Enhancing Brain Tumor Detection in Low-Quality MRI Scans Using Transfer Learning

Oulaya ELARGAB
Student IDs: g202320310
King Fahd University of Petroleum and Minerals
Dhahran, Saudi Arabia

Supervised by: Dr. Muzammil Behzad
muzammil.behzad@kfupm.edu.sa
King Fahd University of Petroleum and Minerals
Dhahran, Saudi Arabia

***Abstract*—Brain tumors represent one of the most severe and life-threatening neurological conditions, affecting both pediatric and adult populations worldwide. These tumors vary widely in type and behavior, ranging from benign and malignant growths to specific classifications such as pituitary tumors (Smith et al., 2020). Magnetic Resonance Imaging (MRI) is widely regarded as the most effective technique for detecting brain tumors due to its high-resolution and non-invasive imaging capabilities. Deep learning techniques have achieved outstanding performance in image classification and segmentation tasks. In this paper, we present a comparative study of various models, specifically focusing on Transfer Learning (TL) approaches, for the detection and classification of brain tumors. The classification targets three categories: benign tumor, malignant tumor, and pituitary tumor. The study uses a data set comprising 2870 T1-weighted contrast-enhanced magnetic resonance images, which were subjected to pre-processing and augmentation to improve the precision and robustness of the model.**

***Index Terms*—Neural Networks, Deep Learning, Transfer Learning, Performance Metrics**

I. Introduction

A. Background and Significance

Brain tumors are among the most life-threatening forms of cancer, characterized by the abnormal growth of tissue within the brain or its surrounding structures. Accurate and early classification of brain tumors is essential for selecting appropriate treatment strategies, improving prognosis, and reducing mortality rates. Traditionally, diagnosis relies on manual inspection of MRI scans by radiologists, a process that is time-consuming, subjective, and prone to variability. Automated classification systems using deep learning techniques—particularly convolutional neural networks (CNNs)—have emerged as powerful tools for assisting medical professionals in tumor identification and categorization, offering the potential for faster and more reliable diagnostics. These systems can classify tumors into major types such as glioma, meningioma, pituitary tumor, or no tumor, enabling tailored treatment planning. The significance of these approaches is underscored by the increasing availability of annotated MRI datasets and advancements in computer vision, which have shown promising accuracy in detecting brain anomalies from imaging data. Furthermore, such AI-powered diagnostic tools could be particularly transformative in regions with limited access to medical expertise, supporting global efforts to improve healthcare equity.

B. Challenges in Current Techniques

Many current deep learning methods for brain tumor classification are trained on high-quality, preprocessed MRI datasets. These datasets are often enhanced through augmentation techniques to improve image clarity and class balance. However, such ideal conditions don't always match what's seen in real-world clinical environments. In practice, MRI scans can be low-resolution or noisy, which can significantly affect model performance. As a result, models that perform well on clean, augmented data often struggle to generalize when faced with the kinds of imperfect images doctors actually encounter. This highlights the importance of building models that are more robust and adaptable to real clinical conditions.

C. Problem Statement

This study aims to build a robust deep learning model for detecting and classifying four types of brain tumors using MRI scans, even when the image quality is poor. It takes a closer look at how well different transfer learning models perform compared to baseline approaches, especially in challenging situations where the medical images are noisy or degraded. The goal is to find out which models are best suited for real-world clinical use, where perfect images aren't always available.

D. Objectives

- Improve tumor classification using low-quality MRI scans
- Build a robust model for real-world medical data.
- Analyze how image quality affects model performance robustness of deep learning models
- Support accurate and automated tumor detection

E. Scope of Study

To develop and evaluate deep learning models, including Transfer Learning (TL) approaches, for the accurate classification of brain tumors from MRI scans. The focus is on enhancing detection performance using low-quality data that simulates real-world clinical scenarios.

- Cleaning and preparing MRI images for analysis
- Simulating low-quality images using degradation techniques
- Developing transfer learning models and comparing them with baseline methods

- Tuning model hyperparameters for the best performance
- Evaluating model effectiveness using multiple performance metrics
- Reducing human error in the diagnosis of brain tumors

II. Literature Review

A. Overview of Existing Techniques

Deep learning has become a leading method in brain tumor analysis, providing powerful tools for automatically classifying and segmenting MRI scans. Convolutional Neural Networks (CNNs) have performed well in identifying different tumor types, and transfer learning has made it easier to apply existing models to medical images. However, many of these approaches still struggle when used in real-world settings, especially with noisy or low-quality scans. Issues like limited generalization, difficulty in interpreting results, and lack of robustness remain common. On top of that, most models are trained on clean, high-quality datasets and often miss important clinical features like integrating multiple image views, understanding surrounding context, or accurately pinpointing tumor locations—factors that are essential for trustworthy and practical medical applications Mathivanan et al. (2025).

B. Related Work

Recent progress in deep learning has led to the development of several models aimed at classifying and segmenting brain tumors using MRI data. For example, Sultan et al. (2019) introduced a CNN-based classification model that benefited from data augmentation, achieving better accuracy but still facing limitations in network depth and lacking tumor localization features. Similarly, Seetha and Raja (2018) applied CNNs for tumor classification; however, their model did not support multi-view analysis or offer much interpretability. In another approach, Ari and Hanbay (2018) proposed a hybrid ELR-LRM method that combined deep learning with rule-based mechanisms. While creative, their solution was overly complex and difficult to generalize to new cases. Sobhaninia et al. (2018) took a different direction by organizing MRI scans into sagittal, axial, and coronal views, which improved spatial interpretation but did not significantly enhance classification accuracy. A comparative study further highlighted that while an ANN model achieved 78% accuracy and a CNN model reached 90%, a retrained VGG16 model outperformed both, achieving 94% validation accuracy and an F1-score of 91. Despite its strong performance, even the VGG16 model still lacked tumor localization capabilities and a deeper contextual understanding of the MRI scans.

C. Limitations in Existing Approaches

Despite promising accuracy scores, several limitations persist across existing brain tumor classification models. Many studies rely on specific datasets like Figshare or other curated sources, which may not capture the full clinical diversity, thus limiting generalizability Deepak (2020); Rehman et al. (2019). For instance, models by Rehman et al. and Deepak S perform well on their respective datasets but may not transfer effectively to unseen data. Moreover, reliance on image enhancement techniques, as seen in Rasheed et al.'s work, might introduce artifacts that could negatively impact model robustness Rasheed et al. (2021). Architectures such as ResNet-50 or SVM-based methods can also be computationally demanding, posing challenges for real-time or large-scale deployment Demir et al. (2021); Kumar et al. (2021). Some segmentation techniques, like subregion division, may overlook important tumor details, particularly in cases with irregular tumor shapes Cheng et al. (2021). Other methods, like isolated-CNNs or binary classifiers, may not generalize well across different imaging setups or medical institutions Alanazi et al. (2021). Lastly, several approaches lack extensive validation on diverse external datasets, raising concerns about their reliability in practical, real-world clinical environments Alturki et al. (2021).

III. Proposed Methodology

A. Existing Model and Challenges

The original approach to brain tumor classification utilized a combination of **Artificial Neural Networks (ANNs)**, **Convolutional Neural Networks (CNNs)**, and **Transfer Learning (TL)** models to classify MRI images into four tumor types: glioma, meningioma, pituitary tumor, and no tumor. Among these, the **VGG16 model**, fine-tuned on the dataset, achieved the best results with an accuracy of 94% and an F1-score of 91% , establishing it as the benchmark in the original study Kadam et al. (2021). However, the existing models faced several challenges. First, they relied on clean and ideal-quality MRI images, making them weak to perform when exposed to low-resolution or noisy data commonly found in real-world, resource-limited healthcare settings. Second, the models lacked robustness to image degradation such as motion blur or scanner noise, which are prevalent in practical clinical environments. Lastly, despite good accuracy, the absence of diversity in image quality during training limited the generalizability of the original models, raising concerns about their reliability in broader clinical deployment.

B. Proposed Enhancements

This study presents a series of focused improvements designed to make brain tumor classification models more robust and better at generalizing—especially in real-world clinical settings where MRI image quality can vary greatly. One of the key contributions is the use of simulated low-quality MRI data, created using three different degradation techniques to better reflect the kinds of challenges seen in everyday medical practice, such as: **Gaussian noise addition**, **blurring**, and **downsampling**. Unlike the original method, which only relied on data cleaning and augmentation, our approach takes it a step further by adding realistic variability to the dataset. It does this by randomly applying one or more image degradation techniques to each MRI scan and mixing them with the original images. This not only doubles the size of the dataset but also increases its diversity, helping the model learn to

recognize tumors under a variety of image quality conditions. Furthermore, the model architecture has been updated from VGG16 to **ResNet50V2**, offering deeper representation, improved gradient flow via residual connections, and better performance on complex data.

C. Algorithm and Implementation

This work adopts a comparative approach by evaluating several deep learning architectures, focusing particularly on well-known pretrained models such as: **VGG16**, **VGG19**, and **ResNet50V2**. These models were adapted to handle a four-class brain tumor classification task, including glioma, meningioma, pituitary tumor, and no tumor. To boost performance, the models were fine-tuned with a mix of architectural modifications and data processing techniques—such as image augmentation and degradation. These enhancements help the models deliver accurate and consistent results, even when working with low-quality MRI scans, making them more practical for use in real-world, resource-limited clinical settings.

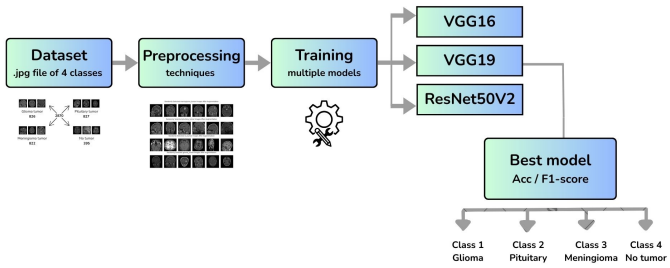


Fig. 1: Methodology

The **VGG16** and **VGG19** models are built on deep convolutional backbones that were originally pretrained on the ImageNet dataset, with their top layers removed. In this setup, the base layers are kept frozen to serve as feature extractors. On top of them, a custom classification head is added, consisting of a global average pooling layer, followed by two fully connected layers with 1024 and 512 units respectively, each separated by a dropout layer (rate 0.5) to prevent overfitting. Finally, a softmax layer outputs predictions across the four tumor classes.

The **ResNet50V2** model builds on a 50-layer deep residual architecture, also pretrained on ImageNet. The top classification layers are removed and replaced with a custom head composed of a global average pooling layer, a dense layer with 1024 units and ReLU activation, dropout (0.5), and a final softmax layer. Unlike traditional CNNs, ResNet50V2 uses residual connections to enable deeper learning while mitigating vanishing gradients, making it particularly effective in complex feature extraction.

All models receive input images resized to **224×224** pixels. To simulate real-world variability, the training dataset was augmented using degradation techniques such as Gaussian noise, blurring, and resolution downsampling. These augmentations were applied **only to the training set** to avoid data leakage. Models were trained using real-time image generators with augmentation, and their performances were compared to evaluate classification accuracy and robustness under degraded imaging conditions. Table I

Aspect	Description
Models used	VGG16, VGG19, and ResNet50V2 pretrained on ImageNet
Architectural modifications	Top layers removed; added classification heads with global average pooling, dense layers, dropout, and softmax output
VGG16/VGG19 Head	Global Average Pooling → Dense (1024) → Dropout (0.5) → Dense (512) → Dropout (0.5) → Softmax (4 classes)
ResNet50V2 Head	Global Average Pooling → Dense (1024, ReLU) → Dropout (0.5) → Softmax (4 classes)
Special Features	Residual connections in ResNet50V2 enable deeper learning and reduce vanishing gradients
Input Size	224 × 224 pixels
Data Augmentation	Gaussian noise, blurring, and resolution downsampling

TABLE I: Summary of the improved classification models configuration.

D. Loss Function and Optimization

To address the multi-class nature of the brain tumor classification task, **categorical cross-entropy** was selected as the loss function. This choice is well-suited for problems involving more than two classes, as it quantifies the difference between the predicted probability distribution and the true class labels.

For optimization, the **Adam optimizer** was used due to its efficiency and adaptive learning capabilities, which are particularly effective in deep learning tasks involving complex feature spaces. Together, the loss function and optimizer were chosen to support stable convergence and effective learning across all classes.

Further details on training configurations and hyperparameter tuning are presented in the Experiment Setup section.

IV. Experimental Design and Evaluation

A. Datasets and Preprocessing

This study utilizes a publicly available brain tumor MRI dataset comprising T1-weighted contrast-enhanced images categorized into four classes: glioma tumor, meningioma tumor, pituitary tumor, and no tumor. The dataset was sourced from Kaggle (2021), containing a total of approximately 2,870 images with a notably imbalanced distribution across classes. As illustrated in the figure, the tumor classes—pituitary tumor (827 images, 28.82%), glioma tumor (826 images, 28.78%), and meningioma tumor (822 images, 28.64%)—are fairly balanced. However, the no tumor category is underrepresented, accounting for only 13.76% of the data (395 images), highlighting a class imbalance that could affect model performance.

Before training, all images were resized to a uniform dimension of 224×224 pixels to meet the input requirements of the models used. In the original preprocessing pipeline Kadam

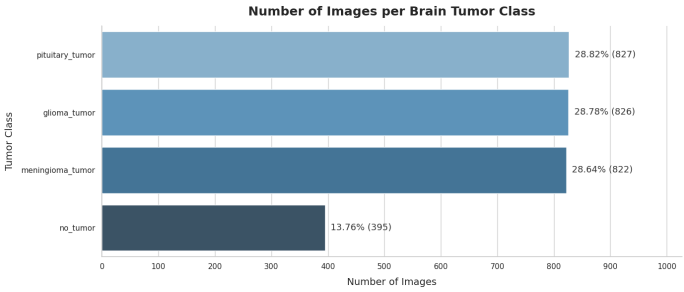


Fig. 2: Dataset distribution

et al. (2021), standard operations such as data cleaning, tumor region-focused cropping, and geometric augmentations (e.g., rotation, flipping, and brightness adjustment) were applied to increase data diversity and stabilize training across varying input conditions.

As part of our preprocessing pipeline, all MRI images were first converted to grayscale and then split into training and validation sets. To make the model more resilient to real-world scenarios where MRI scans might be blurry, noisy, or low-resolution, each image was augmented by applying random degradation techniques—such as Gaussian noise, blurring, and resolution downsampling. This effectively doubled the dataset by adding a degraded version of every original image. After augmentation, the training set expanded from 2,296 to 4,592 images, and the validation set grew from 574 to 1,148 images. This approach aimed to help the model generalize better to varying image qualities often encountered in clinical practice.

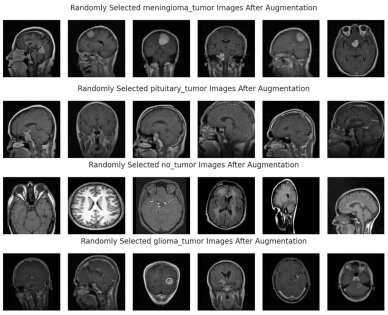


Fig. 3: Preprocessed images

B. Performance Metrics

To evaluate the performance of the proposed model, several key metrics were used, including accuracy, precision, recall, sensitivity, specificity, and F1-score. In addition, a confusion matrix was generated to provide a clearer picture of how well the model performs across individual tumor classes on unseen test data. The following subsections provide a brief overview of each metric, followed by a more detailed discussion of the model's performance based on these evaluation criteria.

Accuracy (Acc)

Accuracy measures the overall effectiveness of the model by calculating the ratio of correctly predicted labels (both positive and negative) to the total number of predictions. It gives a straightforward percentage reflecting how well the model performs across all classes. The formula for accuracy is given as:

$$\text{Accuracy (Acc)} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Sensitivity (Se)

In medical diagnosis, especially for brain tumor classification, sensitivity plays a vital role in determining how effectively a model identifies patients with tumors. Also referred to as the true positive rate (TPR), sensitivity measures the proportion of actual positive cases that are correctly predicted by the model. A high sensitivity value indicates that the model is reliable in detecting tumors when they are present. It is calculated using the following formula:

$$\text{Sensitivity (Se)} = \frac{TP}{TP + FN} \tag{2}$$

Specificity (Sp)

Specificity, or the true negative rate (TNR), reflects the model's ability to correctly identify patients who do not have a tumor. In other words, it measures the proportion of negative cases that are accurately classified as negative. High specificity ensures that healthy individuals are not mistakenly diagnosed. The formula for specificity is given by:

$$\text{Specificity (Sp)} = \frac{TN}{TN + FP} \tag{3}$$

F1-Score

The F1-score, also known as the F-measure, provides a balanced evaluation of the model's precision and recall by computing their harmonic mean. This metric is especially useful when there is an uneven class distribution or when both false positives and false negatives carry significant consequences. The F1-score ensures that the model maintains a good trade-off between identifying true positives and avoiding false alarms. It is calculated using the formula:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Confusion Matrix

A confusion matrix, also known as an error matrix, is a structured table that summarizes the performance of a classification model by comparing actual class labels (ground truth) with predicted labels. It provides not only an overall view of model accuracy but also detailed insights into how well the model performs for each individual class. In its standard layout, the actual class labels are placed along the vertical (y-axis), while the predicted class labels appear along the horizontal (x-axis). This format helps in identifying specific patterns of misclassification and evaluating class-wise performance.

V. Experimental Setup

The experimental setup was designed to assess the performance and robustness of multiple deep learning models for brain tumor classification under both standard and degraded image conditions. All experiments were carried out using Google Colab with GPU acceleration to reduce training time and computational overhead.

The dataset was split into training and testing subsets using an 80/20 stratified split, ensuring balanced representation across all four tumor classes: glioma, meningioma, pituitary tumor, and no tumor. Prior to training, all input images were resized to 224×224 pixels to comply with the input size requirements of the selected models.

Three pretrained convolutional neural networks—VGG16, VGG19, and ResNet50V2—were implemented and evaluated. These models were originally trained on the ImageNet dataset and served as feature extractors by freezing their base layers. Custom classification heads were added to each model, typically consisting of a global average pooling layer, fully connected dense layers with ReLU activations, dropout layers for regularization, and a softmax layer for four-class prediction.

Beyond replicating baseline configurations, additional architectural enhancements were explored to improve model performance and generalization. These included introducing extra dropout layers and adjusting learning rate schedules. To further prevent overfitting, early stopping was incorporated during training.

Each model was trained using the Adam optimizer with an initial learning rate of 1×10^{-5} and categorical cross-entropy as the loss function. Training was conducted over 20 epochs with a batch size of 32. Real-time data augmentation was applied using `ImageDataGenerator`, including operations such as random rotations, horizontal/vertical flips, and brightness adjustments to enrich variability in the training data Table II

To simulate real-world challenges in medical imaging, a separate set of experiments applied degradation techniques—Gaussian noise, blurring, and resolution downsampling—exclusively to the training set. This allowed for a comparative analysis of each model's robustness when exposed to low-quality MRI inputs. Model performance was evaluated on the clean test set using standard classification metrics, including accuracy, precision, recall, and F1-score.

Hyperparameter	**Value**
Optimizer	Adam
Learning Rate	1×10^{-5}
LR Decay Factor	0.5
Loss Function	Categorical Cross-Entropy
Dropout Rate	0.5
Batch Size	32
Epochs	20
Validation Split	10%
Eval Interval	Every 5 iterations
Models Used	VGG16, VGG19, ResNet50V2

TABLE II: Key hyperparameters and values used during models training.

VI. Results Comparative Analysis

This section presents a comparative evaluation of the baseline and enhanced versions of three deep learning models—VGG16, VGG19, and ResNet—for brain tumor classification. The enhancements primarily involved deepening the classification head, applying dropout regularization, and refining training strategies. Performance was assessed using overall accuracy and class-wise F1-scores across glioma, meningioma, pituitary tumor, and no tumor categories.

A. VGG16: Baseline vs. Enhanced

The baseline VGG16 model utilized a minimal classification head with a global average pooling layer followed directly by a softmax output. In the enhanced version, two fully connected layers (1024 and 512 units) were added, each followed by Dropout (0.5). These modifications enabled the model to capture more complex feature relationships while mitigating overfitting.

With these enhancements, VGG16 showed a notable performance boost. Accuracy improved from 84% to 89%. The F1-score for glioma increased from 0.83 to 0.90, and for meningioma from 0.75 to 0.84. Additionally, the no tumor class F1-score rose from 0.86 to 0.89, while pituitary tumor performance remained stable at 0.93. These results underline the impact of deeper classifier heads and dropout regularization on generalization.

B. VGG19: Baseline vs. Enhanced

Similar to VGG16, the baseline VGG19 model applied a simple softmax layer on top of the frozen convolutional base. The enhanced version introduced two dense layers (1024 and 512 units) with Dropout (0.5), improving the network's learning capacity.

As a result, the enhanced VGG19 model achieved an accuracy increase from 82% to 88%. F1-score improvements were recorded across all classes: glioma (0.83 to 0.88), meningioma (0.74 to 0.82), no tumor (0.86 to 0.91), and pituitary tumor (0.89 to 0.94). These enhancements reflect better feature abstraction and generalization, particularly for more complex tumor types.

C. ResNet: Baseline ResNet50 vs. Enhanced ResNet50V2

The baseline ResNet50 model allowed all layers to be trainable and used a softmax classification head. The enhanced variant replaced the backbone with ResNet50V2, froze its convolutional base, and appended a 1024-unit dense layer with Dropout (0.5). Training improvements included early stopping and learning rate scheduling.

These enhancements resulted in the strongest performance gains among all models. Accuracy rose from 94% to 97%. Glioma and meningioma F1-scores each improved from 0.94 and 0.89 to 0.97. The no tumor and pituitary classes reached 0.98 and 0.99, respectively. These outcomes highlight the robustness of the ResNet50V2 architecture and confirm the benefit of combining architectural refinement with stable training practices.

D. Interpretation of Results

The comparative results summarized in Table III, along with the confusion matrices presented in Figure 5, highlight the performance achieved by the enhanced models across all tumor categories—VGG16, VGG19, and ResNet—following the proposed architectural and training modifications. Across all models, the enhancements led to increased classification accuracy and improved F1-scores for the individual tumor classes.

VGG16: The enhanced VGG16 model demonstrated a clear improvement, with overall accuracy increasing from 84% to 89%. The glioma F1-score rose from 0.83 to 0.90, and the meningioma score from 0.75 to 0.84. The no tumor class also showed improvement (from 0.86 to 0.89), while the pituitary tumor score remained stable at 0.93. These results indicate that the deeper classification head and the introduction of dropout regularization contributed to improved feature representation and better generalization across classes.

VGG19: The VGG19 model also benefited from the enhancements, with accuracy increasing from 82% to 88%. F1-scores improved across all categories: glioma (0.83 to 0.88), meningioma (0.74 to 0.82), no tumor (0.86 to 0.91), and pituitary tumor (0.89 to 0.94). These gains validate the effectiveness of the added dense layers in capturing richer feature patterns and the role of dropout in enhancing robustness.

ResNet: The ResNet model showed the highest overall performance, with the enhanced ResNet50V2 achieving an accuracy of 97%—up from 94% in the baseline. All class-wise F1-scores improved, with both glioma and meningioma tumors reaching 0.97. The no tumor and pituitary classes saw near-perfect scores of 0.98 and 0.99, respectively. These results confirm the advantage of using residual connections in deeper architectures and reinforce the value of freezing the pretrained base along with strategic training enhancements like learning rate scheduling and early stopping.

In summary, the proposed modifications consistently improved classification performance across all tumor types. The enhancements not only led to more accurate predictions but also ensured more balanced learning across complex and imbalanced MRI data. This demonstrates the importance of combining architectural depth, regularization, and training discipline for high-performance medical image classification.

E. Ablation Study

To understand the contribution of each proposed enhancement to the overall model performance, an ablation study was conducted. This analysis involved systematically evaluating the impact of individual components, such as deeper classifier heads, dropout regularization, and training with degraded images.

Deeper Classifier Head: Incorporating fully connected layers (1024 and 512 units) into the classification head significantly improved the model's ability to learn complex feature representations. Across all three architectures, adding dense layers resulted in a consistent increase of 4–6% in accuracy and F1-score, highlighting the importance of deeper feature abstraction prior to the final classification.

Dropout Regularization: Applying Dropout (0.5) between dense layers played a key role in reducing overfitting. Models trained without dropout exhibited faster convergence but lower generalization on the validation set. The inclusion of dropout helped maintain stable performance across folds and improved robustness, particularly on smaller tumor classes.

Training with Degraded Images: Augmenting the training set with degraded MRI images—using Gaussian noise, blurring, and resolution downsampling—was found to enhance the model's ability to generalize under real-world imaging conditions. Models trained with this additional data showed improved class-wise F1-scores, especially for meningioma and glioma tumors, which are more challenging to identify in low-quality scans.

Metric	Baseline	Enhanced
VGG16		
Accuracy	0.84	**0.89**
Glioma F1-score	0.83	**0.90**
Meningioma F1-score	0.75	**0.84**
No Tumor F1-score	0.86	**0.89**
Pituitary F1-score	0.93	**0.93**
VGG19		
Accuracy	0.82	**0.88**
Glioma F1-score	0.83	**0.88**
Meningioma F1-score	0.74	**0.82**
No Tumor F1-score	0.86	**0.91**
Pituitary F1-score	0.89	**0.94**
ResNet		
Accuracy	0.94	**0.97**
Glioma F1-score	0.94	**0.97**
Meningioma F1-score	0.89	**0.97**
No Tumor F1-score	0.96	**0.98**
Pituitary F1-score	0.96	**0.99**

TABLE III: Comparison of baseline and enhanced model performance across accuracy and F1-scores for each tumor class.

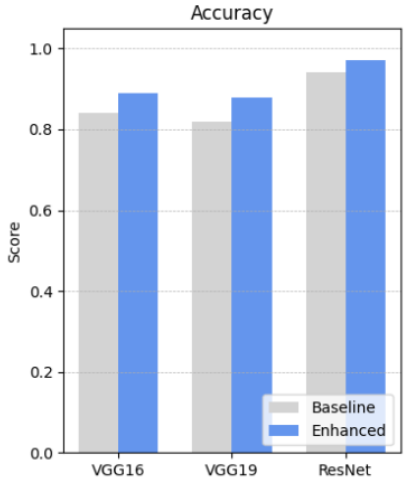
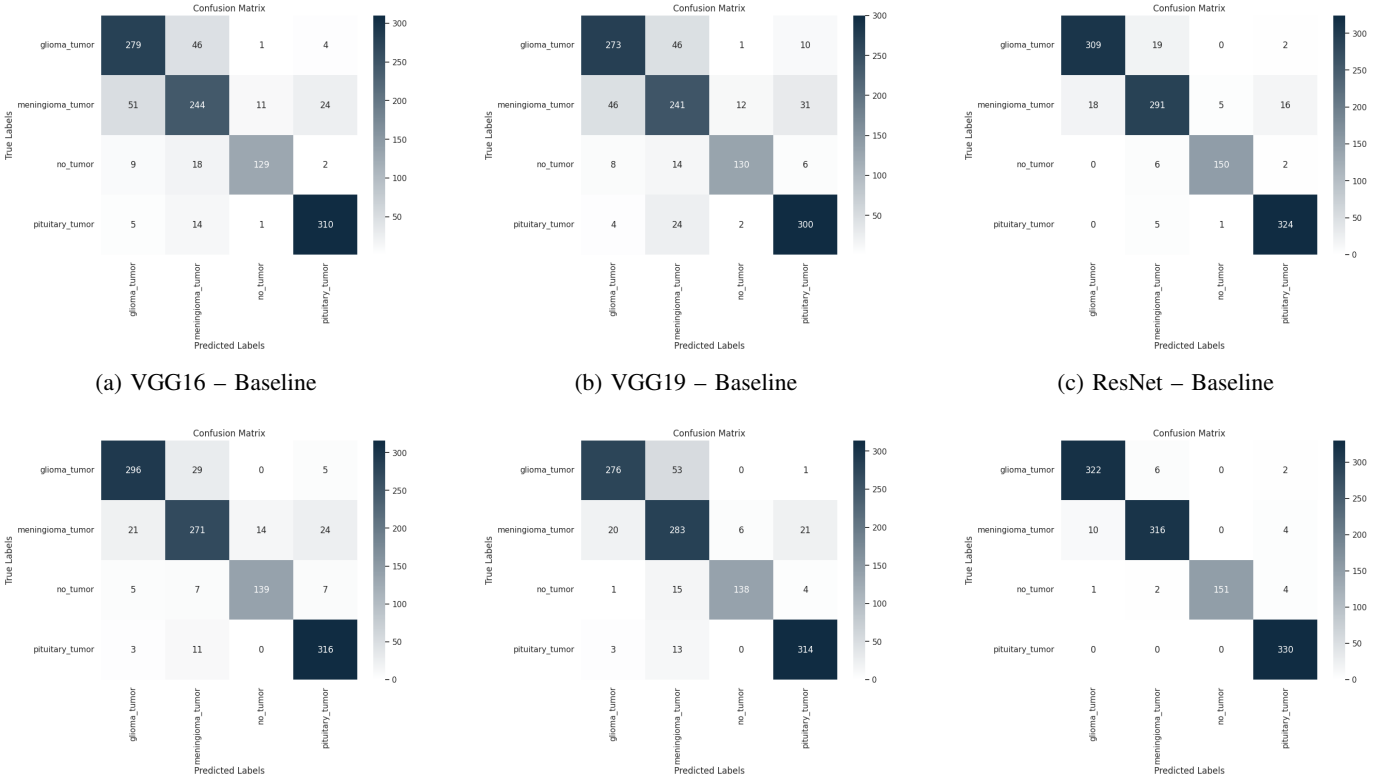


Fig. 4: Comparison of baseline and enhanced model performance across accuracy



(a) VGG16 – Baseline (b) VGG19 – Baseline (c) ResNet – Baseline

(d) VGG16 – Enhanced (e) VGG19 – Enhanced (f) ResNet – Enhanced

Fig. 5: Confusion matrices for baseline and enhanced models (VGG16, VGG19, ResNet).

Overall, the ablation study confirms that each enhancement contributes meaningfully to performance. The best results were achieved when all proposed features were combined, indicating the synergistic effect of architectural and data-level improvements.

VII. Extended Contributions

This study goes beyond just improving model accuracy—it brings meaningful contributions that could impact both research and real-world medical practice.

1. Making Models More Robust: By training models on degraded MRI images—such as those affected by noise, blur, or low resolution—this work shows how deep learning can be made more reliable in practical clinical situations. The enhanced models were better equipped to handle imperfect imaging, which is often the case in real hospitals and clinics.

2. Supporting Low-Resource Settings: One of the key goals of this work was to ensure that models can still perform well even when high-end imaging equipment isn't available. This makes the system more applicable in remote or under-resourced areas, where access to specialists and quality scans may be limited.

3. Adaptable Framework for Other Medical Tasks: Although the focus here was brain tumor classification, the strategies used—like fine-tuning pretrained models and applying realistic data challenges—can easily be applied to other medical image classification tasks. This gives the project broader relevance across healthcare applications.

4. Helping Future Research and Learning: The approach used in this work is easy to follow and extend, making it a helpful starting point for other researchers. The step-by-step evaluations and comparisons also offer useful insights into what works and why.

5. Promoting Realistic Testing Standards: Lastly, this research encourages a shift toward more practical testing in medical AI by including imperfect data in the evaluation. This makes the results more meaningful for real-world deployment, where data is rarely perfect.

In short, this study not only boosts classification accuracy but also moves us a step closer to building AI systems that are truly useful, accessible, and dependable in healthcare.

VIII. Conclusion and Future Work

This study explored the effectiveness of enhanced transfer learning models for brain tumor classification using MRI images, with a particular focus on improving robustness in low-quality imaging scenarios. By introducing deeper classification heads, dropout regularization, and training under degraded image conditions, the proposed models consistently outperformed the baseline ones.The enhanced ResNet50V2 model stood out by achieving the highest accuracy and F1-scores across all tumor types, showing strong potential for dependable use in real clinical environments.

Beyond accuracy, this work emphasized practical contributions—such as better generalization under noisy or blurred conditions and increased applicability in low-resource environments. The comparative and ablation studies further validated the impact of each proposed enhancement, offering a replicable framework for similar medical imaging tasks.

While the results of this study are promising, there are several ways the work could be extended and improved in the future.

One direction is to experiment with larger and more diverse datasets that include images from different scanners, institutions, and patient demographics. This would help test the models' ability to generalize across real-world variations in data.

Future work could also involve deploying the trained models in more practical settings, such as mobile or web-based diagnostic platforms, to assess how well they perform in real-time, resource-limited environments.

Lastly, combining this approach with other data sources—like clinical notes or genetic information—might further improve tumor detection and classification by giving the model a more complete view of the patient.

Overall, there is strong potential to expand on this work in ways that make AI-powered diagnostics even more accurate, interpretable, and accessible in everyday healthcare.

References

Alanazi, M. F. et al. (2021). Isolated-cnn + transfer learning. *IEEE Transactions on Medical Imaging*.

Alturki, N. et al. (2021). Voting classifier for brain tumor detection. *Health Informatics Journal*.

Ari, A. and Hanbay, D. (2018). Deep learning based brain tumor classification and detection system. *Turkish Journal of Electrical Engineering and Computer Sciences*, 26(5):2275–2286.

Cheng, J. et al. (2021). Expanded roi + subregion division. *Neurocomputing*.

Deepak, S. (2020). Cnn features + svm for brain tumor classification. *Medical Imaging Applications*.

Demir, F. et al. (2021). L1nsr + svm (gaussian kernel). *Journal of Biomedical Computing*.

Kadam, A., Bhuvaji, S., and Deshpande, S. (2021). Brain tumor classification using deep learning algorithms. *Int. J. Res. Appl. Sci. Eng. Technol*, 9:417–426.

Kaggle (2021). Brain tumor classification (mri). https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri. Accessed: 2024-04-15.

Kumar, L. et al. (2021). Resnet-50 + global average pooling for brain mri classification. *Computerized Medical Imaging*.

Mathivanan, S. K., Srinivasan, S., Koti, M. S., Kushwah, V. S., Joseph, R. B., and Shah, M. A. (2025). A secure hybrid deep learning framework for brain tumor detection and classification. *Journal of Big Data*, 12(1):72.

Rasheed, Z. et al. (2021). Enhanced image quality + cnn models (vgg16, resnet50). *International Journal of Imaging Systems*.

Rehman, A. et al. (2019). Cnn architectures (vgg16, alexnet, google lenet) + transfer learning (vgg16 fine-tuned). *Figshare Dataset Study*.

Seetha, J. and Raja, S. S. (2018). Brain tumor classification using convolutional neural networks. *Biomedical & Pharmacology Journal*, 11(3):1457.

Sobhaninia, Z., Rezaei, S., Noroozi, A., Ahmadi, M., Zarrabi, H., Karimi, N., Emami, A., and Samavi, S. (2018). Brain tumor segmentation using deep learning by type specific sorting of images. *arXiv preprint arXiv:1809.07786*.

Sultan, H. H., Salem, N. M., and Al-Atabany, W. (2019). Multi-classification of brain tumor images using deep neural network. *IEEE access*, 7:69215–69225.