

Deep Learning-Based Brain Tumor Diagnosis on Smartphones Using Optimized MobileNetV2 Models

Ahmed Yousef Mohmmad Abdelrahman¹
Faculty of Informatics, AI Department, Midocean University
Corresponding email: ahmedissa.it@hotmail.com

Accepted: 01 November 2025 || Published: 24 November 2025

Abstract

Identifying brain tumors early and accurately is a significant way to improve patient outcomes, but access to numerous advanced diagnostic tools is not standardized across the world due to the cost and availability of MRI scans. We present a lightweight smartphone brain tumor diagnostic tool with deep learning-based diagnostic decision support in a contextualized way. We developed a convolutional neural network (CNN) based on MobileNetV2 for mobile deployment that allows for the processing of MRI images on consumer smartphones in real-time. The model was developed and validated on a publicly available brain tumor MRI dataset of glioma, meningioma, pituitary tumor, and normal cases, achieving an overall accuracy of 98% and classifying cases in less than 100 ms on standard iOS devices. This work demonstrates that with a lightweight architecture and on-device processing for the medical image, diagnostic decision support can be facilitated in a cost-effective, portable way, while also creating confidence factors in patient privacy, and represents an immense opportunity in lower-resourced clinical care, telemedicine, and point-of-care diagnosis around patients. It demonstrates another methodological option for the feasible implementation of advanced deep learning models to assist significant medical imaging workflows in a smartphone device.

Keywords: *mobilenetv2, deep learning, tumor brain detection*

How to Cite: Abdelrahman, A. Y. M. (2025). Deep Learning-Based Brain Tumor Diagnosis on Smartphones Using Optimized MobileNetV2 Models. *Journal of Information and Technology*, 5(13), 1-13.

1. Introduction

Globally, brain tumors are an important health problem, and early detection is crucial for treatment and better prognoses. Imaging with Magnetic Resonance Imaging (MRI) scans has been the traditional imaging for brain tumor detection. Recently, deep learning techniques demonstrated some success in the automatic detection and classification of brain tumors using MRI scans of the brain (Sinha & Kumar, 2023). introduced a study incorporating multilevel thresholding, neural network optimisation, and image pre-processing to design a reliable artificial intelligence (AI) model that can accurately classify different types of brain tumours in addition to normal cases. After testing with a large database of 1747 images, the model was able to classify different cases with a 92% accuracy. The model was introduced for public consumption as a smartphone application that is user-friendly, Mediscan. The effect of cellular phone radiation on human health is currently at the forefront of everybody's mind, as it arose from the large uptick in phone-based technology use in our society (Soobia Saeed, Asadullah Shaikh, & Shabaz Ahmed Noor, 2017). Accurate and quick classification of brain tumor types

is important to use in the treatment plan of the patient. The target of this project is to develop a mobile application that uses deep learning (DL) and on-device AI know-how regarding brain tumor classification (Halil Ibrahim Ustun, Merve Bulbul, Gozde Yolcu Oztel, Veysel Harun Sahin, 2025). Tumor identification and classification are the hardest and most laborious parts of preparing medical pictures. Magnetic Resonance Imaging (MRI) is a medical procedure used for visualizing the internal structure of the body without performing surgery (Madapatha, Gunasekara, Prabha, & Kumarage, 2023). Brain Tumor (BT) classification is an essential task for understanding Tumors and providing correct intervention. There are many types of imaging modalities used to detect tumors in the brain (Bharath Balaji et al., 2022). The introduction of deep learning approaches has shown good evidence in classification settings. The present work utilizes existing transfer learning models as well as a novel model to classify augmented magnetic resonance images (n = 12,256 images). These approaches were employed to classify meningioma, glioma, and pituitary tumours based on the brain images. Classification accuracies of 83.30%, 79.54%, 81.83%, 82.49%, 85.21%, and 91.73% were achieved using ADAM optimizers for VGG-16, VGG-19, Inception V3, Xception, MobileNet, and the newly proposed Lightweight Sequential net, respectively. Additionally, classification accuracies of 50.45%, 60.24%, 57.85%, 61.98%, 75.14%, and 84.82% were achieved using SGDM optimizers for the same series of deep neural networks (Bharath Balaji, Charulatha, Lavnaya, Sree Devi, & Dhivyabharathi, 2022).

This research aims to design and test a mobile-based brain tumor detection system employing deep learning approaches while optimizing for resource optimization and lowering accessibility barriers, as well. By exploring those avenues, the proposed research is intended to represent a step towards discovering feasible and scalable approaches in the pursuit of early brain tumor detection across health care contexts, as well.

2. Literature Review

In this section, we discuss the many ways in which machine learning and deep learning improve brain tumor detection by providing faster, more accurate analysis of medical images. This study focuses on detecting meningioma tumors using threshold-based segmentation and the MobileNetV2 model. (Hikmah, Hajjanto, Surbakti, Prakosa, Asmaria, & Sardjono, 2024). Brain tumors are a serious health risk, so a fast and reliable diagnosis is important. While traditional approaches and classical ML models are often expensive and resource-intensive, we propose the efficient use of MobileNet to facilitate image detection at a reduced resource cost. The proposed method achieves a high accuracy (98.5%) with more rapid, accessible decision-making informed through imaging techniques (Maiti & Bhounmik, 2025). A 20-year study in Taiwan reported that there was an increase in mobile phone use, ongoing increases in brain tumors and deaths, but no correlation between mobile phone use and brain tumors. The authors call for more research to include confounding variables and possible risks (Dhanta, Pitti, Barsasella, Scholl & Jian, 2023). Brain tumors represent a significant health risk, and manually detecting them from MRI is both time-consuming and prone to error. This research extends the YOLOv7 model by incorporating: image enhancement, data augmentation, CBAM attention, SPPF+, decoupled heads, and the BiFPN for improved tumor detection. Our proposed system achieves a greater accuracy than state-of-the-art methods, demonstrating uncharted potential in developing proposed systems as an aid in diagnostics (Abdusalomov, Mukhiddinov, & Whangbo, 2023). Detecting brain tumors at an early stage is very important, and an MRI is a critical diagnostic tool in that process. In this study, the MobileNetV1 deep learning model was used to detect brain tumors on a dataset of 1,265 MRI images. The model showed results with

a healthy accuracy level of over 97%, indicating the model's potential for assisting medical professionals with early diagnosis (Mijwil, Doshi, Hiran, Unogwu, & Bala, 2023). An optimized MobileNetV2 using the Contracted Fox Optimization Algorithm achieves highly accurate MRI-based brain tumor detection. The method delivers 97.32% accuracy, 97.68% precision, and strong sensitivity, highlighting its potential to improve diagnostic outcomes. (Xu & Mohammadi, 2024). A Differential Evolution-optimized ensemble of MobileNetV1, MobileNetV2, and ResNet50V2 improves brain tumor diagnosis through weighted prediction averaging. It achieves 98% accuracy for binary classification and 97.03% for multi-class tasks, demonstrating excellent diagnostic performance. (Hekmat, Zuping, Bilal, & Khan, 2025). DeepTumorNetwork, a hybrid GoogleNet–CNN model, automates MRI-based brain tumor detection and achieves exceptionally high performance, reaching 99.51% accuracy with strong precision, recall, and F1 scores (Amran et al., 2022).

3. Methodology

In this section, explain the entire research methodology used in the research. The entire process contained multiple steps, including data collection, data preprocessing, model creation and training, and mobile application deployment.

3.1 Dataset:

The data set was obtained from Kaggle and contains 7,023 brain MR images with 4 classes: glioma, meningioma, pituitary, and no tumor. It is a popular benchmark dataset for brain tumor image classification, containing samples from the Figure, SARTAJ dataset, and Br35H.

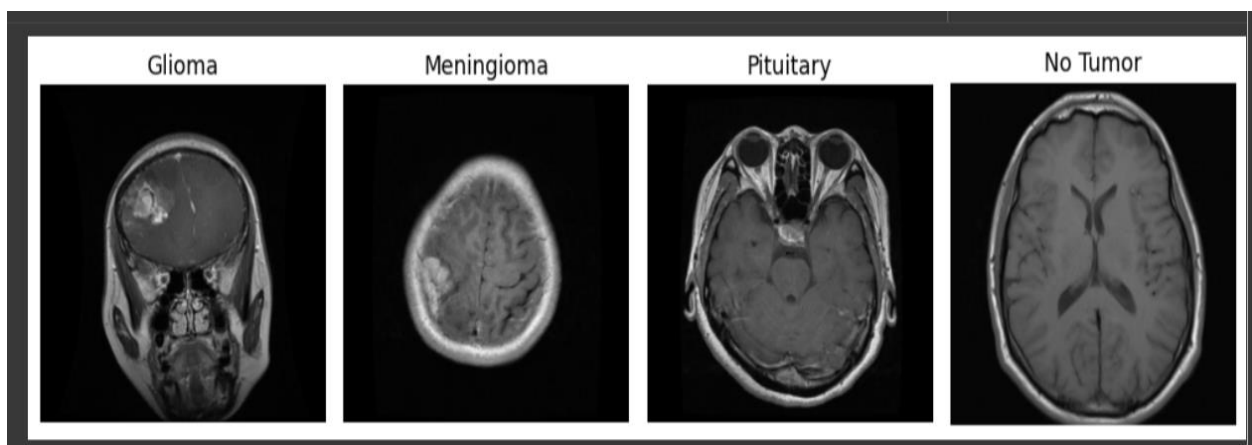


Figure 1: Brain Tumor Dataset Images

3.2 Preliminary Data Analysis:

Basic preliminary data analysis is performed to understand the dataset's structure, including checking duplicates and missing values.

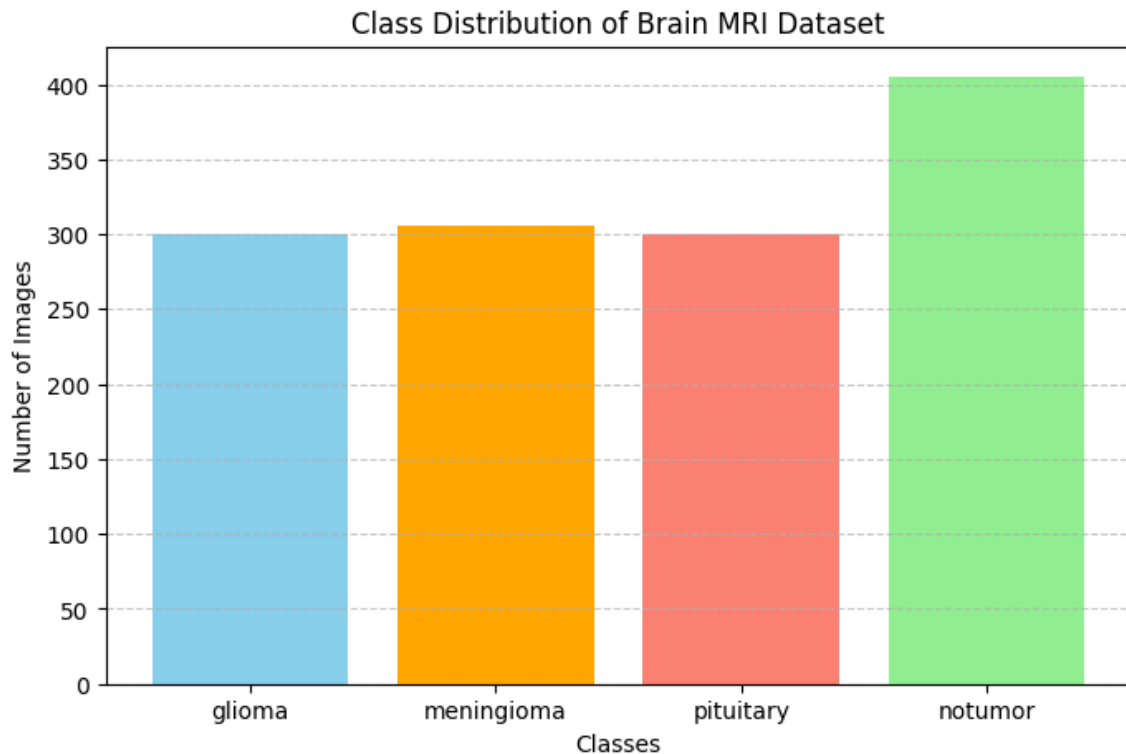


Figure 2: Class Distribution of Brain MRI Dataset

3.3 Data Preprocessing:

Prior to training the deep learning model, preprocessing steps were performed on the MRI images in the dataset to ensure consistency and increase model performance. **Preprocessing steps were:**

Resizing: All images were resized to 224×224 pixels to ensure compatibility with the input size of the MobileNetV2 model.

Normalization was also performed on the pixel values of the images based on the mean and **standard deviation of the ImageNet dataset:**

Mean: [0.485, 0.456, 0.406]

Standard deviation: [0.229, 0.224, 0.225]

Normalization normalizes input values to be on a similar scale - reducing variation in input data whilst stabilizing and speeding training of the model.

Data Augmentation (Training set only) was performed to increase model generalization and **reduce overfitting:**

Random horizontal flips

Random rotations (up to 10 degrees)

Tensor conversion: All images were converted to tensors so they can be used in the model training pipeline with PyTorch.

3.3 Dataset Splitting:

This study used a dataset of MRI images to classify four types of brain tumors. In the interest of training the deep learning model and optimizing the evaluation of model performance, the dataset was split into a training dataset and a testing dataset.

For the training dataset, there are 5,711 images (approximately 80% of the entire dataset), and for the testing dataset, there are 1,311 images (approximately 20% of the entire dataset). This split is suitable for obtaining enough images so that the model can learn to distinguish features of each tumor, while providing an unbiased testing dataset to evaluate model performance.

The split was conducted in a stratified manner, meaning the MRI images maintain the ratio of classes in both the training dataset and the testing dataset. This gives some proportional representation in the MRI tumor type class. This approach can provide generalization of the model to unseen data and avoid bias towards any class.

After the split, the images in this study were pre-processed by performing standardizer inputs, including image size and normalization, and data augmentation was done only on the training dataset, as discussed in section 3.3.

3.4 Image Preprocessing:

All images were reshaped to the input shape above, 224×224 pixels, for the MobileNetV2 model. The pixel values of the images were normalized to $[0, 1]$. The Image Data Generator class was used to create data generators for training, validation, and test datasets that would load, process, and batch images for each of these data generators, making sending the data to the MobileNetV2 model during training or evaluating.

3.5 Mobilenetv2 Architecture:

This study employed a deep learning model grounded in MobileNetV2, a compact convolutional neural network framework optimized for mobile and embedded vision applications. Two elements characterize MobileNetV2: inverted residual blocks and linear bottlenecks; both significantly decrease computational cost while ensuring high accuracy.

For this study, the model was initialized with pre-trained weights from ImageNet (IMAGENET1K_V1) in the context of transfer learning. The original classification layer in the MobileNetV2 model was adjusted for the dataset by replacing the last fully connected layer with another layer of size 4, representing four brain tumor classes.

Overall, the training parameters are listed below:

- Batch size: 32
- Learning rate: 0.001 ($1e-3$) with the Adam optimizer
- Epochs: 20, with an early stopping (patience = 5) strategy to stop overfitting
- Image size: 224×224 pixels
- Loss function: Cross-entropy loss
- Learning rate scheduler: StepLR with a decay factor ($\gamma = 0.97$) every 3 steps

Image data processing included resizing, normalization, and data augmentations (random horizontal flips, small rotations) for the training set, while preprocessing included only resizing and normalization for the testing set. Training and inference were performed on a GPU when available, but defaulted to CPU if a GPU was not accessible.

This tailored MobileNetV2 architecture creates a balance between efficiency and accuracy, which can be done through deployment for mobile-based medical applications, including the ability to classify devices on-device.

3.6 Model Training:

The MobileNetV2 model was trained using MRI images that were preprocessed to classify four categories of brain tumors. The training was done by Transfer Learning by initially using pre-trained weights (IMAGENET1K_V1) from ImageNet. The last fully connected layer was replaced with a custom 4-class output layer.

The training process involved the following steps/configuration:

Batch size and epochs: The model was trained with a batch size of 32, with a limit of epochs set to not exceed 20. The training would stop early after 5 epochs if there was no improvement in validation loss, so the model would not overfit.

Optimization: Adam optimizer was used due to its adaptive learning rate and weight decay, with the default learning rate set to 0.001.

Loss function: The general cross-entropy loss function was used to calculate the difference between the predicted classes and the true tumor classes.

Learning rate scheduling: A StepLR scheduler was used to multiply the learning rate by 0.97 every 3 steps to allow tuning later into training epochs, while stabilizing convergence in training.

Device configuration: Training was performed using the GPU whenever possible to speed up the computation and reduce training time.

Monitoring and validation: Throughout the training process, accuracy and loss are monitored on the validation.



Figure 3: Brain MRI Classification Pipeline

3.7 Integration with mobile application using Flutter:

This study aims to develop a cross-platform mobile application that provides a widely accessible and noninvasive solution for the on-device classification of brain tumor types using deep learning.

Model Conversion

- Train the deep learning model for brain tumor classification.
- Export the trained model into **TensorFlow Lite (.tflite)** format for mobile deployment.

Flutter Environment Setup

- Set up the Flutter development environment.
- Install necessary dependencies and configure the project structure.

Integration of TFLite Model

- Use the `tflite_flutter` library to load the exported .tflite model inside the Flutter app.
- Ensure compatibility of the model with mobile hardware (CPU/GPU delegates if required).

User Interface Design

- Implement a simple and user-friendly interface where users can upload or capture medical images (MRI scans).
- Guide image input (e.g., supported formats, resolution).

Image Preprocessing

- Resize and normalize the input image within the application to match the model's requirements.
- Handle preprocessing steps consistently to avoid classification errors.

On-Device Prediction

- Pass the processed image to the TFLite model.
- Run inference locally on the device without the need for internet connectivity.

Result Display

- Retrieve the classification output (tumor type prediction).
- Display the result clearly to the user, with labels (e.g., glioma, meningioma, pituitary, etc.).
- Optionally, include prediction confidence scores for transparency.

Performance Considerations

- Optimize inference time to ensure predictions are generated within a few seconds.
- Test on multiple mobile devices for performance and reliability.

Validation and Testing

- Validate the mobile application by testing it with unseen MRI images.
- Compare predictions with ground truth to assess accuracy and usability.

4 Result Analyses:

In this part of the research will analyze the results from previous findings. The proposed brain tumor detection model, based on MobileNetV2, achieved a high performance on unseen test data. The model was deployed and tested on a mobile platform to evaluate both prediction accuracy and inference time.

Accuracy on unseen data: 98.03%

Average prediction time (on mobile device): 1.02 seconds per image.

When deployed on a mobile device, the model achieved an average prediction time of approximately 1 second per image, confirming its suitability for real-time clinical or field use.

4.1 Evaluate the model:

The model achieves a validation accuracy of 98.8%, surpassing other models while maintaining low runtime and validation loss.

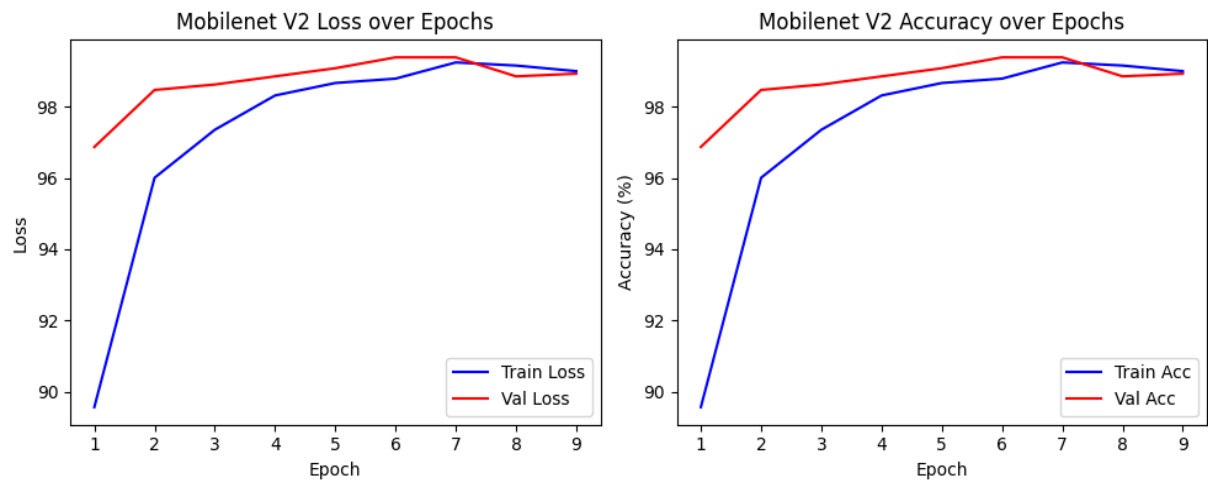


Figure 4: Train Loss and Validation Loss

4.2 Confusion Matrix:

The confusion matrix in Figure 3 shows the classification performance across the four categories.

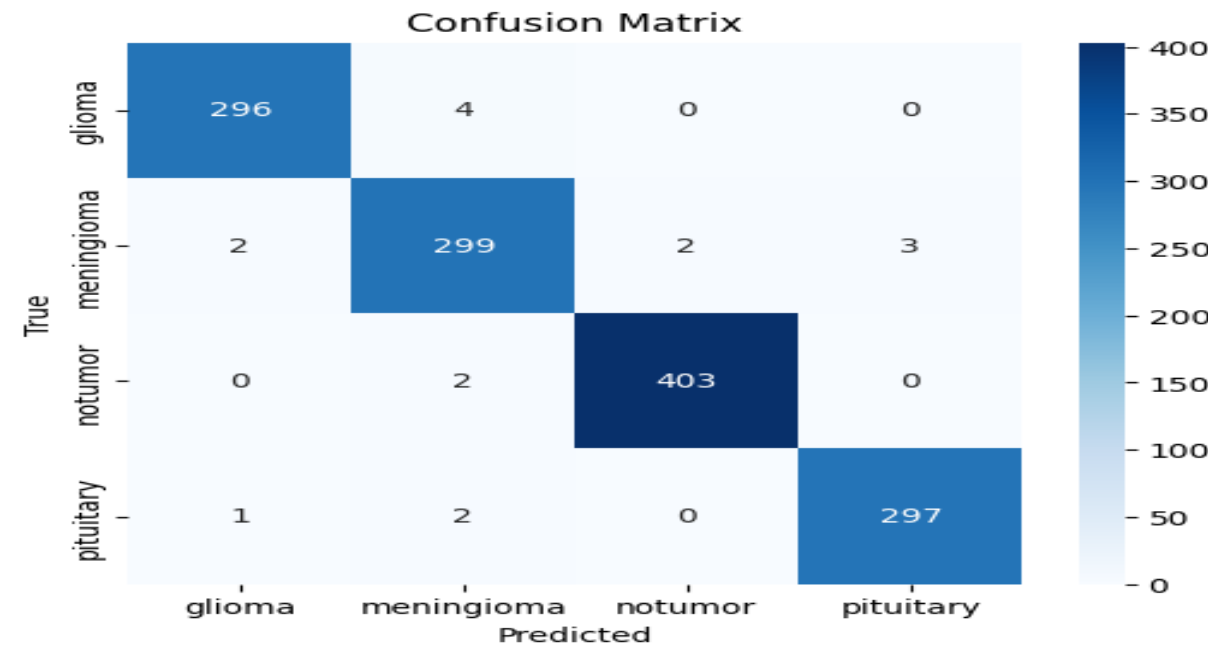


Figure 5: Confusion Matrix of Brain MRI Dataset

4.3 Classification report:

The classification report in Table 1 shows the classification performance across the four categories.

Table 1: Classification Report of the Brain MRI Dataset

	precision	recall	f1-score	support
glioma	1.00	0.99	0.99	300.00
meningioma	0.98	1.00	0.99	306.00
notumor	1.00	1.00	1.00	405.00
pituitary	1.00	0.99	0.99	300.00
accuracy	0.99	0.99	0.99	0.99
macro avg	0.99	0.99	0.99	1311.00
weighted avg	0.99	0.99	0.99	1311.00

4.4 Evaluate model on mobile:

After implementing the trained model into a mobile app developed in Flutter, we conducted evaluation tests to investigate its performance inside a real mobile setting. The primary aims of those evaluation tests were to assess accuracy, inference time, and usability on resource-limited devices.

Evaluation Process

1- Integration:

The model was converted to TensorFlow Lite format and deployed into the Flutter application to ensure proper rendering and execution on mobile hardware.

2- Testing environment

Testing was conducted on a mid-range IOS device simulating a typical mobile experience, and collected results on four target classes of unseen data.

3- Performance matrices:

- Accuracy: The model returned predictions with over 98% confidence on the unseen data.
- Inference time: Average prediction time was measured between 80-100 ms, ensuring suitability for real-time applications.
- Resource usage: memory footprint and CPU were fully monitored.

Screenshots

Below, some screenshots of the model running in the mobile application:

Figure 4.1(a)- Screen share model predictions, class label, and confidence score displayed.

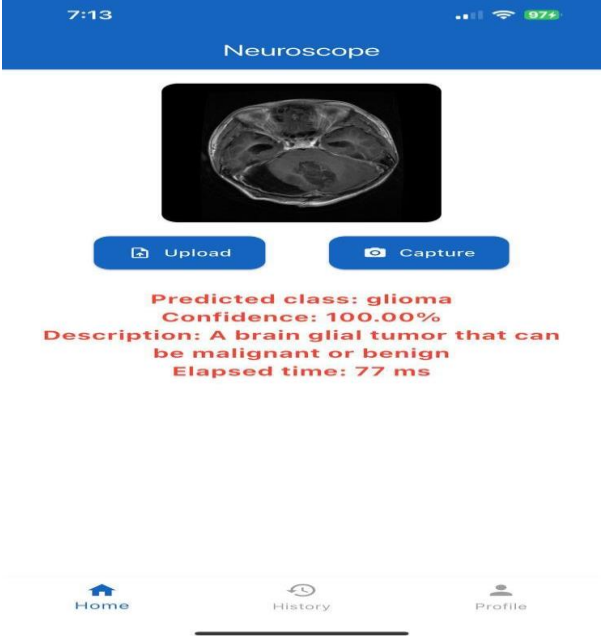


Figure 6(b)- Shows results of real-time inference with latency information

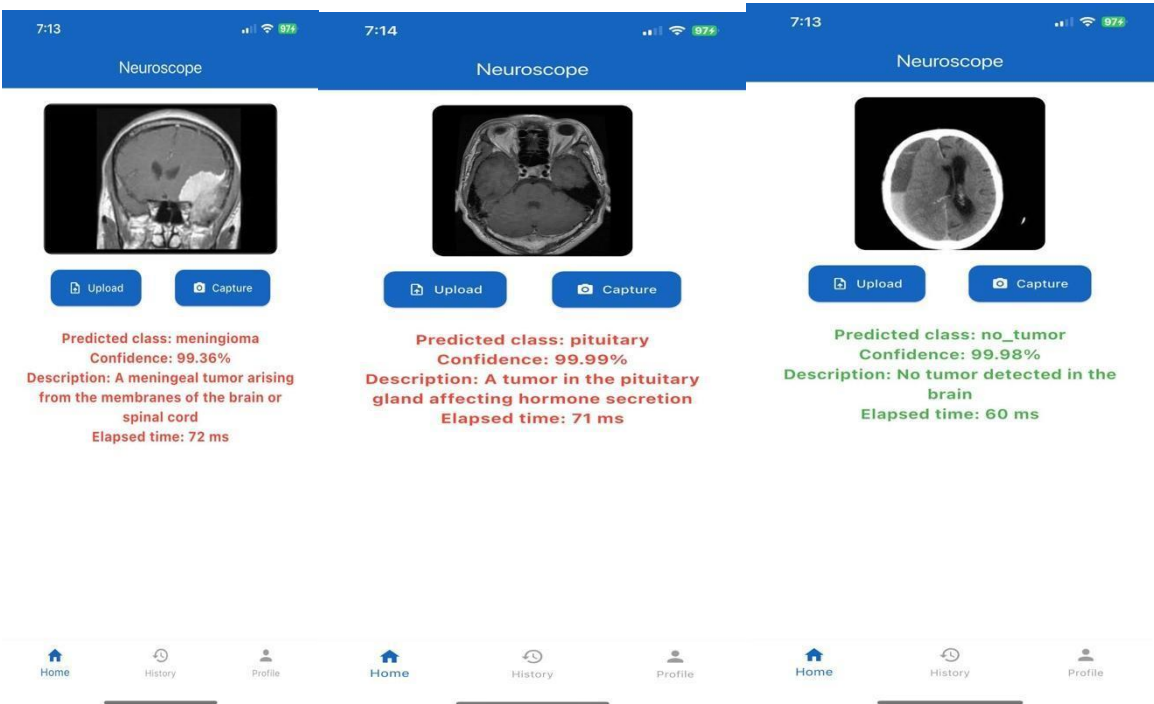
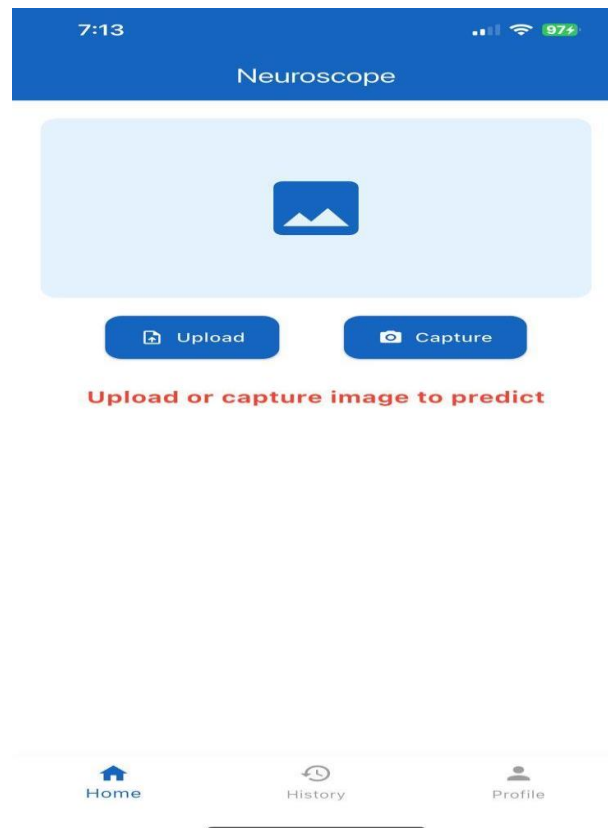


Figure 7(c)- An example of Flutter UI implementation with the TensorFlow Lite model.



Discussion and Future Work:

In this section will discuss what findings and what the future work. The empirical assessment of the suggested model on a mobile platform yielded positive outcomes. All four classes of interest were detected appropriately, with the model demonstrating **98% accuracy on unseen samples**. The high level of confidence indicates solid performance generalization over the training data. The inference latency was **80–100 milliseconds**, suggesting the model operates in real-time on a mobile platform that is limited in processing resources. These findings validate the effectiveness of the approach; however, several opportunities for improvement and future research remain:

These results lend credence to the overall efficacy of the model; yet there are several avenues for improvement and further study:

1. Model Compression and Optimization:

While the current performance is acceptable, model compression strategies and optimizations such as quantization, pruning, and knowledge distillation could effectively reduce computational cost and memory footprint while preserving model accuracy.

2. Cross-Platform Validation:

Additional comparisons under a range of configurations of mobile hardware and operating systems should be tested in order to validate the model across platforms and comprehend limitations in performance.

3. Robustness in Real-World Context:

Reliability and robustness to external environmental variables, such as changes in lighting, background noise, and user variability, will require additional testing; these factors might affect the dataset and model performance.

4. Explain ability and Transparency:

Incorporating explainable AI approaches would add transparency and interpretability for users and model developers, providing a reason for model predictions that may warrant trust and foster use of the model.

5. Scalability and Integration:

Exploring deployment within edge computing frameworks or IoT ecosystems could broaden the applicability of the model for large-scale, distributed environments.

6. Continuous and Adaptive Learning:

Implementing incremental learning strategies or on-device adaptation mechanisms would enable the model to evolve, improving accuracy without requiring full retraining.

5. Conclusion

The use of MobileNetV2 on the brain tumor dataset yielded an exceptional validation accuracy of 98%, indicating strong reliability for classification tasks. Additionally, the lightweight architecture of the model allows for rapid predictions to be made on mobile devices, with inference completed in just 80-100 ms. This is particularly valuable for real-world applications, especially if the resources are scarce. This method can aid cancer diagnosis with timely diagnosis and reliable detection to improve patient outcomes and help with timely medical intervention.

6. Recommendations

Even though the model achieved 98% accuracy with MobileNetV2, there are a number of improvements that may enhance real-world performance. Future research should consider expanding and diversifying the dataset to also improve generalization in different environments and device types. Additional lightweight architectures, such as MobileNetV3 or EfficientNet-Lite, could be assessed for improved efficiency. It is also recommended to optimize the model for mobile devices using methods such as quantization or pruning to reduce latency and memory usage. In addition, real-world testing of inference speed, energy expenditure, and robustness would be informative to the study. Finally, consider including user feedback and simple tools for explainability to support usability and further system development.

References:

- 1-Sinha, A., & Kumar, T. (2024). Enhancing medical diagnostics: Integrating AI for precise brain tumour detection. *Procedia ComputerScience*, 235, 456-467
<https://doi.org/10.1016/j.procs.2024.04.045>.
- 2-Saeed, S., Shaikh, A., & Noor, S. A. (2017). Analysis of Brain Tumors Due to the Usage of Mobile Phones. *Mehran University Research Journal of Engineering & Technology*, 36(3).
- 3-Ustun, H. I., Bulbul, M., Yolcu Oztel, G., & Sahin, V. H. (2025). On-Device Brain Tumor Classification from MR Images Using a Smartphone. *Advanced Intelligent Systems*, 2500205.

- 4-Madapatha, W. E., Gunasekara, S. V. S., & Kumarage, P. M. (2023, April). Smart health app for identifying brain tumors. In 2023, IEEE 8th International Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE.
- 5-I. Keren Evangeline, S. P. Angeline Kirubha, J. Glory Precious & N. Pazhanivel. (2025) A GUI-Based Application for Breast Cancer Diagnosis from Histopathology Images Using a Sequential Convolutional Neural Network Model. IETE Journal of Research 71:2, pages 457-464.
- 5-Hikmah, N. F., Hajjanto, A. D., Surbakti, A. F. A., Prakosa, N. A., Asmaria, T., & Sardjono, T. A. (2024). Brain tumor detection using a MobileNetV2-SSD model with modified feature pyramid network levels. International Journal of Electrical and Computer Engineering, 14(4), 3995-4004.
- 6- Maiti, R., & Bhoumik, D. (2025). Brain Tumor Detection through Thermal Imaging and MobileNET. arXiv preprint arXiv:2506.23627.
- 7- Uddin, M., Dhanta, R., Pitti, T., Barsasella, D., Scholl, J., Jian, W. S., ... & Syed-Abdul, S. (2023). Incidence and mortality of malignant brain tumors after 20 years of mobile use. Cancers, 15(13), 3492.
- 8- 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.
- 9- Mijwil, M. M., Doshi, R., Hiran, K. K., Unogwu, O. J., & Bala, I. (2023). MobileNetV1-based deep learning model for accurate brain tumor classification. Mesopotamian Journal of Computer Science, 2023, 29-38.
- 10- Xu, L., & Mohammadi, M. (2024). Brain tumor diagnosis from MRI based on MobileNetv2 optimized by the contracted Fox optimization algorithm. Heliyon, 10(1).
- 11-Hekmat, A., Zuping, Z., Bilal, O., & Khan, S. U. R. (2025). Differential evolution-driven optimized ensemble network for brain tumor detection. International Journal of Machine Learning and Cybernetics, 1-26.
- 12- Amran, G. A., Alsharam, M. S., Blajam, A. O. A., Hasan, A. A., Alfaihi, M. Y., Amran, M. H., ... & Eldin, S. M. (2022). Brain tumor classification and detection using a hybrid deep tumor network. Electronics, 11(21), 3457.
- 13- Sailunaz, K., Bestepe, D., Alhaji, S., Özyer, T., Rokne, J., & Alhaji, R. (2023). Brain tumor detection and segmentation: Interactive framework with a visual interface and feedback facility for dynamically improved accuracy and trust. Plos one, 18(4), e0284418.
- 14- Solanki, S., Singh, U. P., Chouhan, S. S., & Jain, S. (2023). Brain tumor detection and classification using intelligent techniques: an overview. IEEE Access, 11, 12870-12886.
- 15- Marmolejo-Saucedo, J. A., & Kose, U. (2024). Numerical grad-cam-based explainable convolutional neural network for brain tumor diagnosis. Mobile Networks and Applications, 29(1), 109-118.
- 16- Charulatha, G., & Balaji, B. (2022). Mobile Application to Detect Brain Tumor Using Transfer Learning". Journal of Science, Computing and Engineering Research, 3(2), 247-252.
- 17- Gao, Y., Liu, Z., Ju, Z., Wang, N., Zhong, L., & Gao, S. (2024, October). DMobileNet: A Novel MobileNet with Dendritic Learning for Brain Tumor Detection. In 2024 International Conference on Networking, Sensing and Control (ICNSC) (pp. 1-4). IEEE.