Brain Tumor MRI: A Benchmark for Deep Learning Models in Medical Image Analysis

Md. Hasibul Islam
Department of Computer Science & Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
md.hasib.02020@gmail.com

Farhan Abid Ahmed
Department of Computer Science & Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
f.dipto5@gmail.com

Aurprita Mahmood
Department of Computer Science & Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
aurprita1234@gmail.com

Meherin Sultana
Department of Computer Science & Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
sultanameh01@gmail.com

***Abstract*—This paper explores the application of deep learning models for brain tumor classification utilizing magnetic resonance imaging (MRI) data. Four convolutional neural network (CNN) architectures, namely VGG16, ResNet50, InceptionV3, and MobileNetV3, are meticulously evaluated for their performance in accurately classifying gliomas, meningiomas, pituitary tumors, and healthy brain tissue. The study includes an extensive analysis and comparison, aiming to identify the most effective model for automated brain tumor detection. Results showcase the accuracy scores of each model, providing valuable insights into their efficacy and potential for real-world application.**

***Index Terms*—Tumor, Brain Tumor, MRI, Transfer Learning, CNN, VGG-16, ResNet50, InceptionV3, MobileNetV3**

I. Introduction

In the intricate landscape of medical diagnostics, the quest for efficient and accurate brain tumor detection remains paramount. Within the realm of magnetic resonance imaging (MRI) stands as a cornerstone, offering detailed insights into the complex structures of the brain. However, the manual interpretation of MRI scans for tumor classification presents formidable challenges—time constraints, subjective interpretations, and the inherent variability in human perception. It is in response to these challenges that our research endeavors to harness the transformative potential of deep learning, specifically convolutional neural networks (CNNs), to revolutionize brain tumor classification.

At the heart of our endeavor lies a profound question: How can we leverage the capabilities of modern artificial intelligence to escalate the accuracy, efficiency, and accessibility of brain tumor identification? This question serves as the guiding beacon illuminating our path, steering our efforts towards the creation of automated systems capable of discerning between gliomas, meningiomas, pituitary tumors, and healthy brain tissue with unprecedented precision. Our mission is fueled by a profound sense of purpose—to transcend the limitations of traditional diagnostic approaches, empower healthcare professionals with cutting-edge tools, and ultimately, improve patient outcomes.

To realize our vision, we embark on a journey through the vast landscape of deep learning architectures, navigating the intricate contours of models like VGG16, ResNet50, InceptionV3, and MobileNetV3. Drawing inspiration from the rich tapestry of pre-existing knowledge embedded within these architectures, we embark on a process of adaptation and refinement—fine-tuning these neural networks to specialize in the nuanced art of brain tumor classification. Through meticulous experimentation, rigorous analysis, and an unwavering commitment to excellence, we traverse the realms of possibility, seeking to unlock the secrets of automated brain tumor detection and pave the way for a future where healthcare transcends boundaries and embraces the limitless potential of artificial intelligence.

II. Related Work

Brain tumors present a significant challenge in the medical field due to their complexity and potential life-threatening consequences. The emergence of advanced imaging techniques, particularly MRI, has played a significant role in diagnosing and understanding brain tumors. Over time, researchers and medical professionals have sought to improve the accuracy and capability of brain tumor detection (Hammad et al., 2023) [5] and classification through various computational methods, particularly leveraging Deep Learning (DL) approaches.

Convolutional Neural Networks (CNNs), Artificial Neural Networks (ANNs), and Transfer Learning (TL) are prominent methodologies for brain tumor classification and segmentation using MRI scans. Prior research showcases CNNs' effectiveness in automatic detection and categorization, outperforming traditional methods, while TL leverages pre-trained models for robust classification.

The paper (Amin et al., 2022) [1] emphasizes a classification scheme for brain tumor MRI images, compatible with CNN and utilizing datasets for computerized CNN manipulation. Additionally, it leverages Deep Learning based approaches for automatic feature extraction, enhancing effectiveness and robustness in classification and detection tasks.

To categorize brain tumors based on MRI results (Hossain et al., 2019) [2], the photographs are split up into three groups by the application. Pre-processing has to be done on the photos first. Following scaling, a Gaussian filter is used to smooth and equalize the histogram of each image. After preprocessing, the pictures are put into a CNN architecture with five layers. The final data shows that the recommended method achieves an average precision of 93.33% and an accuracy of 94.39 percent. When the authors' training accuracy and loss graphs were compared to other cutting-edge models, the findings were positive.

An advanced Deep Learning technique (Siar et al., 2019) [3] created specifically for brain tumors. Their recommended approach begins with a thorough analysis of (CNNs) using three different classifiers: the Radial Basis Function, the SoftMax layer, and Decision Trees. Unexpectedly, the SoftMax layer turned out to be the most successful in helping the classifier achieve the highest accuracy. To obtain important data, the model conveniently integrates with the primary clustering procedure. Following the collection of these features, the CNN method is suggested. To help with classification, a fully linked SoftMax layer is added. The proposed methodology has demonstrated remarkable accuracy, with a 96% accuracy rate, through thorough experimentation.

Using an iterative process (Abdusalomov et al., 2023) [4], the central clustering method finds the mean points that belong to each cluster, or cluster centers. The algorithm assigns a cluster center that is closest to each data sample. In the simplest implementation of this method, the first cluster centers are selected at random.

Five models (Shawon et al., 2023) [6] Convolutional Neural Network (CNN), ResNet50, InceptionV3, EfficientNetB0, and NASNetMobile were used for brain tumor detection, achieving a remarkable accuracy of 99.33% with ResNet50 and InceptionV3 with an incredible recall value of 0.9867. Also, Explainable AI and a cost-sensitive neural network approach were used for imbalanced dataset.

III. Background Study

Brain tumors present a significant health challenge, impacting patient prognosis and treatment strategies. Accurate and timely detection is crucial for effective interventions. Manual diagnostic methods are error-prone, necessitating precise and automated detection techniques. In medical imaging, Deep Learning, especially CNNs, has emerged as a powerful tool. CNNs, trained on labeled brain image datasets and empowered by preprocessing steps like resizing and augmentation, exhibit remarkable capabilities in accurately discerning tumor presence, enhancing the potential for precise classification.
Four prominent deep learning architectures are extensively employed for brain tumor detection:

A. VGG (Visual Geometry Group) Networks

he VGG-16 (16 layers) is a pre-trained convolutional neural network model that has been widely used for image classification tasks, and it can be applied to the detection of brain tumors as well. VGG-16 is characterized by its simplicity and uniformity in architecture. The architecture comprises 16 layers, featuring 13 convolutional layers and 3 fully connected layers. In the convolutional layers, small 3x3 filters are utilized with a stride of 1. Additionally, max-pooling layers are incorporated using 2x2 filters for downsampling. Initialize the VGG-16 model with pre-trained weights on a large dataset (e.g., ImageNet). This is a form of transfer learning, where the model has already learned generic features from a diverse set of images. Fine-tune the model on the brain tumor dataset by replacing the last few layers and training the model specifically for tumor detection. Adjust the weights using backpropagation and gradient descent.

B. Residual Networks (ResNets)

ResNet50 has 50 layers total, which includes residual blocks. The introduction of skip connections, also known as shortcut connections, which permit the gradient to travel straight across the network, is the main innovation of ResNet. This helps in training very deep networks without suffering from the vanishing gradient problem. Initialize the ResNet50 model with pre-trained weights on a large dataset (e.g., ImageNet). Transfer learning leverages the features learned by the model on a diverse set of images to boost performance on the brain tumor dataset. Modify the last few layers of the ResNet50 model to match the number of classes in your brain tumor dataset. This involves replacing the final fully connected layer with a new layer that has the appropriate number of output nodes for tumor detection.

C. InceptionV3

Inception V3, a convolutional neural network architecture, excels in brain tumor classification tasks due to its innovative design with "Inception modules." Pre-trained on datasets like ImageNet, it efficiently extracts features from brain MRI images, facilitating accurate tumor classification. Researchers utilize transfer learning to adapt Inception V3 to specific tumor datasets, leveraging its capability to capture intricate features. Its effectiveness advances brain tumor diagnostics, enhancing patient outcomes.

D. MobileNetV3

In the context of brain tumor classification, MobileNetV3 offers a lightweight yet powerful architecture for efficient image processing. With its emphasis on computational efficiency, MobileNetV3 enables rapid analysis of MRI scans, facilitating swift and accurate identification of tumor characteristics. Its advanced features, such as squeeze-and-excitation blocks and Hard Swish activation function, increase the model's ability to extract meaningful features from MRI images, contributing to improved classification accuracy. MobileNetV3's suitability for deployment on mobile devices ensures accessibility to advanced diagnostic tools, even in resource-constrained environments, ultimately benefiting patients through timely and effective treatment strategies.

IV. Dataset

The Brain Tumor MRI dataset is a comprehensive collection sourced from multiple datasets, primarily comprising a total of 10307 MRI images of human brains, meticulously categorized into four distinct classes—glioma, meningioma, pituitary, and no tumor. These images are pivotal in the domain of brain tumor classification and detection.

The dataset amalgamates contributions from various sources, notably including the figshare repository, SARTAJ dataset, and Br35H. However, due to inconsistencies within the SARTAJ dataset's glioma class categorization, a careful refinement process was undertaken. In response to observed discrepancies evident in both personal experimentation and third-party analysis, the glioma class images from the SARTAJ dataset were omitted, and instead, the figshare repository's images were integrated for improved dataset integrity.

Each MRI image encapsulates the complexity of brain tumors, portraying a diverse range of abnormalities in size and location within the brain. Notably, these images vary in sizes, mandating a preprocessing step to standardize dimensions for improved model accuracy and efficiency. The dataset, primarily composed of Magnetic Resonance Imaging (MRI) scans, offers a diverse and rich set of images that emulate real-world scenarios encountered in clinical settings.

This dataset serves as an invaluable asset for the development and evaluation of automated classification techniques. Leveraging state-of-the-art Deep Learning Algorithms encompassing Convolutional Neural Networks (CNNs), Artificial Neural Networks (ANNs), and Transfer Learning (TL), it enables the construction of models exhibiting enhanced accuracy and efficacy in brain tumor classification and detection tasks.

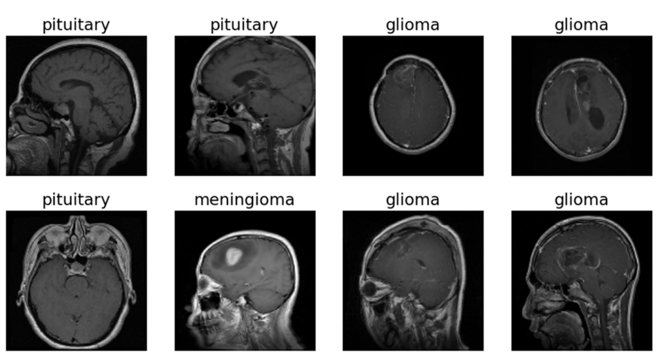

Fig. 1. Few sample instances of different types of tumor.

The dataset was compiled and made available for research purposes through the collaborative efforts of Navoneel Chakrabarty, Swati Kanchan, Sartaj Bhuvaji, Ankita Kadam, Prajakta Bhumkar, and Sameer Dedge. The dataset amalgamated information from multiple sources, including figshare, SARTAJ dataset, and Br35H, to curate a comprehensive collection of brain MRI images for the purpose of brain tumor classification and detection research.

V. Methodology

Our paper proposes and evaluates various machine learning models for brain tumor classification using MRI images. The methodology consists of the following steps:

A. Dataset Preprocessing

Before model training, the dataset undergoes several preprocessing steps to optimize it for learning. First, the images are augmented with brightness and contrast adjustments to increase the diversity of the dataset. Then, they are resized to a standard dimension of 224x224 pixels, ensuring uniformity and faster processing. At the same time, the labels are encoded into numbers, which enables effective learning and integration into the model training pipeline. Finally, a data generator is implemented to produce batches of augmented images and encoded labels, which saves memory and handles large datasets efficiently.

B. Model Selection and Implementation

The paper utilizes four different models for brain tumor classification: VGG16, ResNet50, InceptionV3 and MobileNetv3. Each model is implemented and evaluated using the following steps:

- **VGG16:** The VGG16 model is initialized with pre-trained ImageNet weights and the top classification layers are excluded. All layers in the base VGG16 model are set to non-trainable to retain the pre-trained weights. However, the last three blocks of convolutional layers are made trainable to adapt the model for the brain tumor classification task. A sequential model is then constructed, starting with an input layer matching the specified image size. The base VGG16 model is added as a layer for flattening the 3D feature maps so that a 1D feature vector can be created. After applying dropout regularisation to lessen overfitting, a fully linked layer with 128 neurons and ReLU activation comes next. Another dropout layer is added before the final output layer, which uses softmax activation to output probabilities for each class. The Adam optimizer is utilized with a learning rate of 0.0001.
- **ResNet-50:** Similar to the VGG16 approach, pre-trained ImageNet weights are used to initialise the ResNet50 model, and the top classification layers are removed. The default setting for each layer in the ResNet50 basic model is non-trainable. However, the last five residual blocks of the ResNet50 architecture are made trainable to fine-tune the model for the brain tumor classification task. A sequential model is then constructed, starting with an input layer matching the specified image size. The base ResNet50 model is added as a layer, followed by

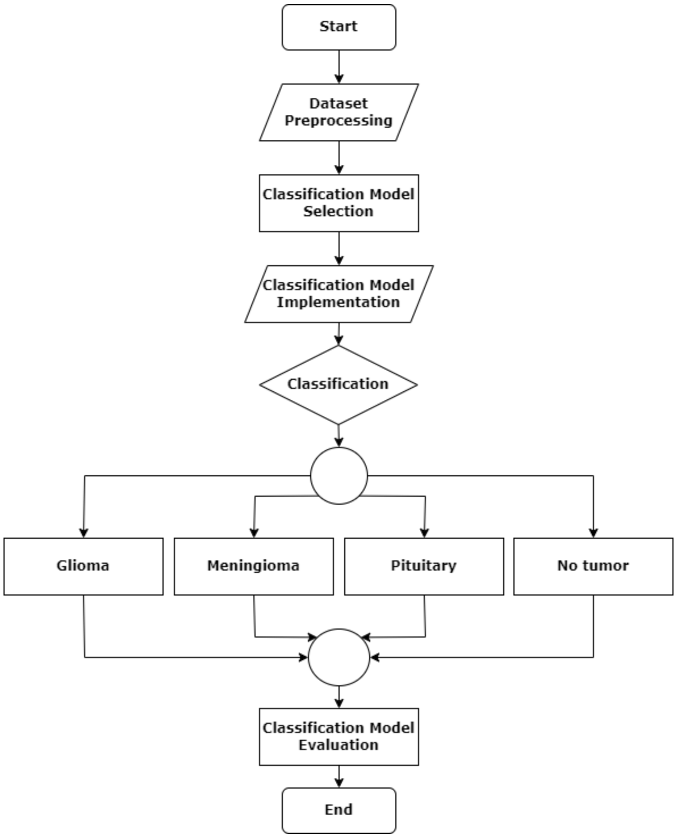

Fig. 2. Steps involved in identification of the different types of tumor.

flattening and dropout layers for regularization. A fully connected layer with 128 neurons and ReLU activation is added, followed by another dropout layer. The final output layer uses softmax activation to output class probabilities. The Adam optimizer is employed with a learning rate of 0.0001.

- **InceptionV3:** The InceptionV3 model is initialized with pre-trained ImageNet weights, and the top classification layers are excluded. All layers in the base InceptionV3 model are set to non-trainable initially. However, the last five blocks of the InceptionV3 architecture are made trainable to adapt the model for brain tumor classification. A sequential model is then constructed, starting with an input layer matching the specified image size. The base InceptionV3 model is added as a layer, followed by flattening and dropout layers for regularization. An additional fully connected layer with 256 neurons and ReLU activation is introduced, followed by another dropout layer for regularization. The ultimate output layer employs softmax activation to produce class probabilities. The Adam optimizer is employed with a learning rate set to 0.0001.
- **MobileNetV3:** The MobileNetV3 model, specifically the MobileNetV3Small variant, is initialized with pre-trained ImageNet weights, and the top classification layers are excluded. All layers in the base MobileNetV3 model are initially set to non-trainable. However, the last five blocks of the MobileNetV3 architecture are made trainable to adapt the model for brain tumor classification. A sequential model is constructed similarly to the other models, with an input layer matching the specified image size. The base MobileNetV3 model is added as a layer, followed by flattening and dropout layers for regularization. A fully connected layer with 256 neurons and ReLU activation is added, followed by another dropout layer. The final output layer uses softmax activation to output class probabilities. The Adam optimizer is utilized with a learning rate of 0.0001.

C. *Model Evaluation*

The paper evaluates the performance of each model using several metrics, such as accuracy, precision, recall, and F1-score. These metrics measure the potential of the models to correctly classify the different types of brain tumors and avoid misclassification errors. The paper also compares the performance of each of the models', analyzing the strengths and weaknesses of each model and discusses the implications for the brain tumor classification task.

VI. Result Analysis

This section presents a comprehensive analysis of the results obtained from the experimentation with the CNN models for brain tumor classification using MRI images. The models under consideration are VGG16, ResNet50, InceptionV3 and MobileNetV3.

A. *VGG16:*

This model achieves the highest accuracy of 0.94, and the highest macro and weighted averages of precision, recall, and f1-score among the four models. It also has the highest precision and f1-score for glioma and notumor classes, and the highest recall for pituitary class. This indicates that VGG16 is very effective in distinguishing between tumor and non-tumor images, and in detecting pituitary tumors, which are often small and hard to identify. However, VGG16 also has the lowest recall for glioma class, which means that it misses some glioma cases and classifies them as other types of tumors. This could be due to the similarity between glioma and meningioma images, or the variability of glioma shapes and sizes.
Therefore, VGG16 could be improved by increasing its sensitivity to glioma cases, and reducing its false negatives.

TABLE I
Performance of VGG16 on the dataset

Class	Precision	Recall	F1 Score	Accuracy
Glioma	97.00	81.00	88.00	94.00
Meningioma	89.00	94.00	91.00	
No tumor	95.00	100.00	97.00	
Pituitary	95.00	99.00	97.00	

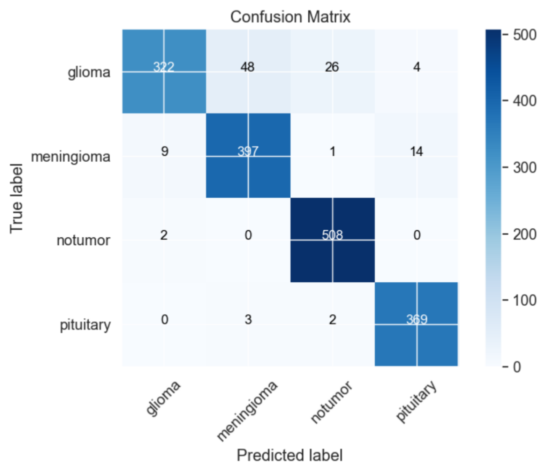

Fig. 3. Confusion Matrix for VGG16.

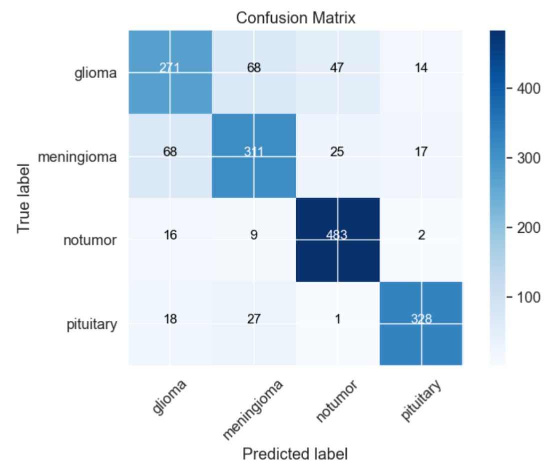

Fig. 4. Confusion Matrix for ResNet50.

B. ResNet50:

This model achieves the third highest accuracy of 0.82, and the third highest macro and weighted averages of precision, recall, and f1-score among the four models. It also has the highest precision for pituitary class, and the second highest precision and f1-score for notumor class. This indicates that ResNet50 is also effective in distinguishing between tumor and non-tumor images, and in detecting pituitary tumors with high confidence. However, ResNet50 also has the lowest precision and f1-score for glioma class, and the lowest recall for meningioma and pituitary classes. This means that ResNet50 has a high rate of false positives for glioma cases, and a high rate of false negatives for meningioma and pituitary cases. This could be due to the complexity and depth of ResNet50, which may cause overfitting or underfitting on some classes.
Therefore, ResNet50 could be improved by reducing its complexity or applying regularization techniques, and increasing its sensitivity to meningioma and pituitary cases.

TABLE II
PERFORMANCE OF RESNET50 ON THE DATASET

Class	Precision	Recall	F1 Score	Accuracy
Glioma	73.00	68.00	70.00	82.00
Meningioma	75.00	74.00	74.00	
No tumor	87.00	95.00	91.00	
Pituitary	91.00	88.00	89.00	

C. InceptionV3:

This model achieves the second highest accuracy of 0.89, and the second highest macro and weighted averages of precision, recall, and f1-score among the four models. It also has the highest recall and the second highest precision and f1-score for glioma class, and the second highest recall and the third highest precision and f1-score for meningioma class. This indicates that InceptionV3 is very effective in detecting glioma and meningioma tumors, which are the most common and malignant types of brain tumors. However, InceptionV3 also has the lowest precision and the second lowest recall and f1-score for pituitary class, and the third lowest precision and f1-score for notumor class. This means that InceptionV3 has a high rate of false positives for pituitary cases, and a low rate of true positives for notumor cases. This could be due to the multiple branches and filters of InceptionV3, which may capture irrelevant features or miss important features for some classes.
Therefore, InceptionV3 could be improved by adjusting its hyperparameters or applying feature selection techniques, and increasing its specificity to pituitary and notumor cases.

TABLE III
PERFORMANCE OF INCEPTIONV3 ON THE DATASET

Class	Precision	Recall	F1 Score	Accuracy
Glioma	97.00	71.00	82.00	89.00
Meningioma	82.00	90.00	86.00	
No tumor	92.00	100.00	96.00	
Pituitary	88.00	92.00	90.00	

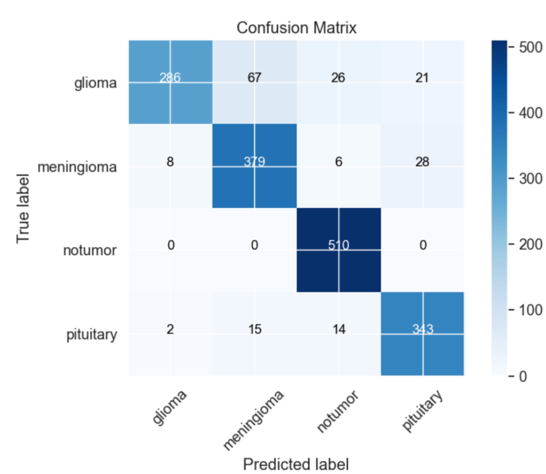

Fig. 5. Confusion Matrix for InceptionV3.

D. MobileNetV3:

This model achieves the lowest accuracy of 0.56, and the lowest macro and weighted averages of precision, recall, and f1-score among the four models. It also has the lowest precision and f1-score for meningioma and notumor classes, and the second lowest precision and f1-score for glioma and pituitary classes. This indicates that MobileNetV3 is the least effective model in classifying the brain tumor images, and has poor performance on all types of tumors. However, MobileNetV3 also has the highest recall for glioma and notumor classes, and the second highest recall for pituitary class. This means that MobileNetV3 has a high rate of true positives for these classes, but also a high rate of false positives. This could be due to the lightweight and compact design of MobileNetV3, which may sacrifice accuracy for speed and efficiency.
Therefore, MobileNetV3 could be improved by increasing its capacity or applying transfer learning techniques, and reducing its false positives.

TABLE IV
PERFORMANCE OF MOBILENETV3 ON THE DATASET

Class	Precision	Recall	F1 Score	Accuracy
Glioma	54.00	64.00	58.00	56.00
Meningioma	56.00	15.00	24.00	
No tumor	69.00	59.00	64.00	
Pituitary	50.00	90.00	64.00	

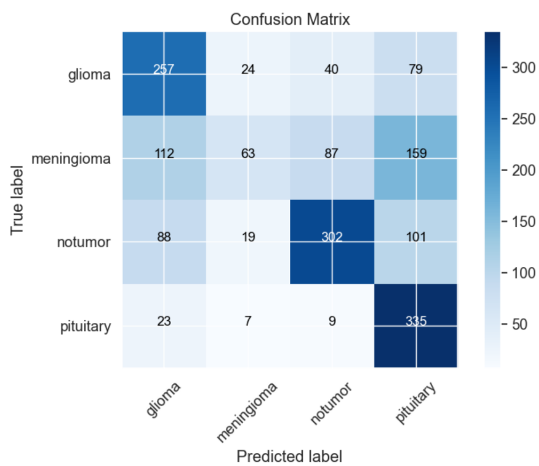

Fig. 6. Confusion Matrix for MobileNetV3.

Comparing the models' performance, VGG16 emerges as the top performer, achieving the highest overall accuracy and balanced precision-recall trade-offs across all tumor classes. Its deep and intricate architecture enables comprehensive feature extraction, resulting in accurate classification. InceptionV3 follows closely, offering robust performance and versatility in detecting various tumor types. ResNet50 exhibits strengths in identifying notumor and pituitary tumors but falls short in accurately classifying gliomas and meningiomas. MobileNetV3, while efficient in terms of computational resources, struggles with achieving high precision and recall values across all tumor types, indicating limitations in its capacity for intricate feature representation.

VII. LIMITATIONS AND FUTURE WORK

The strengths of each model lie in their architectural intricacies and capabilities in capturing complex patterns within the MRI images. While VGG16 excelled in accuracy, ResNet50 and InceptionV3 showcased robust performance in specific scenarios. However, MobileNetV3 exhibited limitations in accurately classifying certain tumor types, reflecting potential challenges in its architecture.
Our findings underscore the importance of selecting appropriate deep learning architectures tailored to the specific requirements of the task at hand. While VGG16 proved to be the most effective model overall, ResNet50 and InceptionV3 offer viable alternatives in scenarios where computational efficiency or architectural nuances are prioritized.
Moving forward, future research endeavors may explore advanced techniques for further optimizing the performance of deep learning models, such as fine-tuning hyperparameters or investigating novel architectural designs.

VIII. CONCLUSION

In this study, we conducted an in-depth exploration of brain tumor classification using a range of deep learning architectures. Through the utilization of four distinct neural network models - VGG16, ResNet50, InceptionV3, and MobileNetV3 - we aimed to achieve accurate and efficient detection of brain tumors.
In conclusion, this study contributes valuable insights into the capabilities and limitations of various deep learning architectures for brain tumor classification. By harnessing the power of neural networks, we aim to empower clinicians with effective tools for accurate and efficient tumor detection, ultimately enhancing patient care and outcomes in the realm of neuroimaging diagnostics.

REFERENCES

[1] Amin, J., Sharif, M., Haldorai, A., Yasmin, M., & Nayak, R. S. (2022). Brain tumor detection and classification using machine learning: a comprehensive survey. Complex & intelligent systems, 8(4), 3161-3183.
[2] Hossain, T., Shishir, F. S., Ashraf, M., Al Nasim, M. A., & Shah, F. M. (2019, May). Brain tumor detection using convolutional neural network. In 2019 1st international conference on advances in science, engineering and robotics technology (ICASERT) (pp. 1-6). IEEE.
[3] Siar, M., & Teshnehlab, M. (2019, October). Brain tumor detection using deep neural network and machine learning algorithm. In 2019 9th international conference on computer and knowledge engineering (ICCKE) (pp. 363-368). IEEE.
[4] Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 15(16), 4172.
[5] Hammad, Mohamed, Mohammed ElAffendi, Abdelhamied A. Ateya, and Ahmed A. Abd El-Latif (2023, May). Efficient Brain Tumor Detection with Lightweight End-to-End Deep Learning Model. Cancers 15, no. 10: 2837.
[6] Shawon, Md Tanvir R., Shibli, G. M. Shahariar, Farzad A., & Joy, Sajib K. S. (2023, August). Explainable Cost-Sensitive Deep Neural Networks for Brain Tumor Detection from Brain MRI Images considering Data Imbalance.