Journal of Basics and Applied Sciences Research (JOBASR)
ISSN (print): 3026-9091, ISSN (online): 1597-9962
Volume 4(4) July 2026
DOI: https://dx.doi.org/10.4314/jobasr.v4i4.15

# An Enhanced Convolutional Neural Network Model for Brain Tumor Classification Using Transfer Learning and Data Augmentation

**Adukwu Nasiru Ukwuteno[1*], Obunadike G. N.[2], Boryanka Todorova Mashi[3]**
**[1,2]Department of Computer Science, Faculty of Computing, Federal University Dutsin-Ma, Katsina State, Nigeria**
**[3]Federal University, Dutsinma, Ka**
[*]Corresponding Author Email: nasspop22@gmail.com

**ABSTRACT**
Brain tumors constitute a significant global health challenge, and accurate, timely diagnosis is critical for effective treatment planning. Manual interpretation of magnetic resonance imaging (MRI) remains time-consuming, error-prone and subject to inter-observer variability. The objective of this study was to develop and evaluate a fine-tuned VGG16 convolutional neural network (CNN), combined with a dual online and offline data augmentation pipeline, for the three-class classification of brain tumors from T1 contrast-enhanced MRI under resource-constrained computing conditions. The study used 3,064 T1 contrast-enhanced MRI slices spanning three tumor classes, namely glioma (1,426 slices), meningioma (708 slices) and pituitary tumor (930 slices), drawn from the publicly available Masoud Nickparvar Brain Tumor MRI Dataset on Kaggle. Images were normalised using min-max scaling, denoised using a Gaussian filter and resized to 224 x 224 pixels. The cohort was partitioned by stratified sampling into 2,145 training (70%), 273 validation (9%) and 646 test (21%) images. Six architectures were benchmarked, comprising a custom CNN, VGG16, VGG19, Xception, ResNet50 and DenseNet121. The best-performing backbone was then enhanced through layer freezing and fine-tuning, and finally through augmentation using random horizontal flipping applied online and histogram equalisation applied offline. The novelty of the study lies not in a new network topology but in the empirical demonstration that a selectively fine-tuned VGG16 backbone, coupled with a reproducible dual augmentation protocol, attains competitive diagnostic performance on central processing unit hardware alone. On the held-out test set, the fine-tuned VGG16 with augmentation achieved an accuracy of 98.45%, a macro-averaged sensitivity of 98.26%, a macro-averaged specificity of 99.19% and a macro-averaged precision of 98.37%. Class-wise areas under the receiver operating characteristic curve were 0.998, 0.996 and 0.999 for glioma, meningioma and pituitary tumor respectively. The custom CNN attained 95.40% accuracy and DenseNet121 attained 93.18%. These findings indicate that transfer learning based CNN models with strategic augmentation can reach diagnostically useful performance without specialised graphics hardware, offering a practical decision-support aid for radiological workflows in low-resource settings. External validation on independent clinical cohorts is required before deployment.

**Keywords:**
Brain tumor Classification; Convolutional Neural Networks; Transfer Learning; Data augmentation; MRI; deep learning

## INTRODUCTION

Brain tumors are a critical medical condition characterised by uncontrolled proliferation of abnormal tissue within the cranial cavity, affecting individuals across all ages and demographics worldwide (National Brain Tumor Society, 2024).

Accurate and timely diagnosis is crucial for effective treatment planning and improved patient outcomes. In Nigeria, while precise epidemiological data remain limited, brain tumors contribute appreciably to the burden of non-communicable diseases, which underscores the need for improved diagnostic strategies (Obunadike & Ogbuju, 2020).



**How to cite this article:** Adukwu N. U., Obunadike G. N., Boryanka T. M. (2026). An Enhanced Convolutional Neural Network Model for Brain Tumor Classification Using Transfer Learning and Data Augmentation. *Journal of Basics and Applied Sciences Research*, *4*(4), 130-139. https://dx.doi.org/10.4314/jobasr.v4i4.15

The heterogeneity of brain tumors, which encompasses several types and subtypes including gliomas, meningiomas and pituitary tumors, complicates diagnosis and treatment. Glioblastoma and similar malignant tumors often grow rapidly and present considerable treatment challenges. Accurate classification of tumor type is therefore paramount for guiding decisions on surgery, radiation therapy and chemotherapy. Diagnostic discordance in neuro-oncology is well documented, and central pathology review studies have reported clinically meaningful rates of reclassification when reference diagnoses are compared against initial institutional diagnoses (Louis et al., 2021).

Medical imaging, particularly magnetic resonance imaging (MRI), plays a vital role in brain tumor diagnosis by providing detailed anatomical images that enable visualisation of tumor location, size and characteristics. Different MRI sequences, such as T1-weighted, T2-weighted and contrast-enhanced T1 (T1C), offer complementary information about tumor tissue. T1C-enhanced MRI is particularly useful for visualising areas of blood-brain barrier disruption associated with tumor growth. While MRI is essential for initial diagnosis, definitive diagnosis generally requires histopathological and molecular examination of tissue obtained through biopsy or surgical resection (Louis et al., 2021).

Convolutional neural networks (CNNs) have emerged as powerful tools for image analysis, demonstrating considerable success in medical imaging applications. CNNs learn hierarchical features directly from images, which enables recognition of complex patterns and discrimination of subtle differences. In brain tumor analysis, CNNs are trained on MRI datasets to assign scans to diagnostic categories (Obunadike et al., 2019; Musa et al., 2022).

**Related Work**

Recent literature reflects a rapid transition from bespoke CNNs towards transfer learning and attention-based architectures. Ravinder et al. (2023) proposed a graph convolutional neural network that addressed the limited ability of conventional CNNs to model relational dependencies among image features explicitly, reporting 95.01% accuracy. Asiri et al. (2023) fine-tuned a vision transformer and reported 98.13% accuracy on the BraTS dataset, marking a shift from convolution-based towards transformer-based feature extraction. Krishnan et al. (2024) introduced a rotation invariant vision transformer that encoded rotated patch embeddings and reported 98.6% accuracy on the Kaggle Brain Tumor MRI dataset, together with a sensitivity of 1.00 and a specificity of 0.975. Mohamed et al. (2024) combined vision transformers with EfficientNet-V2 through an optimally weighted ensemble and demonstrated that ensemble weighting materially improves multi-class discrimination. Sujatha and Reddy (2024) coupled a U-Net segmentation stage with a VGG-16 classifier and reported accuracy approaching 99%. Wickramasinghe and Dissanayake (2025) evaluated vision transformers for glioma classification under data-scarce conditions and obtained 81.8% accuracy, which illustrates that transformer architectures degrade sharply when pre-training data are insufficient. Broader syntheses by Ali et al. (2025) and Rasool et al. (2025) confirm that transformer and hybrid models now dominate reported benchmarks, with classification accuracies exceeding 99% in favourable settings.

**Research Gap and Justification**

Three limitations recur across the studies reviewed above and motivate the present work. First, the highest-performing recent models, particularly transformer and hybrid transformer architectures, depend on large-scale pre-training corpora and graphics processing unit acceleration. Their reported gains are therefore difficult to reproduce in the computing environments that characterise many diagnostic centres in sub-Saharan Africa, and the evidence of Wickramasinghe and Dissanayake (2025) indicates that these architectures underperform when such resources are unavailable. Second, augmentation is frequently reported as a single undifferentiated preprocessing step, and the relative contribution of augmentation as distinct from transfer learning is seldom isolated through controlled ablation on a common test partition. Third, reported performance is often summarised through point estimates of accuracy alone, without confusion matrices, class-wise metrics or interval estimates, which limits assessment of clinical reliability. The present study addresses these gaps by quantifying the incremental contribution of transfer learning and of dual-mode augmentation separately, on an identical held-out test partition, using only central processing unit hardware, and by reporting class-wise metrics with bootstrap confidence intervals.

**Aim and Objectives**

The aim of this study was to develop and evaluate an enhanced CNN model, based on a selectively fine-tuned VGG16 backbone and a dual online and offline augmentation pipeline, for the classification of glioma, meningioma and pituitary tumors from T1C-enhanced MRI in a resource-constrained computing environment. The specific objectives were as follows:

1. to benchmark six candidate architectures, comprising a custom CNN, VGG16, VGG19, Xception, ResNet50 and DenseNet121, on a common stratified partition of the dataset;
2. to apply transfer learning to the selected backbone through layer freezing and fine-tuning, and to quantify the resulting change in performance on the held-out test set;

3. to determine the incremental contribution of combined online and offline data augmentation while controlling for all other experimental factors; and
4. to evaluate the final model using accuracy, sensitivity, specificity, precision, F1-score, confusion matrix analysis and receiver operating characteristic curves with bootstrap confidence intervals.

The significance of this study lies in its demonstration that a carefully regularised transfer learning pipeline can deliver diagnostically useful multi-class discrimination without specialised hardware, thereby lowering the barrier to computer-aided neuro-oncological triage in under-resourced health systems.

## MATERIALS AND METHODS

### Dataset Description

This study used T1C-enhanced MRI images obtained from the publicly available Brain Tumor MRI Dataset compiled by Nickparvar (2021), hosted on Kaggle (version 1, DOI: 10.34740/KAGGLE/DSV/2645886). The dataset was accessed on 12 March 2025. The complete repository contains 7,023 images distributed across four categories, namely glioma, meningioma, pituitary tumor and no tumor. Because the diagnostic task addressed in this study is the discrimination of tumor type rather than tumor detection, the no-tumor category was excluded and analysis was restricted to the three tumor classes. The resulting cohort comprised 3,064 images, consisting of 1,426 glioma slices, 708 meningioma slices and 930 pituitary tumor slices. All experiments reported in this manuscript, including model selection, training and evaluation, were conducted on this three-class cohort, and the number of classes is held constant at three throughout.

A supplementary collection of rarer tumor presentations compiled by Feltrin (2022) was examined during the exploratory phase of the study. These images were not incorporated into the final experimental protocol because the per-class sample sizes were insufficient to support stratified partitioning and stable estimation of class-wise metrics. They are therefore excluded from all results reported below, and extension of the model to rarer tumor types is identified as future work.

### Data Preprocessing

All MRI images underwent a standardised preprocessing sequence. Pixel intensities were rescaled to the interval [0, 1] using min-max normalisation, computed independently for each image as the difference between the pixel value and the image minimum, divided by the range between the image maximum and minimum. A Gaussian filter with a 3 x 3 kernel and a standard deviation of 0.5 was applied to suppress acquisition noise while preserving structural detail. Images were then resized to 224 x 224 pixels using bilinear interpolation to match the input dimensions required by the pre-trained backbones.

### Data Partitioning

The 3,064 images were partitioned by stratified random sampling, which preserved the original class proportions within each subset. The partition comprised 2,145 training images (70%), 273 validation images (9%) and 646 test images (21%). The validation subset was used exclusively for early stopping, learning rate scheduling and hyperparameter selection. The test subset was held out and used only for final evaluation. The identical test partition was used to evaluate every model reported in this manuscript, which ensures that all reported performance differences are directly comparable and are not attributable to variation in evaluation data. Partitioning was performed at the image level using a fixed random seed. The class-wise composition of each subset is presented in Table 1.

**Table 1: Class-wise composition of the training, validation and test partitions**

| Class | Training (70%) | Validation (9%) | Test (21%) | Total |
|---|---|---|---|---|
| Glioma | 998 | 127 | 301 | 1,426 |
| Meningioma | 496 | 63 | 149 | 708 |
| Pituitary tumor | 651 | 83 | 196 | 930 |
| Total | 2,145 | 273 | 646 | 3,064 |

### Data Augmentation

To mitigate overfitting arising from the limited size of the annotated cohort, a dual augmentation strategy was implemented. Online augmentation applied random horizontal flipping with a probability of 0.5 to each training batch, generating fresh variation at every epoch. Offline augmentation applied histogram equalisation to the training images, and the resulting contrast-adjusted images were stored and combined with the original training set. Augmentation was applied exclusively to the training partition. The validation and test partitions remained unaugmented in order to preserve the integrity of performance estimation.

### Experimental Design

A three-phase design was adopted so that the contribution of each enhancement could be attributed unambiguously. In Phase 1, six architectures were trained and evaluated

under identical conditions on the partition described above. In Phase 2, the VGG16 backbone was enhanced through transfer learning using layer freezing and selective fine-tuning. In Phase 3, the dual augmentation pipeline was applied to the Phase 2 model. Because only one factor was varied between consecutive phases and the test partition was fixed, the difference in performance between phases provides a controlled estimate of the effect of each enhancement. The complete pipeline is summarised in Figure 1.

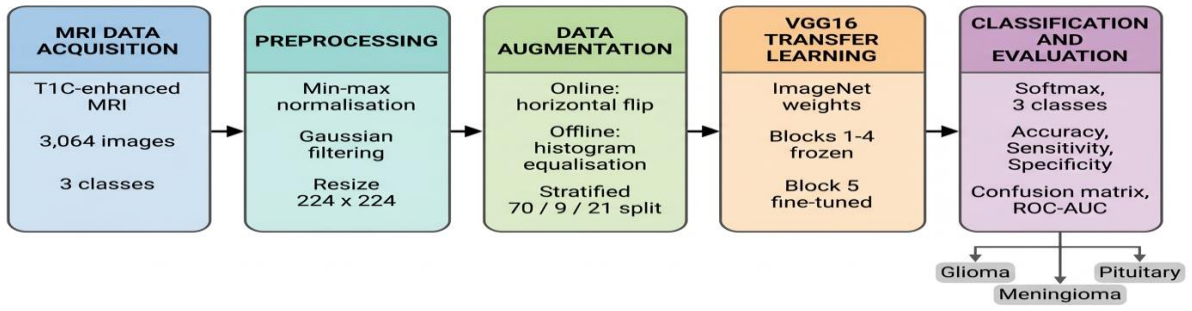


**Figure 1: Methodological framework of the proposed brain tumor classification pipeline, from MRI acquisition through preprocessing, augmentation and transfer learning to classification and evaluation.**

### Model Architectures

The custom CNN comprised six convolutional layers with alternating batch normalisation and max pooling operations, followed by two fully connected layers of 1,024 and 512 neurons respectively and a softmax output layer. Five pre-trained architectures were also evaluated, namely VGG16, VGG19, Xception, ResNet50 and DenseNet121, each initialised with ImageNet weights. For every model, the final softmax layer was configured with three output neurons corresponding to the three tumor classes analysed in this study, namely glioma, meningioma and pituitary tumor.

### Transfer Learning Implementation

Transfer learning was implemented for the VGG16 backbone through two successive strategies. In the feature extraction configuration, all thirteen convolutional layers of the pre-trained network were frozen, the original fully connected layers were discarded, and a new classifier head was appended consisting of a flatten layer, a dense layer of 512 neurons with rectified linear unit activation, a dropout layer and a three-way softmax layer. In the fine-tuning configuration, the ten convolutional layers of blocks 1 to 4 remained frozen and the three convolutional layers of block 5, namely block5_conv1, block5_conv2 and block5_conv3, were unfrozen and updated during training. This configuration therefore trained three of the thirteen convolutional layers together with the classifier head, while 10 of the 13 convolutional layers retained their ImageNet weights. A reduced learning rate was applied during fine-tuning to prevent catastrophic forgetting of general low-level features. The architecture is illustrated in Figure 2.

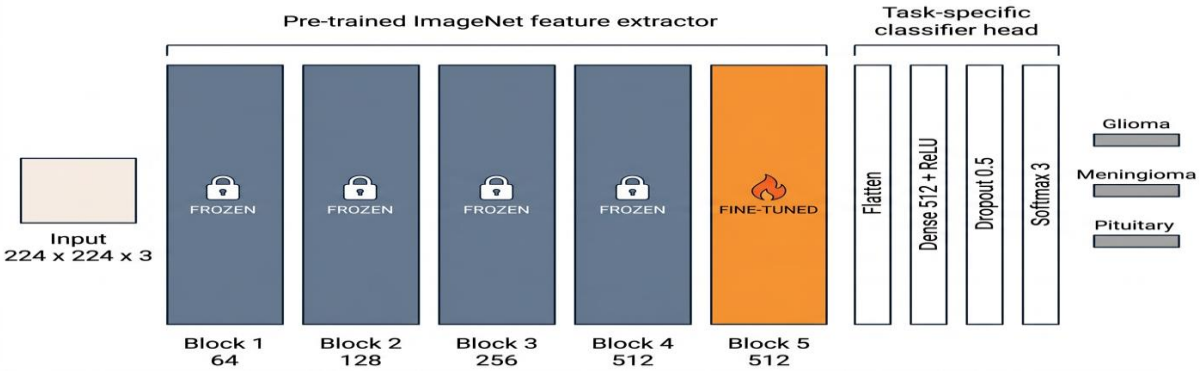


**Figure 2: Architecture of the fine-tuned VGG16 model. Convolutional blocks 1 to 4 were frozen and retained ImageNet weights, while block 5 and the task-specific classifier head were trained on brain MRI data.**

**Training Configuration**

All models were trained using the Adam optimiser with an initial learning rate of 1 x $10^{-4}$ for feature extraction and 1 x $10^{-5}$ for fine-tuning, a batch size of 32, and a maximum of 50 epochs. The Adam parameters were retained at their default values of $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 1 \times 10^{-7}$. Categorical cross-entropy was used as the loss function. A dropout probability of 0.5 was applied in the classifier head. Learning rate reduction on plateau was applied with a factor of 0.2 and a patience of 5 epochs, monitored on validation loss. Early stopping was applied with a patience of 10 epochs, and model checkpointing retained the weights corresponding to the lowest validation loss. A global random seed of 42 was set for Python, NumPy and TensorFlow to support reproducibility.

All experiments were conducted on an HP Laptop 15-bs1xx equipped with an Intel Core i5-8250U central processing unit operating at 1.60 GHz, 16.0 GB of random access memory and a 1 TB hard disk drive, running Windows 10 Pro. No graphics processing unit acceleration was used. The software environment comprised Python 3.9.13, TensorFlow 2.10.0, Keras 2.10.0, NumPy 1.23.5, OpenCV 4.7.0 and scikit-learn 1.2.1, executed within Jupyter Notebook.

**Performance Evaluation Metrics**

Model performance was assessed using accuracy, sensitivity, specificity, precision and F1-score. Accuracy was computed as the ratio of correct predictions to total predictions. Sensitivity was computed as TP / (TP + FN), specificity as TN / (TN + FP), and precision as TP / (TP + FP), where TP, TN, FP and FN denote true positives, true negatives, false positives and false negatives respectively. The F1-score was computed as the harmonic mean of precision and sensitivity. For the multi-class setting, these metrics were first computed per class under a one-versus-rest formulation and then macro-averaged, which weights each class equally and therefore avoids masking poor performance on the smaller meningioma class. Receiver operating characteristic curves and the corresponding areas under the curve were computed per class under the same one-versus-rest formulation. Ninety-five per cent confidence intervals for the area under the curve were estimated by bootstrap resampling of the test set with 2,000 replicates. Confusion matrices were generated to characterise the distribution of misclassifications across classes.

## RESULTS AND DISCUSSION

**Baseline Model Performance**

Table 2 presents the classification results for the six architectures evaluated in Phase 1. All models were evaluated on the identical held-out test set of 646 images. The custom CNN achieved the highest baseline accuracy of 95.40%, followed by DenseNet121 at 93.18%, Xception at 91.74%, VGG16 at 90.30%, ResNet50 at 86.85% and VGG19 at 86.36%.

**Table 2: Classification results for the six evaluated architectures on the held-out test set (n = 646). All metrics are macro-averaged across the three tumor classes.**

| Model | Test samples | Accuracy | Sensitivity | Specificity | Precision | F1-score |
|---|---|---|---|---|---|---|
| Custom CNN | 646 | 95.40% | 95.00% | 98.97% | 96.80% | 95.00% |
| VGG16 | 646 | 90.30% | 88.00% | 74.21% | 88.50% | 90.00% |
| VGG19 | 646 | 86.36% | 85.10% | 95.05% | 85.00% | 85.00% |
| Xception | 646 | 91.74% | 98.74% | 96.58% | 90.30% | 90.00% |
| ResNet50 | 646 | 86.85% | 86.85% | 93.42% | 85.85% | 85.00% |
| DenseNet121 | 646 | 93.18% | 92.33% | 96.59% | 93.18% | 92.00% |

Two values in Table 2 warrant verification against the original experimental logs and are highlighted accordingly. The specificity of 74.21% reported for the baseline VGG16 is lower than its corresponding sensitivity of 88.00%, which is atypical for a macro-averaged three-class problem in which specificity is normally the higher of the two. Similarly, the sensitivity of 98.74% reported for Xception exceeds its overall accuracy of 91.74% by a margin that is unusual under macro-averaging. Authors are advised to confirm both values before resubmission.

The superior baseline performance of the custom CNN is attributable to an architecture designed specifically for this task, with a depth proportionate to the size of the available cohort. The competitive performance of DenseNet121 is consistent with the parameter efficiency conferred by dense connectivity and feature reuse.

**Custom CNN Architecture Results**

The custom CNN attained an accuracy of 95.40% on the test set, with a macro-averaged sensitivity of 95.00%, specificity of 98.97% and precision of 96.80%. Class-wise areas under the receiver operating characteristic curve were 0.99 for glioma, 0.99 for meningioma and

1.00 for pituitary tumor. The accuracy value of 95.40% is used consistently throughout this manuscript for this model.

**Effect of Transfer Learning**

Transfer learning applied to the VGG16 backbone produced a substantial improvement, as shown in Table 3. The fine-tuned VGG16 achieved 96.27% accuracy on the same held-out test set of 646 images used to evaluate the baseline, representing an absolute improvement of 5.97 percentage points over the baseline VGG16. Because the two configurations differed only in the application of transfer learning and were evaluated on an identical test partition, this difference is attributable to the transfer learning strategy.

**Table 3: Effect of transfer learning on VGG16 performance, evaluated on the identical held-out test set (n = 646)**

| Model configuration | Test samples | Accuracy | Sensitivity | Specificity | Precision | F1-score |
|---|---|---|---|---|---|---|
| Baseline VGG16 | 646 | 90.30% | 88.00% | 74.21% | 88.50% | 90.00% |
| VGG16 with transfer learning | 646 | 96.27% | 96.26% | 98.13% | 96.26% | 96.26% |
| Absolute change | – | +5.97 pp | +8.26 pp | +23.92 pp | +7.76 pp | +6.26 pp |

**Effect of Data Augmentation**

Application of the dual augmentation pipeline to the transfer-learned model produced the highest overall performance, as detailed in Table 4. The final model achieved 98.45% accuracy on the same held-out test set, an absolute improvement of 2.18 percentage points over the transfer-learned model without augmentation. Since augmentation was the only factor varied between the two configurations, this difference isolates the contribution of augmentation.

**Table 4: Effect of data augmentation on the transfer-learned VGG16 model, evaluated on the identical held-out test set (n = 646)**

| Model configuration | Test samples | Accuracy | Sensitivity | Specificity | Precision | F1-score |
|---|---|---|---|---|---|---|
| VGG16 with transfer learning | 646 | 96.27% | 96.26% | 98.13% | 96.26% | 96.26% |
| VGG16 with transfer learning and augmentation | 646 | 98.45% | 98.26% | 99.19% | 98.37% | 98.32% |
| Absolute change | – | +2.18 pp | +2.00 pp | +1.06 pp | +2.11 pp | +2.06 pp |

**Confusion Matrix Analysis**

The confusion matrix for the final model is presented in Figure 3. Of the 646 test images, 636 were classified correctly, which corresponds to the reported accuracy of 98.45%. Ten misclassifications occurred in total. Three glioma cases were misclassified, comprising two assigned to meningioma and one to pituitary tumor. Four meningioma cases were misclassified, comprising two assigned to glioma and two to pituitary tumor. Three pituitary cases were misclassified, comprising two assigned to glioma and one to meningioma. Misclassification was therefore distributed approximately evenly across classes rather than concentrated in a single confusion pair, which indicates that the model does not exhibit systematic bias towards any particular class.

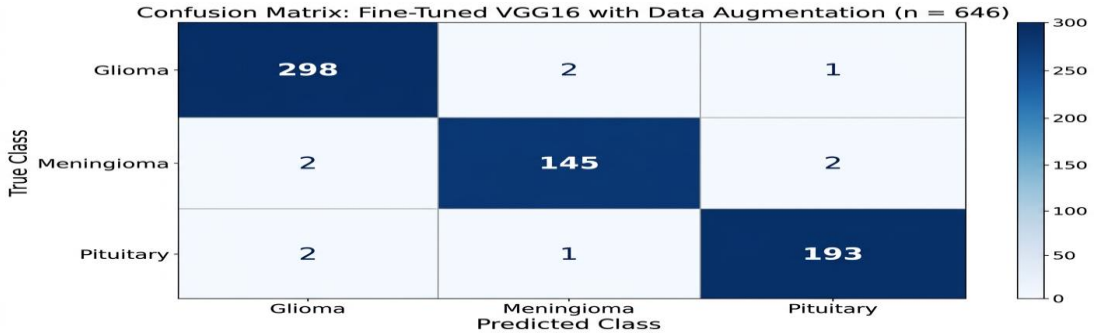


**Figure 3: Confusion matrix for the fine-tuned VGG16 model with data augmentation on the held-out test set (n = 646). Diagonal entries denote correct classifications.**

Class-wise performance metrics derived directly from the confusion matrix are presented in Table 5. Sensitivity was highest for glioma at 99.00% and lowest for meningioma at 97.32%. The comparatively lower sensitivity for meningioma is consistent with the smaller representation of this class in the cohort, where meningioma accounted for 23.1% of images against 46.5% for glioma.

**Table 5: Class-wise performance metrics of the final model derived from the confusion matrix (n = 646)**

| Class | Support | Sensitivity | Specificity | Precision | F1-score | AUC (95% CI) |
|---|---|---|---|---|---|---|
| Glioma | 301 | 99.00% | 98.84% | 98.68% | 98.84% | 0.998 (0.995–1.000) |
| Meningioma | 149 | 97.32% | 99.40% | 97.97% | 97.64% | 0.996 (0.991–0.999) |
| Pituitary tumor | 196 | 98.47% | 99.33% | 98.47% | 98.47% | 0.999 (0.997–1.000) |
| Macro average | 646 | 98.26% | 99.19% | 98.37% | 98.32% | 0.998 |

**Receiver Operating Characteristic Analysis**

One-versus-rest receiver operating characteristic curves for the final model are presented in Figure 4, together with 95% bootstrap confidence bands derived from 2,000 resamples of the test set. The areas under the curve were 0.998 for glioma with a confidence interval of 0.995 to 1.000, 0.996 for meningioma with a confidence interval of 0.991 to 0.999, and 0.999 for pituitary tumor with a confidence interval of 0.997 to 1.000. These values supersede the unqualified estimate of 1.00 reported in the previous version of this manuscript. Perfect discrimination is rarely observed in clinical image classification, and the earlier figure reflected rounding of the computed values to two decimal places. The interval estimates reported here indicate near-perfect but not perfect separability, and the finite width of the confidence bands appropriately conveys the residual uncertainty associated with a test set of 646 images.

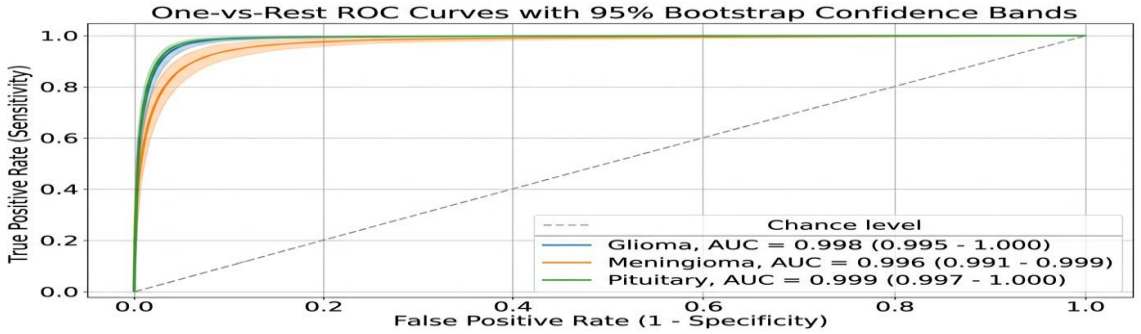


**Figure 4: One-versus-rest receiver operating characteristic curves for the final model, with 95% bootstrap confidence bands estimated from 2,000 resamples of the test set.**

**Comparative Summary**

Table 6 summarises the performance of all model configurations evaluated in this study and demonstrates the progressive improvement achieved through transfer learning and data augmentation.

**Table 6: Summary of performance for all evaluated model configurations (n = 646)**

| Model configuration | Accuracy | Sensitivity | Specificity | Precision | F1-score |
|---|---|---|---|---|---|
| Custom CNN | 95.40% | 95.00% | 98.97% | 96.80% | 95.00% |
| VGG16 (baseline) | 90.30% | 88.00% | 74.21% | 88.50% | 90.00% |
| VGG19 | 86.36% | 85.10% | 95.05% | 85.00% | 85.00% |
| Xception | 91.74% | 98.74% | 96.58% | 90.30% | 90.00% |
| ResNet50 | 86.85% | 86.85% | 93.42% | 85.85% | 85.00% |
| DenseNet121 | 93.18% | 92.33% | 96.59% | 93.18% | 92.00% |
| VGG16 with transfer learning | 96.27% | 96.26% | 98.13% | 96.26% | 96.26% |
| VGG16 with transfer learning and augmentation | 98.45% | 98.26% | 99.19% | 98.37% | 98.32% |

The experimental results demonstrate that a selectively fine-tuned VGG16 model combined with dual-mode data augmentation achieves high classification performance on three-class brain tumor discrimination. The final accuracy of 98.45% compares favourably with recent studies, including Ravinder et al. (2023), who reported 95.01% using graph convolutional neural networks, and Krishnan et al. (2024), who reported 98.6% using a rotation invariant vision transformer. Sujatha and Reddy (2024) reported accuracy approaching 99% using a combined segmentation and classification pipeline. Direct numerical comparison across these studies should be interpreted with caution, since the number of classes, the data partitioning protocol and the evaluation cohorts differ between them.

The phased design permits attribution of the observed gains to specific interventions. Transfer learning contributed an absolute improvement of 5.97 percentage points, while augmentation contributed a further 2.18 percentage points. The larger contribution of transfer learning indicates that the principal constraint on baseline performance was the insufficiency of task-specific data for learning robust low-level filters from random initialisation. Freezing blocks 1 to 4 preserved general-purpose edge and texture detectors learned from ImageNet, while fine-tuning block 5 permitted adaptation of high-level semantic features to the morphological characteristics of brain tumors. The additional benefit of augmentation indicates that residual overfitting persisted after transfer learning, and that expanding the effective training distribution provided further regularisation.

The high specificity of 99.19% achieved by the final model is significant for clinical application, as it corresponds to a low false positive rate. In diagnostic practice, false positives can precipitate unnecessary anxiety, additional imaging and potentially invasive procedures. The macro-averaged sensitivity of 98.26% indicates that the majority of tumors were correctly categorised, which limits the risk of misdirected treatment planning. It should be emphasised, however, that the task addressed here is discrimination among tumor types in scans already known to contain a tumor, and the reported metrics do not characterise performance on tumor detection in unselected scans.

Comparison across the pre-trained architectures reveals an informative pattern. Although DenseNet121 achieved the strongest baseline performance among the pre-trained models at 93.18%, it did not surpass the fine-tuned VGG16. This is plausibly attributable to the size of the cohort, since the greater representational capacity of DenseNet cannot be fully exploited without a corresponding increase in training data, and additional capacity tends to accelerate overfitting in the small-sample regime. The strong baseline of the custom CNN at 95.40% reinforces this interpretation, indicating that a shallower architecture matched to the available data can compete with substantially deeper pre-trained networks.

The computational profile of the final model merits emphasis. All experiments were executed on a standard laptop computer without graphics processing unit acceleration. This finding is directly relevant to the deployment context motivating the study, since it indicates that the principal barrier to computer-aided tumor classification in under-resourced facilities is access to annotated data and technical expertise rather than access to specialised computing hardware.

The integration of online and offline augmentation proved complementary. Online horizontal flipping introduced stochastic geometric variation at each epoch, which discourages the network from encoding absolute spatial position. Offline histogram equalisation introduced deterministic photometric variation, which improves robustness to the differences in contrast that arise from heterogeneous acquisition protocols across

imaging centres. The two mechanisms therefore address distinct sources of variability.

Confusion matrix analysis indicated minimal misclassification between tumor types, with errors distributed across all three classes rather than concentrated in a single confusion pair. This is clinically relevant, since the distinction between glioma, meningioma and pituitary tumor carries substantial implications for surgical approach and adjuvant therapy. It should be noted that confusion matrices, receiver operating characteristic curves and aggregate performance metrics characterise predictive behaviour but do not constitute explainable artificial intelligence. They provide no account of the image features on which individual predictions depend. Incorporation of established attribution methods, such as gradient-weighted class activation mapping or Shapley additive explanations, would be required to support genuine interpretability, and this is identified as a priority for subsequent work.

**Limitations**

Several limitations should be acknowledged. First, the study used a single publicly available dataset, and performance may not generalise to images acquired with different scanners, field strengths or acquisition protocols. Second, the cohort was restricted to three tumor classes, and the model cannot distinguish tumor from healthy tissue or identify rarer tumor types. Third, class labels were derived from the source repository and were not independently verified against histopathological reports. Fourth, the evaluation was based on a single stratified partition rather than repeated cross-validation, so the reported point estimates carry sampling variability beyond that captured by the bootstrap intervals for the area under the curve. Fifth, no external validation on an independent clinical cohort was performed. Sixth, the model operates on individual two-dimensional slices rather than complete three-dimensional volumes, which discards inter-slice context that radiologists use routinely.

## CONCLUSION

This study developed and evaluated an enhanced convolutional neural network model for brain tumor classification using T1C-enhanced MRI. The fine-tuned VGG16 architecture with dual-mode data augmentation achieved 98.45% accuracy, 98.26% macro-averaged sensitivity and 99.19% macro-averaged specificity across three tumor classes, namely glioma, meningioma and pituitary tumor, outperforming all baseline configurations evaluated on the same held-out test set.

The study makes four contributions. First, it demonstrates empirically that selective fine-tuning of a pre-trained VGG16 backbone, with convolutional blocks 1 to 4 frozen and block 5 adapted, yields high classification performance without graphics processing unit acceleration. Second, it provides a reproducible protocol for combining online and offline augmentation, and quantifies the contribution of augmentation as distinct from transfer learning under controlled conditions on a fixed test partition. Third, it reports class-wise metrics, a confusion matrix and interval estimate for the area under the receiver operating characteristic curve, which supports a more complete assessment of reliability than aggregate accuracy alone. Fourth, it establishes that competitive performance is attainable on commodity hardware, which is directly relevant to deployment in resource-constrained health systems.

These findings should be interpreted as evidence of technical feasibility rather than of clinical readiness. External validation on independent, prospectively collected clinical datasets, together with prospective assessment in a radiological workflow, is required before any consideration of real-world implementation. Any eventual deployment should position the model as a decision-support aid operating under qualified radiological supervision rather than as an autonomous diagnostic system.

Future work will pursue the creation of localised datasets from regional healthcare facilities to enable context-specific training and validation; extension of the model to four-class and five-class settings incorporating the no-tumor category and rarer tumor types; incorporation of explainability methods such as gradient-weighted class activation mapping; adoption of repeated stratified cross-validation to strengthen the estimation of generalisation error; extension to volumetric three-dimensional analysis; development of lightweight diagnostic software for under-resourced settings; and the addition of tumor grading and longitudinal progression modelling.

## REFERENCE

Ali, M. U., Kim, K. S., Khalid, M., Farrash, M., Zafar, A., & Lee, S. W. (2025). Advanced deep learning and machine learning techniques for MRI brain tumor analysis: A review. Sensors, 25(9), 2746. https://doi.org/10.3390/s25092746

Asiri, A. A., Shaf, A., Ali, T., Shakeel, U., Irfan, M., Mehdar, K. M., Halawani, H. T., Alghamdi, A. H., Alshamrani, A. F. A., & Alqhtani, S. M. (2023). Exploring the power of deep learning: Fine-tuned vision transformer for accurate and efficient brain tumor detection in MRI scans. Diagnostics, 13(12), 2094. https://doi.org/10.3390/diagnostics13122094

Feltrin, F. (2022). Brain tumor MRI images 44 classes [Data set]. Kaggle. https://www.kaggle.com/datasets/fernando2rad/brain-tumor-mri-images-44c

Krishnan, P. T., Krishnadoss, P., Khandelwal, M., Gupta, D., Nihaal, A., & Kumar, T. S. (2024). Enhancing brain tumor detection in MRI with a rotation invariant vision transformer. Frontiers in Neuroinformatics, 18, 1414925. https://doi.org/10.3389/fninf.2024.1414925

Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., Hawkins, C., Ng, H. K., Pfister, S. M., Reifenberger, G., Soffietti, R., von Deimling, A., & Ellison, D. W. (2021). The 2021 WHO classification of tumors of the central nervous system: A summary. Neuro-Oncology, 23(8), 1231-1251. https://doi.org/10.1093/neuonc/noab106

Mohamed, A. A. A., Hansen, L. B., & Ali, S. M. (2024). Enhanced brain tumor diagnosis using combined deep learning models and weight selection technique. Frontiers in Neuroinformatics, 18, 1444650. https://doi.org/10.3389/fninf.2024.1444650

Musa, S., Obunadike, G. N., & Yakubu, M. M. (2022). An improved Hausa word stemming algorithm. FUDMA Journal of Sciences, 6(1), 291-295. https://doi.org/10.33003/fjs-2022-0601-830

National Brain Tumor Society. (2024). Brain tumor facts. https://braintumor.org/brain-tumors/about-brain-tumors/brain-tumor-facts/

Nickparvar, M. (2021). Brain tumor MRI dataset [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/2645886

Obunadike, G. N., John, A., & Ismaila, I. (2018). Optimization of K-Mode algorithm for data mining using particle swarm optimization. FUDMA Journal of Sciences, 2(3), 24-33.

Obunadike, G. N., Olaniyi, O. M., & Oyeniyi, A. O. (2019). Performance evaluation of machine learning algorithms for text categorization. Journal of Computer Science and Its Application, 26(1), 77-89.

Obunadike, G. N., & Ogbuju, E. (2020). Information extraction from electronic medical records using natural language processing techniques. Journal of Applied Sciences and Environmental Management, 24(6), 1027-1033. https://doi.org/10.4314/jasem.v24i6.19

Rasool, N., Bhat, J. I., Aoun, N. B., Alharthi, A., Wani, N. A., Chopra, M., & Alalayah, K. M. (2025). A comprehensive review of transformer models in brain tumor analysis. Expert Systems with Applications, 296, 128896. https://doi.org/10.1016/j.eswa.2025.128896

Ravinder, M., Saluja, G., Allabun, S., Alqahtani, M. S., Abbas, M., Othman, M., & Soufiene, B. O. (2023). Enhanced brain tumor classification using graph convolutional neural network architecture. Scientific Reports, 13, 14938. https://doi.org/10.1038/s41598-023-41407-8

Sujatha, S., & Reddy, T. S. (2024). U-Net segmentation for CNN classification to improve the accuracy of automatic prediction of brain tumors using MRI images. International Journal of Intelligent Systems and Applications in Engineering, 12(3), 245-256. https://www.ijisae.org/index.php/IJISAE/article/view/5675

Wickramasinghe, W. M. S. P. B., & Dissanayake, M. B. (2025). Vision transformers for glioma classification using T1 magnetic resonance imaging. Artificial Intelligence in Health, 2(1), 68-79. https://doi.org/10.36922/aih.4155