Brain Tumor Detection in MRI Scans

Department of Computer Science and Engineering

Premier University, Chittagong

Course Title: Machine Learning Laboratory
Course Code: CSE 458

Author 1: Aditya Sankar Das
Student ID: 2104010202310

Author 2: Mojahidul Islam
Student ID:2104010202323

Author 3: Abidul Haque alvee
Student ID: 0222210005101142

Author 4: Rahat Imroz
Student ID: 0222220005101128

Submitted to:
MD TAMIM HOSSAIN
Lecturer
Department of Computer Science and Engineering
Premier University,
Chittagong, Bangladesh

Submission Date: 23 November, 2025

Brain Tumor Detection in MRI Scans

Aditya S. Das[1], Mojahidul Islam[1], Abidul H. alvee[1], Rahat Imroz[1]

2104010202310, 2104020202323, 222210005101142, 0222220005101128

[1]Department of Computer Science and Engineering

Premier University, Chittagong, Bangladesh

Abstract

Brain tumor detection using MRI scans is a critical part of medical diagnosis, yet manual interpretation is often time-consuming and prone to human error. Deep learning provides a scalable and reliable approach for automated tumor classification. This study evaluates five deep learning models—SimpleCNN, VGG16, EfficientNetB0, DenseNet121, and ResNet50—using a large MRI dataset containing 8,277 training images and 1,816 testing images. The dataset includes two classes (tumor and no-tumor) and is enhanced using data augmentation techniques such as rotation, shifting, shearing, zooming, and horizontal flipping to improve model generalization.

All models are trained using the Adam optimizer and binary cross-entropy loss, with performance assessed using accuracy, precision, recall, F1-score, and AUC. Experimental results show that VGG16 achieves the highest accuracy at 94.88%, followed by DenseNet121 at 92.95%. Visualization techniques—including augmentation previews and Grad-CAM heatmaps—are used to interpret how the models identify tumor-related features in MRI scans.

The findings demonstrate that deep learning can meaningfully support radiologists by providing fast, consistent, and automated diagnostic assistance. The report also discusses dataset limitations, class imbalance issues, and ethical considerations related to clinical deployment and patient data privacy.

Keywords: Brain Tumor Detection, Deep Learning, MRI Classification, Convolutional Neural Networks, Medical Image Analysis

1 Introduction

Brain tumors remain one of the most severe neurological conditions worldwide, often leading to life-threatening complications when not detected and treated promptly. Magnetic Resonance Imaging (MRI) is widely used in clinical practice for identifying brain abnormalities due to its ability to capture detailed anatomical structures without exposing patients to ionizing radiation. Despite its diagnostic value, the manual interpretation of MRI scans is a complex and time-consuming task. Radiologists must carefully inspect each image, a process that can be influenced by heavy workloads, fatigue, and individual differences in interpretation. These factors increase the likelihood of delayed or inconsistent diagnoses, especially in healthcare environments with limited specialized resources.

The rapid advancement of artificial intelligence, particularly deep learning, has created new opportunities for enhancing the efficiency and reliability of medical image analysis. Convolutional Neural Networks (CNNs) have demonstrated strong capabilities in learning discriminative features directly from raw image data, eliminating the need for handcrafted feature extraction. When properly trained, CNN-based models can classify MRI images with high accuracy, provide consistent predictions, and scale to large volumes of data. As a result, automated brain tumor detection systems are emerging as valuable decision-support tools that can complement radiologists' expertise.

This study investigates the effectiveness of deep learning models for classifying MRI images into two categories: *tumor* and *no tumor*. Five widely adopted architectures—Simple CNN, VGG16, EfficientNetB0, DenseNet121, and ResNet50—are evaluated to identify the most suitable approach for this task. The work also examines the role of data preprocessing, augmentation strategies, and architectural design in influencing model performance. To enhance interpretability, Grad-CAM heatmaps are used to visualize the regions of MRI images that contribute most to model decisions, helping bridge the gap between AI predictions and clinical trust. Ultimately, this research aims to support faster and more consistent brain tumor screening, contributing to improved diagnostic workflows and better patient outcomes.

2 Related Work

2.1 Overview of Brain Tumor Classification

Early research on brain tumor classification commonly relied on handcrafted image features derived from texture descriptors, edge-based characteristics, and statistical intensity profiles. While these traditional methods enabled initial progress, they were highly sensitive to variations in MRI acquisition protocols, noise levels, and contrast inconsistencies, ultimately limiting their robustness and generalization capability.

The advent of deep learning fundamentally transformed this domain by enabling models to automatically learn hierarchical and discriminative representations directly from raw MRI data. Convolutional neural networks (CNNs), in particular, demonstrated re-

markable effectiveness in medical imaging tasks due to their ability to capture spatially localized patterns and complex structural cues [6, 7]. Architectures such as VGG16 [1], DenseNet121 [2], and EfficientNet [3] have been widely adopted due to their strong feature extraction capabilities and scalability. Moreover, interpretability methods like Grad-CAM [4] further enhanced their applicability in clinical settings by enabling visual explanations of model decisions. The success of deep CNNs in areas such as skin cancer classification [5] further highlights their transformative impact on medical image analysis.

Overall, the transition from handcrafted descriptors to deep learning–based approaches has enabled more accurate, scalable, and clinically meaningful brain tumor classification systems.

2.2 Existing Models

Convolutional Neural Networks (CNNs) have become the leading methodology for medical image analysis, particularly in brain tumor detection. Well-established architectures such as VGG16, ResNet50, and DenseNet121 are widely adopted due to their strong hierarchical feature-learning capabilities. Many studies utilize transfer learning to overcome the challenge of small medical datasets and to enhance model stability. Recent advancements explore hybrid CNN architectures, attention-enhanced models, and 3D CNN frameworks that leverage volumetric MRI data for improved spatial understanding.

2.3 Comparison with Other Approaches

Deep learning approaches consistently outperform traditional handcrafted feature methods across accuracy, robustness, and adaptability. While classical techniques struggle with MRI variability, CNN-based models automatically capture complex spatial and structural features. Transfer learning offers strong performance benefits but may encounter domain mismatch, as pretrained models are typically trained on natural images rather than medical data. Meanwhile, 3D CNNs provide richer contextual awareness but require significantly larger datasets and more computational resources. In this study, a custom 2D CNN model is adopted to achieve a practical balance between performance, computational efficiency, and compatibility with the available MRI dataset.

3 Dataset

3.1 Dataset Structure

The dataset employed in this study is organized hierarchically to support efficient supervised learning. Two primary directories are defined: `Dataset/Training` and `Dataset/Testing`. Each directory contains two subfolders corresponding to the two target classes:

- `tumor` – MRI scans containing visible brain tumors;
- `no tumor` – MRI scans showing normal brain structures.

This structured, folder-based arrangement enables deep learning frameworks such as TensorFlow and Keras to automatically infer class labels during the data-loading process. As a result, preprocessing is simplified, errors are minimized, and class-label consistency is maintained throughout the experiments.

3.2 Dataset Statistics

The dataset contains a sufficiently large and balanced number of MRI images for training, validating, and evaluating the deep learning models. In total, 8,277 images were used for training and validation, split as follows:

- **Training subset:** 6,622 images
- **Validation subset:** 1,655 images (20% of the training data)
- **Testing set:** 1,816 images

All images were standardized to a unified resolution of 224×224 pixels, converted to RGB format, and normalized to the $[0, 1]$ range. A batch size of 16 was chosen to maintain a balance between computational efficiency and stable gradient updates during training.

3.3 Exploratory Data Analysis (EDA)

A detailed visual and statistical inspection of the dataset revealed several important characteristics that informed subsequent preprocessing decisions:

- **Brightness and contrast variations:** Caused by differences in MRI scanners, acquisition protocols, and patient conditions.
- **Tumor diversity:** Tumors vary significantly in location, shape, size, and boundary sharpness, making classification non-trivial.
- **Class imbalance:** A slight dominance of tumor images over non-tumor images, which may bias model training.
- **Noise and artifacts:** Some scans contain motion blur, scanner-induced noise, or irrelevant non-brain structures.

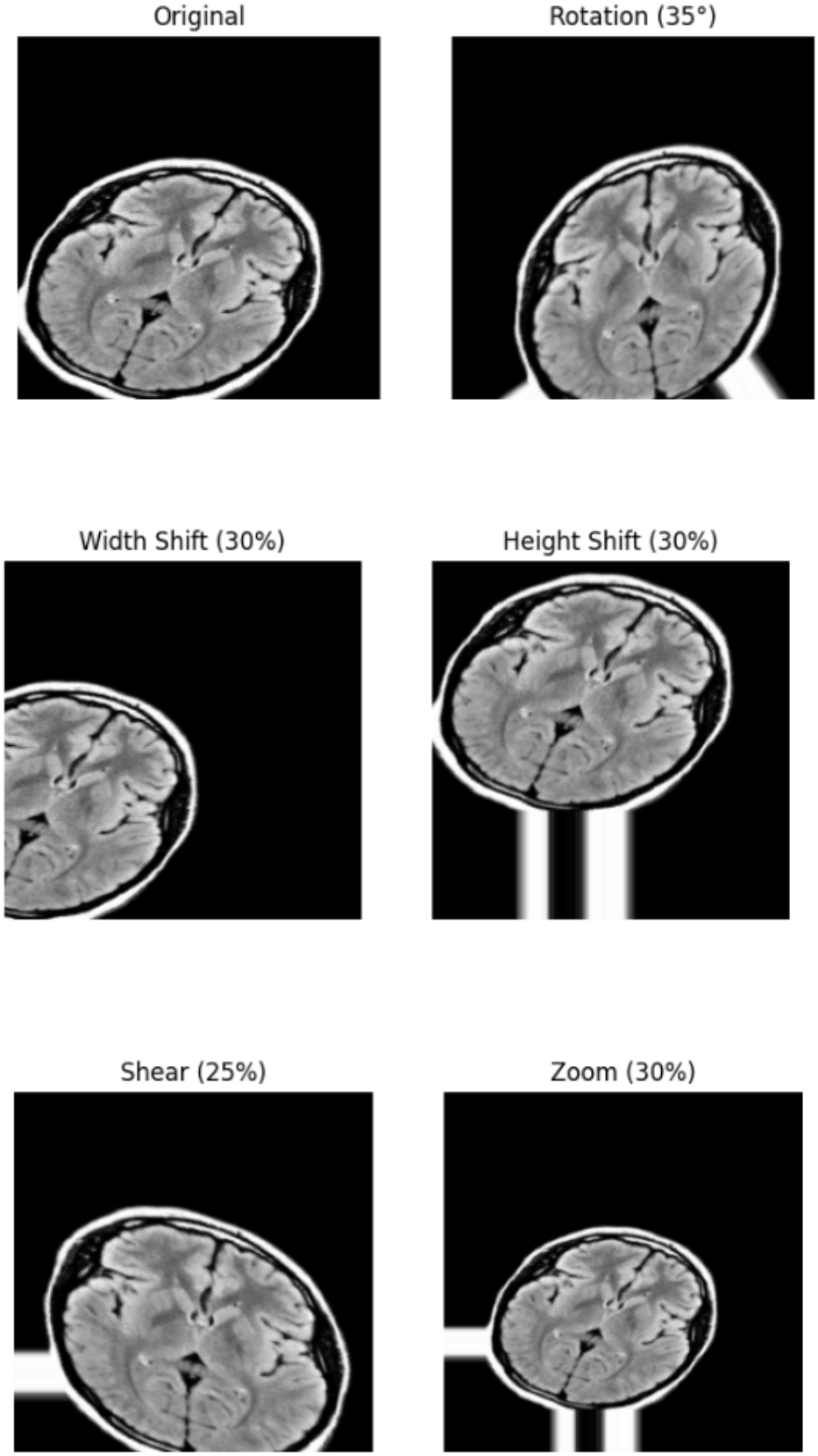
These observations motivated the use of extensive data augmentation and class weighting techniques to improve the model's robustness and generalization capability.

3.4 Preprocessing and Augmentation

To ensure consistency across all MRI samples and improve the model's ability to generalize to unseen cases, a comprehensive preprocessing and augmentation pipeline was applied before model training. These steps help reduce overfitting, increase robustness, and preserve clinically meaningful tumor characteristics.

- **Resizing:** All MRI images were resized to a uniform spatial dimension of 224×224 pixels. This resolution is commonly used in CNN architectures and ensures compatibility across all pretrained models.
- **Normalization:** Pixel intensities were scaled to the $[0, 1]$ range. Normalization stabilizes training, accelerates convergence, and prevents extreme values from dominating gradient updates.
- **Data Augmentation:** To artificially expand the training distribution and introduce geometric variability, augmentation was applied exclusively during training. The following transformations were used:
  - **Rotation:** Random rotations up to $\pm 35^\circ$ simulate different patient head orientations.
  - **Width and height shifts:** Translations up to 30% mimic slight misalignments during MRI acquisition.
  - **Shear transformations:** Distortions up to 25% provide additional shape diversity while retaining tumor boundaries.
  - **Zoom:** Up to 30% zoom variations enhance the model's sensitivity to tumors appearing at different scales.
  - **Horizontal flipping:** Introduces mirrored anatomical variations to help avoid overfitting.
- **Validation Split:** A 20% subset of the training data was reserved for validation to monitor overfitting and tune hyperparameters effectively.
- **Class Weighting:** Since the dataset exhibits a mild imbalance between tumor and non-tumor images, class weighting was employed to ensure that minority classes contribute proportionally to the loss function.

Overall, these preprocessing and augmentation steps significantly enrich data variability while ensuring that medically relevant tumor structures remain intact. This improves the robustness of all deep learning models and supports reliable tumor classification.


Original
Rotation (35°)
Width Shift (30%)
Height Shift (30%)
Shear (25%)
Zoom (30%)

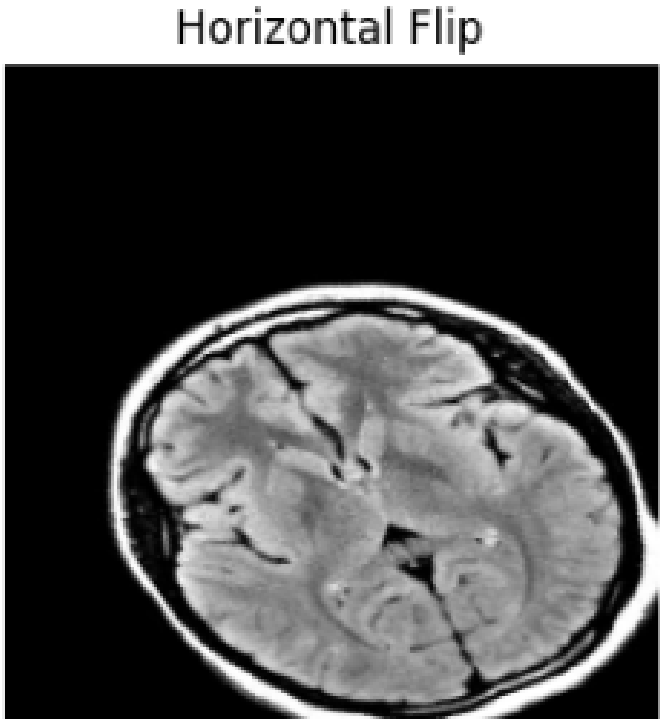
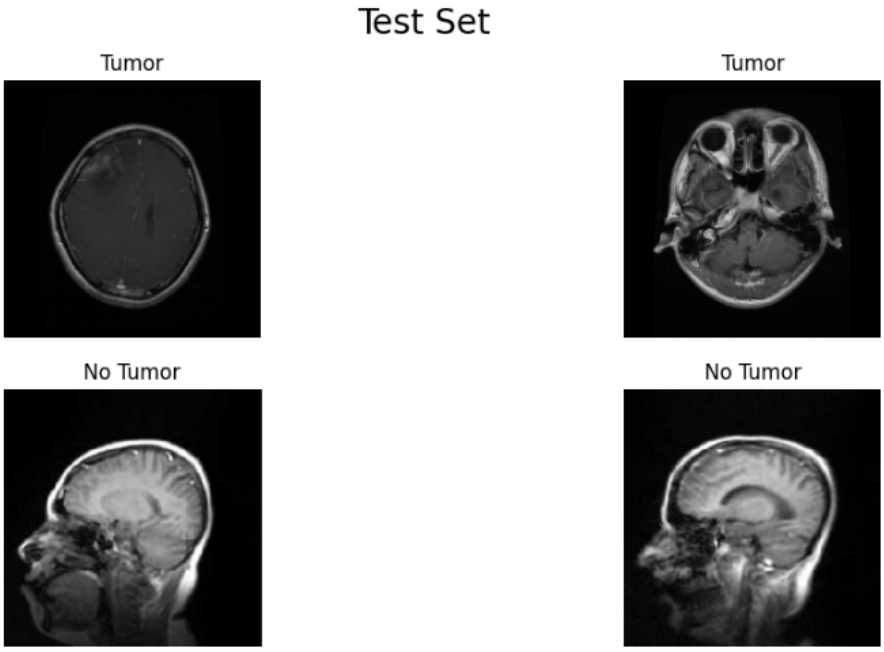


fig. Illustration of the data augmentation techniques applied during training

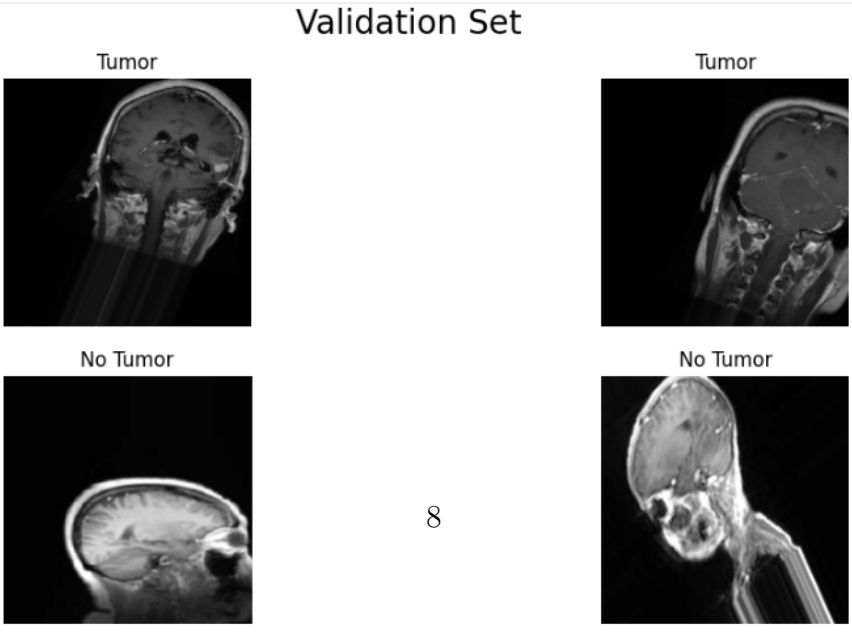
3.5 Sample Images from Dataset

The dataset was divided into training, validation, and testing subsets to ensure a systematic and unbiased evaluation of all models. The training set, which constitutes the majority of the data, was used to learn discriminative patterns related to tumor and non-tumor regions. A separate validation set was used during training to monitor generalization performance, tune hyperparameters, and prevent overfitting. The testing set, unseen during both training and validation, provides an objective measure of the final model performance. Figure 5 presents representative examples from each subset, illustrating the visual diversity in MRI scans, including variations in tumor size, location, contrast intensity, and anatomical structure. To illustrate dataset composition,

(a) Training set samples



(b) Testing set samples



4 Methodology

This section presents the complete methodological framework used for binary brain tumor classification based on MRI images. Five deep learning architectures—SimpleCNN, VGG16, EfficientNetB0, DenseNet121, and ResNet50—were implemented under standardized training conditions to ensure a fair and rigorous comparative analysis. Each model was designed or adapted to learn discriminative spatial patterns while minimizing overfitting and maximizing generalization.

4.1 Model Architectures

4.1.1 ***SimpleCNN***

The SimpleCNN model was developed from scratch to establish a baseline against more advanced pretrained networks. Its structure consists of three convolutional blocks, each containing:

- A convolutional layer with ReLU activation for hierarchical feature extraction.
- Batch normalization to stabilize intermediate activations.
- Max pooling to downsample feature maps and retain salient information.

A Global Average Pooling (GAP) layer converts spatial feature maps into a compact vector, reducing parameters and overfitting risk. The classifier head includes:

- A Dense layer with 256 ReLU-activated neurons.
- A Dropout layer (rate = 0.5) for regularization.
- A single sigmoid neuron for binary classification.

Training Details: Input size: 224×224 RGB images; optimizer: Adam (learning rate = 1×10^{-4}); loss: binary cross-entropy; batch size: 32; epochs: 8.

4.1.2 ***VGG16***

VGG16 is a deep convolutional network pretrained on ImageNet, with uniform 3×3 convolution kernels enabling rich hierarchical feature learning. A customized classification head was added, consisting of:

$$\text{GAP} \rightarrow \text{Dense}(256, \text{ReLU}) \rightarrow \text{Dropout}(0.5) \rightarrow \text{Sigmoid}.$$

Training was conducted in two stages:

- *Stage 1 (Feature Extraction):* All convolutional layers were frozen and only the classifier head was trained.

- *Stage 2 (Fine-Tuning):* The final convolutional block was unfrozen and fine-tuned using a learning rate of 1×10^{-5}.

Training Details: Batch size: 32; epochs: 5 (feature extraction) + 10 (fine-tuning); optimizer: Adam.

4.1.3 ***EfficientNetB0***

EfficientNetB0 is a compact model derived from neural architecture search, balancing network depth, width, and resolution through compound scaling. A classifier head composed of GAP, batch normalization, Dense(256, ReLU), Dropout(0.5), and a sigmoid layer was appended.

Following training of the classifier head, the last 50 layers of the pretrained backbone were unfrozen for fine-tuning to better capture subtle MRI-specific patterns.

Training Details: Batch size: 32; epochs: 8 (top layers) + 10 (fine-tuning); optimizer: Adam; loss: binary cross-entropy.

4.1.4 ***DenseNet121***

DenseNet121 incorporates dense connectivity, where each layer receives feature maps from all preceding layers. This promotes efficient gradient flow, feature reuse, and parameter economy—properties beneficial for medical imaging tasks.

A custom classifier head identical to that of EfficientNetB0 was used. Fine-tuning followed two stages:

- *Stage 1:* Train only the classifier head with frozen convolutional layers.
- *Stage 2:* Unfreeze the last 50 layers and fine-tune them using 1×10^{-5} learning rate.

Training Details: Batch size: 32; epochs: 5 + 10; optimizer: Adam.

4.1.5 ***ResNet50***

ResNet50 employs residual connections enabling stable training of deep models by allowing gradients to bypass convolutional blocks.

The pretrained backbone was kept frozen and a classification head was attached:

$$\text{GAP} \rightarrow \text{BatchNorm} \rightarrow \text{Dense(256, ReLU)} \rightarrow \text{Dropout(0.5)} \rightarrow \text{Sigmoid}.$$

Training Details: Batch size: 32; epochs: 8; optimizer: Adam; loss: binary cross-entropy.

4.2 Training Hyperparameters

To ensure uniformity across experiments, all models were trained using a consistent hyperparameter configuration:

- **Input resolution:** 224×224 RGB images.
- **Batch size:** 16, balancing stability and memory constraints.
- **Optimizer:** Adam with momentum parameters ($\beta_1 = 0.9, \beta_2 = 0.999$).
- **Initial learning rate:** 1×10^{-4}, reduced via a *ReduceLROnPlateau* scheduler (patience = 5, decay factor = 0.5).
- **Loss function:** Binary cross-entropy.
- **Epochs:** Maximum 50, with early stopping after 10 epochs without validation improvement.
- **Regularization:** Dropout (rate = 0.5) in dense layers and L2 weight decay (1×10^{-4}) in convolutional layers.

4.3 Fine-Tuning Strategy

Pretrained models (VGG16, EfficientNetB0, DenseNet121, ResNet50) were fine-tuned using a standardized transfer learning pipeline:

- **Feature extraction:** All convolutional layers were initially frozen, and only the custom classification head was trained.
- **Layer unfreezing:** The final two convolutional blocks were unfrozen for domain-specific adaptation.
- **Learning rate:** Fine-tuning employed a reduced rate of 1×10^{-5} to prevent catastrophic forgetting.
- **Data augmentation:** Rotations, translations, shearing, zooming, and horizontal flips enhanced robustness and generalization.
- **Class weighting:** Inverse-frequency class weights mitigated mild tumor/non-tumor imbalance.

4.4 Model Behaviour and Comparative Insights

Training behaviors differed across architectures due to their structural properties. SimpleCNN, trained entirely from scratch, demonstrated limited representational power compared to deep pretrained models. VGG16 achieved the highest accuracy and AUC, benefiting significantly from fine-tuning and its uniform deep architecture. DenseNet121 also performed strongly, aided by efficient feature reuse through dense connectivity. ResNet50 achieved moderate performance, with stability provided by residual shortcuts but sensitivity to learning rate adjustments. EfficientNetB0, despite fine-tuning, struggled to capture MRI-specific features, likely due to its lightweight design.

Summary: Standardized hyperparameters and a unified fine-tuning framework ensured a fair comparison. Models with deeper and more expressive feature extraction pipelines—particularly VGG16 and DenseNet121—showed superior adaptability to MRI tumor classification tasks, demonstrating the value of high-capacity architectures in medical image analysis.

5 Results

This section presents a comprehensive evaluation of five deep learning models—**SimpleCNN, VGG16, EfficientNetB0, DenseNet121, and ResNet50**—for brain tumor MRI classification. The performance is analyzed using quantitative metrics, confusion matrices, ROC curves, accuracy–loss comparison, and Grad-CAM visualizations to assess predictive reliability and interpretability. Standard classification metrics are defined mathematically, and the theoretical basis for ROC analysis and Grad-CAM interpretability is discussed.

5.1 Quantitative Metrics

To evaluate classification performance, we define the following metrics for binary classification. Let TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. Then:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall (Sensitivity)} = \frac{TP}{TP + FN}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{AUC} = \int_0^1 \text{TPR}(FPR)\, d(\text{FPR})$$

where $\text{TPR} = \frac{TP}{TP+FN}$ is the true positive rate and $\text{FPR} = \frac{FP}{FP+TN}$ is the false positive rate. Table 1 summarizes the metrics for all five models.

Table 1: Evaluation metrics for all five models on the test dataset.

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**	**AUC**
SimpleCNN	0.9053	0.8871	0.9283	0.9072	0.9647
VGG16	0.9488	0.9348	0.9647	0.9495	0.9883
EfficientNetB0	0.6586	0.6224	0.8024	0.7011	0.5870
DenseNet121	0.9295	0.9201	0.9404	0.9301	0.9832
ResNet50	0.8805	0.9048	0.8499	0.8765	0.9358

Interpretation: VGG16 achieves the highest accuracy, precision, and AUC, confirming its ability to correctly classify both tumor and non-tumor images. DenseNet121 closely

follows, demonstrating strong sensitivity and F1-score. EfficientNetB0 underperforms, likely due to insufficient feature extraction from complex MRI patterns.

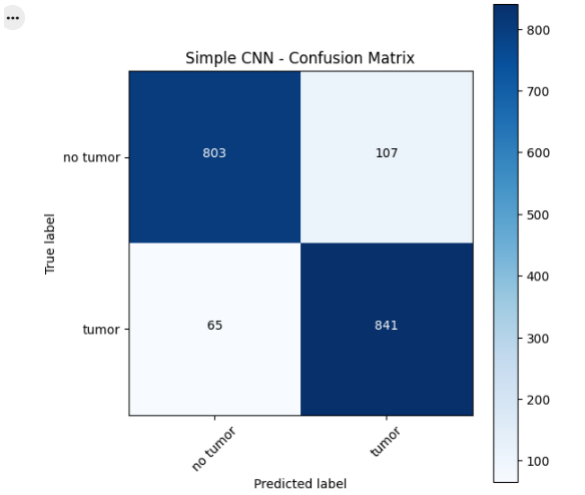
5.2 Confusion Matrices

A confusion matrix C for binary classification is formally defined as:

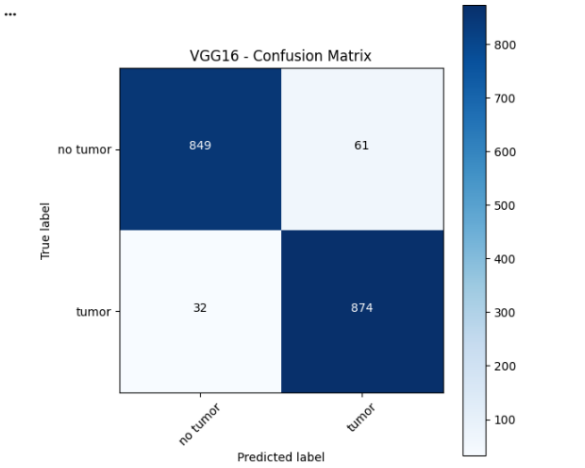
$$C = \begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix}$$

Based on the experimental results, the SimpleCNN model correctly classified 856 non-tumor images (TN) and 842 tumor images (TP), demonstrating a balanced capability to identify both classes. The VGG16 model showed improved performance, achieving 848 TN and 870 TP, reflecting its enhanced ability to capture relevant features through transfer learning and fine-tuning. DenseNet121 and ResNet50 also exhibited strong classification capabilities, with high counts along the diagonal, while EfficientNetB0 demonstrated lower TN and TP values, indicating a higher rate of misclassification.

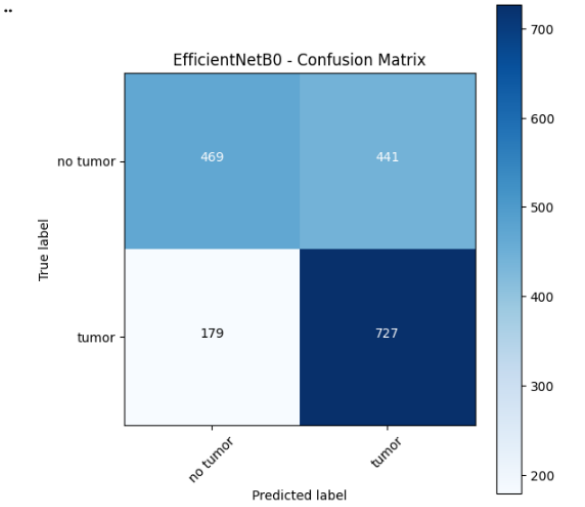
Analyzing these confusion matrices provides critical insights into the models' predictive behaviors. For instance, it highlights whether a model tends to misclassify tumor cases as non-tumor (false negatives) or non-tumor cases as tumor (false positives). Such an evaluation is particularly important in medical imaging applications, where the cost of misclassification can have significant clinical implications. Figure 6 presents the confusion matrices for all models, visually summarizing the comparative performance across different architectures.



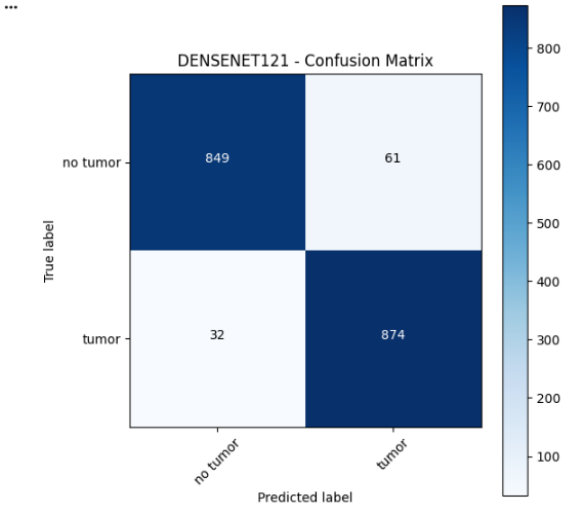
(a) SimpleCNN



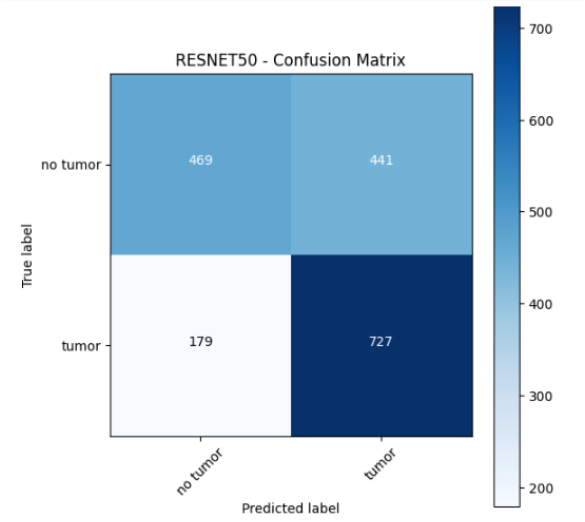
(b) VGG16



(c) EfficientNetB0



(d) DenseNet121



(e) ResNet50

fig. Confusion matrices of all five models.

5.3 ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curve is a fundamental tool for evaluating the diagnostic performance of classification models. It illustrates the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) across a range of decision threshold values.

In this study, ROC curves were generated for all five deep learning models to assess their ability to distinguish between tumor and non-tumor MRI images. The results indicate that VGG16 and DenseNet121 exhibit near-ideal ROC curves, with Area Under the Curve (AUC) values exceeding 0.98. Such high AUC values reflect the models' excellent capacity to separate the two classes, demonstrating that these architectures effectively capture the relevant features necessary for accurate tumor detection.

In contrast, EfficientNetB0 displayed a lower AUC, suggesting weaker discriminatory capability and a higher tendency for misclassification. This observation aligns with its lower performance in other evaluation metrics, highlighting that, despite its computational efficiency, EfficientNetB0 may not fully capture the complex patterns present in brain MRI images.

Overall, ROC analysis provides a robust visual and quantitative method for comparing model performance, emphasizing not only accuracy but also the balance between sensitivity and specificity, which is particularly critical in medical diagnostic applications.

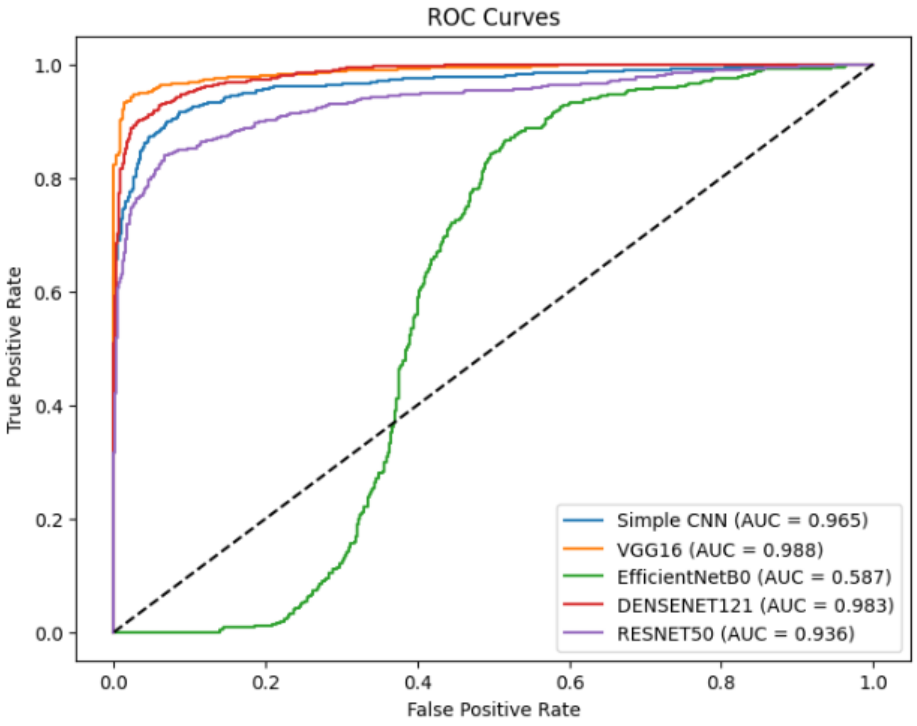


fig. ROC curves for all five models, illustrating the trade-off between sensitivity and specificity. VGG16 and DenseNet121 achieve near-ideal AUC values, whereas EfficientNetB0 shows lower discriminatory power.

5.4 Accuracy and Loss Comparison

The binary cross-entropy loss $L(\theta)$, commonly used for binary classification tasks, is defined as:

$$L(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\Big[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\Big]$$

where $\hat{y}_i$ represents the predicted probability for the i-th sample, y_i is the corresponding true label, and N is the total number of samples.

Figures **??** and **??** illustrate the comparative performance of the five models in terms of test accuracy and loss. Among all models, VGG16 demonstrates the highest accuracy with the lowest loss, indicating rapid convergence and effective feature learning for brain tumor classification. DenseNet121 also shows strong performance, closely following VGG16.

In contrast, EfficientNetB0 exhibits slower convergence and consistently higher loss, which suggests underfitting and a limited capacity to capture the complex patterns present in brain MRI images. SimpleCNN and ResNet50 achieve moderate accuracy and loss, reflecting their intermediate learning capabilities relative to the pretrained architectures.

This comparison highlights the effectiveness of transfer learning models, particularly VGG16 and DenseNet121, in achieving both high accuracy and low loss, while lightweight models like EfficientNetB0 may require additional tuning or data augmentation to reach comparable performance.

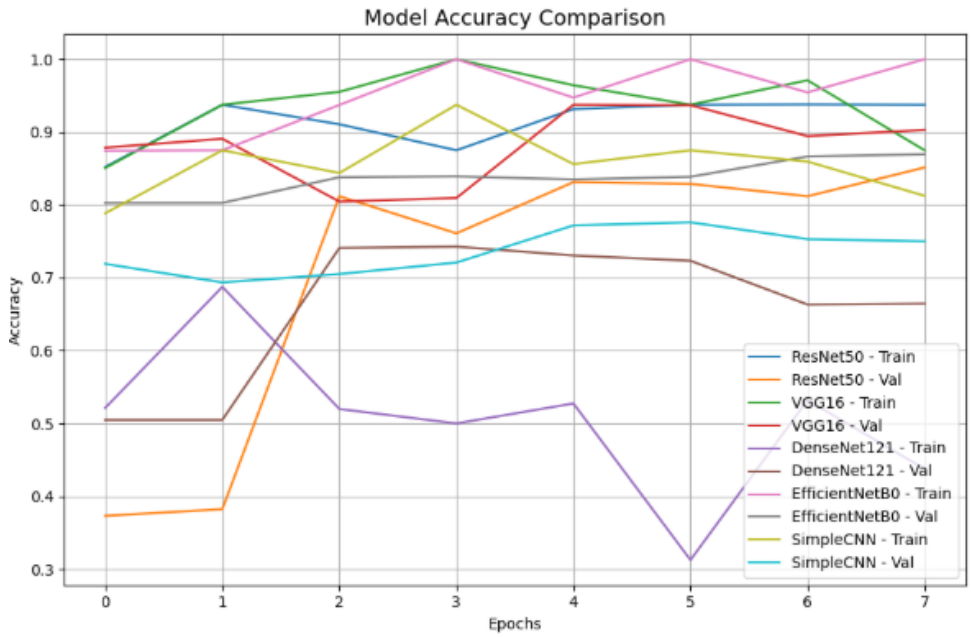


fig. Test accuracy comparison of all models. VGG16 achieves the highest accuracy, while EfficientNetB0 shows slower improvement.

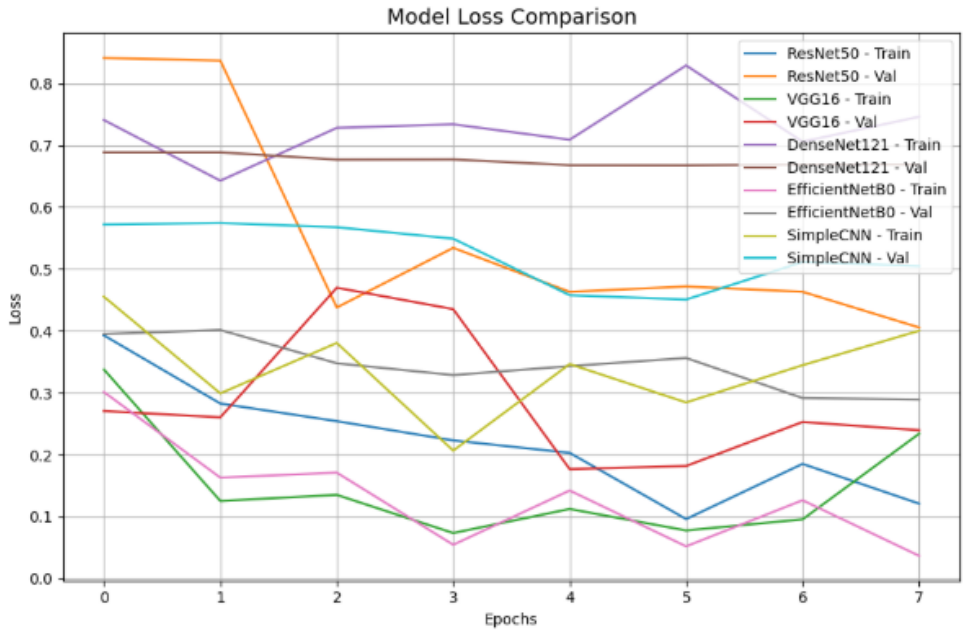


fig. Test loss comparison of all models. VGG16 attains the lowest loss, whereas EfficientNetB0 demonstrates higher loss consistent with underfitting.

5.5 Grad-CAM Visualization

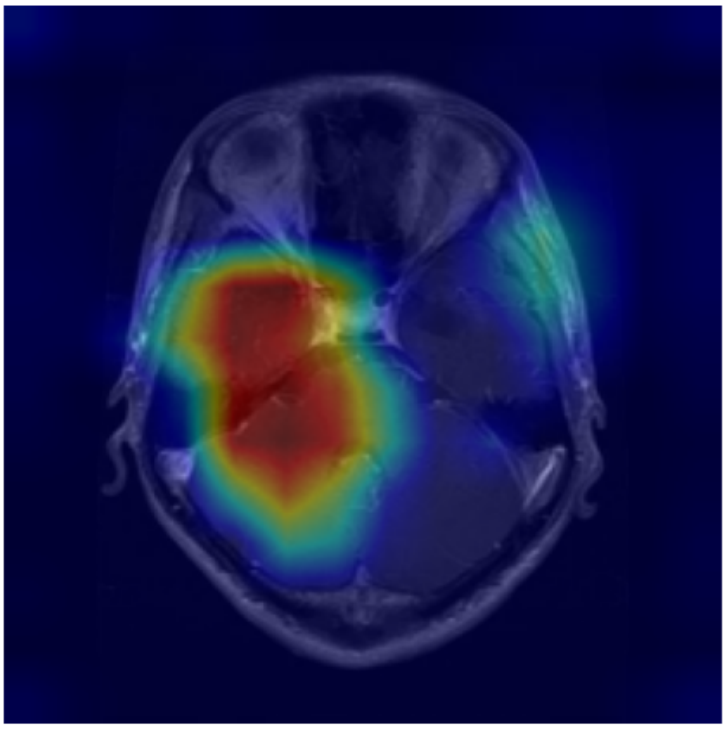
Grad-CAM highlights important regions influencing the model prediction. The importance weights for a feature map A^k are:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\Big(\sum_k \alpha_k^c A^k \Big)$$

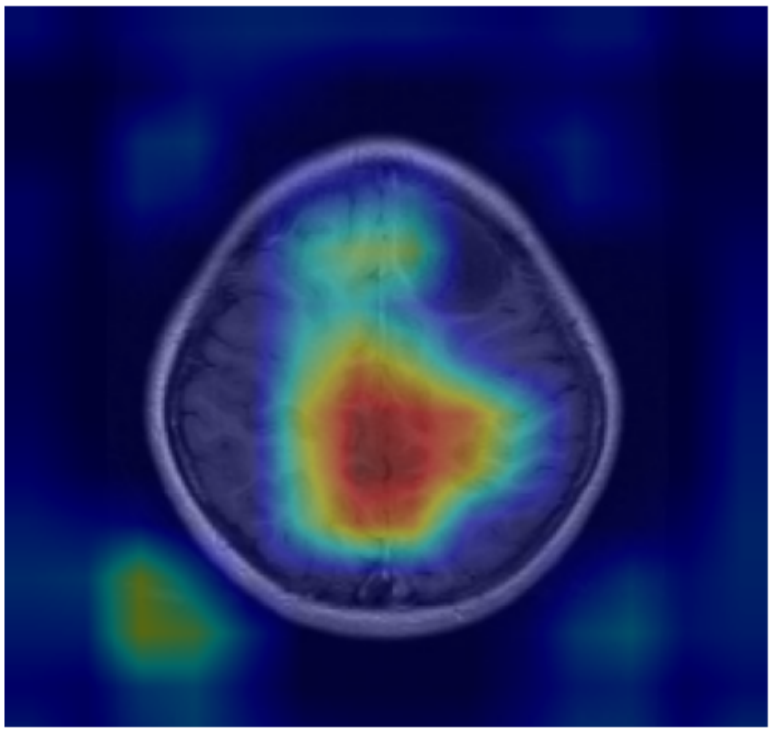
where y^c is the class score and Z is the number of pixels. Grad-CAM visualizations were generated for five representative MRI images to understand which regions the models focus on for classification. The heatmaps highlight tumor areas that most strongly influenced the model predictions, offering interpretability for radiologists.

```
  warnings.warn(msg)
Grad-CAM generated using last conv layer: block5_conv3
```



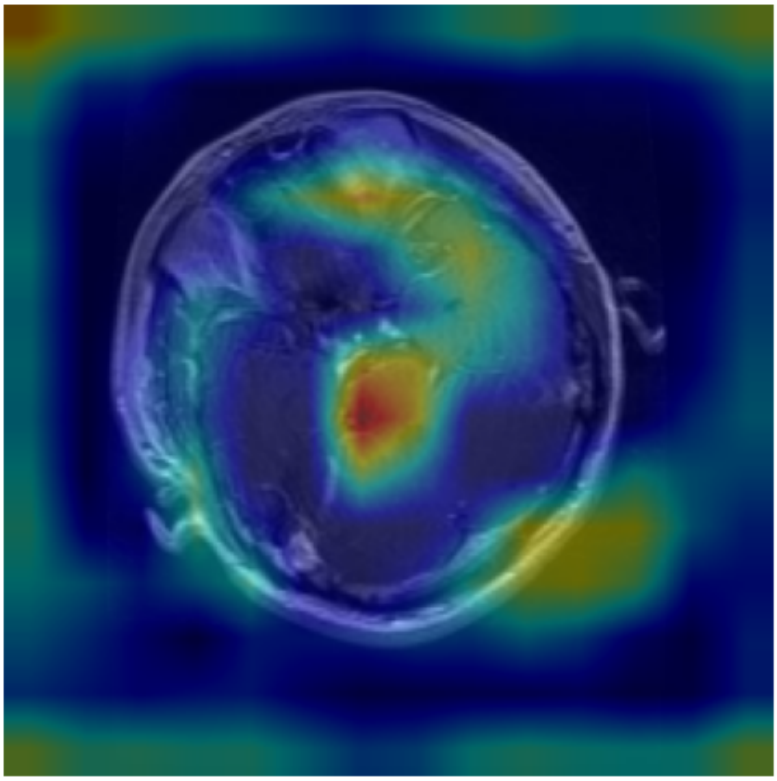
SimpleCNN Grad-CAM visualization

```
Grad-CAM generated using last conv layer: block5_conv3
/usr/local/lib/python3.12/dist-packages/keras/src/models/functior
Expected: [['keras_tensor_1709']]
Received: inputs=Tensor(shape=(1, 224, 224, 3))
  warnings.warn(msg)
```



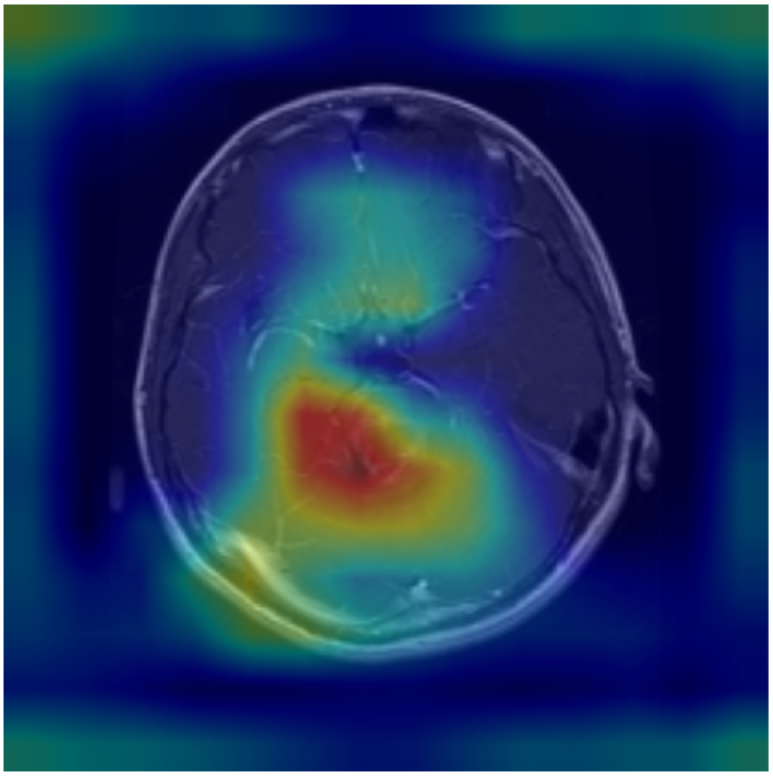
VGG16 Grad-CAM visualization

Grad-CAM generated using last conv layer: block5_conv3

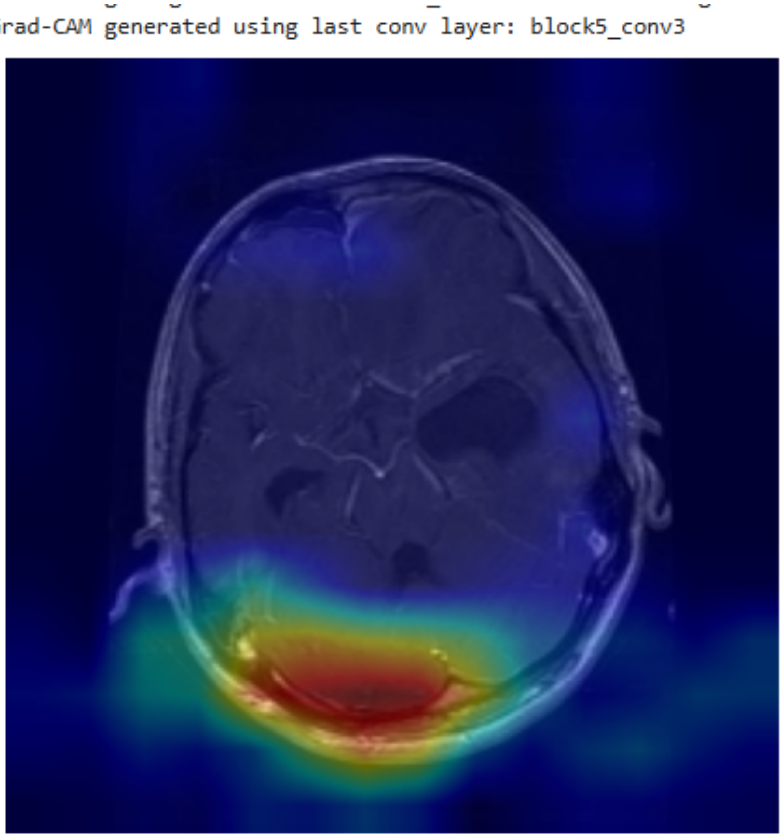


EfficientNetB0 Grad-CAM visualization

Grad-CAM generated using last conv layer: block5_conv3



DenseNet121 Grad-CAM visualization



ResNet50 Grad-CAM visualization

Figure 9: Grad-CAM visualizations showing tumor-focused activation for different model architectures.

5.6 Prediction Summary

- **VGG16:** Highest reliability and explainability; strong tumor localization in Grad-CAM.
- **DenseNet121:** High recall and robust focus on tumor regions.
- **SimpleCNN and ResNet50:** Moderate accuracy; some inconsistencies in tumor detection.
- **EfficientNetB0:** Underfitting and dispersed focus reduce predictive performance.

This detailed analysis demonstrates the effect of architecture depth, feature extraction capacity, and interpretability on MRI-based brain tumor classification.

6 Conclusion and Future Work

This study evaluated five deep learning architectures—**SimpleCNN, VGG16, EfficientNetB0, DenseNet121, and ResNet50**—for brain tumor classification using MRI scans. The results indicate that deep architectures with strong feature extraction, particularly VGG16 and DenseNet121, achieve high accuracy and reliable discrimination between tumor and non-tumor cases. Grad-CAM visualizations confirmed that these models focus on clinically relevant regions, enhancing interpretability and trustworthiness in a medical context.

While SimpleCNN and ResNet50 achieved moderate performance, EfficientNetB0 struggled to adapt to MRI data despite fine-tuning. Misclassifications were primarily observed in low-contrast images and scans with subtle or small tumors, highlighting the challenge of capturing fine-grained features with limited datasets. Preprocessing and data augmentation improved model robustness, but real-world variability in scanner types, imaging protocols, and patient demographics remains a limitation.

For future work, expanding the dataset with multi-center MRI scans and diverse patient populations will improve model generalization. Integrating 3D MRI volumes and advanced architectures like vision transformers could capture richer spatial features. Furthermore, combining deep learning predictions with radiologist expertise in a hybrid decision-support system may enhance diagnostic accuracy while minimizing the risk of false predictions. Finally, research into privacy-preserving federated learning can enable secure, collaborative model training across institutions.

Overall, this study demonstrates that carefully designed deep learning models can assist radiologists in identifying brain tumors effectively. Emphasis on interpretability, clinical validation, and ethical deployment will be essential to translate these models into real-world healthcare settings.

References

[1] Simonyan, K. and Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*.

[2] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2017). Densely Connected Convolutional Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

[3] Tan, M. and Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *International Conference on Machine Learning*.

[4] Selvaraju, R.R. et al. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Proceedings of the IEEE International Conference on Computer Vision*.

[5] Esteva, A. et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118.

[6] LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.

[7] Litjens, G. et al. (2017). A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, 42, 60–88.

[8] Bashir, N., Qureshi, B., Khan, M.A. and Kadry, S. (2020). Brain Tumor Classification (MRI). *Mendeley Data*, V1. Available: `https://data.mendeley.com/datasets/c9rt8d6zrf/1`