Indonesian Journal of Electrical Engineering and Computer Science
Vol. 43, No. 1, July 2026, pp. 281~298
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v43.i1.pp281-298

MobileNetV2 with transfer learning for brain tumor classification

Aziz Srai
Department of Computer Science, Faculty of Sciences and Techniques Al Hoceima, Abdelmalek Essaadi University, Tétouan, Morocco

Article Info

Article history:

Received Mar 6, 2026
Revised Jun 2, 2026
Accepted Jun 27, 2026

Keywords:

Artificial intelligence
Deep learning
Fine tuning
Neuro-oncology
Transfer learning

ABSTRACT

This document presents a deep learning-based approach for the automatic classification of brain tumors from magnetic resonance imaging (MRI) images, using the lightweight MobileNetV2 model combined with transfer learning and fine-tuning techniques. The study aims to address the constraints of resource-limited medical environments, where model speed and lightweight design are as important as accuracy. The dataset used is a fusion of three public databases (Figshare, SARTAJ, and BR35H), comprising 4,480 images for training and 1,600 for testing, divided into four classes: glioma, meningioma, pituitary tumor, and healthy brain. The methodology includes several steps: resizing the images to 224×224 pixels, normalizing pixel values between 0 and 1, augmenting the data through random rotations, shifts, and zooms to avoid overfitting, and then extracting features using MobileNetV2 pre-trained on ImageNet. The strategy adopted comprises two phases: first, transfer learning where only the layers added at the top of the model are trained for 10 epochs, then fine-tuning consisting of unfreezing the last 20 layers of the base model and retraining them with a reduced learning rate for 5 epochs. The results obtained show an overall accuracy of 86.56% after fine-tuning, with a macro-mean area under the ROC curve (AUC) of 0.9645, indicating excellent discriminatory power. The confusion matrix reveals that the "no tumor" class achieves perfect performance (400 out of 400), while the "meningioma" class remains the most difficult to classify, often confused with gliomas and pituitary tumors. Compared to more resource-intensive models like VGG16, ResNet50, or EfficientNet, MobileNetV2 offers an optimal balance between performance and lightweight design, with a significantly lower number of parameters, making it particularly well-suited to resource-constrained environments. The authors conclude that this approach provides reliable diagnostic support for radiologists, accelerating tumor detection without replacing medical expertise. Future directions include clinical validation on multi-center data, integration of an automated segmentation step, and exploration of newer architectures such as attention mechanisms.

Corresponding Author:

Aziz Srai
Department of Computer Science, Faculty of Sciences and Techniques Al Hoceima
Abdelmalek Essaadi University
Tétouan, Morocco
Email: a.srai@uae.ac.ma

1. INTRODUCTION

The brain is made up of several billion cells: neurons, which create, direct, and control information, and glial cells, which support neurons and enable their development and nutrition, and contribute

significantly to brain tissue homeostasis. Figure 1 presents four magnetic resonance imaging (MRI) brain image classes: glioma, meningioma, pituitary tumor, and no tumor. There are several types of glial cells (astrocytes, oligodendrocytes, microglia, ependymal cells), each with specific roles. Throughout life, brain cells can acquire mutations in their deoxyribonucleic acid (DNA) due to aging, environmental factors, and genetics, thus transforming into cancerous cells that escape cellular control mechanisms. Brain tumors are subdivided into two distinct types:

- Primary tumors: these originate in the brain mainly from glial cells that make up the nervous system [1]. They can be benign (two-thirds of primary tumor cases) or malignant [2].
- Secondary tumors, also called brain metastases, originate from a pre-existing cancer located in another organ and migrate via the bloodstream to brain tissue [3]. The cancers most frequently associated with the development of brain metastases are lung cancer, breast cancer, and melanoma [4]. Secondary tumors, or brain metastases, are the most common type of brain tumor, occurring 3 to 10 times more frequently than primary brain tumors [5].
- Gliomas: Glioma is the general name given to the tumor group encompassing primary brain tumors originating from glial cells [6]. There are different types of gliomas associated with the three types of glial cells: astrocytomas (for astrocytes), oligodendrogliomas (for oligodendrocytes), and ependymomas (for ependymal cells) [7].
- Meningioma: A meningioma is a tumor that develops from the meninges, the membranes that cover the brain and spinal cord [8]. It accounts for approximately 35% of primary brain tumors, making it the most common [9]. Meningiomas are benign in 80% of cases (grade I), but atypical (grade II) and malignant anaplastic (grade III) forms also exist [10]. They occur more frequently in women, due to hormone receptors [11]. Most meningiomas remain asymptomatic and are discovered incidentally during an imaging examination [12]. When they become symptomatic, they cause seizures, headaches, or focal neurological deficits [13]. Through mass effect, they can compress a brain region, a cranial nerve, or a blood vessel [14]. On MRI, a typical meningioma is a well-defined, extra-axial mass, often onion-shaped [15].
- Pituitary: A pituitary tumor, or pituitary adenoma, is a benign tumor that develops in the pituitary gland, the "conductor" of hormones [16]. The pituitary gland is located in the sella turcica, a small bony cavity at the base of the skull [17]. Pituitary adenomas are very common, found in approximately 10% of the population at autopsy [18]. Most are microscopic and completely asymptomatic. They can become symptomatic through two mechanisms: hormonal excess or local compression. So-called functional adenomas secrete hormones in excessive amounts. Prolactin is the hormone most often found in excess, causing a syndrome in women (galactorrhea, menstrual irregularities) and in men (decreased libido, impotence). Excess growth hormone in adults causes acromegaly, with enlargement of the hands, feet, and facial features. Excess cortisol (Cushing's disease) leads to weight gain, stretch marks, hypertension, and skin fragility. Non-functioning adenomas do not secrete hormones and are revealed by their mass effect. Compression of the optic chiasm, which crosses just above the pituitary gland, causes bitemporal hemianopsia (loss of peripheral vision in both eyes) [19].

In summary, these three tumors differ in their origin, behavior, and prognosis, but share the need for precise imaging and multidisciplinary management.

On the other hand, the introduction of MRI in the field of oncology was initiated by physician Raymond Damalian in 1971 when he discovered that tumor tissue has a different relaxation time than healthy tissue. Since then, several imaging techniques have emerged, notably diffusion-weighted imaging (DWI), created in 1986 by physician and physicist Denis Le Bihan, which allows for the exploration of the brain's anatomical connectivity by measuring a signal sensitive to the movement of water molecules [20]. In 1990, the work of neuroscientist and biophysicist Seiji Ogawa brought about considerable progress with the development of functional MRI, showing the brain "in action." This magnetic susceptibility imaging technique exploits the oxygenation properties of blood, also known as the BOLD effect – Blood-Oxygen Level-Dependent. Finally, the 2000s marked a new advance in MRI with the use of contrast agents having magnetic susceptibility for the study of vascularization, the best known and used being gadolinium chelate, allowing in particular the assessment of vascular permeability [21]. In summary, MRI and artificial intelligence play an important role today in the field of oncology.

In this article, we used deep learning as the primary algorithm for brain tumor detection [22]. Brain tumor detection typically occurs in two stages: image preprocessing and classification [23]. For classification, we used a pre-trained MobileNetV2 model. We then employed fine-tuning to monitor the performance of our model. Two important concepts in deep learning are used in this document: transfer learning and fine-tuning. Figure 2, for example, presents an image prediction using transfer learning, and Figure 3 shows an example of a prediction using fine-tuning. The remainder of the article is organized as

follows: Section 2 describes the methodology followed in this document, Section 3 describes the experimental results obtained, section 4 concludes the article and proposes research perspectives.

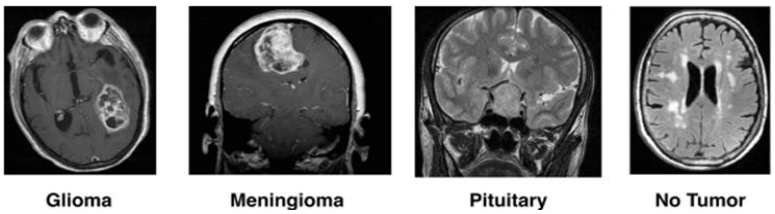

Figure 1. Four (4) types of MRI images of brain tumor

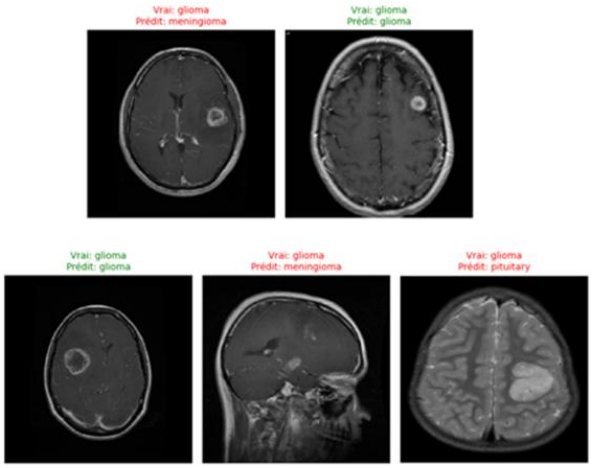

Figure 2. Example of a prediction without fine-tuning, using only transfer learning

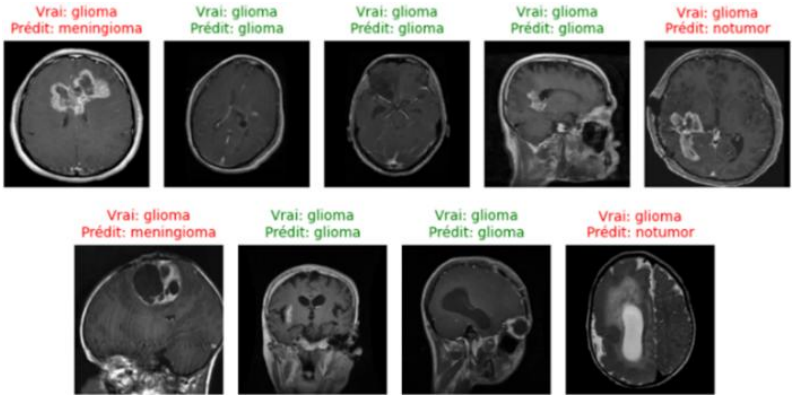

Figure 3. Example of a prediction using Fine-tuning

In other words, and before presenting the results of this scientific article, we must explain our choice of the MobileNetV2 model. Indeed, the choice of MobileNetV2 as a high-performance approach is based on its excellent balance between accuracy, speed, and lightweight nature, making it particularly well-suited to constrained environments such as mobile or embedded systems. Unlike more resource-intensive models such as VGG16 or ResNet50, or EfficientNet, EfficientNetV2, ConvNeXt, vision transformer (ViT), Swin Transformer, RegNet, MobileViT, and MobileNetV3, MobileNetV2 uses separable depth convolutions and inverse linear bottleneck blocks, drastically reducing the number of parameters and the computational load without significantly sacrificing accuracy. It is highly resistant to overfitting on small datasets thanks to its efficient structure and allows for rapid fine-tuning. Compared to architectures like InceptionV3 or DenseNet121, it offers significantly faster inferences and a smaller memory footprint, while maintaining high

performance for tasks such as classification and detection. This energy and computational efficiency make it superior in real-time environments or on limited hardware. The contribution of this article can be summarized in the following points:

- We apply MobileNetV2 with transfer learning and fine-tuning to a multi-class brain tumor MRI dataset.
- We provide a detailed comparative analysis with existing models (EDCNN, InceptionV3, VGG16 ensemble, BrainNet, Robust CNN+U-Net, generic CNN).
- We analyze misclassification causes (especially meningioma vs. pituitary and glioma) and discuss the practical advantages of MobileNetV2 for resource-constrained environments.
- We report not only accuracy but also sensitivity, specificity, precision, F1-score, confusion matrices, ROC curves, and efficiency considerations.

2. METHOD

The methodology section explains the data (dataset, preprocessing), model findings, and evaluation metrics used in this study. Every step is crucial to achieving high accuracy in brain tumor classification.

2.1. Dataset description

For this research, we used a freely accessible MRI image dataset called Msoud [1], downloadable from the Kaggle platform. It comprises images divided into four categories: glioma, meningioma, pituitary tumor, and healthy brain illustrated in Figure 1. The data are organized into three main subsets: training, validation, and testing. The initial splitting of the dataset was performed by the authors, with 4,480 images for training, 1,120 images for validation, and 1,600 images for testing, each class being proportionally represented. This predefined partitioning was implemented by the creators of the original dataset to ensure fair evaluation and avoid any risk of cross-contamination between the sets. The exact distribution between the training, validation, and test data was retrospectively analyzed to ensure adequate balance Figures 4-6. No data leakage was observed, as the partitioning was fixed by the original authors.

Figure 4. Example of a prediction percentile distribution of train data

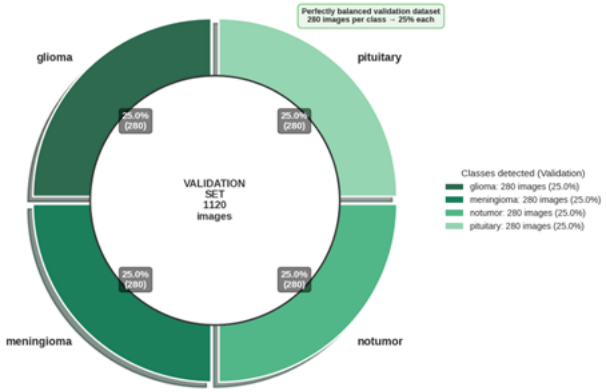

Figure 5. Example of a prediction percentile distribution of validation data

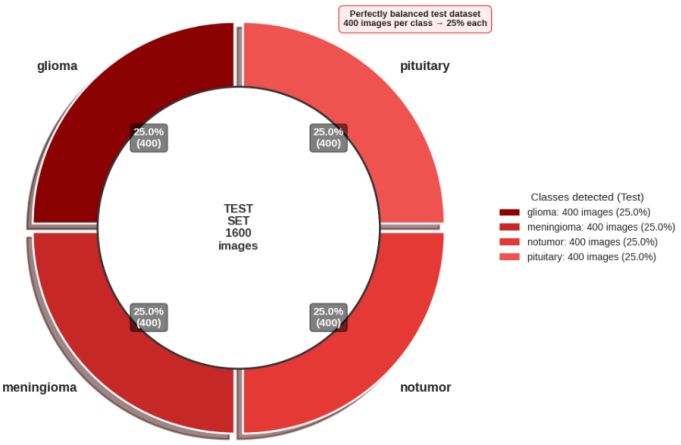

Figure 6. Example of a prediction percentile distribution of test data

2.2. Preprocessing

We have prepared a medical image processing pipeline (brain MRI) for a convolutional neural network (MobileNetV2), using augmentation techniques to prevent overfitting and improve generalization. We first defined several essential parameters for preparing the training data for this model dedicated to analyzing medical brain images. The variable IMG_SIZE is set to 224 by 224 pixels, a standard resolution suitable for most convolutional neural network architectures such as ResNet or VGG, and particularly well-suited to our subject, MobileNetV2. The batch size, BATCH_SIZE, is set to 32, meaning that 32 images will be processed simultaneously in each training iteration, a good compromise between gradient stability and memory usage. Finally, the parameter NUM_CLASSES is set to 4, corresponding to the four diagnostic categories that the model must distinguish: glioma, meningioma, pituitary tumor, and no tumor. Next, the code creates an augmented image generator, a crucial technique for artificially enriching the training set and improving model robustness. Each loaded image is first normalized with a rescale factor of 1:255, converting pixel values from the range of 0 to 255 to the range of 0 to 1, thus accelerating learning convergence. The generator then applies various random transformations: rotation limited to 20 degrees in either direction, horizontal and vertical shifts of up to 10 percent of the image's width or height, zooms of up to 10 percent, and random horizontal flips. All these variations simulate the differences in orientation and position that can be encountered in clinical practice. Finally, validation is set at 20 percent, meaning that the generator will automatically reserve this proportion of images to evaluate model performance during training, with the remaining 80 percent being used for actual learning. We summarize the approach to data preprocessing and augmentation in the following points:

- Resizing: to 224×224 pixels (MobileNetV2 input size).
- Normalization: pixel values scaled to [0,1].
- Data augmentation (training only): random rotation (±20°), horizontal/vertical shifts (±10%), zoom (±10%), horizontal flips.
- Validation split: 20% of training data used for validation.

2.3. MobileNetV2 Architecture

The first layer represents a 224x224 pixel input image with 3 color channels (RGB), followed by a second layer specifying the use of the MobileNetV2 architecture as the feature extractor. The third layer performs a GlobalAveragePooling2D that transforms the feature maps into a feature vector. The fourth layer is a dense layer of 128 neurons with ReLU activation, followed by a 30% dropout layer to regularize the model and prevent overfitting. The sixth layer is a dense layer of 4 neurons with Softmax activation, which ultimately produces probabilities for the four classes: Glioma, Meningioma, Pituitary Tumor, and Healthy.

MobileNetV2 uses depthwise separable convolutions and inverted linear bottlenecks, reducing parameters and computational cost. Our custom head : GlobalAveragePooling2D → Dense (128, ReLU) → Dropout (30%) → Dense (4, Softmax). After presenting the architecture of the MobileNetV2 model, we present below a flow diagram of the study proposed in this document.

Figure 7 shows the proposed MobileNetV2 architecture, consisting of feature extraction, GlobalAveragePooling2D, Dense, Dropout, and Softmax layers for four-class classification. Figure 8 presents the overall experimental workflow, from data preparation and model training to evaluation and final results.

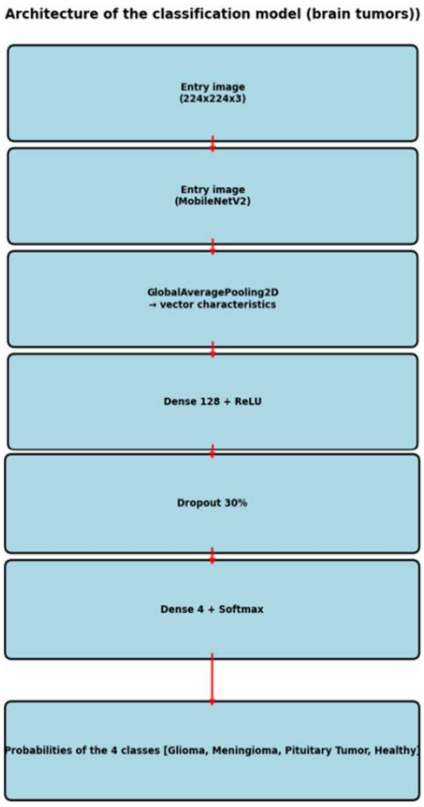

Figure 7. Functional details of the basic MobileNetv2 architecture

Figure 8. Complete flow diagram

2.4. Operational model

The overall architecture of the operational model for the MobileNetV2 that is utilized for the classification of brain tumors based on MRI images is presented in Figure 9. The procedure can be divided into three primary stages: data pre-processing, MobileNetV2 architecture, and final classification identification.

- Data augmentation: Data augmentation was applied to the training set to improve model generalization by artificially expanding it. Techniques used included horizontal and vertical flips, as well as random rotations within a ±20° range, to simulate variations in image acquisition position.
- Image Resizing: All of the images are resized so that they conform to the specifications that MobileNetV2 requires for input (for example, 224 by 224 pixels).
- Normalisation: Images are normalised to ensure that the intensities of each pixel are consistent for effective learning. As input, the images that have been preprocessed are introduced into the MobileNetV2 model.
- In the second stage of the process, the MobileNetV2 architecture is made up of several convolutional layers that are interspersed with batch normalisation, ReLU activations, and max pooling layers. One of the most well-known characteristics of the MobileNetV2 model is its effectiveness in feature extraction through the utilisation of depthwise separable convolutions:

a) Feature Extraction: Extracting features at multiple levels is what convolutional layers are all about.
b) Batch Normalization: Batch Normalization is a process that normalizes the outputs of layers to guarantee stable and effective training.
c) ReLU6 activation: The ReLU activation increases the amount of non-linearity while preserving the numerical predictability.
d) Max Pooling: Max Pooling Layers can reduce the spatial dimensions while still preserving the essential characteristics.
e) Dense layers: To pass on the features that were extracted from these layers to the dense layers, they are first flattened.

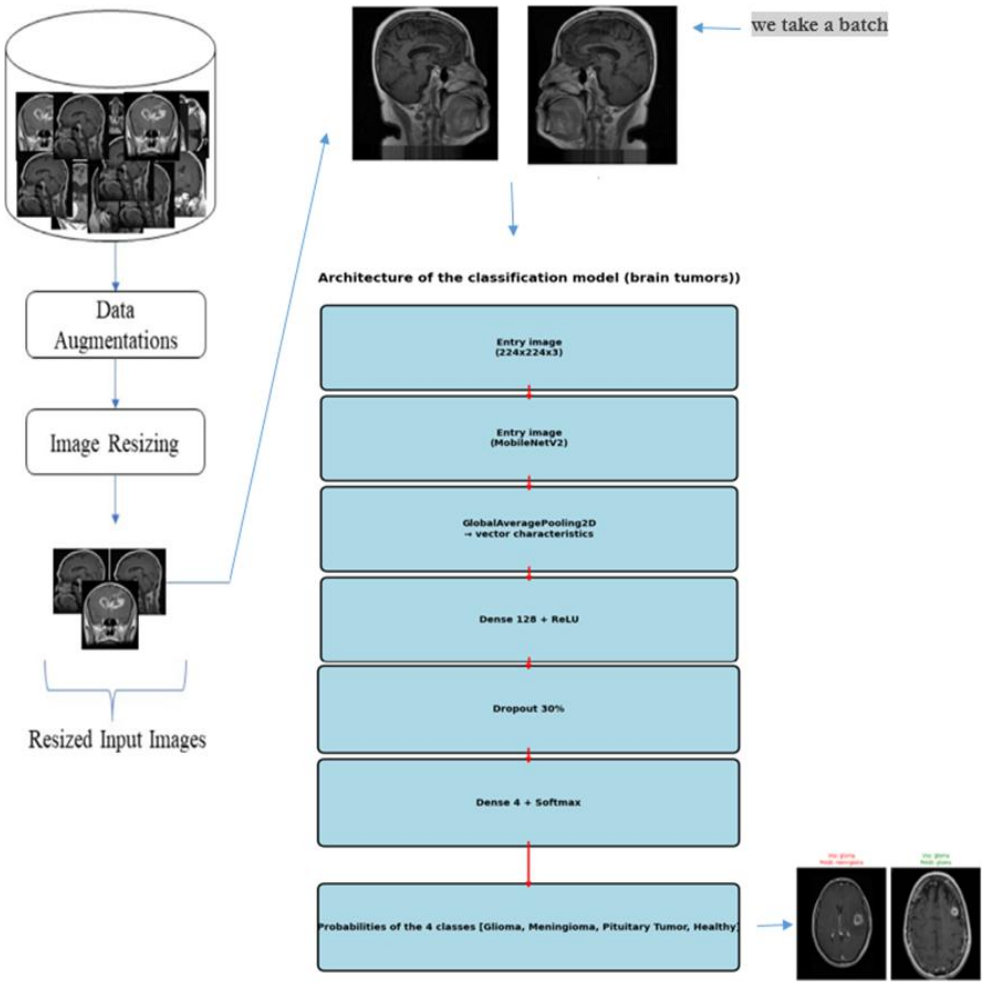

Figure 9. Detailed overview of the MobilenetV2 architecture

2.5. Transfer learning and fine-tuning strategy

– Transfer learning: We loaded MobileNetV2 pre-trained on ImageNet, froze all base layers, and trained only the newly added top layers for 10 epochs (optimizer: Adam, learning rate: 0.001, batch size: 32).
– Fine-tuning: We unfroze the last 20 layers of the base model and retrained with a lower learning rate (0.0001) for 5 epochs to adapt features to the medical domain.

This strategy avoids overfitting and leverages general features from natural images.

2.6. Evaluation metrics

In this study, the performance of the proposed brain tumor classification model is analyzed using multiple evaluation metrics. These metrics are defined and explained further in this document. We used accuracy, precision, recall (sensitivity), F1-score, specificity, confusion matrix, and ROC-AUC. These are critical in medical imaging to balance false positives and false negatives.

2.6.1. Accuracy

Accuracy refers to the proportion of samples that a model classifies correctly. To compute it, we define true positives (TP) and true negatives (TN) as the samples correctly classified, while false positives (FP) and false negatives (FN) correspond to misclassified samples. In a binary medical classification task for example, distinguishing healthy individuals from sick patients a TP occurs when a sick patient is correctly identified as sick, and a TN when a healthy person is correctly identified as healthy. Conversely, a FP corresponds to a healthy person mistakenly classified as sick, and a FN to a sick patient incorrectly predicted as healthy. The formula below shows how to calculate the accuracy score on a given dataset:

$$accuracy = \frac{(\mathrm{TP}+\mathrm{TN})}{(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN})}$$

This metric can be computed during the training phase, and Figure 10 illustrates a typical example. Unlike the loss function, which usually decreases, the accuracy curve follows a logarithmic increase.

As with the loss curves, comparing the accuracy achieved on the training data with that obtained on the test data makes it possible to detect overfitting or underfitting.

However, this accuracy metric has certain limitations, as it provides only a global score. Returning to the example of binary classification in the medical field, accuracy alone does not indicate how well the model identifies healthy subjects or, conversely, sick patients. To address this, we can examine the sensitivity and specificity scores. Sensitivity measures the model's ability to correctly detect sick patients, while specificity reflects its ability to correctly identify healthy individuals. These two metrics allow the model to be fine-tuned according to the specific needs of the user. The equations below detail how these two scores are calculated.

$$\text{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

$$\text{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$$

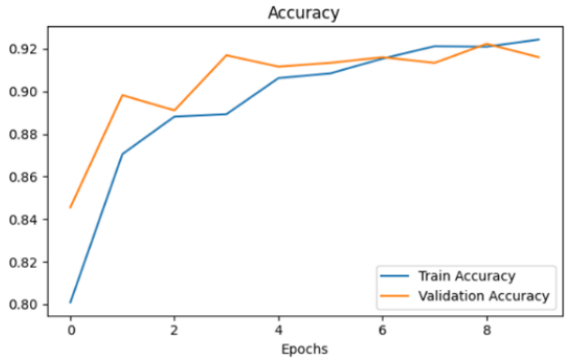

Figure 10. Accuracy learning curve on the training and test dataset

2.6.2. Precision

To minimize false positives, precision evaluates a model's positive predictive accuracy by calculating the proportion of correctly predicted positive samples among all samples predicted as positive. Its mathematical expression is:

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$$

In the context of tumor classification, precision plays a crucial role in reducing false-positive diagnoses. For instance, mistakenly identifying a tumor on a standard brain scan can lead to unnecessary medical interventions and significant distress for the patient. A model with high precision is therefore more reliable when predicting the presence of a tumor.

2.6.3. Recall/sensitivity

Recall, also commonly referred to as sensitivity, measures the model's ability to correctly identify positive samples. It is defined by the following formula:

$$\text{Recall} = \text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}}$$

In medical imaging, recall is arguably the most critical metric to consider, since failing to detect a tumor (resulting in a false negative) can have fatal consequences for a patient. A high recall ensures that the model captures nearly all genuine tumor cases, making it just as important as precision. This analysis places strong emphasis on recall because the inherent cost of missing a tumor is far greater and much more severe than that of a false positive.

2.6.4. F1-score

The F1-score computes the weighted average of precision and recall using their harmonic mean, combining both metrics into a single value. It can be expressed as follows:

$$\text{F1} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision+Recall}}$$

In classification tasks such as brain tumor detection, where datasets often suffer from class imbalance, the F1-score provides a more meaningful assessment of model performance. This metric is particularly useful when there is a trade-off between precision and recall, as it gives equal weight to both. A high F1-score indicates that the model not only correctly identifies true positives (e.g., actual tumors) but also avoids generating a large number of false positives.

2.6.5. Specificity

Specificity is a key evaluation metric in medical diagnostic systems. It quantifies the proportion of correctly identified negative cases TN relative to all actual negative cases. Specificity is formally defined as:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}}$$

High specificity is especially crucial in medical image classification to prevent false positives, which can lead to unnecessary clinical procedures, patient anxiety, and higher healthcare expenses.

3. RESULTS AND DISCUSSION

The performance of the proposed system was evaluated on the Google Colab platform based on the factor of the evaluation matrix and the model performance.

3.1. Evaluation metrics

The performance of the proposed model is evaluated using various metrics, including accuracy, precision, recall, and the F1 score. Our model is compared to a set of models Table 1 [24]. The comparison of our model with existing models, according to these different metrics, is presented in Figures 11 and 12 [25].

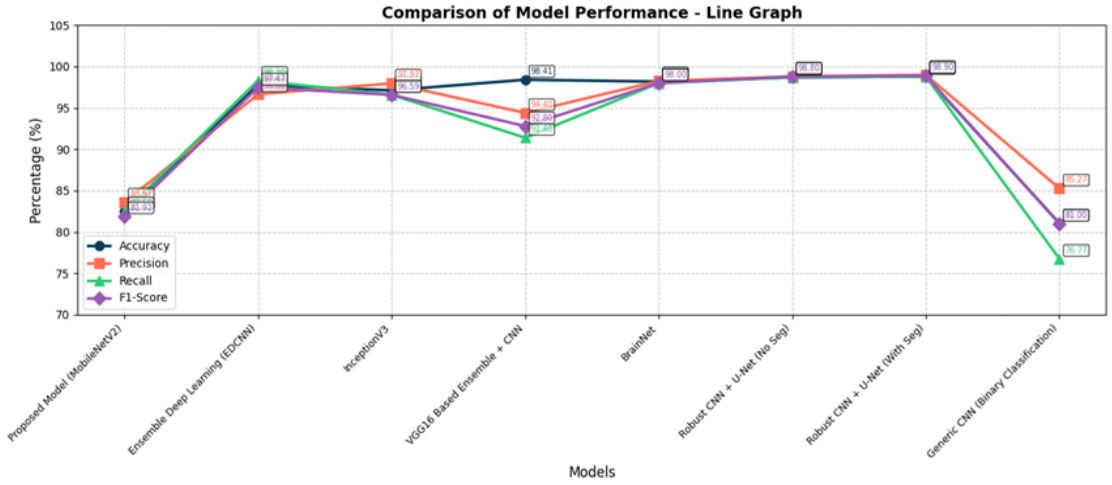

Figure 11. Comparison of model performance, line graph

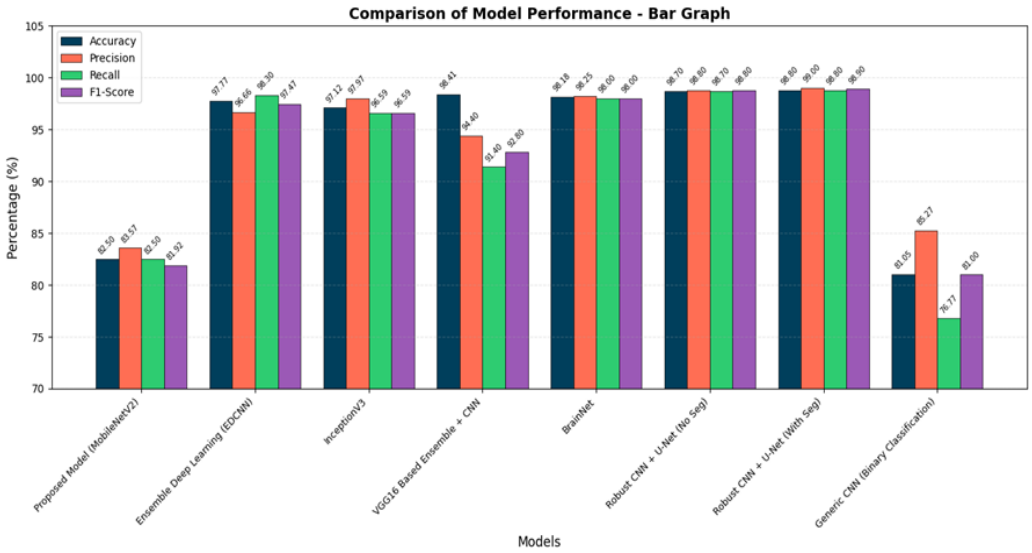

Figure 12. Comparison of model performance, bar graph

Table 1. Comparison of the accuracy of our model with other existing models

Model	F1-Score (%)	Accuracy (%)
Proposed model (MobileNetV2)	86.0	86.56
Ensemble deep learning (EDCNN)	97.0	97.77
InceptionV3	96.0	97.12
VGG16 based ensemble+CNN	92.0	98.41
BrainNet	98.0	98.18
Robust CNN+U-Net (no segmentation)	98.0	98.70
Robust CNN+U-Net (with segmentation)	98.0	98.80
Generic CNN (binary classification)	81.0	81.05

3.2. Model performance

The model was trained on 4,480 MRI images divided into four classes: glioma, meningioma, pituitary tumor, and no tumor. After initial training, validation accuracy reached 88%; after fine-tuning, test accuracy was 86.56% Table 1. The macro-average AUC was 0.9645 Figure 25, indicating excellent discriminability.

In this study, and to support our choice of the MobileNetV2 model as the best model for the classification and detection of brain tumors, we developed several models, including a CNN model and others such as VGG16, VGG19, ResNet50, Xception, and EfficientNet. We monitored the performance of each model, sometimes changing the number of epochs for some. We confirmed that the MobileNetV2 model provides the best performance with a simple architecture [26]. The performance of the models used in this study is presented in the graphs, using the CNN model architecture. Figure 13 compares the precision achieved by each model, while Figure 14 illustrates the overall performance trend across different architectures. Figure 15 presents the architecture of the proposed CNN model used for comparison, and Figure 16 compares the accuracy and F1-score of all models before and after fine-tuning.

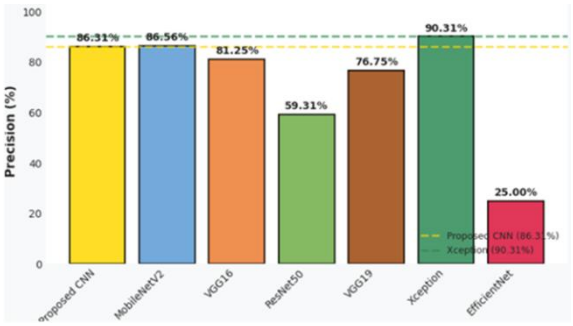

Figure 13. precision by model

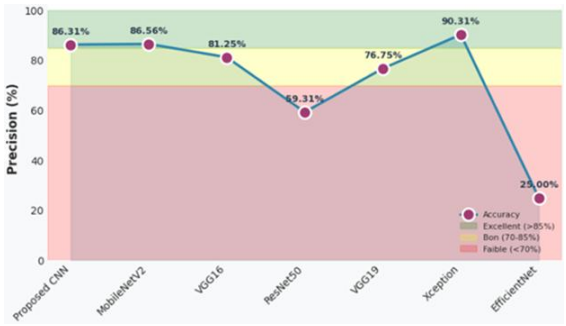

Figure 14. performance trend

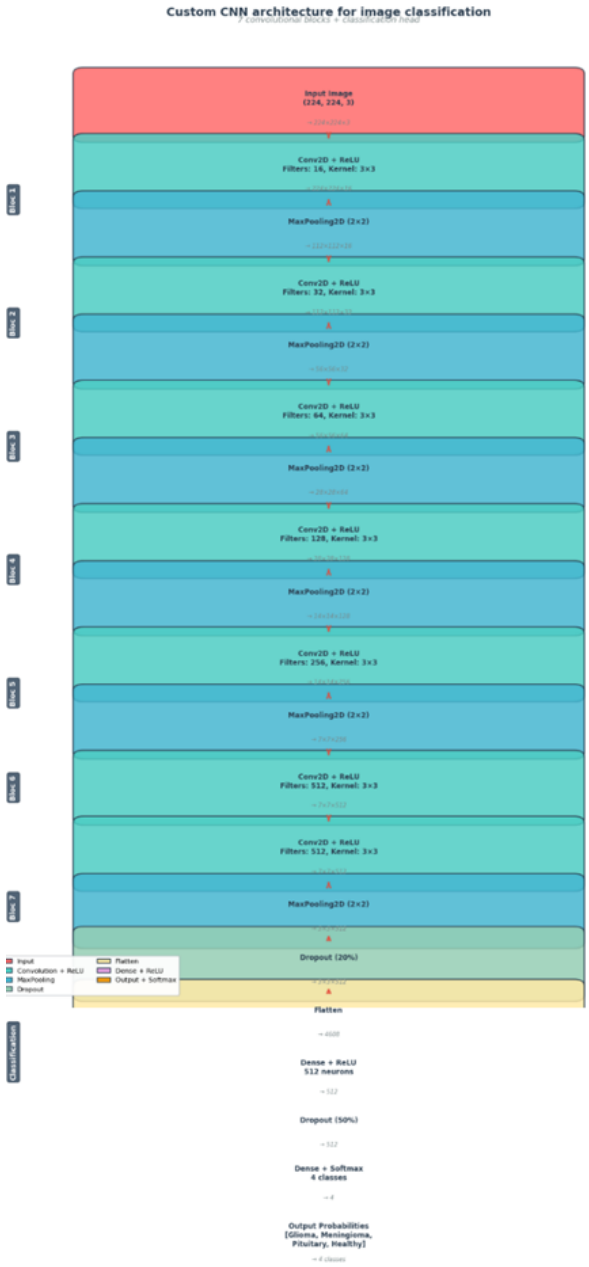

Figure 15. Proposed CNN model architecture

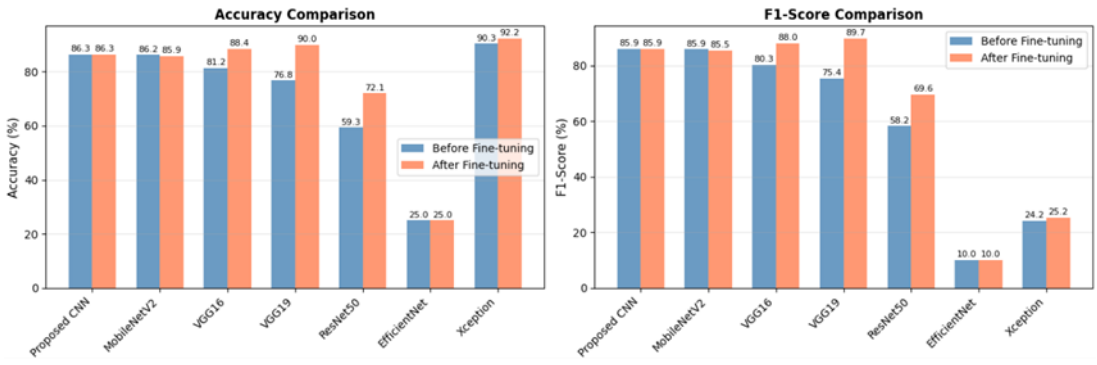

Figure 16. Model performance comparison

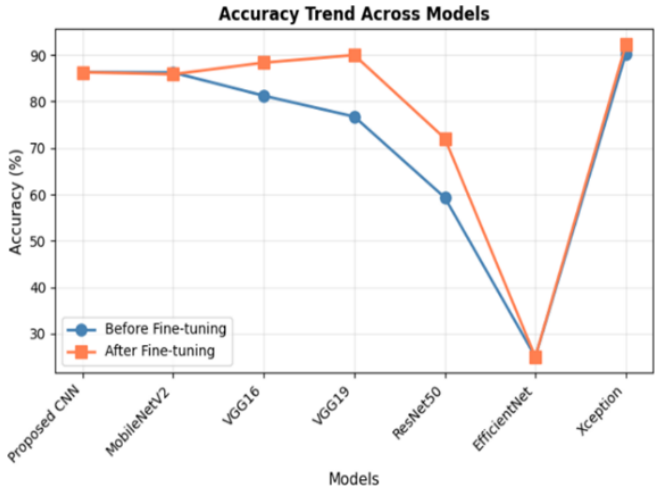

Figure 17. Accuracy trend across models

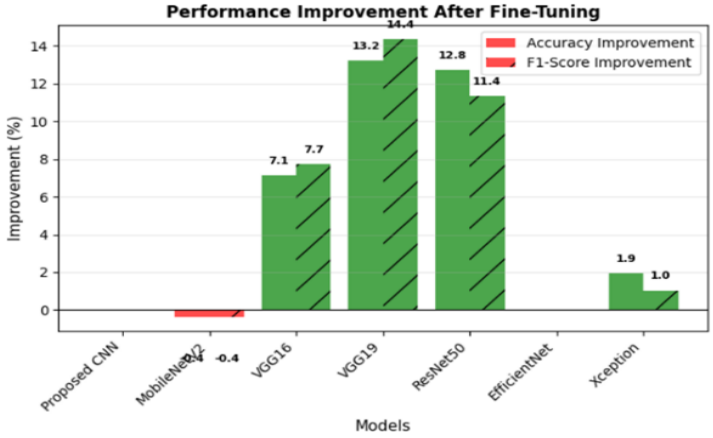

Figure 18. Performance improvement after fine-tuning

3.2.1. Training and validation accuracy

The progression of the training and validation accuracy across all epochs is shown in Figure 19, demonstrating consistent improvement in the model's ability to classify brain tumor images accurately. According to Figure 19, the x-axis represents epochs, while the y-axis indicates accuracy, which appears to range from 0 to 1. The blue curve, corresponding to Training Accuracy, starts at around 0.6 and rises rapidly to approximately 0.95 at epoch 8. The orange curve, representing Validation Accuracy, starts at a slightly lower level (around 0.55) and follows a similar trend, reaching nearly 0.93 at the end. Both curves increase sharply between epochs 0 and 2, then their increase gradually slows. Training accuracy remains consistently greater than or equal to validation accuracy across all epochs. At epoch 8, the two curves converge, with a very small difference (around 0.02), indicating good generalization of the model. No significant overfitting is observed, as the two curves evolve in parallel and remain close. This trend suggests that the model learns effectively without excessively memorizing training data.

3.2.2. Training and validation loss

This line graph illustrates the evolution of the loss function of a machine learning model over training epochs, as shown in Figure 20 which presents the training and validation loss curves. From 0 to 8 in steps of 2 the x-axis represents the epochs, while the y-axis indicates the loss, which appears to decrease on a scale (generally from 0 to 1 or higher). The blue curve, representing the loss on the training set (Train Loss), starts at a high value (around 0.7) and decreases rapidly to around 0.1 at epoch 8. The orange curve, representing the loss on the validation set (Validation Loss), starts at a slightly higher level (around 0.8) and follows a similar decline, reaching approximately 0.15 at the end. Both curves drop sharply between epochs 0 and 2, and then their decline gradually slows. The training loss remains consistently lower than the validation

loss across all epochs, which is expected. From epoch 4 onward, the difference between the two curves stabilizes and remains very small (approximately 0.03 to 0.05), indicating the absence of significant overfitting. The convergence of the two curves toward low values suggests that the model learns correctly and generalizes well to unseen data.

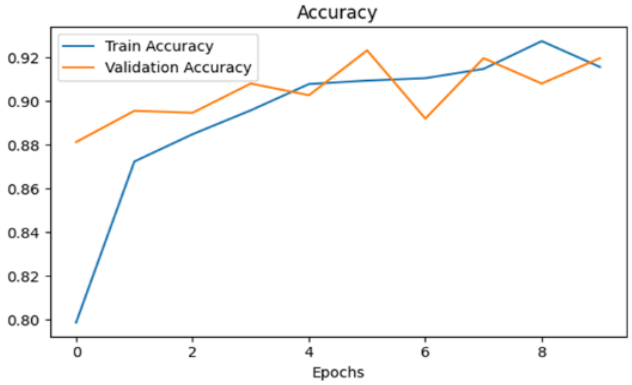

Figure 19. Training and validation accuracy over all epochs

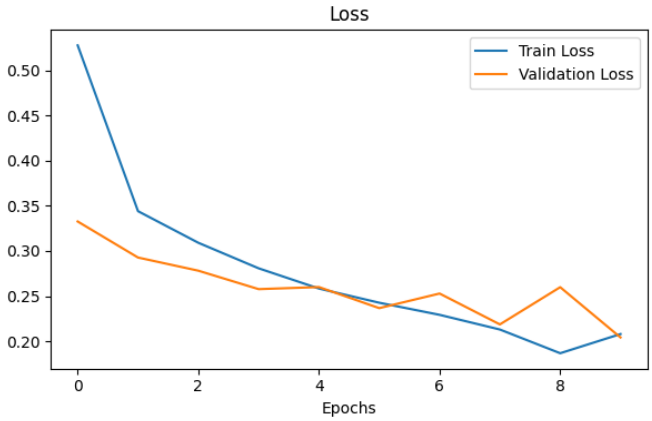

Figure 20. Training and validation loss over all epochs

3.2.3. Training and validation accuracy with fine-tuning techniques

This line graph, shows in Figure 21, illustrates the training and validation accuracy curves with fine-tuning over all training epochs the evolution of the accuracy of a machine learning model as a function of training epochs, here ranging from 0 to 4 in increments of 0.5. The x-axis represents the epochs, while the y-axis indicates the accuracy, which ranges from approximately 0.750 to 0.925. Only one curve is visible (Train Accuracy), as the accuracy on the validation set is not shown or is absent from this graph. The blue curve starts at an accuracy of approximately 0.775 at epoch 0, then increases gradually but unevenly throughout the training. Plateaus and slight fluctuations are observed: the accuracy stagnates around 0.800 between epochs 0.5 and 1.0, then progresses to approximately 0.875 at epoch 2.5. After a slight dip around epoch 3.0, accuracy rises again to reach its maximum of approximately 0.915 at epoch 4.0. The curve is not perfectly smooth, suggesting some variability in learning.

3.2.4. Training and validation loss with fine-tuning techniques

This line graph, this line graph, shown in Figure 22, illustrates the training and validation loss curves with fine-tuning over all training epoch illustrates the evolution of the loss function of a machine learning model over epochs, from 0 to 4 in steps of 0.5. The x-axis represents the epochs, while the y-axis indicates the loss, the scale of which is not explicitly given but appears to decrease in a classic manner. Two curves are present: the blue curve for the loss on the training set (Train Loss) and the orange curve for the loss on the validation set (Validation Loss). Both curves start at similar values, probably around 0.5 to 0.6, then decrease rapidly between epochs 0 and 1. From epoch 1 onward, the training loss continues to decrease

gradually and stabilizes at a low value (around 0.1) by epoch 4, while the validation loss follows a similar trend but remains slightly higher. A very small difference is observed between the two curves over most of the training period, indicating good generalization of the model without significant overfitting. The convergence of the two losses towards low and stable values suggests that the training is effective and that the model has accurately captured the underlying structures of the data.

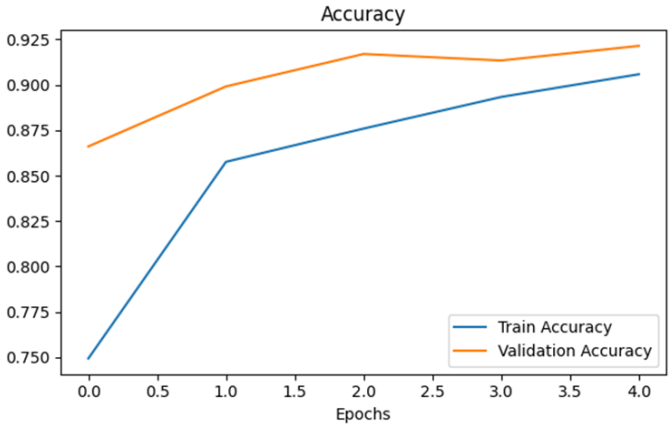

Figure 21. Training and validation accuracy over all epochs using fine tuning techniques

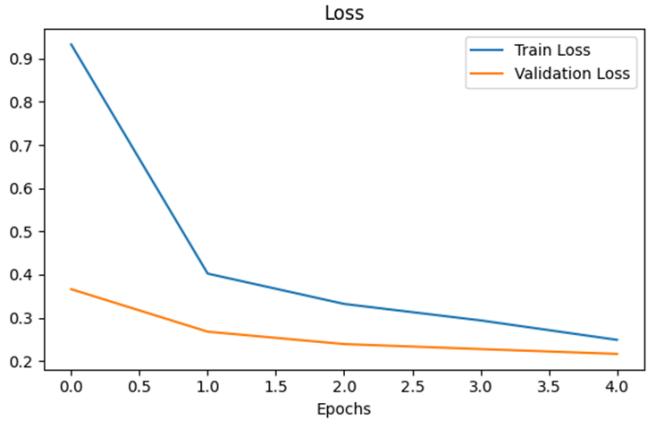

Figure 22. Training and validation loss over all epochs with Fine-Tuning techniques

3.2.5. Confusion matrix and normalized confusion matrix heatmap

To evaluate the performance of the proposed brain model, a confusion matrix was calculated, the results of which are summarized in the figures below. These figures provide a detailed overview of the model's effectiveness in classifying tumors into four categories: glioma, meningioma, no tumor, and pituitary tumor. Figures 23 and 24 present the normalized confusion matrix heatmaps before and after fine-tuning, highlighting the classification performance for each class. Figure 25 shows the multi-class ROC curves, demonstrating the model's strong discriminative capability across all classes. We have shown all the results before and after fine-tuning.

This confusion matrix allows us to evaluate the performance of a classification model on four types of brain tumors (glioma, meningioma, no tumor, pituitary adenoma). The rows represent the actual classes (ground truth) and the columns the classes predicted by the model. The main diagonal groups the correct predictions: thus, 276 gliomas are correctly identified, while 69 are confused with meningiomas, 20 are labeled as "no tumor," and 35 as pituitary adenomas. For the "meningioma" class, 254 diagnoses are correct, but 22 are classified as gliomas, 14 as "no tumor," and, most notably, 110 as pituitary adenomas, indicating significant confusion with the latter category. The "no tumor" class shows excellent results with 392 correct classifications and only 2, 3, and 3 errors, respectively. Finally, the "pituitary adenoma" class achieved the

highest score with 398 correct predictions, only one error for glioma, one for meningioma, and none for "no tumor." A color bar (dark red for 350, white for 100, dark gray for 50) illustrates the frequency range. Overall, the model excels for "no tumor" and "pituitary adenoma," but shows a notable weakness in distinguishing meningiomas from pituitary adenomas.

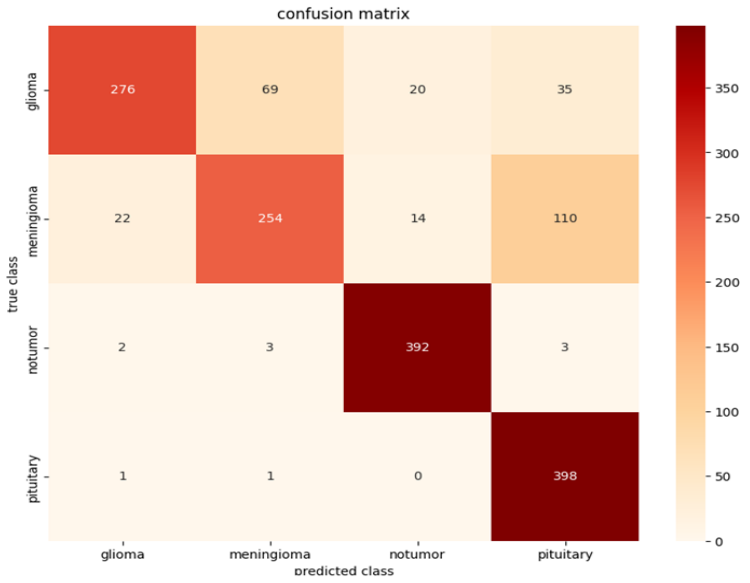

Figure 23. Normalized confusion matrix heatmap of the model without fine-tuning

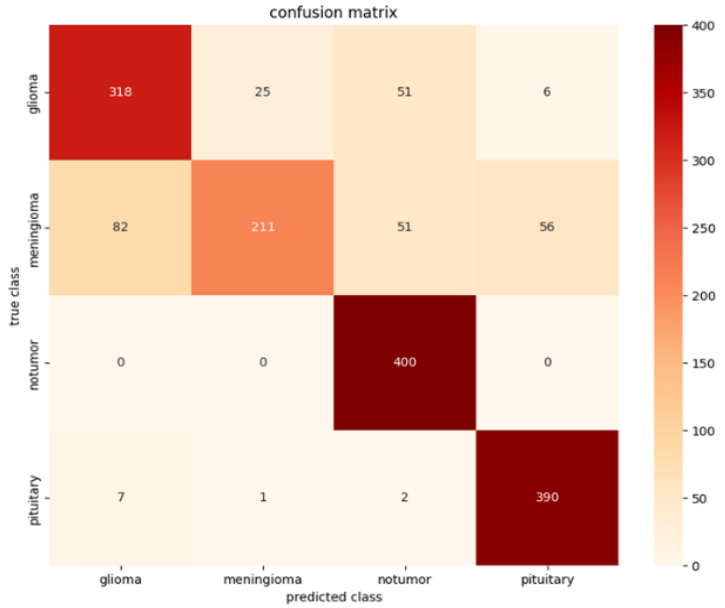

Figure 24. Normalized confusion matrix heatmap of the model with fine-tuning

This confusion matrix evaluates the performance of a medical classification model in distinguishing four types of brain lesions: glioma, meningioma, not tumor, and pituitary adenoma. Along the main diagonal, we observe the correct classifications: 318 gliomas correctly identified, 211 meningiomas, 400 "not tumor" (a perfect score), and 390 pituitary adenomas. The most significant errors concern glioma, with 25 cases confused with meningioma, 51 with "not tumor," and 6 with pituitary adenoma. Meningioma is the most misclassified: only 211 correct predictions out of 400, with 82 errors classified as glioma, 51 as "not tumor," and 56 as pituitary adenoma. The "no tumor" class is perfectly recognized (no errors in either entry or

output), while the pituitary adenoma makes only 7 errors in glioma, 1 in meningioma, and 2 in "no tumor." A color bar ranging from white (0) to dark brown (400) allows visualization of the frequency range. Overall, the model excels for cases without tumors and pituitary adenomas, but shows a notable weakness in distinguishing meningiomas from other classes, particularly gliomas and pituitary adenomas.

Before fine-tuning Figure 23, significant confusion occurred between meningioma and pituitary adenoma (110 misclassifications). After fine-tuning Figure 24, this confusion reduced, but meningioma remained the hardest class (only 211 correct out of 400). Possible reasons:

– Meningiomas can appear similar to pituitary tumors in certain MRI planes.
– Class imbalance or intrinsic variability in the dataset. The “no tumor” class achieved perfect classification after fine-tuning (400/400).

This graph illustrates the ROC curves for a multi-class classification with four distinct classes, where each curve represents the trade-off between the true positive rate (TPR) and the false positive rate (FPR). The areas under the curve (AUC) are excellent for all classes: 0.921 for class 0, 0.945 for class 1, 0.997 for class 2, and 0.995 for class 3, all well above the reference diagonal of the random classifier (AUC = 0.5). Classes 2 and 3 achieve near-perfect performance, while classes 0 and 1 remain very strong. The macro-mean AUC, calculated at 0.9645, confirms that the model discriminates very well overall across all categories, with only a slight margin for improvement for classes 0 and 1.

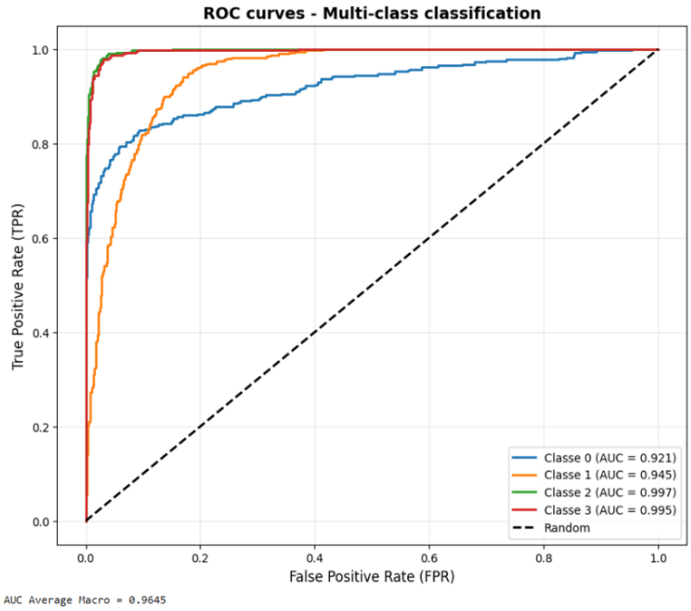

Figure 25. Graph presents the ROC (Receiver Operating Characteristic) curves for a classification task involving the four classes (Class 0 to 3)

4. CONCLUSION

In this article, we propose a deep learning-based approach for the automatic classification of brain tumors from MRI images, using the lightweight MobileNetV2 model combined with transfer learning and fine-tuning techniques. The study was conducted on a public dataset comprising three reference databases (Figshare, SARTAJ, and BR35H), covering four classes: glioma, meningioma, pituitary tumor, and no tumor.

The preprocessing steps (resizing, normalization, and data augmentation through rotation and deformation) improved the model's robustness and prevented overfitting. The training curves show good generalization, with a validation accuracy of approximately 88% after initial training and 87.5% after fine-tuning, with no significant difference from the training accuracy. The confusion matrices before and after fine-tuning reveal a notable improvement in discrimination between classes, particularly for meningiomas and cases without tumors. The model achieves an overall accuracy of 86.56% across all test data, with a macro average AUC of 0.9645, confirming its potential as a reliable diagnostic aid.

Compared to more resource-intensive models such as VGG16 or ResNet50, MobileNetV2 offers an excellent balance between performance and lightweight design, making it deployable on modest hospital

infrastructures. Residual weaknesses (confusion between gliomas and meningiomas) can be mitigated by using more balanced data or networks based on attention mechanisms.

For future work, we plan to evaluate this model on images from different MRI scanners, integrate an automatic segmentation step, and perform clinical validation of the tool. Furthermore, the application of other transfer learning architectures (DenseNet, etc.) will be explored to further improve classification accuracy. Artificial intelligence, through its speed and consistency, does not replace the doctor but provides him with systematic analysis and valuable advice, thus accelerating the diagnosis of brain tumors and improving patient care.

FUNDING INFORMATION

Authors state that no funding was involved in the preparation of this manuscript.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Aziz Srai	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

The dataset that supports the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset.

REFERENCES

[1] M. Nickparvar, "Brain tumor MRI dataset," Kaggle. Accessed: May 22, 2025. [Online]. Available: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset
[2] J. Cheng, "Brain tumor dataset," Figshare. Accessed: May 22, 2025. [Online]. Available: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427/5
[3] A. Hamada, "Brain tumor detection," Kaggle. Accessed: May 22, 2025. [Online]. Available: https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection
[4] H. ZainEldin *et al.*, "Brain tumor detection and classification using deep learning and sine-cosine fitness grey wolf optimization," *Bioengineering*, vol. 10, no. 1, p. 18, Dec. 2023, doi: 10.3390/bioengineering10010018.
[5] B. Lei *et al.*, "Neuroimaging retrieval via adaptive ensemble manifold learning for brain disease diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1661–1673, Jul. 2019, doi: 10.1109/JBHI.2018.2872581.
[6] K. Gasmi *et al.*, "Enhanced brain tumor diagnosis using combined deep learning models and weight selection technique," *Frontiers in Neuroinformatics* , vol. 18, Nov. 2024, doi: 10.3389/fninf.2024.1444650.
[7] S. Padmakala and S. Uma Maheswari, "Deep learning-based MRI analysis for brain tumor detection: insights from ResNet and DenseNet models," in *2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, IEEE, Nov. 2024, pp. 1620–1626. doi: 10.1109/ICECA63461.2024.10800894.
[8] S. M. Rasa *et al.*, "Brain tumor classification using fine-tuned transfer learning models on magnetic resonance imaging (MRI) images," *Digital Health*, vol. 10, Jan. 2024, doi: 10.1177/20552076241286140.
[9] K. Gencer and G. Gencer, "Hybrid deep learning approach for brain tumor classification using EfficientNetB0 and novel quantum genetic algorithm," *PeerJ Computer Science*, vol. 11, p. e2556, Jan. 2025, doi: 10.7717/PEERJ-CS.2556.
[10] A. Hamza and R. Damaševičius, "Deep learning for brain tumor segmentation and classification: a systematic review of methods and trends," *Computers, Materials and Continua*, vol. 86, no. 1, pp. 1–41, 2026, doi: 10.32604/cmc.2025.069721.
[11] N. Netshamutshedzi, R. Netshikweta, J. C. Ndogmo, and I. C. Obagbuwa, "A systematic review of the hybrid machine learning models for brain tumour segmentation and detection in medical images," *Frontiers in Artificial Intelligence*, vol. 8, Sep. 2025, doi: 10.3389/frai.2025.1615550.
[12] K. Enda *et al.*, "Transfer learning strategies for pathological foundation models: a systematic evaluation in brain tumor classification." Apr. 08, 2025. [Online]. Available: http://arxiv.org/abs/2501.11014.

[13] M. A. Khan *et al.*, "Transfer learning for accurate brain tumor classification in MRI: a step forward in medical diagnostics," *Discover Oncology*, vol. 16, no. 1, p. 1040, Jun. 2025, doi: 10.1007/s12672-025-02671-4.
[14] S. Wang and M. Li, "Comparative analysis of machine learning techniques on the BraTS dataset for brain tumor classification," *Frontiers in Oncology*, vol. 15, Sep. 2025, doi: 10.3389/fonc.2025.1596718.
[15] M. S. Alam *et al.*, "Automatic human brain tumor detection in mri image using template-based k means and improved fuzzy c means clustering algorithm," *Big Data and Cognitive Computing*, vol. 3, no. 2, pp. 1–18, May 2019, doi: 10.3390/bdcc3020027.
[16] A. Raza *et al.*, "A hybrid deep learning-based approach for brain tumor classification," *Electronics*, vol. 11, no. 7, p. 1146, Apr. 2022, doi: 10.3390/electronics11071146.
[17] Z. Rasheed *et al.*, "Brain tumor classification from MRI Using image enhancement and convolutional neural network techniques," *Brain Sciences*, vol. 13, no. 9, p. 1320, Sep. 2023, doi: 10.3390/brainsci13091320.
[18] L. P. Clarke *et al.*, "MRI segmentation: methods and applications," *Magnetic Resonance Imaging*, vol. 13, no. 3, pp. 343–368, Jan. 1995, doi: 10.1016/0730-725X(94)00124-L.
[19] K. M. Iftekharuddin, J. Zheng, M. A. Islam, and R. J. Ogg, "Fractal-based brain tumor detection in multimodal MRI," *Applied Mathematics and Computation*, vol. 207, no. 1, pp. 23–41, 2009, doi: 10.1016/j.amc.2007.10.063.
[20] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on MRI brain tumor segmentation," *Magnetic Resonance Imaging*, vol. 31, no. 8, pp. 1426–1438, Oct. 2013, doi: 10.1016/j.mri.2013.05.002.
[21] P. R. Kumar, K. Bonthu, B. Meghana, K. S. Vani, and P. Chakrabarti, "Multi-class brain tumor classification and segmentation using hybrid deep learning network model," *Scalable Computing: Practice and Experience*, vol. 24, no. 1, pp. 69–80, Apr. 2023, doi: 10.12694/scpe.v24i1.2088.
[22] N. Chakrabarty, "Brain MRI images for brain tumor detection: dataset used in the study "brain tumor analysis using deep learning and VGG-16 Ensembling learning approaches," Kaggle. Accessed: Jun. 25, 2024. [Online]. Available: https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection
[23] S. S. Shouma, M. R. Miah, S. Ahmed, M. A. Islam, R. Raiyan, and M. S. Ahammad, "Automated classification of brain tumor based on MRI images using deep learning," in *2025 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON)*, IEEE, Nov. 2025, pp. 669–674. doi: 10.1109/BECITHCON69222.2025.11504051.
[24] M. M. Sherif, "Brain Tumor Dataset," Kaggle. Accessed: May 22, 2025. [Online]. Available: https://www.kaggle.com/datasets/mohamedmetwalysherif/braintumordataset
[25] P. Kumar, "Brain MRI," Kaggle. Accessed: May 22, 2025. [Online]. Available: https://www.kaggle.com/datasets/pradeep2665/brain-mri
[26] P. T. V. Huong, L. D. Thinh, T. T. Oanh, T. X. Bach, H. Q. Huy, and T. A. Vu, "A novel approach for detecting diabetic retinopathy using two-stream CNNs model," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 41, no. 1, pp. 200–209, 2026, doi: 10.11591/ijeecs.v41.i1.pp200-209.

BIOGRAPHIES OF AUTHORS

Prof. Aziz Srai He obtained his doctorate in engineering sciences, specialization in computer engineering, at the Mohammadia School of Engineers of the Mohammed V University of Rabat in 2021. His doctoral research focused on the development of models based on deep learning for the detection and classification of brain tumors. He is currently a lecturer at the Faculty of Science and Technology of Al Hociema, Abdelmalek Essaâdi University in Morocco where he teaches programming, machine learning and deep learning. His research interests include applications of artificial intelligence in neuro-oncology. He can be contacted at email: a.srai@uae.ac.ma.