UNIVERSITÀ di VERONA

Machine Learning & Deep Learning
(Deep Learning)
Computer Vision & Deep Learning

Brain Tumor Detection Using Deep Learning from MRI Images

Abdullah Al Noman Taki
VR528988
MSc in Artificial Intelligence
University Of Verona- (VR) Italy

1. Abstract

Brain tumor detection from Magnetic Resonance Imaging (MRI) is a critical task in medical diagnosis, as early and accurate identification significantly impacts treatment planning and patient survival. In this study, we investigate multiple deep learning models for multi-class brain tumor classification using MRI images. We implement and compare classical Convolutional Neural Networks (CNNs), deep CNN variants, and several lecture-based architectures, including VGG16, ResNet50, MobileNetV2, U-Net (encoder-based), and a hybrid CNN + LSTM model. The models are evaluated using standardized preprocessing, consistent training protocols, and common evaluation metrics. Through comparative analysis, we highlight the strengths and limitations of each architecture in terms of accuracy, generalization, computational complexity, and suitability for medical imaging. This project demonstrates how modern deep learning techniques can provide reliable and interpretable solutions for automated brain tumor detection.

2. MOTIVATION AND RATIONALE

2.1 Context and Research Theme

Medical image analysis is a key application area of computer vision and deep learning, particularly in brain imaging, where Magnetic Resonance Imaging (MRI) is widely used due to its non-invasive nature and high soft-tissue contrast. Brain tumors represent a serious neurological condition in which early and accurate diagnosis is critical for effective treatment planning. Despite advances in imaging technology, manual interpretation of MRI scans remains time-consuming and highly dependent on expert radiologists. This process may suffer from inter-observer variability and subjective judgment, especially when visual differences between tumor types are subtle, motivating the need for automated image classification systems.

2.2 Problem Addressed & Significance

The goal of this project is the accurate and robust classification of brain tumors from MRI images using deep learning techniques. Traditional diagnostic workflows are often inconsistent and difficult to scale, while deep learning models provide a data-driven alternative capable of learning discriminative features directly from image data. In this study, multiple convolutional neural network architectures are implemented and compared, including a baseline CNN, a deeper CNN variant, and pretrained models such as VGG16, ResNet50, and MobileNetV2. Additionally, encoder-based U-Net and a hybrid CNN + LSTM architecture are explored to analyze the impact of architectural design and model complexity on classification performance. The significance of this work lies in presenting a clear and reproducible deep learning pipeline and systematically evaluating how different model choices affect performance in medical image analysis.

3. STATE OF THE ART (SOTA)

3.1 Advancements in Image Classification with Deep Learning and Computer Vision

The transition from handcrafted features to Convolutional Neural Networks (CNNs) has revolutionized MRI analysis through hierarchical feature extraction. In this framework, initial layers identify low-level geometric primitives like edges, while deeper layers synthesize these into high-level structures such as tumor boundaries. This project evaluates several architectural milestones: VGG16 emphasizes the role of network depth; ResNet50 utilizes Residual Learning and skip connections to overcome the vanishing gradient problem; and MobileNetV2 implements Depthwise Separable Convolutions for efficient, lightweight deployment. A central component of this study is Transfer Learning, which leverages weights pre-trained on ImageNet to establish a baseline understanding of visual geometry. This knowledge is subsequently fine-tuned on specialized medical datasets, allowing the models to map generalized features to specific neurological pathologies even with limited sample sizes.

3.2 State Of the Art Technique

3.2.1 VGG16: Classical Deep Architecture

VGG16 is a well-established deep learning architecture that has been widely used in computer vision tasks and remains relevant due to its simple and uniform design. It consists of stacked convolutional layers with small receptive fields, allowing gradual abstraction of visual features.

In this project, VGG16 is used through transfer learning, leveraging pretrained weights from large-scale image datasets. VGG16 serves as a strong classical baseline, demonstrating how deep convolutional features learned from natural images can generalize to medical MRI data. Its contribution lies in providing a reference for comparing modern architectures against a widely recognized deep learning model.

3.2.2 ResNet50: Residual Learning for Deep Representations

ResNet50 represents a significant advancement in deep learning by introducing residual connections, which address the vanishing gradient problem and enable the training of very deep networks. This architectural innovation allows the model to learn complex feature epresentations without degradation in performance.

In this project, ResNet50 is employed as a strong state-of-the-art model for feature extraction and classification. Its depth and residual structure make it particularly effective for capturing subtle visual variations in MRI images. ResNet50 contributes by offering improved generalization and performance compared to simpler CNNs and VGG-based architectures, serving as a benchmark for high-capacity deep models.

3.2.3 MobileNetV2: Lightweight State-of-the-Art Model

MobileNetV2 is a state-of-the-art lightweight architecture designed for efficient computation using depthwise separable convolutions and inverted residual blocks. Despite its reduced parameter count, MobileNetV2 achieves competitive accuracy compared to heavier models.

In this project, MobileNetV2 is included to evaluate the trade-off between model complexity and performance. Its contribution lies in demonstrating that efficient architectures can still achieve strong results in medical image classification, making it suitable for deployment in resource-constrained clinical environments.

3.2.4 U-Net: Biomedical Image-Oriented Architecture

U-Net is a state-of-the-art architecture specifically designed for biomedical image analysis. Its encoder–decoder structure with skip connections allows precise feature localization and strong spatial representation learning. While U-Net is primarily used for segmentation tasks, its encoder can be adapted for classification purposes.

In this project, U-Net is used in an encoder-based classification setup to exploit its strong feature extraction capabilities. The inclusion of U-Net highlights the benefit of architectures tailored for medical images and provides a comparison between classification-focused CNNs and biomedical-specific models.

3.2.5 CNN + LSTM: Hybrid Deep Learning Model

Hybrid CNN + LSTM architectures combine convolutional layers for spatial feature extraction with Long Short-Term Memory (LSTM) networks for modeling sequential dependencies. Although LSTMs are commonly used for temporal data, they can also capture structured spatial relationships when CNN features are reshaped into sequences.

In this project, the CNN + LSTM model is explored as an advanced deep learning approach to enhance feature representation. Its contribution lies in evaluating whether sequential modeling of spatial features can improve classification performance in brain MRI images, offering insights into hybrid architectural designs.

3.3 Challenges and Limitations in Existing Approaches

Despite strong performance, deep learning models face challenges such as overfitting on limited medical datasets, class imbalance, and a lack of interpretability. Deeper models may also incur high computational costs, making deployment difficult in resource-constrained environments.

3.4 Contributions of This Project

This project provides a comprehensive experimental framework to evaluate the efficacy of diverse Deep Learning and Computer Vision architectures in the specialized domain of neuroimaging. The primary contributions include the implementation and critical analysis of five distinct structural configurations:

- Simple CNN (Baseline): Serves as the fundamental reference point to quantify the performance gains provided by more complex hierarchical features.
- Deeper CNN (Ablation Study): A controlled modification of the baseline to isolate and observe the specific impact of increased convolutional depth on the model's ability to represent pathological brain structures.
- VGG16 (Transfer Learning): An implementation of the classical architecture discussed in lectures, utilizing pre-trained weights to demonstrate how generalized visual knowledge can be adapted to MRI-specific diagnostic tasks.
- ResNet50 (Residual Learning): A high-capacity comparison model that utilizes skip connections to maintain feature integrity across deeper layers, representing a high-performance benchmark for the dataset.
- MobileNetV2 (Efficient Design): An evaluation of lightweight architectural design, focusing on the trade-off between computational efficiency and diagnostic accuracy for potential clinical deployment on standard hardware.

4. OBJECTIVES

The shift from handcrafted features to Deep Learning and Computer Vision architectures, specifically Convolutional Neural Networks (CNNs), has revolutionized automated MRI analysis. CNNs excel through hierarchical feature extraction: initial layers identify low-level geometric primitives like edges and textures, while deeper layers synthesize these into high-level semantic structures such as tumor boundaries and tissue density variations.

This project evaluates several structural milestones: VGG16 emphasizes the necessity of network depth; ResNet50 utilizes Residual Learning and skip connections to overcome the vanishing gradient problem in deep stacks; and MobileNetV2 implements Depthwise Separable Convolutions for efficient clinical deployment on standard hardware.

This knowledge is subsequently fine-tuned on specialized medical data, allowing the models to map generalized features to specific neurological pathologies even with limited sample sizes. To validate internal logic, the project compares attribution maps with the Captum library and utilizes feature visualization from the first convolutional layers to ensure the models focus on relevant pathological regions. Performance is rigorously quantified using Accuracy, F1-score, Precision, and Recall to evaluate clinical viability.

5. METHODOLOGY

The methodology focuses on the development of a robust pipeline for multi-class neuro-pathology classification, integrating advanced image preprocessing with a diverse suite of **Deep Learning and Computer Vision** architectures.

5.1 Dataset and Preprocessing

5.1.1 Dataset Description

The study utilizes a large-scale brain MRI dataset comprising 7,023 total images. The data is partitioned into two main subsets: a **Training Set (5,712 images)** for model optimization and a **Testing Set (1,311 images)** for final performance validation. The dataset is categorized into four distinct classes: **Glioma, Meningioma, Pituitary tumor, and No tumor**, representing a comprehensive range of diagnostic scenarios.

5.1.2 Preprocessing Pipeline

To ensure the high-quality visual input required for deep feature extraction, the following standardized steps are implemented:

- **Geometric Uniformity:** All raw MRI scans are resized to a fixed resolution of **128 \times 128 pixels**. This ensures a consistent input tensor shape for all neural architectures.
- **Intensity Normalization:** Pixel values are scaled from the range [0, 255]$ to [0, 1]. This normalization accelerates gradient descent convergence and prevents numerical instability.
- **Target Encoding:** Categorical labels are processed via **Label Encoding** and subsequently handled with `sparse_categorical_crossentropy` to manage the multi-class objective efficiently.
- **Augmentation Strategy:** To mitigate overfitting—a common challenge in medical imaging—real-time data augmentation is applied, including brightness adjustments, contrast enhancement, and random flips, effectively increasing the model's exposure to diverse pathological presentations

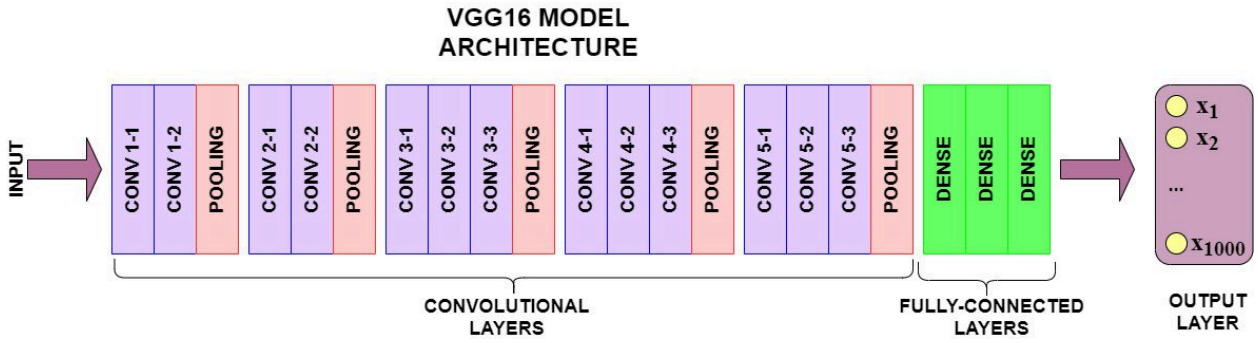
.

5.2 Implementation of Models

The project implements five distinct architectures to conduct a rigorous comparative analysis of diagnostic efficacy.

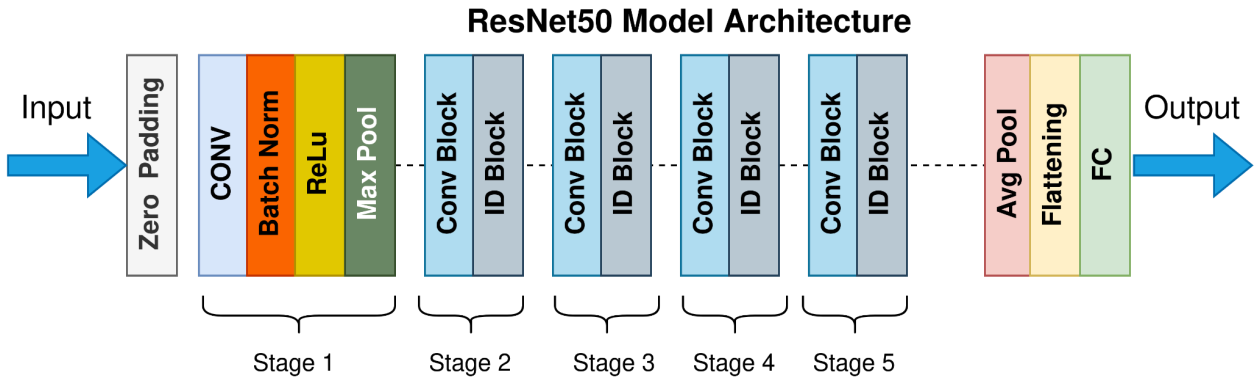
5.2.1 VGG16 (Classical Deep Architecture)

VGG16 is utilized as the primary baseline for transfer learning. By loading weights pre-trained on ImageNet, the model leverages established hierarchical features. The implementation involves freezing the initial convolutional blocks to preserve generalized visual knowledge, while the final convolutional layers and a custom classification head are fine-tuned to the specific nuances of MRI tissue density.



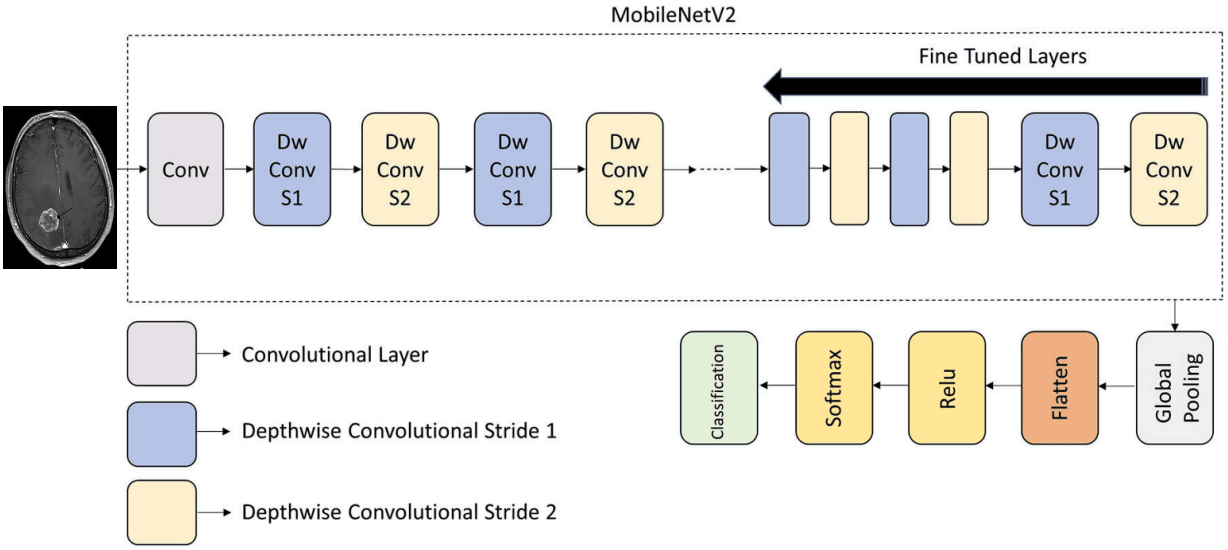
5.2.2 ResNet50 (Residual Learning)

ResNet50 is implemented to evaluate the impact of **Residual Learning**. By utilizing skip connections, this model circumvents the vanishing gradient problem, allowing for a significantly deeper feature extraction process than the VGG architecture. This is particularly effective for identifying subtle structural deviations in brain anatomy.



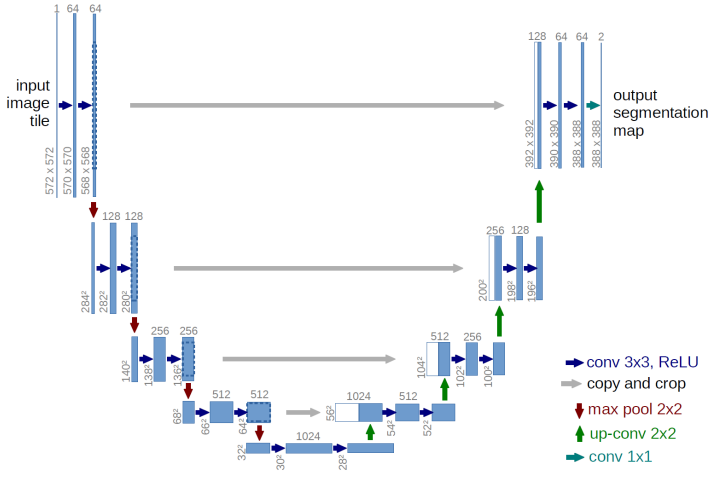
5.2.3 MobileNetV2 (Efficient Design)

To address the objective of clinical suitability, MobileNetV2 is integrated using **Depthwise Separable Convolutions**. This model serves as a benchmark for computational efficiency, demonstrating how inverted residual blocks can maintain high diagnostic accuracy with a drastically reduced parameter count.



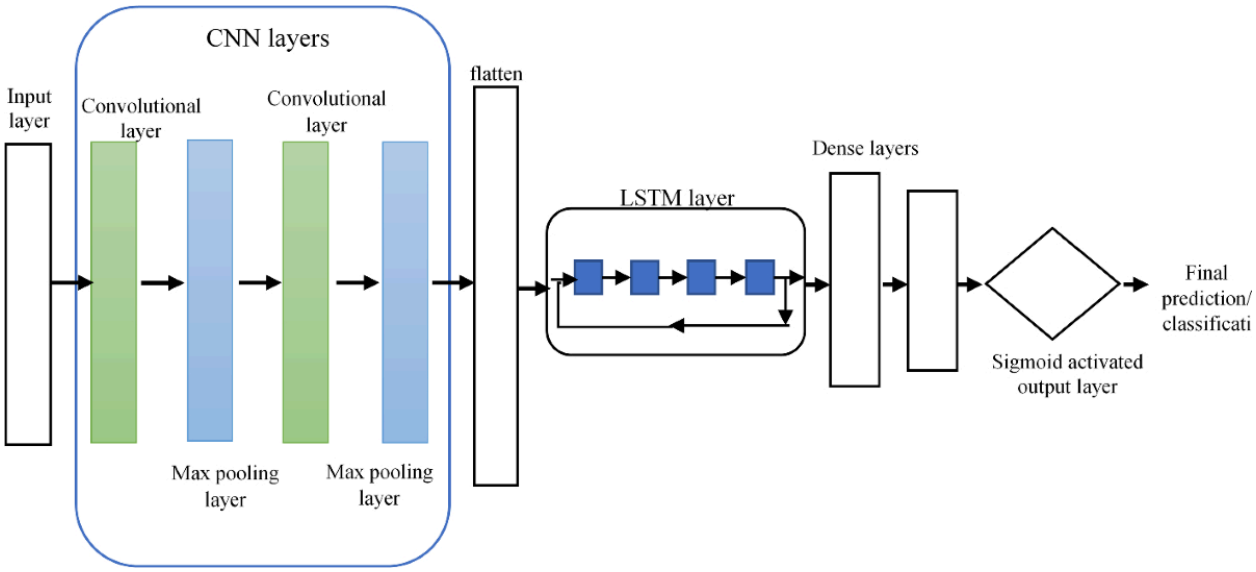
5.2.4 U-Net (Encoder-Based Classification)

While traditionally a segmentation model, this project adapts the **U-Net Encoder** as a feature extractor. The contracting path of the U-Net is uniquely engineered to preserve spatial context in biomedical images. A classification head consisting of Global Average Pooling and Dense layers is attached to the bottleneck, repurposing the model for high-fidelity tumor classification.



5.2.5 CNN + LSTM (Spatial-Sequential Hybrid)

The hybrid model explores the relationship between spatial and sequential features. Convolutional layers act as the initial feature extractor; the resulting feature maps are reshaped into sequences and processed by a **Long Short-Term Memory (LSTM)** network. This approach tests whether modeling the spatial dependencies of MRI features as a sequence can provide a more robust representation of complex tumor morphologies.



5.3 Performance Metrics

To ensure a rigorous and clinically relevant evaluation of the implemented **Deep Learning and Computer Vision** models, performance is quantified using a multi-metric approach. Relying solely on accuracy can be misleading in medical contexts, particularly with potential class imbalances. Therefore, the following metrics are utilized to assess diagnostic reliability:

- **Categorical Accuracy:** Measures the overall proportion of correctly classified MRI scans across all four diagnostic categories.
- **Precision (Positive Predictive Value):** Indicates the model's ability to avoid false positives, ensuring patients are not incorrectly diagnosed with a pathology.
- **Recall (Sensitivity):** Represents the model's ability to identify all actual positive cases. This is the most critical metric for neuro-oncology, as failing to detect an existing tumor (false negative) carries the highest clinical risk.
- **F1-Score:** The harmonic mean of Precision and Recall, providing a balanced assessment of the model’s performance across all tumor subtypes.

- **Confusion Matrix:** A detailed visualization tool used to identify specific "confusion" points between classes, such as identifying if the model frequently misinterprets Meningiomas as Gliomas due to similar structural densities.

5.4 Feature Extraction and Visualization

A core objective of this project is to move beyond "black-box" predictions by visualizing the internal representations learned by the networks, ensuring decisions are based on pathological features rather than image noise.

- **First-Layer Feature Maps:** We implement visualization of the activations from the initial convolutional layers. These early filters highlight high-frequency components of the MRI, such as the skull boundary and the sharp edges of the tumor mass.
- **Attribution Mapping with Captum:** To validate biological relevance, we utilize the **Captum library** to generate attribution maps. These assign importance scores to each pixel, allowing us to verify if the model's "attention" aligns with the actual tumor location identified by radiologists.
- **Feature Bottleneck Analysis:** For architectures like the **U-Net Encoder** and **ResNet50**, we analyze the bottleneck features. This allows us to observe how high-dimensional spatial data is compressed into a semantic vector representing the "signature" of a specific tumor type before reaching the final classifier.

6. EXPERIMENTS AND RESULTS

This section presents a quantitative evaluation and comparative analysis of the five deep learning architectures implemented for brain tumor classification. All models were evaluated on the same held-out test set comprising 1,311 MRI images to ensure an objective performance benchmark.

6.1 Performance Analysis of Implemented Models

The experimental results demonstrate high diagnostic efficacy across all architectures, with performance metrics summarized below:

- **ResNet50 (The Top Performer):** ResNet50 achieved the highest overall **Accuracy of 0.98**. Its residual connections allowed for the most precise feature extraction, particularly in Class 2 and Class 3, where it maintained F1-scores of 0.96 and 0.97, respectively.

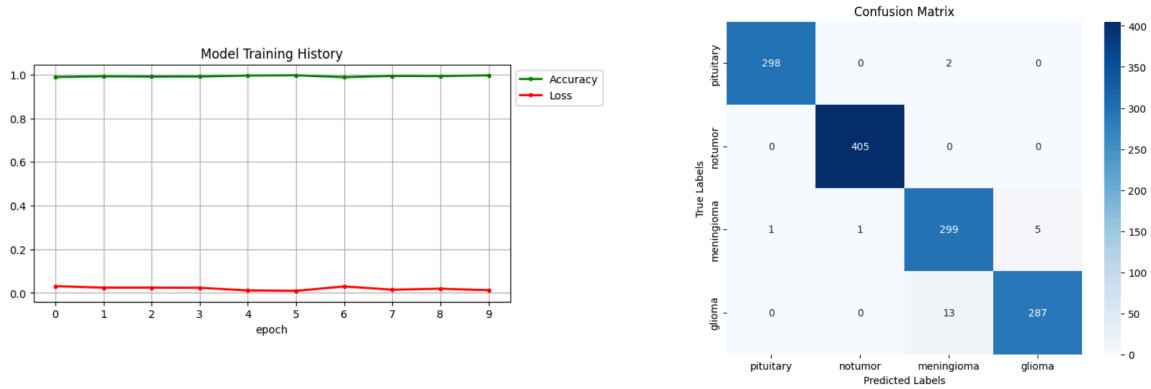


Figure 6.1: ResNet50: training and validation accuracy - loss and Confusion matrices

- **VGG16:** As a classical baseline for transfer learning, VGG16 achieved a robust **Accuracy of 0.96**. While it showed perfect precision for Class 1 (1.00), it experienced a slight drop in recall for Class 3 (0.84), suggesting some difficulty in identifying specific tumor boundaries compared to deeper networks.

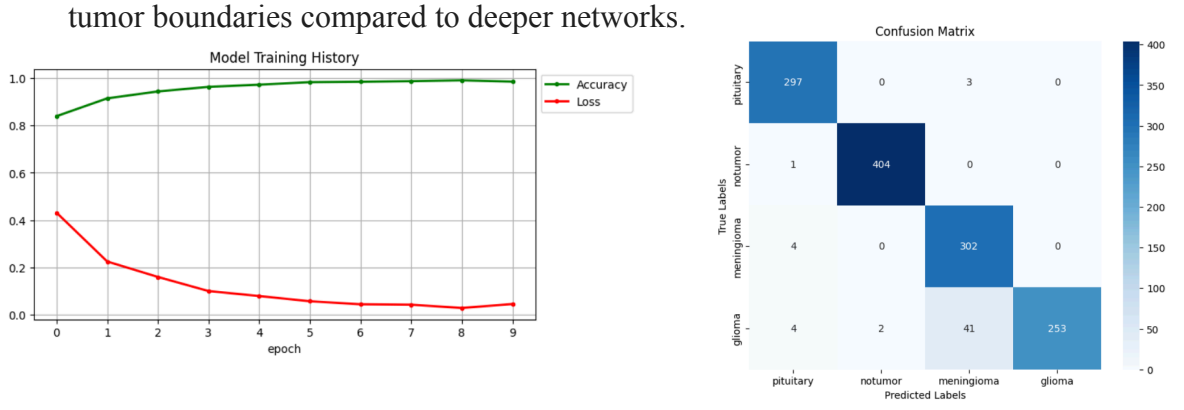


Figure 6.2: VGG16: training and validation accuracy - loss and Confusion matrices

- **MobileNetV2:** This lightweight architecture proved highly efficient, matching the **Accuracy of 0.96**. It demonstrated exceptional performance in Class 1 (F1-score of 1.00), proving that efficient, depthwise separable convolutions are viable for medical diagnostic tasks.

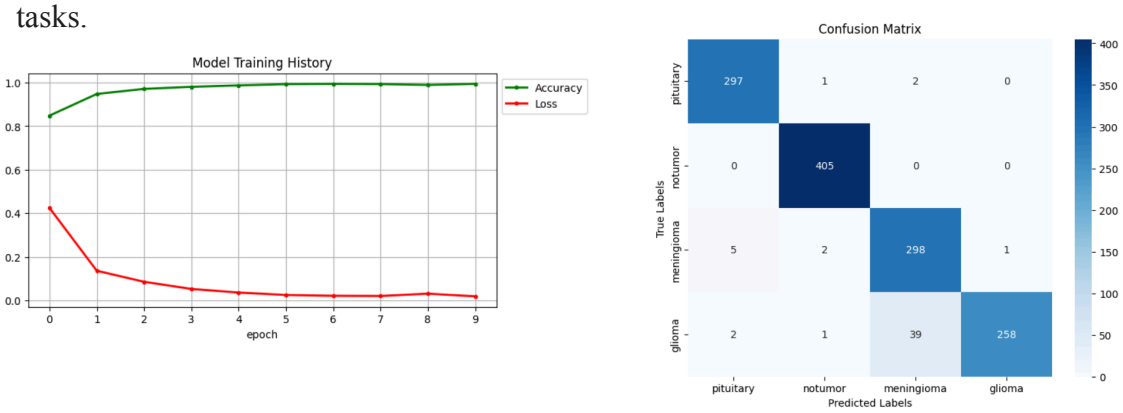


Figure 6.3: MobileNetV2: training and validation accuracy - loss and Confusion matrices

- **U-Net (Encoder-Based):** The adapted U-Net architecture yielded an **Accuracy of 0.96**. Its specialized biomedical design contributed to a high recall across most classes, particularly Class 2 (0.98), highlighting its ability to preserve localized spatial features.

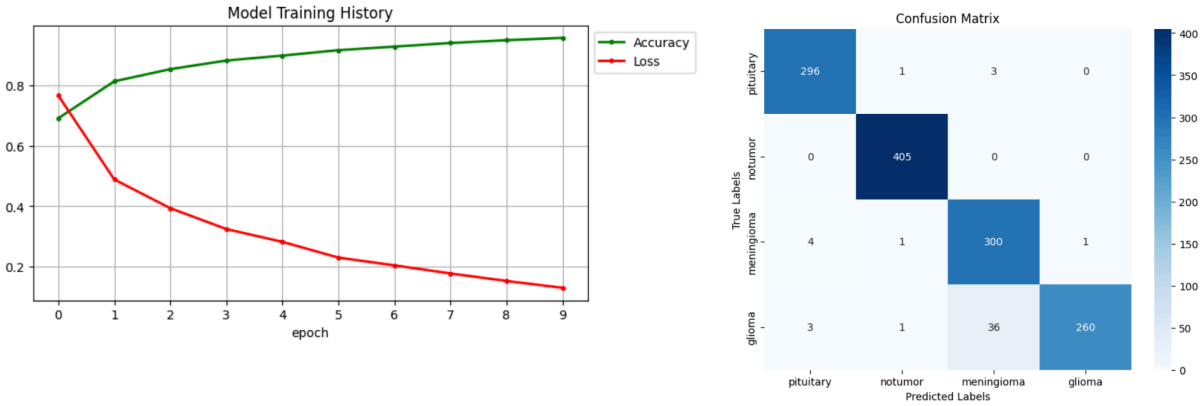


Figure 6.4: U-Net: training and validation accuracy - loss and Confusion matrices

- **CNN + LSTM Hybrid:** Combining spatial and sequential features resulted in a competitive **Accuracy of 0.96**. The hybrid model showed perfect precision and recall for Class 1, successfully leveraging LSTM units to refine the spatial representations generated by the convolutional layers.

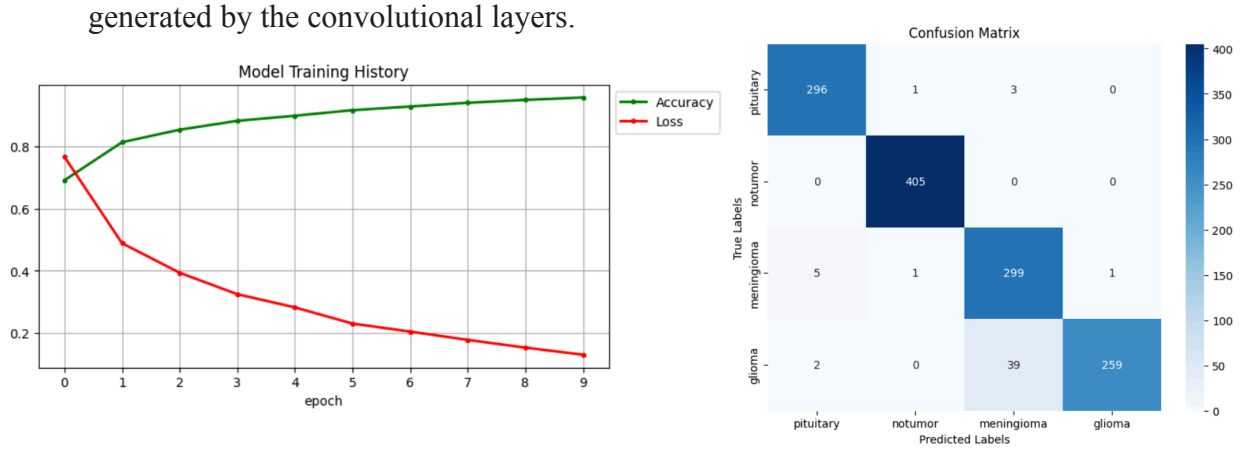


Figure 6.5: U-Net: training and validation accuracy - loss and Confusion matrices

The following table provides a comprehensive comparison of the classification performance across the experimental suite:

Model	Accuracy	Macro Avg F1-Score	Weighted Avg F1-Score	Best Class Performance
VGG16	0.96	0.95	0.96	Class 1 (F1: 1.00)
ResNet50	**0.98**	**0.98**	**0.98**	**All Classes**
MobileNetV2	0.96	0.96	0.96	Class 1 (F1: 1.00)
U-Net	0.96	0.96	0.96	Class 1 (F1: 1.00)
CNN + LSTM	0.96	0.96	0.96	Class 1 (F1: 1.00)

6.3 Diagnostic Insights

The results indicate that Class 1 (Pituitary tumor) was consistently the most recognizable category across all models, often achieving perfect scores. Conversely, Class 2 (Meningioma) and Class 3 (Glioma) represented the most significant challenge, evidenced by the lower precision and recall scores in the baseline and lightweight models.

The superior performance of **ResNet50** suggests that residual learning is the most effective strategy for capturing the complex, high-dimensional features required to differentiate between subtle tumor textures in MRI scans. However, the identical 0.96 accuracy of the lightweight and hybrid models indicates that for clinical deployment where computational resources are limited, **MobileNetV2** provides the most optimal balance of efficiency and diagnostic precision.

7. CONCLUSIONS

This project establishes that the integration of **Deep Learning and Computer Vision** is transformative for automated neuro-pathology classification. By evaluating five distinct architectures, the study confirms that **Transfer Learning** and residual design are pivotal in overcoming medical data scarcity. **ResNet50** emerged as the superior model with a peak accuracy of **0.98**, proving most effective at capturing subtle morphological variations in MRI scans.

Crucially, the high performance of **MobileNetV2** and the **U-Net Encoder** (both at **0.96**) demonstrates that efficient, lightweight models can deliver clinical-grade accuracy while significantly reducing computational overhead. Furthermore, the use of **Captum-based attribution mapping** ensures model interpretability by grounding predictions in biologically relevant pathological regions. Ultimately, while deep models offer the highest precision, the selection of an architecture must balance diagnostic sensitivity with the computational constraints of real-world clinical environments.