: https://doi.org/10.1007/s11042-024-19333-2.