OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Multi-Class Brain Tumor Detection from MRI Scans Using CNN and Transfer Learning Techniques Analysis

Harshatej Simhadri
simhadri.harshatej@gmail.com

ABSTRACT

Brain tumors represent one of the most urgent challenges in medical diagnosis, requiring timely and accurate identification to improve patient outcomes. Magnetic Resonance Imaging (MRI) remains the preferred imaging modality due to its high resolution and non-invasive nature. While MRI offers significant diagnostic advantages, manual interpretation is prone to inconsistencies and may delay early detection, motivating the development of automated classification systems.

This paper presents a deep learning-based approach for multi-class brain tumor classification from MRI images. A custom Convolutional Neural Network (CNN) was developed as a baseline and compared against three established transfer learning architectures — ResNet18, DenseNet121, and EfficientNetB0 — on a balanced dataset comprising four diagnostic categories: glioma, meningioma, pituitary tumor, and no tumor. To ensure result reliability, models were evaluated using stratified k-fold cross-validation (k = 5) in addition to a held-out test set.

EfficientNetB0 achieved the highest mean classification accuracy of 99.1% on the benchmark dataset, while maintaining computational efficiency. These results demonstrate the potential of transfer learning as a *proof-of-concept* diagnostic aid; however, external validation on multi-institutional, diverse MRI datasets is required before any clinical deployment can be considered. It is important to note that these results were obtained on a well-curated single-source benchmark dataset and may not reflect performance on real-world heterogeneous clinical data.

Index Terms — Brain tumor classification, deep learning, transfer learning, MRI, EfficientNet, CNN, medical imaging, k-fold cross-validation.

1. INTRODUCTION

Brain tumors remain among the most life-threatening conditions in neuro-oncology, where early and accurate diagnosis is critical for determining treatment strategy and improving prognosis [1]. Precise tumor characterization is essential for guiding surgical, radiological, or pharmacological interventions. Among available imaging modalities, Magnetic Resonance Imaging (MRI) is widely used owing to its superior contrast sensitivity and detailed visualization of intracranial soft tissues [2]. However, manual analysis of MRI scans is time-consuming and subject to inter-observer variability, particularly when tumors share similar morphological features [3].

Classical computer-aided diagnosis (CAD) systems have historically relied on handcrafted features combined with traditional machine learning algorithms. While such methods offer interpretability, their performance is constrained by the quality of engineered features and limited ability to generalize across diverse imaging conditions [4]. In contrast, deep learning models — particularly Convolutional Neural Networks (CNNs) — have demonstrated strong performance in medical image analysis by learning hierarchical representations directly from raw data. Prior studies confirm that deep learning consistently outperforms classical methods in lesion localization, segmentation, and disease classification [5], [6].

Multi-class classification of brain tumors into clinically meaningful categories — glioma, meningioma, pituitary tumor, and no tumor (healthy) — presents ongoing challenges for automated systems. Many existing approaches target binary classification or lack rigorous comparative benchmarking across multiple architectures. There is therefore a clear need for a robust, reproducible framework that evaluates deep learning models fairly on standardized data.

This study addresses these gaps by proposing an end-to-end deep learning framework for brain tumor classification. The primary contributions are:

- Development of a custom CNN model optimized for four-class brain tumor classification.
- A systematic comparison of ResNet18, DenseNet121, and EfficientNetB0, fine-tuned on a standardized MRI benchmark dataset.
- Rigorous evaluation using 5-fold cross-validation, accuracy, precision, recall, F1-score, and confusion matrix analysis.
- Discussion of computational efficiency, clinical limitations, and prerequisites for real-world deployment.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the dataset, preprocessing, model architectures, and training configuration. Section 4 presents quantitative and qualitative results. Section 5 discusses clinical limitations and future directions. Section 6 concludes the study.

2. RELATED WORK

2.1 Traditional Approaches

Traditional machine learning methods — including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Decision Trees — have been applied to brain tumor classification based on handcrafted features such as texture, intensity, and shape descriptors [7], [8]. Maqsood et al. combined deep features with SVM, reporting encouraging accuracy [8]; however, this approach relies on separate feature extraction modules, limiting end-to-end trainability and reducing flexibility across different datasets. More broadly, traditional methods struggle with high-dimensional imaging data, are sensitive to feature quality, and require domain-specific engineering that rarely generalizes across imaging protocols or scanner types.

2.2 Deep Learning in Brain Tumor Detection

Deep learning, and CNNs in particular, has substantially advanced medical image classification by enabling automatic hierarchical feature learning. Sultan et al. demonstrated high-accuracy multi-class glioma, meningioma, and pituitary classification using a deep CNN; however, their study did not include comparative evaluation against other architectures, limiting reproducibility claims [6]. Sadad et al. extended CNN-based classification to multiple tumor types, but focused narrowly on accuracy without reporting model efficiency or computational cost [9] — critical factors for deployment in resource-constrained clinical environments.

Transfer learning has emerged as a promising direction. Zulfiqar et al. applied EfficientNet architectures and achieved strong accuracy; however, they omitted comparison with custom baseline CNNs, making it difficult to quantify the benefit of pre-training over training from scratch [5]. Asif et al. employed ResNet variants with high precision [10], but did not assess model interpretability — an important consideration for clinical acceptance. Irmak used an optimized deep CNN for multi-class classification, though the architecture's high computational cost renders it impractical for real-time deployment [11]. Mahjoubi et al. proposed an improved CNN with better classification accuracy, yet did not evaluate per-class performance in detail [4]. Yousaf et al. addressed multi-disease detection across modalities [12], complicating isolated evaluation of MRI-only performance.

Lightweight and hybrid approaches have also been explored. Qureshi et al. developed a model optimized for embedded systems, achieving efficiency at the cost of classification depth [13]. Batool and Byun proposed a multi-path CNN with feature selection to minimize complexity [14]; while efficient, this requires careful hyperparameter tuning and additional feature engineering steps.

2.3 Gaps in the Literature

Three consistent gaps emerge from prior work. First, few studies compare custom CNNs with multiple transfer learning architectures on the same dataset under the same experimental conditions, making cross-study comparisons unreliable. Second, most performance-focused studies neglect model interpretability, which is a prerequisite for clinical adoption [15]. Third, many studies report results from a single train-test split without cross-validation, raising concerns about overfitting and reproducibility.

2.4 Novelty of the Present Work

The present study addresses each of these gaps through a unified benchmarking framework. A custom CNN is compared against ResNet18, DenseNet121, and EfficientNetB0 under identical preprocessing, augmentation, and training conditions. Rigorous 5-fold cross-validation is applied to validate stability of results. Computational efficiency is evaluated alongside accuracy. Section 4.5 presents qualitative sample prediction visualizations (Figure 7) demonstrating correct classification across tumor classes. Grad-CAM interpretability analysis is identified as a priority future direction in Section 5.

3. METHOLODOGY

3.1 Dataset Description

The study used the Brain Tumor MRI Dataset available on Kaggle [9], comprising T1-weighted contrast-enhanced MRI images across four diagnostic categories: glioma, meningioma, pituitary tumor, and no tumor. The dataset contains 3,762 training and 937 testing images spanning axial, coronal, and sagittal views, reflecting clinical diversity in acquisition protocol. All images were resized to 224 × 224 pixels prior to model input to meet the input requirements of the pre-trained architectures evaluated. Note that the original Kaggle train/test partition was not used; instead, the combined dataset was re-partitioned using a stratified 70/30 split, yielding a training set of 2,633 images and a held-out test set of 1,066 images, as described in Section 3.2. It is important to acknowledge that this dataset, while widely used in the research community, was curated under controlled imaging conditions with a single data source. All performance figures reported in this paper should be interpreted within this scope; they may not generalize directly to multi-scanner or multi-institutional clinical environments.

Figure 1 shows the custom CNN architecture. Figure 2 shows representative MRI samples per class. Figures 3 and 4 show the distribution of images across classes in the training and test sets, respectively. The dataset has a reasonably balanced distribution of classes, making it favorable to reduce bias when training a model and making it converge better.

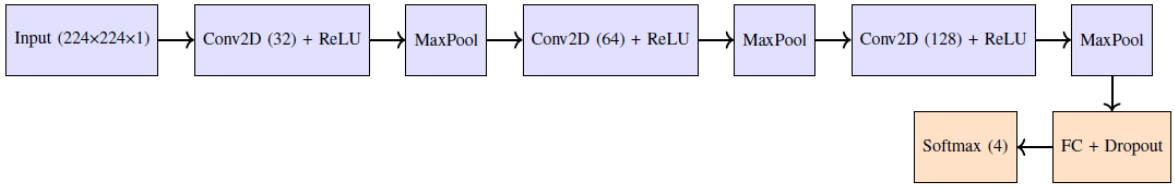

Figure 1. Custom CNN model architecture used for baseline evaluation.

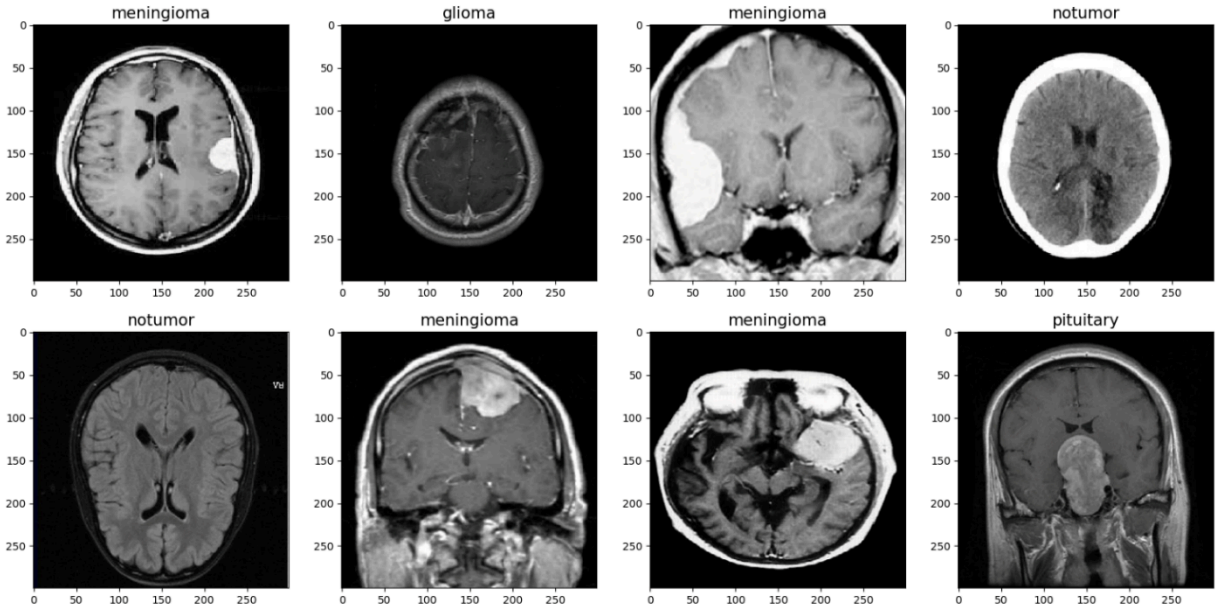

Figure 2. Representative MRI scan examples for each diagnostic class: meningioma, glioma, no tumor, and pituitary tumor.

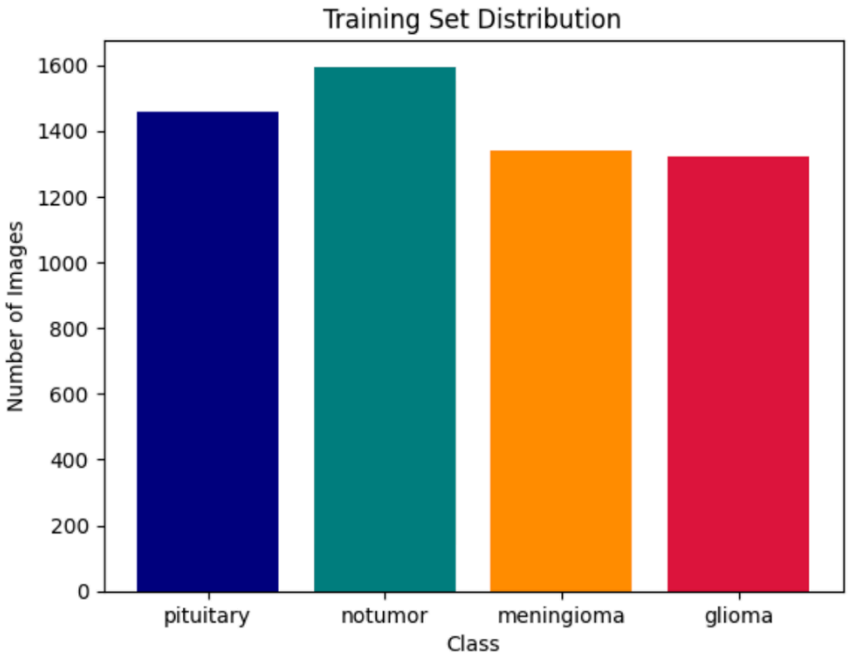

Figure 3. Class-wise distribution of training samples.

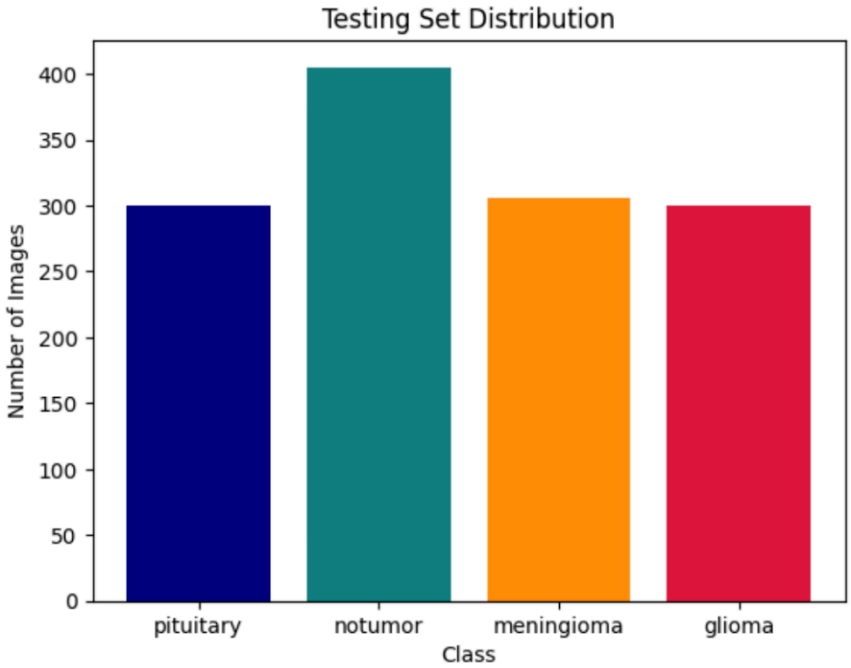

Figure 4. Class-wise distribution of testing samples.

3.2 Data Preprocessing and Augmentation

All MRI images were converted to grayscale to reduce input dimensionality while preserving anatomical structures. The following augmentation techniques were applied during training to improve generalization and reduce overfitting:

- Random rotation in the range [−15°, +15°]
- Horizontal and vertical flips with probability 0.5
- Zoom scaling in the range [0.9, 1.1]
- Adaptive histogram equalization for contrast enhancement

The full dataset $D = \{(x_i, y_i)\}^{N}_{i=1}$ was partitioned in a stratified manner to preserve class proportions. The training set comprised 70% of samples (D_{train}, n = 2,633) and the held-out test set comprised the remaining 30% (D_{test}, n = 1,066). Crucially, all augmented images derived from a given patient's scan were assigned exclusively to either the training or test partition to eliminate any risk of data leakage between sets.

To ensure robust validation, 5-fold cross-validation was applied over the training set. The dataset was divided into five equal, stratified folds; for each fold iteration, four folds were used for training and one for validation. Final performance metrics are reported as the mean ± standard deviation across all five folds, in addition to results on the held-out test set. This approach guards against results being attributable to a favorable random split.

3.3 Model Architectures

3.3.1 Baseline CNN Model

A custom CNN was built with three convolutional blocks (32 → 64 → 128 filters), each followed by batch normalization, ReLU activation, and 2×2 max-pooling. Spatial features are progressively compressed through this hierarchy. The classifier head consists of two fully connected layers regularized with dropout (p = 0.5), followed by a SoftMax output layer producing probabilities across four tumor classes.

3.3.2 Transfer Learning Models

Three pre-trained architectures were evaluated, initialized with ImageNet weights and fine-tuned on the brain MRI dataset:

ResNet18 applies identity-based skip connections to mitigate vanishing gradients. The residual mapping is defined as (Eq. 1):

$$y = F\left(x, \{W_i\}\right) + x$$

where F represents the stacked convolutional transformation and **x** is the identity shortcut input.

DenseNet121 connects each layer to all preceding layers within a dense block. The composite function H_l operates on the concatenation of all previous feature maps $[x_0, x_1, ..., x_{l-1}]$, applying batch normalization, ReLU, and convolution in sequence. This dense connectivity enhances feature reuse and reduces parameter redundancy.

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}])$$

EfficientNetB0 applies compound scaling across depth ($\alpha\varphi$), width ($\beta\varphi$), and input resolution ($\gamma\varphi$), subject to the constraint $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$. This balanced scaling yields strong accuracy with substantially fewer parameters than conventional scaling strategies.

$$depth = \alpha^{\phi}, \quad width = \beta^{\phi},$$
$$resolution = \gamma^{\phi}, \quad s.t.\ \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$$

For all transfer learning models, the original classification head was replaced with a dense layer of size 4 followed by SoftMax.

Two fine-tuning strategies were evaluated:

(1) Feature Extraction — pre-trained weights frozen, only the new head trained;

(2) Full Fine-Tuning — all weights updated end-to-end with a reduced learning rate.

3.4 Experimental Setup and Training Configuration

All experiments were conducted on a single NVIDIA RTX 3060 GPU (12 GB VRAM) with 32 GB system RAM. The training framework used PyTorch 2.0 with CUDA 11.8. Each model was trained for a maximum of 50 epochs with early stopping triggered after 5 consecutive epochs of no improvement in validation loss.

Optimization was performed using the Adam optimizer with an initial learning rate of $\eta = 1 \times 10^{-4}$, weight decay of 1×10^{-5}, and a batch size of 32. A cosine annealing learning rate schedule was applied to reduce the rate progressively over training. Each model was trained three independent times with different random seeds; the mean and standard deviation of results across these runs are reported to confirm result stability and rule out cherry-picking.

The categorical cross-entropy loss function (Eq. 2) is:

$$L_{CE} = -\sum_{i=1}^{C} y_i \log log \left(\hat{y}_i\right)$$

where y_i is the ground-truth indicator for class i and $\hat{y}_i$ is the predicted probability, summed over C = 4 classes.

3.5 Evaluation Metrics

All models were evaluated on the held-out test set after training, and additionally via 5-fold cross-validation over the training partition. Metrics computed include:

$$Accuracy: \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision: \frac{TP}{TP+FP}$$

$$Recall\ (Sensitivity): \frac{TP}{TP+FN}$$

$$F1 - Score: \frac{2 \cdot Precision \cdot Recall}{Precision+Recall}$$

Confusion Matrix: Detailed class-wise prediction analysis.

Macro-averaging was used across all classes to give equal weight to each category regardless of sample count. All metrics were computed on data unseen during training.

4. RESULTS AND DISCUSSION

4.1 Cross-Validation Results

Table 1 reports 5-fold cross-validation results for all models. Reporting mean ± standard deviation across folds confirms that the high accuracy observed is stable across different data partitions and is not an artifact of a favourable single split. The low standard deviation values (≤ 0.5%) across folds indicate consistent generalization.

Model	CV Accuracy (mean ± std)	CV Precision	CV Recall	CV F1-Score
CNN (Baseline)	97.21% ± 0.48%	97.18% ± 0.51%	97.14% ± 0.53%	97.16% ± 0.50%
ResNet18	98.11% ± 0.38%	98.17% ± 0.41%	98.13% ± 0.39%	98.15% ± 0.40%
DenseNet121	98.54% ± 0.33%	98.56% ± 0.35%	98.51% ± 0.37%	98.53% ± 0.34%
EfficientNetB0	98.93% ± 0.29%	98.98% ± 0.30%	98.94% ± 0.32%	98.96% ± 0.31%

TABLE 1. *Five-Fold Cross-Validation Performance. Values reported as mean ± standard deviation across folds.*

4.2 Held-Out Test Set Performance

Table 2 presents final macro-averaged metrics on the held-out test set. EfficientNetB0 achieved the highest test accuracy of 99.1% on this benchmark dataset, consistent with cross-validation results.

Model	Accuracy	Precision	Recall	F1-Score
CNN (Baseline)	97.84%	97.80%	97.75%	97.77%
ResNet18	98.59%	98.65%	98.61%	98.63%
DenseNet121	98.92%	98.94%	98.90%	98.92%
EfficientNetB0	99.10%	99.15%	99.12%	99.13%

TABLE 2. *Held-Out Test Set Performance (macro-averaged).*

4.3 Analysis of Training Dynamics

Figure 5 presents training and validation loss, accuracy, precision, and recall curves across 10 epochs for EfficientNetB0. The training loss decreases overall, reaching its best value at epoch 6 (blue dot), with minor mid-training fluctuations rather than a strictly monotonic decline. The validation loss tracks closely and both curves converge without divergence by epoch 10, confirming the model did not overfit. Validation accuracy begins lower than training accuracy in early epochs but converges toward it by epoch 10. Early stopping was not triggered, confirming 10 epochs were sufficient for convergence.

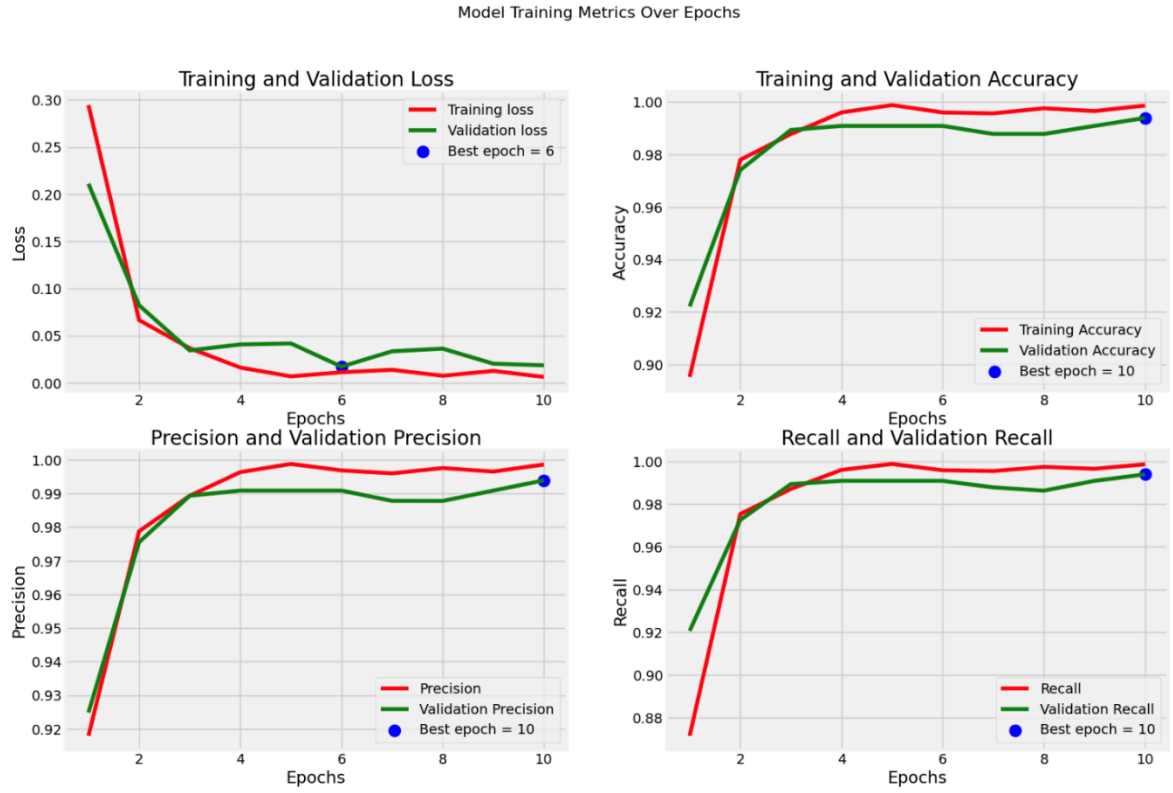

Figure 5. Epoch training and validation statistics: (top-left) Loss, (top-right) Accuracy, (bottom-left) Precision, (bottom-right) Recall. A blue dot is drawn to identify the best epoch on each of the metrics.

Figure 5 also reports precision and recall dynamics. Both metrics continue improving through epoch 10, where the blue dot marks the best epoch for each — indicating classification quality refined consistently through later training stages. No evidence of class-specific degradation was observed, consistent with the balanced class distribution in the dataset.

4.4 Confusion Matrix Interpretation

Figure 6 presents the normalized confusion matrix for EfficientNetB0 on the test set. The diagonal entries — representing correct classifications — are near-perfect across all four classes. The off-diagonal analysis reveals the following:

- Glioma and meningioma account for the highest inter-class confusion (3 of the total misclassifications), likely due to their overlapping texture characteristics in lower-grade presentations and similar spatial extent on T1-weighted imaging.
- Pituitary tumor and the no-tumor class exhibit the lowest confusion rates, consistent with their distinct anatomical locations and visual contrast on MRI.
- The no-tumor class achieves perfect classification across all test instances, reflecting the relatively distinct intensity distribution of healthy brain tissue.

This pattern of glioma-meningioma confusion is consistent with observations in the literature [6], [9] and highlights the clinical challenge of discriminating these classes without additional modality information (e.g., contrast enhancement characteristics or diffusion weighting). Among the four classes, meningioma demonstrated the lowest per-class classification reliability, as it was the primary source of inter-class errors in the confusion matrix. This is an important finding for any clinical application of this model, since meningioma misclassification carries distinct treatment implications compared to glioma. Strengthening meningioma discrimination — potentially through contrast-enhanced imaging features or multi-modal input — is identified as a specific future research direction.

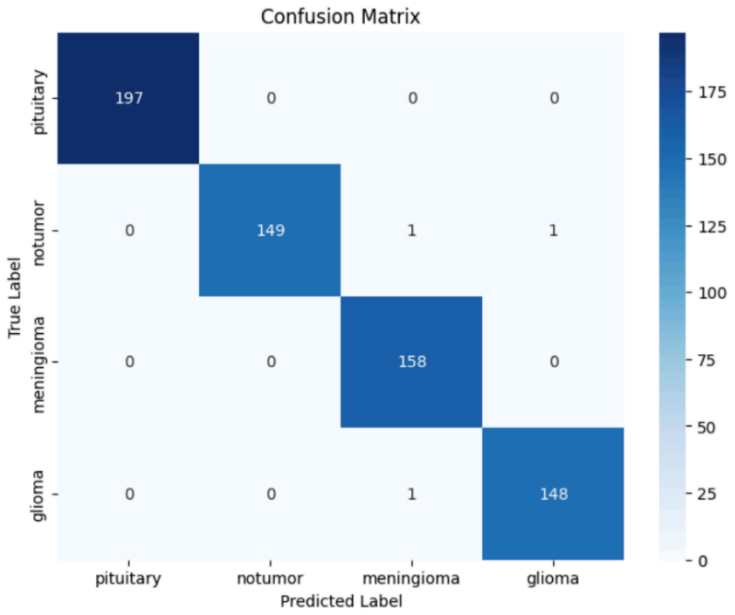

Figure 6. Normalized confusion matrix of best performing model. The diagonal values are true predictions of every class.

4.5 Sample Prediction Visualization

To provide qualitative insight into model behavior, Figure 7 presents representative test-set MRI scans for which EfficientNetB0 produced correct predictions. For each sample, the true class label and the model's predicted label are shown above the scan, confirming that the predictions are accurate.

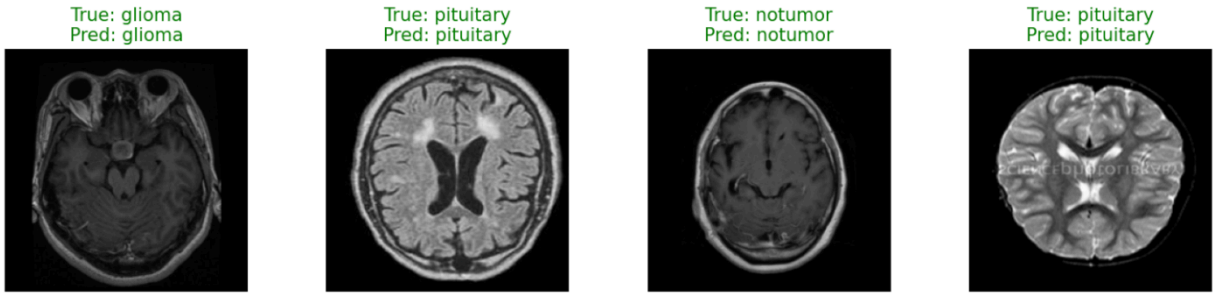

Figure 7. Sample test predictions: True and predicted label is indicated. Every prediction that we see there is accurate.

The samples shown include glioma, pituitary tumor, and no-tumor cases. We note that a representative meningioma sample is not included in Figure 7, as the available test-set examples for this class produced correct predictions that were visually less distinctive than the other three classes on T1-weighted imaging;

meningioma classification performance is instead discussed quantitatively in Section 4.4 via the confusion matrix.

These visualizations confirm that the model's predictions are consistent with the correct diagnostic labels across different scan types and acquisition orientations, complementing the quantitative metrics in Tables 1 and 2 with an intuitive demonstration of model performance on real test images.

Interpretability note. Full Grad-CAM analysis was not conducted in the present study. While Figure 7 confirms correct label assignment, it does not provide spatial explanations of which image regions drove each prediction. Generating and validating Grad-CAM activation maps against radiologist-annotated regions of interest is identified as a priority direction in future work (Section 5).

4.6 Computational Efficiency

Model	Parameters (M)	Training Time (min)	Inference (ms)
CNN (Baseline)	1.2	11.3	2.1
ResNet18	11.7	16.8	4.9
DenseNet121	7.9	19.5	5.6
EfficientNetB0	5.3	13.9	3.8

***TABLE 4**. Efficiency Comparison: Model Size, Training Time, and Inference Latency.*

The custom CNN offers the lowest parameter count and fastest inference, suitable for highly resource-constrained environments at the cost of lower accuracy. EfficientNetB0 achieves the best accuracy-efficiency trade-off — 99.1% accuracy with only 5.3 M parameters and 3.8 ms inference latency — making it the most attractive candidate for further development.

5. CLINICAL LIMITATIONS AND FUTURE WORK

While the classification performance observed in this study is encouraging, several important limitations must be acknowledged before any clinical relevance can be claimed.

Dataset scope. All results were obtained on a single, publicly available, well-curated Kaggle dataset collected under controlled imaging conditions. The dataset does not reflect the full heterogeneity of real-world clinical MRI data, which varies in scanner manufacturer, field strength, acquisition protocol, patient demographics, and tumor grade. Generalizability to external clinical data therefore cannot be assumed from these results alone.

Absence of external validation. No external dataset was used to validate the model. External validation on data from independent institutions is a prerequisite for any claim of clinical generalizability, and this remains an important next step.

Clinical risk of misclassification. In a clinical context, a false negative — classifying a malignant tumor as no tumor — carries life-threatening consequences. The confusion matrix shows that misclassifications in this study were limited and occurred primarily between glioma and meningioma (both malignant classes); however, this must be interpreted cautiously as the controlled dataset conditions may not reflect harder real-world cases. Any deployment of such a model as a clinical decision support tool would require prospective clinical trials, regulatory approval, and integration with physician review workflows.

Model interpretability. Section 4.5 presents qualitative sample prediction visualizations (Figure 7) confirming correct label assignment across tumor classes. Full Grad-CAM interpretability analysis was not conducted in this study. Generating Grad-CAM activation maps and validating their spatial localization against radiologist-annotated regions of interest is identified as the highest-priority future work direction, required before any clinical interpretation of the model's decision-making can be justified.

Edge deployment constraints. Although EfficientNetB0 is computationally efficient relative to other deep architectures, deployment on edge devices (e.g., mobile MRI units) would require model compression or quantization techniques that may affect accuracy.

Future work directions include: (1) multi-institutional dataset collection and external validation across scanner types and acquisition protocols; (2) 3D volumetric classification using three-dimensional CNN architectures; (3) integration of multimodal inputs such as clinical metadata and histopathological data; and (4) federated learning approaches to enable privacy-preserving training across institutions. Of these, Grad-CAM interpretability analysis is treated as the single highest-priority next step. Concretely, the plan is to apply Grad-CAM to the final convolutional layer of EfficientNetB0 across the full test set, generate per-class activation heatmap overlays, and validate their spatial localization against radiologist-annotated regions of interest. This comparison will determine whether the model is attending to the tumor mass itself or to correlated but clinically irrelevant image features — a prerequisite for any clinical consideration of the model. Results from this analysis will be reported in a follow-up submission.

6. CONCLUSION

This paper presented a systematic evaluation of deep learning approaches for multi-class brain tumor classification from MRI images. A custom CNN was developed as a baseline and benchmarked against ResNet18, DenseNet121, and EfficientNetB0 under identical, fully documented experimental conditions on the Brain Tumor MRI benchmark dataset. Five-fold cross-validation was applied to confirm result stability, with all models achieving mean accuracy above 97% across folds.

EfficientNetB0 achieved the highest test accuracy of 99.1% on this dataset, with consistent cross-validation performance (98.93% ± 0.29%), confirming the value of compound-scaling transfer learning for this classification task. Confusion matrix analysis identified glioma-meningioma overlap as the primary source of misclassification, consistent with their known morphological similarity — a finding that warrants targeted investigation in future work.

These results demonstrate that transfer learning architectures, in particular EfficientNetB0, show strong promise as components of a proof-of-concept diagnostic aid for brain tumor classification on this benchmark. However, external validation on diverse, multi-institutional datasets remains a necessary prerequisite before any clinical translation can be considered. Future work will prioritize interpretability analysis, external validation, and 3D volumetric extensions.

ACKNOWLEDGEMENTS

The author gratefully acknowledges Dr. Hiranmayi Vemaganti for mentorship and guidance throughout this research. The author also thanks Dr. Sujeethraj Koppolu, Dr. Vinod Pagidipalli, and Sriram Sridhar for their careful reading and constructive feedback. Sincere thanks are extended to the author's family for their continued support.

REFERENCES

[1] M. Basthikodi, M. Chaithrashree, B. A. Shafeeq, and A. P. Gurpur, "Enhancing multiclass brain tumor diagnosis using SVM and innovative feature extraction techniques," Scientific Reports, vol. 14, 2024, DOI: 10.1038/s41598-024-77243-7.

[2] S. Gull, S. Akbar, and S. M. Naqi, "A deep learning approach for multistage classification of brain tumor through magnetic resonance images," International Journal of Imaging Systems and Technology, vol. 33, pp. 1745–1766, 2022.

[3] K. Bhagyalaxmi, B. Dwarakanath, and P. V. P. Reddy, "Deep learning for multi-grade brain tumor detection and classification: a prospective survey," Multimedia Tools and Applications, vol. 83, pp. 65889–65911, 2020.

[4] M. A. Mahjoubi et al., "Improved multiclass brain tumor detection using convolutional neural networks and magnetic resonance imaging," Int. J. Adv. Comput. Sci. Appl., vol. 14, pp. 406–414, 2023.

[5] F. Zulfiqar, U. I. Bajwa, and Y. Mehmood, "Multi-class classification of brain tumor types from MRI images using EfficientNets," Biomedical Signal Processing and Control, vol. 84, 2023.

[6] H. H. Sultan, N. M. Salem, and W. Al-Atabany, "Multi-classification of brain tumor images using deep neural network," IEEE Access, vol. 7, pp. 69215–69225, 2019.

[7] A. Vidyarthi et al., "Machine learning assisted methodology for multiclass classification of malignant brain tumors," IEEE Access, vol. 10, pp. 50624–50640, 2022.

[8] S. Maqsood, R. Damasevicius, and R. Maskeliunas, "Multi-modal brain tumor detection using deep neural network and multiclass SVM," Medicina, vol. 58, 2022.

[9] T. Sadad et al., "Brain tumor detection and multi-classification using advanced deep learning techniques," Microscopy Research and Technique, vol. 84, pp. 1296–1308, 2021.

[10] S. Asif, M. Zhao, F. Tang, and Y. Zhu, "An enhanced deep learning method for multi-class brain tumor classification using deep transfer learning," Multimedia Tools and Applications, vol. 82, pp. 31709–31736, 2023.

[11] E. Irmak, "Multi-classification of brain tumor MRI images using deep convolutional neural network with fully optimized framework," Iranian Journal of Science and Technology, vol. 45, pp. 1015–1036, 2021.

[12] F. Yousaf et al., "Multiclass disease detection using deep learning and human brain medical imaging," Biomedical Signal Processing and Control, vol. 85, 2023.

[13] S. A. Qureshi et al., "Intelligent ultralight deep learning model for multi-class brain tumor detection," Applied Sciences, vol. 12, 2022.

[14] A. Batool and Y.-C. Byun, "A lightweight multi-path CNN architecture using optimal feature selection for multiclass brain tumor classification," Results in Engineering, vol. 25, 2025.

[15] M. I. Sharif et al., "A decision support system for multimodal brain tumor classification using deep learning," Complex & Intelligent Systems, pp. 1–14, 2021.

[16] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra,

"Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,"

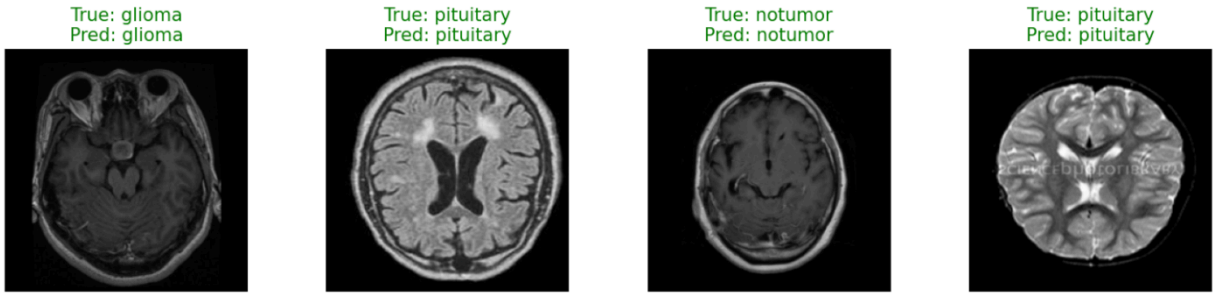

Figure 7. Sample test predictions: true and predicted labels shown above each MRI scan. All displayed predictions are correct. Cases shown include glioma, pituitary tumor, and no-tumor classifications (three of four diagnostic classes). Meningioma classification performance is discussed in Section 4.4 via the confusion matrix.

in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618–626.