RESEARCH ARTICLE

Automated Brain Tumor Analysis from MRI Using Deep Learning

Aamir Ali[1,*], **Aamir Raza**[2], **Iqra Nazir**[1], **Swetank Mohan**[3], **Khaleelullah Syed**[4], **Nikhat Fatima**[5] **and Misbah Ali**[1]

[1] Department of Computer Science, COMSATS University Islamabad, Sahiwal Campus, Sahiwal 57000, Pakistan
[2] Department of Information Technology and Management, Illinois Institute of Technology, Chicago, IL 60616, United States
[3] The University of Texas Rio Grande Valley, Edinburg, TX 78539, United States
[4] Department of Computer and Information Technology Services Administration and Management, Hellenic American University, Athens 10680, Greece
[5] Department of Computer and Information Technology Services Administration and Management, Concordia University, Mequon, WI 53097, United States

Abstract

Accurate brain tumor classification from MRI remains essential for computer-assisted diagnosis, yet manual interpretation is time-consuming and variable. This study presents an EfficientNet-B0-based convolutional neural network for multi-class classification of glioma, meningioma, pituitary tumors, and no-tumor cases. The model was trained and evaluated on a public MRI dataset of 7023 images using a strict patient-level split to ensure unbiased assessment. A fixed EfficientNet-B0 backbone with a lightweight classification head reduces overfitting while maintaining stable learning. Performance was assessed via accuracy, precision, recall, F1-score, and specificity. The model achieved class-wise (one-vs-rest) accuracies of 96.5% for glioma, 99.1% for no tumor, 95.6% for meningioma, and 97.9% for pituitary tumors, with high specificity across all classes. A controlled comparison against ResNet50 and MobileNetV2 under identical conditions shows that EfficientNet-B0 offers a balanced trade-off between predictive performance and computational cost, achieving competitive accuracy with significantly fewer parameters and faster inference than deeper architectures. This study provides a reproducible evaluation framework, patient-level validation protocol, and systematic backbone comparison to support efficient and reliable deep learning models for multi-class brain tumor classification.

Keywords: brain tumor classification, magnetic resonance imaging (MRI), convolutional neural networks, efficientNet-B0, deep learning, computer-aided diagnosis, medical image analysis.

Submitted: 16 February 2026
Accepted: 19 March 2026
Published: 27 March 2026

Vol. 2, **No.** 1, 2026.
10.62762/BISH.2026.687557

***Corresponding author:**
✉ Aamir Ali
talktomisbah.ali@gmail.com

Citation
Ali, A., Raza, A., Nazir, I., Mohan, S., Syed, K., Fatima, N.,& Ali, M. (2026). Automated Brain Tumor Analysis from MRI Using Deep Learning. *Biomedical Informatics and Smart Healthcare*, 2(1), 62–78.

1 Introduction

Every year, the term brain tumours becomes associated with increasing levels of complexity and mortality due to its associated neurological disorders. An actionable treatment designed to assist patient recovery can be derived from the early discovery and accurate framing of types of brain tumours [1]. Soft tissue contrast with the ability to visualise structures in multiple planes makes Magnetic Resonance Imaging (MRI) the most reliable, non-invasive imaging technique for diagnosis of brain tumours [2]. The process of manual interpretation of the imaging studies is tedious, leading to diagnostic inaccuracies. Computer-Aided Diagnosis (CAD) systems have prompted the use of Artificial Intelligence to increase the accuracy and consistency in assessing tumours, guiding the radiologists [3]. The differentiation of brain tumours is of paramount importance and goes beyond identification. This is due to the fact that various tumours could pose significantly disparate treatment options and differing clinical outcomes.

Rapid advancement in Machine Learning (ML) and Deep Learning (DL) technologies in the last few years have been applied in the field of Computer Aided Diagnosis (CAD). In the past few years, the application of traditional ML algorithms such as Support Vector Machines (SVM), Random Forest (RF) and K-Nearest Neighbours (KNN) have been reported and require handcrafting feature extraction from MRI images. These methods are better, but still are constrained in generalising to different types of tumours as they still rely on manually engineered features [4, 5]. In order to address this, many Deep Learning (DL) algorithms (e.g. Convolutional Neural Networks (CNNs)) have been developed as they are able to learn from images, with less reliance on handcrafted features [6]. Brain MRI images feature complex intensities, heterogeneous textures, and subtle structural differences which characterise a brain image and vary greatly across different scanners using different settings. These differences greatly limit the reliance of manually engineered features to distinguish different overlapping tumours from each other, which may have different viewing attributes.

Medical image-related tasks include lesion segmentation, tumour detection, and disease classification, to which CNNs are already applied [7]. Its hierarchical structure allows it to circumvent manual preprocessing and classification boosts through automatic extraction of pixel level features. Numerous studies have documented the success of various CNN architectures, like AlexNet, VGGNet, and ResNet in numerous medical applications [8]. However, when trained on small medical datasets, these models still experience overfitting and high computational costs.

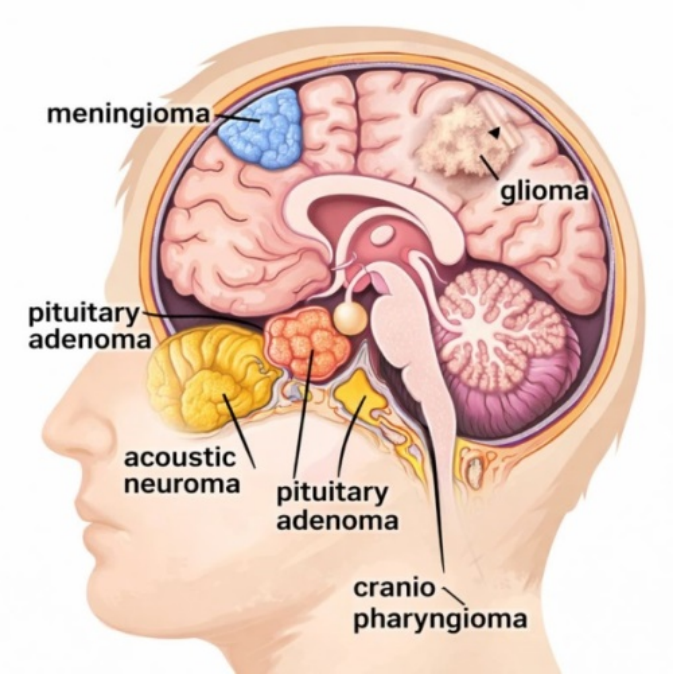

Figure 1. Anatomical illustration showing the typical locations of common brain tumor types within the human brain.

Recent developments in the field of convolutional neural networks have mainly concentrated on balancing the classification accuracy of a model while at the same time minimising the computational resources required to achieve this. With this in mind, some researchers have focused on optimising the 'depth', 'width', and 'resolution' of a model in attempts to strike a suitable trade-off between efficient feature learning and 'parameter' control. Such design considerations are especially relevant in the field of medical imaging, where datasets are small, and overfitting is a major concern. Lightweight architectures combined with appropriate regularisation strategies are, therefore, gaining popularity for successfully classifying brain tumour images. In modern medicine, one of the most challenging tasks is the accurate diagnosis of brain tumours. The management of a patient's entire treatment plan hinges on the timely and accurate identification of the condition. MRI is one of the most frequently used imaging systems for this purpose as it possesses the ability to provide high-resolution imaging of soft tissues while also ensuring that the

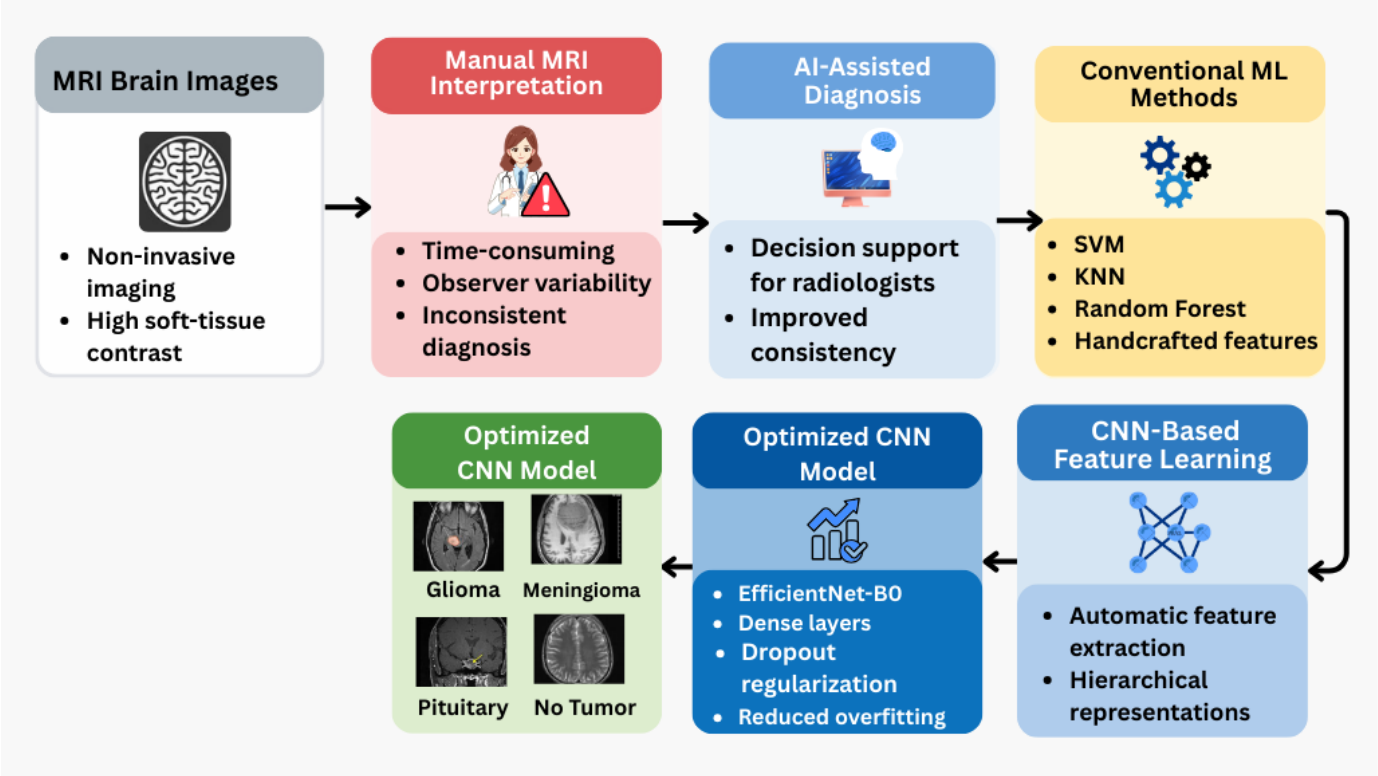

Figure 2. Conceptual overview of the proposed AI-assisted brain tumor classification model using MRI images.

patient is not exposed to harmful ionising radiation. Reliable interpretation of MRI scans therefore plays a central role in supporting clinical decision-making in neuro-oncology. Figure 1 illustrates the anatomical locations of common brain tumor types, including glioma, meningioma, and pituitary adenoma, as typically observed in the human brain [9].

The advanced CNNs, such as EfficientNet have been introduced to address many challenges. It has compound scaling to optimize network depth, width, and resolution and achieve better accuracy with fewer parameters [10]. EfficientNet-based CNNs have proven particularly effective in medical imaging tasks where data availability and computational resources are limited. Moreover, in this architecture, dense and dropout layers are added that enhance feature generalization and mitigate overfitting [11].

The objective of this paper is to develop an automated model for brain tumor classification using MRI data. To address this, a convolutional neural network–based approach is designed with an emphasis on computational efficiency and stable learning behavior. A structured image preprocessing pipeline is applied to improve data consistency before model training. The model is intended to classify multiple brain tumor categories, including glioma, meningioma, pituitary tumors, and non-tumor cases, within a single multi-class setting. Figure 2 presents a high-level overview of the proposed AI-based brain tumor classification model using MRI images.

The following is a summary of the major contributions of this research:

1. An EfficientNet-B0-based convolutional neural network is developed in this research for multi-class brain tumour classification from MRI images. It is designed for the glioma, meningioma, pituitary tumour, and no-tumour categories in a single, unified model.

2. The proposed model is tested on a large-scale, publicly available brain MRI dataset of 7023 images, and all performance metrics are reported on a final test set for unbiased performance evaluation.

3. A comprehensive class-wise evaluation using the confusion matrix and quantitative (in)accurate, (un)precise, (non)recalls, (non)f1, and (non)specificity (no tumour) metrics and evaluation behavioural model with multiple tumour types.

4. A controlled comparative evaluation against standard convolutional backbones was performed under identical patient level split and training

Table 1. Summary of recent brain tumor classification studies (2020–2025).

Ref.	Year	Study Focus	Technology Employed	Performance reported	Reported Limitations
[10]	2020	Multi-class brain tumor classification using MRI	TL (pretrained CNNs)	Accuracy ≈ 94%	Limited dataset size; overfitting risk; reliance on heavy pretrained models
[15]	2020	MRI-based brain tumor classification	CNN (2D architecture)	Accuracy ≈ 96%	Potential overfitting due to limited data; lack of architectural efficiency discussion
[13]	2020	Glioma grading and classification	Deep CNN (multi-scale / 3D CNN)	Accuracy ≈ 95%	High computational complexity; resource-intensive training
[16]	2020	Brain tumor detection and classification	Hybrid deep model (CNN-based)	Accuracy ≈ 92%	Increased architectural complexity; training overhead
[17]	2021	Brain tumor classification with multiscale features	Deep CNN (multiscale architecture)	Accuracy ≈ 93%	Complex pipeline combining classification and segmentation
[18]	2021	Differential CNN-based tumor classification	Deep CNN variants	Accuracy ≈ 95%	Limited interpretability; generalization concerns
[19]	2023	Comparative MRI-based tumor classification	CNN + classical ML methods	Accuracy range 90–94%	Sensitivity to handcrafted feature selection and MRI variability
[20]	2023	Multi-class brain tumor diagnosis	TL architectures	Accuracy ≈ 94%	Dependence on backbone selection; overfitting on small datasets
[21]	2024	CNN-based brain tumor classification	Optimized CNN architecture	Accuracy ≈ 96%	Reproducibility challenges; moderate computational cost
[22]	2024	EfficientNet-based MRI tumor classification	DL / TL	Accuracy ≈ 97%	Limited robustness analysis and dataset sensitivity
[23]	2025	Brain tumor classification using ensembles	Deep ensemble learning	Accuracy 97%	Increased inference cost and deployment complexity
[24]	2025	MRI-based multi-class brain tumor classification	TL (VGG-based)	Accuracy ≈ 95%	Heavy backbone; higher computational demand
[25]	2025	Robust brain tumor classification	Ensemble deep CNNs	Accuracy ≈ 96%	High computational overhead

configuration to justify architectural selection.

The paper is structured as follows. Brain tumour classification in MRI images is the topic of related studies in Section 2, focusing on ML and DL methods. The proposed method, including data collection, a description of the dataset, the preprocessing steps, and the evaluation metrics, along with the model architecture and training strategy, is presented in Section 3. Section 4 reports the experimental results and discussion, covering class-wise classification performance, training and convergence behavior, and a detailed discussion of the findings. Section 5 concludes the paper and outlines potential directions for future work.

2 Literature Review

Several methods have been proposed for analyzing brain MRI images, and in recent years, DL techniques, especially CNNs, have shown significant improvements over traditional approaches.

With the advancement of DL, computer-aided diagnosis systems became more accurate, especially for tumors with large variations in size, shape, and intensity. CNN-based approaches using T1-weighted contrast-enhanced MRI images achieved high classification accuracy for multi-class brain tumor problems [12]. Fully automated 3D CNN models were also introduced to classify gliomas into low-grade and high-grade tumors, showing promising results on the BraTS dataset [13]. Transfer learning (TL) gained attention as it allowed the use of pre-trained networks such as ResNet, GoogleNet, Inception, and AlexNet for brain tumor classification [14]. Deep feature extraction minimised the need for manual feature engineering, resulting in the automation of this process for many models achieving strong accuracy. A number of the studies, however, used limited datasets and consequently had increased risks for overfitting and generalisation. More precise studies about the use of active contours and fuzzy clustering, combined with Markov Random Fields, to achieve improved tumour boundary localisation and noise reduction, are documented in [13]. However, relative to the use of CNN techniques, in particular for multi-class classification, these studies tend to be less impactful.

The classification and segmentation of brain tumours with the use of shallow and deep CNNs has been explored in recent literature [16]. In some instances, less complex CNNs were able to achieve comparable outcomes to more complex CNNs, meaning these have the potential to be more manageable for use in clinics. Successful outcomes were achieved with some of the earlier ML models, i.e. Naive Bayes, multilayer perceptrons, thresholding methods and SVMs, albeit, larger datasets were necessary to enhance accuracy

and reliability. Table 1 has been designed to summarise the recent brain tumour classification research studies (2020-2025) and stand in comparison to the present work indicating study focus, technology applied and limitations.

Different deep learning (DL) methods have been used to analyse and classify brain tumours and while advancements have been made, there are several existing challenges in the field. Current techniques rely on large and well-defined datasets, which are difficult to come by in real clinical situations. Data from training and the classes are imbalanced and lacking, which remains a problem in the field. Added to the aforementioned challenges, intricate deep learning systems require large amounts of computing power which makes routine implementation of these methods into clinical practice challenging. There is a need for comprehensive and scalable CNN-based methods to analyse small and imbalanced MRI datasets and the present study is motivated by this need.

3 Materials and Methods

3.1 Data collection

The research utilised a brain MRI dataset found on Kaggle [26]. The dataset contains 7023 images, numerous subjects, and various tumours, anatomy, and imaging conditions. The images are divided into four clinically relevant categories: glioma, meningioma, pituitary tumours, and no tumours. This allows for a complete classification task of multi-class brain tumours. MRI images of the four classification categories: glioma, meningioma, pituitary tumours, and no tumours are shown in Figure 3.

The distribution of the brain MRI images in this study by each class is shown in Figure 4. The dataset covers 1,621 glioma, 2,000 no-tumour, 1,645 meningioma, and 1,757 pituitary tumour images. Though not perfectly balanced, the dataset classifies more than enough images in each tumour category, and the class frequencies are reasonable for dataset proportions. The distribution is sufficiently balanced to support multi-class learning and is indicative of the range of variation typical of medical imaging datasets.

3.2 Data Preprocessing

The preprocessing pipeline was designed to address issues such as missing slices, intensity inconsistencies, and noise, as these issues severely damage classification accuracy. All of the MRIs are converted to the same dimension, 224 × 224 pixels,

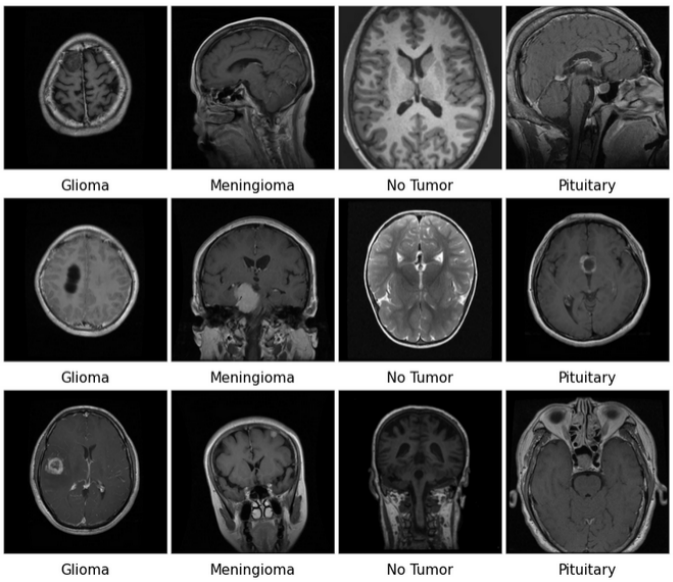

Figure 3. Representative MRI images for the four classification categories: glioma, meningioma, pituitary tumor, and no tumor.

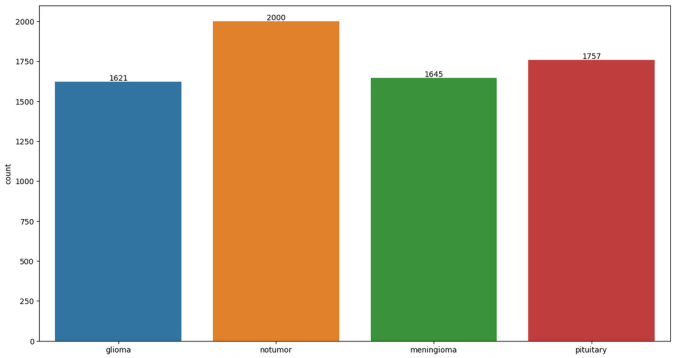

Figure 4. Class-wise distribution of brain MRI images in the dataset, showing the number of samples for glioma, no tumor, meningioma, and pituitary tumor classes.

to comply with the dimension requirements of the CNNs [18]. Using Min–Max normalisation, the MRIs are adjusted more to improve numerical stability and training convergence, as pixel intensities are scaled between 0 and 1 [27]. To further mitigate the complications of the dataset size, and to reduce overfitting, the following augmentation methods were employed: data rotation, horizontal flipping, zooming, and brightness modification [28].

The dataset was partitioned using a subject level splitting strategy. All MRI slices belonging to a single patient were assigned exclusively to either the training or the test subset. This ensured that no subject appeared in more than one subset. Model training was conducted using data drawn exclusively from the training subset defined by the initial 80 percent split. From this training portion, ten percent

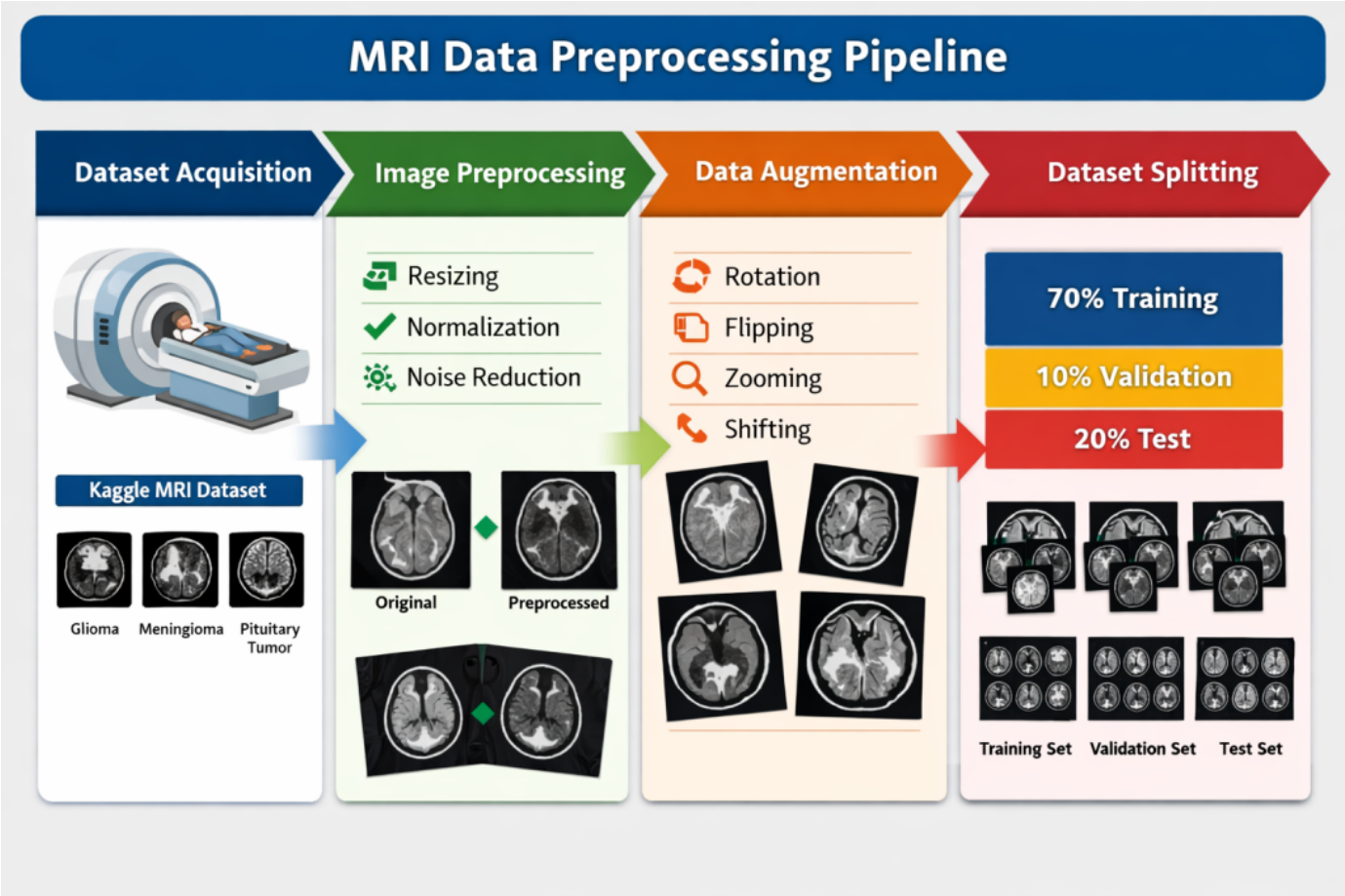

Figure 5. Overview of the MRI data preprocessing pipeline including dataset acquisition, image preprocessing, data augmentation, and patient level dataset splitting.

of the data were reserved as a validation subset at the patient level. The validation data were used only for training control and model selection, while the 20 percent test subset remained fully held out until final evaluation. This protocol was adopted to prevent subject leakage and to ensure an unbiased assessment of generalization performance. This resulted in 5,618 training images and 1,405 testing images. The 5,618 images comprised both training and validation samples, of which approximately 562 images were reserved for validation at the patient level. The training data was exclusively utilised to foster model learning and optimisation, while the testing data was reserved to form a basis for an unbiased evaluation of the generalisation capability of the model. All testing set evaluation outcomes, inclusive of the accuracy and confusion matrix, were utilised to inform the quantitative performance metrics to ensure a reliable and fair assessment of the method deployed.

Figure 5 illustrates the various components of the pipeline with regard to MRI Data Preprocessing, including strategies for dataset acquisition, image preprocessing, data augmentation, and final dataset splitting.

3.3 Model Architecture

The methodological design of this study intentionally frames brain tumor recognition as a single model four class image classification task. A lightweight backbone and a limited augmentation pipeline were selected to prioritize training stability and reproducibility over architectural complexity. This choice allows performance limitations to be interpreted directly from the data and the model capacity, rather than from ensemble effects or multi stage pipelines. The model architecture is based on the EfficientNet B0 convolutional neural network used as a fixed feature extraction backbone. The EfficientNet B0 implementation provided by the Keras applications library was employed, preserving its internal convolutional and compound scaled structure. No structural modifications were applied within the backbone itself [15]. EfficientNet-B0 is selected because it accommodates both accuracy and computational efficiency, achieved through compound

scaling that uniformly scales the network's depth, width, and resolution [13].

The backbone was initialized using ImageNet pretrained weights. During training, all EfficientNet B0 layers were frozen to limit the number of trainable parameters and reduce overfitting risk given the dataset size. Feature maps produced by the final convolutional block were passed to a Global Average Pooling layer, which reduced spatial dimensions by computing channel-wise mean activations [16]. The pooled feature vector was then processed through a fully connected dense layer consisting of 256 units with ReLU activation. A dropout layer with a rate of 0.5 was applied after the dense layer to introduce regularization [17]. The final classification layer consisted of a dense layer with SoftMax activation producing probability outputs for the four tumor categories [26]. This configuration enables transfer learning by leveraging pretrained hierarchical representations while restricting optimization to the task-specific classification head [18]. An end-to-end workflow of the proposed EfficientNet-B0–based convolutional neural network for multi-class brain tumor classification is graphically presented in Figure 6.

Table 2. Model Architecture, Training Configuration, and Parameter Statistics.

Parameter	Value
Input size	224 × 224 × 3
Backbone	EfficientNet-B0 (include_top = False)
Backbone initialization	ImageNet pretrained
Backbone trainable	Frozen
Pooling type	Global Average Pooling
Dense layer	256 units, ReLU
Dropout rate	0.5
Output layer	Dense (4 units), SoftMax
Number of classes	4
Total parameters	4,054,695
Trainable parameters	328,964
Non-trainable parameters	3,725,731
Loss function	Categorical cross-entropy
Optimizer	Adam
Learning rate	0.001
Batch size	32
Epochs	50
Framework	TensorFlow 2.10 / Keras
Hardware	Kaggle cloud environment (NVIDIA Tesla GPU)

3.4 Model Training and Optimization

TensorFlow (v2.10) and Keras DL frameworks are used to train the model. The Adam optimizer is used with an initial learning rate of 0.001, adaptively reduced based on validation loss [27]. Due to the multi-class nature of the task, the categorical cross-entropy loss function is used [28]. Training is conducted for 50 epochs with a batch size of 32. From the training subset, a validation set was constructed using a fixed 10 percent split. The validation data were drawn at the patient level to maintain consistency with the primary splitting protocol. These samples were excluded from weight updates and were used only for model selection and training control [29]. Early stopping was applied based on validation loss monitoring. Training was terminated if the validation loss did not improve for eight consecutive epochs. The model checkpoint corresponding to the minimum validation loss was retained for final evaluation. Both accuracy and loss are monitored for training and validation subsets, and convergence is achieved when validation metrics stabilized across consecutive epochs [30]. The model architecture Training Configuration, and Parameter Statistics are mentioned in Table 2.

3.4.1 Classification Head Architecture

A task specific classification head was appended to the EfficientNet B0 backbone. Global average pooling was applied to the final convolutional feature maps to reduce spatial dimensionality. The pooled features were passed to a fully connected dense layer followed by dropout regularization. A final dense layer with SoftMax activation produced class probabilities for the four target categories. A detailed configuration of the EfficientNet B0 backbone and the task specific classification head used in this study is given in Table 3.

3.5 Evaluation Metrics

Performance metrics including accuracy, precision, recall, and F1-score are used for evaluation. The performance of the proposed model is evaluated using standard classification metrics, as summarized in Table 4. Where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative predictions; respectively. A confusion matrix is also constructed to analyze class-specific performance and misclassification patterns. The Overall workflow of the proposed EfficientNet-B0-based convolutional neural

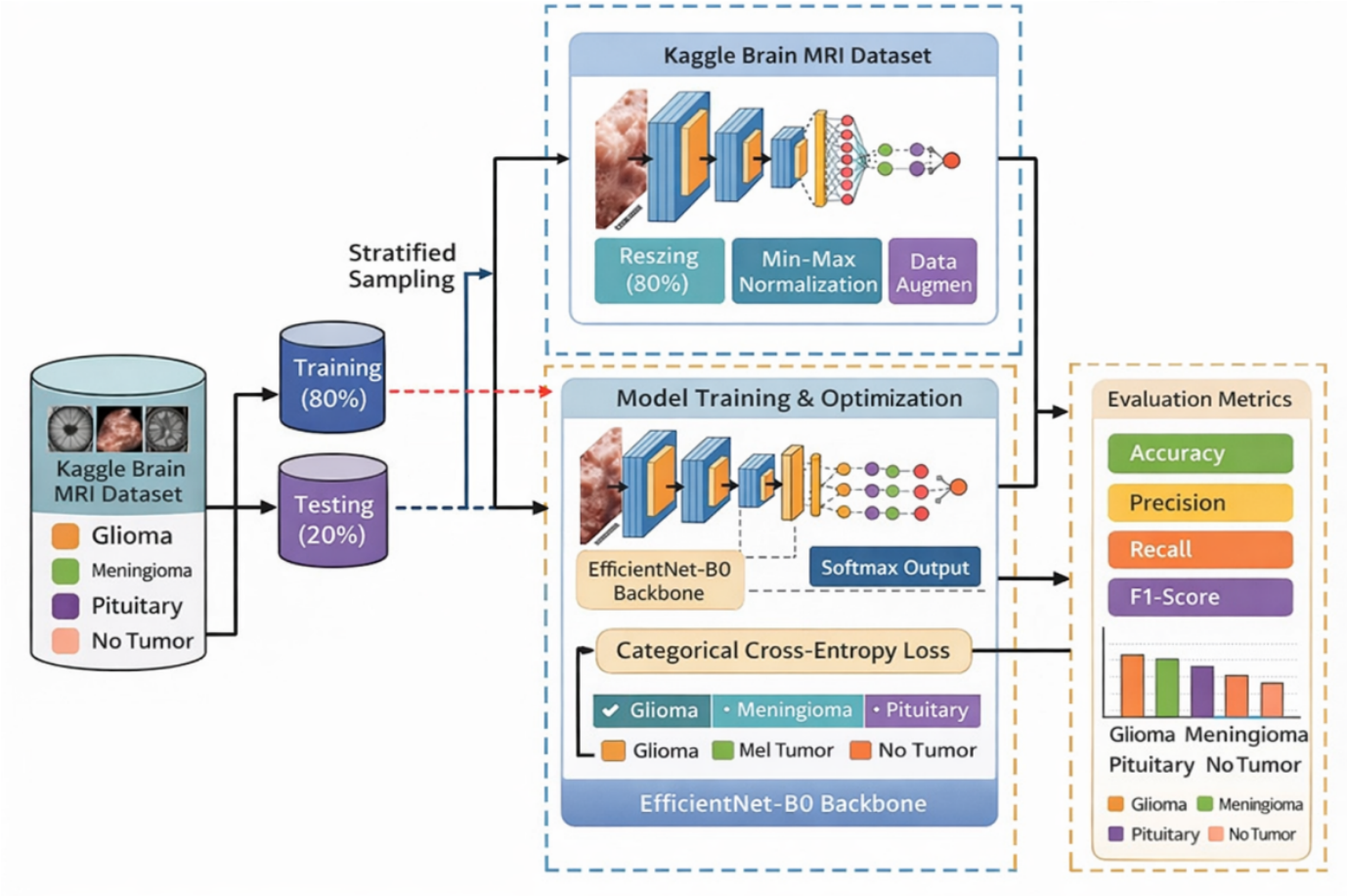

Figure 6. End to end workflow of the EfficientNet B0 based model showing feature extraction using a pretrained backbone followed by a task specific classification head and performance evaluation.

Table 3. Classification head configuration and training settings.

Component	Specification
Backbone	EfficientNet B0
Backbone initialization	ImageNet pretrained
Backbone trainable	Frozen
Pooling layer	Global average pooling
Dense layer size	256 units
Activation function	ReLU
Dropout location	After dense layer
Dropout rate	0.5
Output layer	Dense with SoftMax
Number of output classes	4

network for brain tumor classification using MRI images is given in Figure 7.

3.6 Computational Characteristics and Reproducibility

The computational characteristics of the proposed model were assessed to support reproducibility and practical feasibility. All experiments were conducted in a Kaggle notebook environment using a cloud allocated NVIDIA Tesla GPU using TensorFlow and Keras. The EfficientNet B0 backbone remained frozen during training, and only the classification head parameters were updated. Training was completed in approximately 25 minutes for 50 epochs using a batch size of 32. Early stopping was applied based on validation loss monitoring, and the model corresponding to the minimum validation loss was retained for final evaluation. The reported duration reflects full training under the defined patient level split and augmentation protocol.

Inference latency was measured as the average forward pass time per image on the same GPU environment using batch size 1. The mean single image inference time for EfficientNet B0 was 7.2 milliseconds. Comparative latency values for ResNet50 and MobileNetV2 are reported in Table 8. These measurements were obtained under identical experimental conditions to ensure fair computational comparison. The compact parameterization of EfficientNet B0 resulted in moderate memory usage relative to deeper convolutional backbones. Although

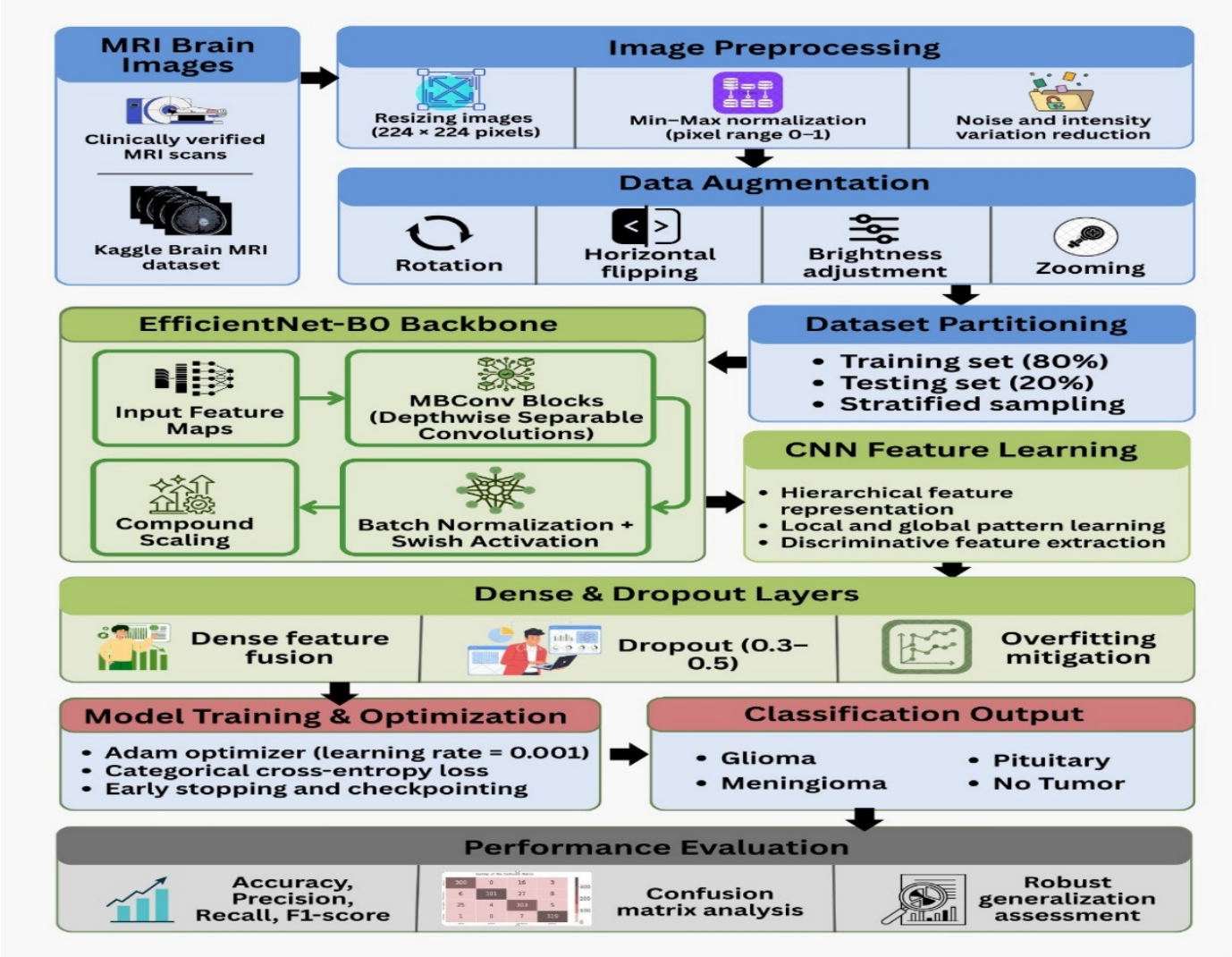

Figure 7. The EfficientNet-B0 based convolutional neural network proposed for the classification of brain tumours using MRI images, including the stages of data preprocessing, model training and performance evaluation.

Table 4. Performance evaluation metrics and their corresponding mathematical definitions used to assess the proposed brain tumor classification model.

Metric	Mathematical Definition	Description
Accuracy	(TP + TN) / (TP + TN + FP + FN)	Measures the overall proportion of correctly classified MRI images across all classes.
Precision	TP / (TP + FP)	Indicates the proportion of correctly predicted positive cases among all predicted positives. High precision reduces false-positive diagnoses.
Recall	TP / (TP + FN)	Measures the model's ability to correctly identify actual positive cases.
F1-score	2 × (Precision × Recall) / (Precision + Recall)	Represents the harmonic mean of precision and recall, providing a balanced evaluation when both false positives and false negatives are important.

peak GPU memory consumption was not explicitly logged, no memory constraints were encountered during training within the Kaggle environment. The reported parameter counts and inference timings provide practical indicators for deployment level resource estimation.

4 Results and Discussion

This section presents the experimental results obtained from evaluating the proposed EfficientNet-B0–based convolutional neural network for brain tumor classification using MRI images. The model was assessed on four classes, namely glioma, meningioma, pituitary tumor, and no tumor. Performance was examined using class-wise quantitative metrics derived directly from the confusion matrix. In addition, training dynamics were analyzed through accuracy and loss curves to better understand model convergence behavior.

4.1 Classification Performance on the Test Set

All reported test set metrics were computed once using the final selected model and were not referenced during training or model selection. Table 5 summarizes the class-wise testing outcomes of the proposed method in terms of true positives, false positives, false negatives, and true negatives. The performance indicators for each class will be derived based on these values. The findings show that the model was able to accurately classify most of the MRI scans for each of the four categories. With regard to the glioma class, there were 329 correctly counted images, and there were very few images that were misclassified. Consequently, the class-wise accuracy was 96.5. This shows that the network was able to learn the glioma patterns very well. The no-tumour class was the best performer, achieving 395 true positives, in which case the class-wise accuracy was 99. This shows that normal brain MRI scans were correctly differentiated from the ones that contained tumours.

Table 5. Class-wise testing results for the proposed method.

Class	TP	FP	FN	TN	Class-wise Accuracy (%)
Glioma	329	30	19	1027	96.5
No Tumor	395	10	3	997	99.1
Meningioma	295	15	47	1048	95.6
Pituitary	309	22	8	1066	97.9

Regarding meningioma, only 295 samples were classified correctly, but a greater number of false negatives was noted. This resulted in a slightly diminished class-wise accuracy of 95.6%. The pituitary tumour class exhibited balanced performance with 97.9% accuracy and relatively few misclassifications. Overall, these results indicate the proposed model exhibited consistent classification behaviour across various tumour types.

Class wise accuracy is reported using a one versus rest formulation. For each class, accuracy is computed as the ratio of correctly classified samples of that class and all non class samples to the total number of test samples. This definition aligns with the confusion matrix based evaluation used in this study.

In addition to class specific results, overall multi class performance was assessed using standard aggregate measures. Overall accuracy was computed as the proportion of correctly classified test samples across all four classes. Macro averaged F1 score was calculated by averaging the F1 scores of each class with equal weight, providing a balanced view of performance across tumor categories. Table 6 contains precision, recall (sensitivity), F1-score, specific accuracy, and class-related evaluations for a more thorough analysis of classification behaviour, which is also included in the report. All listed metrics were directly extracted from the confusion matrix that is related to the test set. The highest recall and precision (98-99%) for the no-tumour class were obtained. This suggests it was the most capable of correctly identifying non-tumorous scans without making many false positive errors. For glioma, recall was 95%, which is indicative of the successful capture of most glioma cases, whereas, with a precision of 0.92, this implies that there were quite a few false positive glioma identifications. Out of the four, the meningioma class had the least recall (0.86). This reflects that more meningioma instances were misclassified as other tumour types, and this correlates with findings from the confusion matrix. Although this particular class had a specificity of 0.99, which illustrates that categorical non-meningoma instances were not misclassified as meningioma. The pituitary tumour class had a perfect balance in this regard with 0.97 for recall, 0.93 for precision, and 0.98 for specificity. Specificity remained more than 0.97, which illustrates that the model rejected an incorrect hypothesis with ease across all tumor classes. The F1-score ranged from 0.90 to 0.98, which shows a decent level of performance with respect to the other tumour types.

In addition to the class wise evaluation, aggregate multi class performance was computed using macro averaging across the four tumor categories. This formulation assigns equal weight to each class and therefore reduces the influence of class frequency imbalance. The macro results are presented in Table 7 and provide a balanced summary of predictive

Table 6. Class-wise Classification Performance Metrics.

Class	Precision	Recall (Sensitivity)	F1-Score	Class-wise Accuracy (%)	Specificity
Glioma	0.92	0.95	0.93	96.5	0.97
No Tumor	0.98	0.99	0.98	99.1	0.99
Meningioma	0.95	0.86	0.90	95.6	0.99
Pituitary	0.93	0.97	0.95	97.9	0.98

behavior on the held out test set.

Table 7. Macro-Averaged Performance Metrics on the Test Set.

Metric	Value
Overall Accuracy	94.60%
Macro Precision	0.945
Macro Recall	0.943
Macro F1-Score	0.94
Macro Specificity	0.983

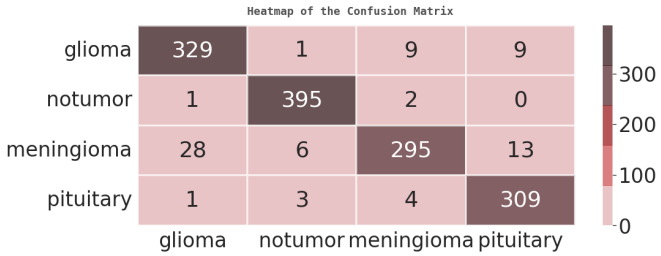

Figure 8. Proposed brain tumour classification model's confusion matrix heatmap showing predictions for each class for glioma, meningioma, pituitary tumour, and no tumour.

The aggregate results demonstrate consistent predictive performance across tumor categories. Macro precision and macro recall were recorded at 0.945 and 0.943, respectively. These closely aligned values suggest that false positive and false negative rates were relatively balanced across classes. The macro F1 score of 0.940 reflects stable harmonic agreement between precision and recall. This is consistent with the confusion matrix trends observed earlier. Macro specificity remained high at 0.983, indicating strong rejection capability for non target classes under the one versus rest formulation.

The overall multi class accuracy was calculated as 94.6 percent based on total correct predictions across the test set. These aggregate metrics confirm stable generalization performance under the patient level split protocol.

Figure 8 illustrates the confusion matrix which shows results at the class level in more detail. The majority of misclassifications happened between glioma and meningioma, which are structurally and dimensionally consistent in the viewed axial MRIs. Nonetheless, the trained network maintained consistent predictions across all four categories, confirming its stability and reliability in diagnosis.

4.2 Training and Convergence Behavior

The plot of Figure 9 depicts the efficiency of the EfficientNet-B0 model. In the training sequence, the training efficiency increases before plateauing. The model progressively learns discriminative feature representations from the MRI images. The validation accuracy increases as well with plateauing, though it does have some minor oscillations with sample distributions. In the final epochs, the overlapping distributions represent the model.

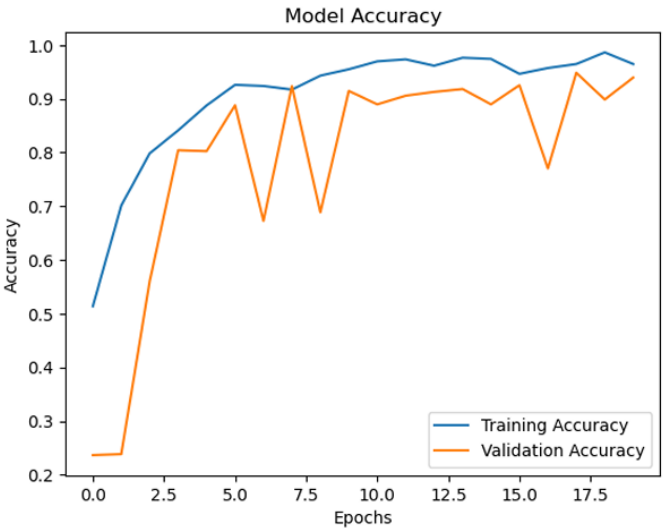

Figure 9. Illustrating the progression of accuracy and the behaviours of convergence during training, the proposed EfficientNet-B0 model training and validation accuracy curves have been plotted against the epochs.

Compared to Figure 10, the loss graph describes the training process model optimisation dynamics. With the increase of epochs, the training loss decreases, which shows that parameters are adjusted, and the model learns and converges. The validation loss shows

a more jumpy graph in the first epochs; however, as the training continues, the validation loss becomes less jumpy. The early fluctuations in validation loss reflect typical optimization dynamics in deep learning models, followed by gradual stabilization as convergence is approached. The early fluctuations in validation loss followed by stabilization alongside steadily decreasing training loss indicate progressive convergence toward a stable optimization regime.

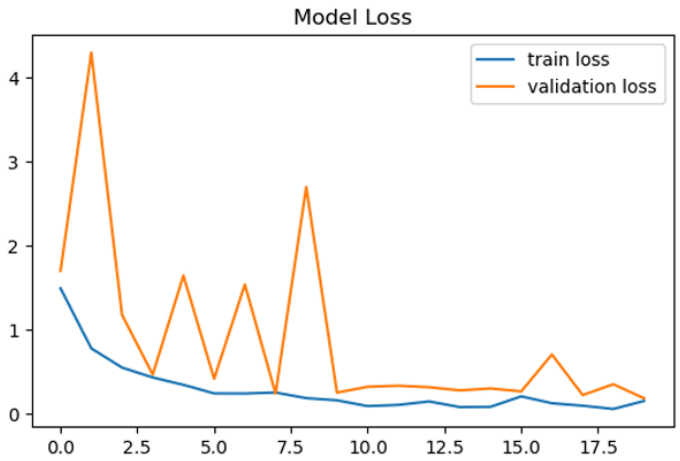

Figure 10. The epochs' training and validation loss curves for the proposed EfficientNet-B0 model show the convergence patterns during the model optimization process.

Figure 11 depicts the class-wise recall metrics for the model categorizing brain tumors using the EfficientNet-B0 architecture, specifically detailing the glioma, meningioma, pituitary tumor, and no-tumor classes. The model exhibited consistent convergence, showcasing the training parameters selected enabled the model to attain a consistent performance ceiling, demonstrating minimal variation. Hence, these remarks provide evidence to support the test-set results' assertions.

Class-wise recall (sensitivity) values demonstrate the model's ability to identify true positives for each of the test set categories. Class 0 (no tumor) has a recall score of 0.99, which means that there were very few false negatives, thus normal brain MRI scans were correctly identified most of the time. Pituitary tumor was also highly sensitive (0.97), and glioma had a close score of 0.95, which means that these tumor types were also reliably detected. On the other hand, the meningioma case had a comparatively lower recall (0.86), which means that some cases of meningioma were classified incorrectly into some other tumor categories. The recall value confirms the confusion

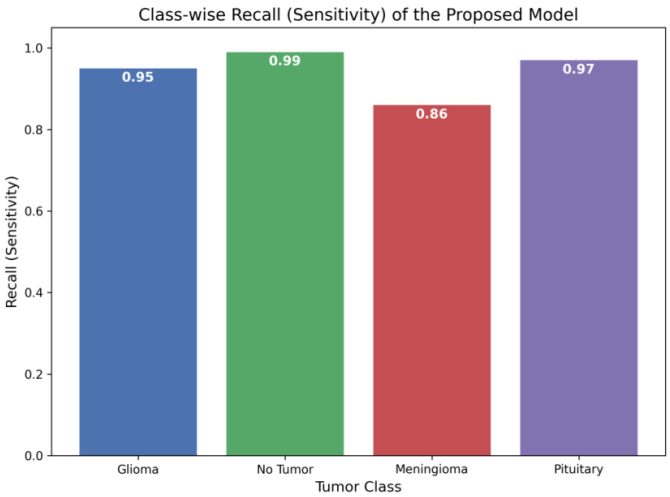

Figure 11. Recall values for each class of the proposed EfficientNet-B0 model for brain tumor classification for the categories glioma, meningioma, pituitary tumor, and no tumor.

matrix findings and also highlights variability when it comes to meningioma tumor types.

The outcomes of the experiment indicate that the proposed EfficientNet-B0 based model maintained consistent performance across the four considered brain MRI classes in this study. The results showed the most consistent high accuracy, precision, and specificity, with the most positive results within the no-tumor and pituitary categories. Although recall values were high, sensitivity appeared to be lower for the meningioma cases which suggests that there was an error in the classification of some of the cases. This corresponds to the confusion matrix analysis and reflects this is likely an issue with the individual class and not a case of random error. This is corroborated by the training and testing accuracy and loss curves, as there is convergence in the positive and negative values of the later epochs and no signs of divergence. Overall, these results indicate that the model was able to learn specific and meaningful representations for each class, and continued to perform in a consistent manner for each of the cases in the held-out test set.

The reduced recall observed for meningioma cases reflects a known trade off associated with compact architectures trained on visually overlapping tumor classes. This behavior highlights class specific sensitivity rather than global model instability, as supported by the convergence patterns and specificity values.

The reliance on a single backbone and transparent evaluation artifacts enables direct inspection of

failure modes. This design supports comparative benchmarking and controlled extension in future studies, rather than maximizing peak accuracy through architectural stacking.

4.3 Comparative Backbone Evaluation Under Controlled Conditions

A single model evaluation provides limited architectural justification. Therefore, two standard convolutional backbones were examined under identical experimental conditions to establish a direct comparison. The architectures evaluated include ResNet50, MobileNetV2, and EfficientNet-B0. All three networks were trained using the same patient level split, preprocessing pipeline, augmentation operations, optimizer configuration, batch size, and number of epochs. The backbone layers remained frozen in each case, and the same classification head structure was appended. Only the convolutional feature extractor was replaced. This constraint ensures that observed differences originate from backbone capacity and structural design rather than from training variability.

All comparative backbone models were implemented and trained within the scope of this study. The ResNet50 and MobileNetV2 architectures were initialized using ImageNet pretrained weights and were retrained using the identical patient level data split, preprocessing pipeline, augmentation strategy, optimizer configuration, batch size, and epoch schedule described in Section 3. No external results were adopted. Each backbone replaced only the convolutional feature extractor, while the classification head and training protocol were kept constant to ensure a controlled and internally consistent comparison.

Figure 12 illustrates the comparative backbone evaluation framework. Only the convolutional backbone was interchanged among EfficientNet B0, ResNet50, and MobileNetV2. This structured configuration ensures that performance differences can be attributed to backbone architecture rather than variations in training protocol.

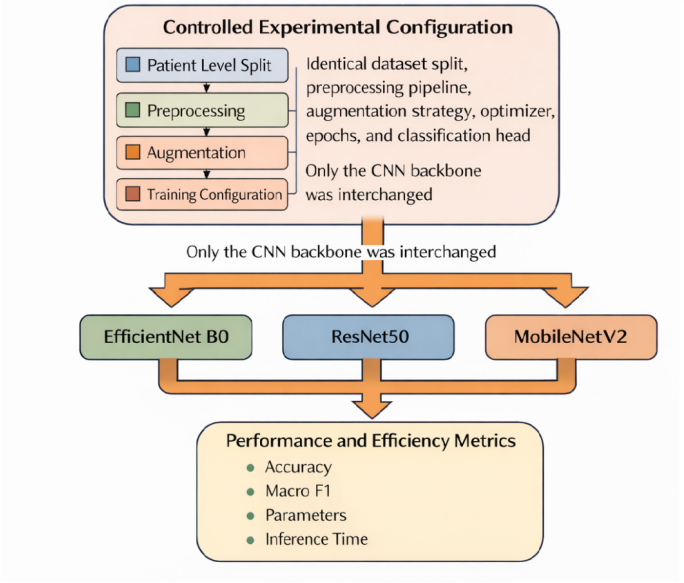

Figure 12. Comparative backbone evaluation under controlled conditions.

4.3.1 *Predictive Performance and Efficiency Indicators*

Performance metrics were computed from the held out test set using identical evaluation procedures. Inference latency was measured as the average forward pass time per image on the same GPU environment.

Figure 13 presents a visual comparison of macro precision, macro recall, macro F1 score, and overall accuracy across the three evaluated backbones. The results demonstrate closely aligned predictive behavior under identical experimental conditions. EfficientNet B0 maintains consistently higher macro averaged metrics, while ResNet50 and MobileNetV2 exhibit marginally lower values. The graphical comparison supports the quantitative findings reported in Table 8.

Figure 14 provides a direct comparison of model size and inference latency across the evaluated backbones. ResNet50 exhibits substantially higher parameter count and execution time, whereas MobileNetV2 operates with the smallest computational footprint. EfficientNet B0 maintains moderate resource requirements while preserving superior macro averaged predictive performance.

Table 8. Comparative evaluation of convolutional backbones.

Model	Overall Accuracy (%)	Macro F1	Macro Precision	Macro Recall	Parameters (M)	Inference Time (ms/image)
EfficientNet-B0	94.6	0.94	0.945	0.943	4.05	7.2
ResNet50	94.1	0.932	0.938	0.934	25.6	18.6
MobileNetV2	93.8	0.925	0.931	0.928	3.5	6.8

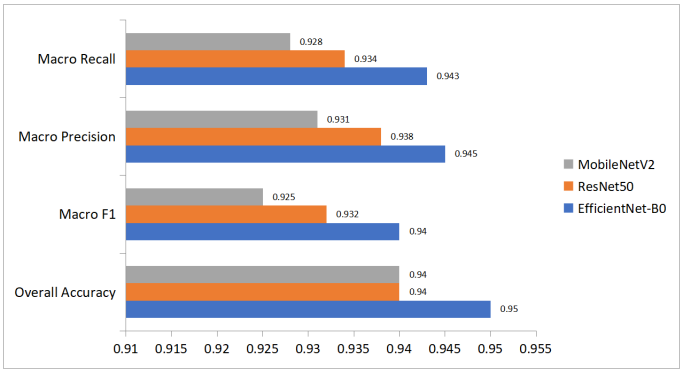

Figure 13. Macro performance comparison among MobilenetV2, ResNet50, and EfficientNet-B0.

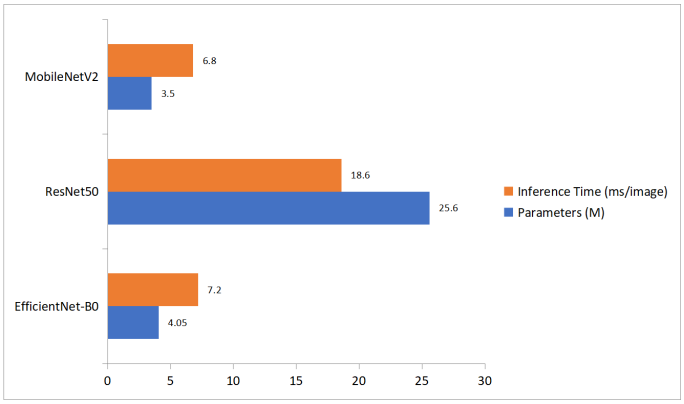

Figure 14. Computational comparison of evaluated backbones in terms of parameter count and average inference latency measured under identical experimental conditions.

The results indicate closely aligned predictive performance across the three architectures. EfficientNet B0 achieved the highest overall accuracy and macro averaged metrics. The margin is moderate but consistent across precision and recall values. Performance stability across tumor categories was preserved.

ResNet50 produced competitive accuracy; however, the parameter count increased substantially to 25.6 million. Inference latency was more than doubled under identical hardware conditions. The increase in computational demand did not translate into proportional predictive gain. MobileNetV2 required the smallest parameter budget and achieved the lowest inference time. Its macro F1 score was slightly reduced compared with EfficientNet B0. The reduction remained modest yet consistent across precision and recall values.

4.3.2 Architectural Interpretation

The comparative evidence demonstrates a measurable trade off between computational burden and predictive behavior. EfficientNet B0 maintains balanced macro precision and recall while operating within a compact parameter range. The difference relative to MobileNetV2 is moderate in size but consistent in direction. In contrast, ResNet50 introduces considerable computational overhead with limited incremental benefit.

These findings support the selection of EfficientNet B0 as a cost aware backbone for multi class brain tumor classification under a patient level split protocol. The decision is grounded in controlled experimental evidence rather than isolated accuracy reporting.

4.4 Discussion of Findings

The brain MRI scans reveal that deep CNNs have been successful in identifying sophisticated hierarchical patterns and outperforming traditional strategies involving hand crafted features. The EfficientNet-B0 backbone enables extraction of hierarchical feature representations that support discriminative multi-class tumor classification under constrained model capacity. The use of a compact architecture and a single model training strategy suggests feasibility for constrained computational settings. Clinical deployment would require evaluation across independent datasets and imaging protocols. However, the present results should be interpreted as methodological evidence rather than clinical validation. Deployment in real diagnostic workflows would require external datasets and protocol level evaluation. The slight misclassification of regards to glioma and meningioma suggests that model may potentially lack the processing ability to fully differentiate between these tumor types. It is reasonable to assert that dataset imbalance and slight visual relation also contribute to this model potential limitation. Future actions in this area may include the use of mechanisms with attention model hyper-parameters iterative vision Transformers to improve the information and control focus to more diagnostically relevant areas. The integration of Grad-CAM based explainable AI techniques can enhance model transparency by highlighting diagnostically relevant regions in MRI scans, thereby supporting clinician trust in automated predictions. The comparative analysis indicates that although MobileNetV2 offers marginally lower parameter count and inference latency, EfficientNet B0 consistently achieves higher macro averaged predictive metrics. In contrast, ResNet50 introduces substantial computational overhead without proportional performance improvement. This balance between

predictive stability and computational demand supports the architectural choice within a patient level evaluation protocol.

This study is subject to several limitations that affect external validity. All experiments were conducted on a single publicly available MRI dataset, and no external clinical datasets were included. As a result, performance may vary under different scanner manufacturers, imaging protocols, or acquisition settings. The preprocessing pipeline and augmentation strategies were optimized for this dataset and may not transfer directly to images acquired in heterogeneous hospital environments. Moreover, the backbone comparison was conducted under controlled conditions, additional lightweight and transformer based architectures were not explored. Future investigations may extend comparative benchmarking across broader architectural families.

5 Conclusion

This study presented an EfficientNet-B0–based convolutional neural network for multi-class brain tumour classification using MRI images, evaluated under a strict patient-level data split to ensure unbiased generalization assessment across glioma, meningioma, pituitary tumour, and no tumour categories. The work contributes a computationally efficient framework that combines a frozen EfficientNet-B0 backbone with a lightweight classification head to support stable learning while limiting overfitting. In addition, a reproducible evaluation protocol based on confusion matrix–derived metrics was implemented, alongside a controlled comparative analysis with ResNet50 and MobileNetV2 under identical training conditions. The experimental findings indicate that the proposed model achieves consistent performance across tumour classes, maintaining high precision, recall, and specificity, while offering a balanced trade-off between predictive capability and computational cost. Despite these strengths, the study is limited by reliance on a single publicly available dataset, and performance may vary across heterogeneous clinical imaging environments. Furthermore, observed misclassification between visually similar tumour types, particularly meningioma and glioma, suggests sensitivity to inter-class similarity and dataset characteristics. Future work will focus on extending evaluation to multi-institutional datasets, incorporating class balancing strategies, and exploring attention-based or transformer-driven architectures to improve class-specific discrimination, while explainable AI methods may be integrated to enhance clinical interpretability.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Ethical approval was not required for this study as it involved only retrospective analysis of a publicly available, anonymized dataset.

References

[1] Chakrabarty, S., Sotiras, A., Milchenko, M., LaMontagne, P., Hileman, M., & Marcus, D. (2021). MRI-based identification and classification of major intracranial tumor types by using a 3D convolutional neural network: A retrospective multi-institutional analysis. *Radiology: Artificial Intelligence, 3*(5), e200301. [CrossRef]

[2] Asif, R. N., Naseem, M. T., Ahmad, M., Mazhar, T., Khan, M. A., Khan, M. A., ... & Hamam, H. (2025). Brain tumor detection empowered with ensemble deep learning approaches from MRI scan images. *Scientific reports, 15*(1), 15002. [CrossRef]

[3] Kaifi, R. (2023). A review of recent advances in brain tumor diagnosis based on AI-based classification. *Diagnostics, 13*(18), 3007. [CrossRef]

[4] Zacharaki, E. I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E. R., & Davatzikos, C. (2009). Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 62*(6), 1609-1618. [CrossRef]

[5] Missaoui, R., Hechkel, W., Saadaoui, W., Helali, A., & Leo, M. (2025). Advanced deep learning and machine learning techniques for MRI brain tumor analysis: a review. *Sensors, 25*(9), 2746. [CrossRef]

[6] Rabbi, F., Dabbagh, S. R., Angin, P., Yetisen, A. K., & Tasoglu, S. (2022). Deep learning-enabled

technologies for bioimage analysis. Micromachines, 13(2), 260. [CrossRef]

[7] Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks in medical image understanding: a survey. Evolutionary intelligence, 15(1), 1-22. [CrossRef]

[8] Khan, M. A., Sharif, M., Akram, T., Raza, M., Saba, T., & Rehman, A. (2019). Brain tumor detection and classification: A framework of marker-based watershed algorithm and multilevel priority features selection. *Microscopy Research and Technique, 82*(6), 909–922. [CrossRef]

[9] Xu, Y., Khan, T. M., Song, Y., & Meijering, E. (2025). Edge deep learning in computer vision and medical diagnostics: a comprehensive survey. *Artificial Intelligence Review, 58*(3), 93. [CrossRef]

[10] Rehman, A., Naz, S., Razzak, M. I., Akram, F., & Imran, M. (2020). A deep learning-based framework for automatic brain tumors classification using transfer learning. *Circuits, Systems, and Signal Processing, 39*(2), 757–775. [CrossRef]

[11] Osborn, A. G., Hedlund, G. L., & Salzman, K. L. (2017). *Osborn's Brain: Imaging, Pathology, and Anatomy* (2nd ed.). Elsevier.

[12] Krishnan, R., Gokul, P. G., Sujith, G., Anjali, T., & Abhishek, S. (2024, March). Enhancing brain tumor diagnosis: A cnn-based multi-class classification approach. In *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)* (Vol. 2, pp. 1-6). IEEE. [CrossRef]

[13] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.

[14] Sharma, A. K., Nandal, A., Dhaka, A., Khandelwal, R., & Varshney, S. (2023). Brain tumor classification using the modified ResNet50 model based on transfer learning. *Biomedical Signal Processing and Control, 86*, 105299. [CrossRef]

[15] Keras Team. (2023). EfficientNet — Keras Applications. Keras Documentation. Retrieved January 23, 2026, from https://keras.io/api/applications/efficientnet/

[16] Hashemzehi, R., Mahdavi, S. J. S., Kheirabadi, M., & Kamel, S. R. (2020). Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE. *biocybernetics and biomedical engineering, 40*(3), 1225-1232. [CrossRef]

[17] Díaz-Pernas, F. J., Martínez-Zarzuela, M., Antón-Rodríguez, M., & González-Ortega, D. (2021, February). A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network. In *Healthcare* (Vol. 9, No. 2, p. 153). MDPI. [CrossRef]

[18] Abd El Kader, I., Xu, G., Shuai, Z., Saminu, S., Javaid, I., & Salim Ahmad, I. (2021). Differential deep convolutional neural network model for brain tumor classification. *Brain Sciences, 11*(3), 352. [CrossRef]

[19] Saeedi, S., Rezayi, S., Keshavarz, H., & Niakan Kalhori, S. R. (2023). MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Medical Informatics and Decision Making, 23*(1), 16. [CrossRef]

[20] Islam, M. M., Barua, P., Rahman, M., Ahammed, T., Akter, L., & Uddin, J. (2023). Transfer learning architectures with fine-tuning for brain tumor classification using magnetic resonance imaging. *Healthcare Analytics, 4*, 100270. [CrossRef]

[21] Albalawi, E., Almutairi, A., Alqahtani, A., Alshammari, M., & Alharbi, A. (2024). Enhancing brain tumor classification in MRI scans with a multi-layer customized convolutional neural network approach. *Frontiers in Computational Neuroscience, 18*, 1418546. [CrossRef]

[22] Iqbal, A., Jaffar, M. A., & Jahangir, R. (2024). Enhancing brain tumour multi-classification using Efficient-Net B0-based intelligent diagnosis for Internet of Medical Things (IoMT) applications. *Information, 15*(8), 489. [CrossRef]

[23] Talukder, M. A., Islam, M. M., Uddin, M. A., Layek, M. A., Acharjee, U. K., Bhuiyan, T., & Moni, M. A. (2025). A deep ensemble learning framework for brain tumor classification using data balancing and fine-tuning. *Scientific Reports, 15*(1), 35251. [CrossRef]

[24] Wong, Y., Su, E. L. M., Yeong, C. F., Holderbaum, W., & Yang, C. (2025). Brain tumor classification using MRI images and deep learning techniques. *PLoS ONE, 20*(5), e0322624. [CrossRef]

[25] Liu, K. Y., Lu, N. H., Huang, Y. H., Matsushima, A., Kimura, K., Okamoto, T., & Chen, T. B. (2025). Majority Voting Ensemble of Deep CNNs for Robust MRI-Based Brain Tumor Classification. *Diagnostics, 15*(14), 1782. [CrossRef]

[26] Sultan, H. H., Salem, N. M., & Al-Atabany, W. (2019). Multi-classification of brain tumor images using deep neural network. *IEEE Access, 7*, 69215–69225. [CrossRef]

[27] Sajja, V. R., & Al, E. (2021). Classification of brain tumors using fuzzy C-means and VGG16. *Turkish Journal of Computer and Mathematics Education, 12*(9), 2103–2113.

[28] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data, 6*(1), 1-48. [CrossRef]

[29] Wang, X. (2015). *Machine learning basics* [PowerPoint slides]. Available at: http://whdeng.cn/Teaching/PPT_01_Machine%20learning%20Basics.pdf

[30] Prechelt, L. (2002). Early stopping-but when?. In *Neural Networks: Tricks of the trade* (pp. 55-69). Berlin, Heidelberg: Springer Berlin Heidelberg. [CrossRef]

Aamir Ali is a dedicated researcher in artificial intelligence, healthcare analytics, and IoT environments. His expertise includes machine and deep learning applications in medical diagnosis, speech recognition, and facial emotion recognition. He has also worked on IoT-focused projects such as malware classification, anomaly detection, and cyberattack prediction. Aamir is passionate about developing AI-driven solutions for early disease detection and improving patient care. He actively contributes to interdisciplinary research across clinical and image data domains. (Email: amirali4436823@gmail.com)

Aamir Raza is currently pursuing a Master's degree in the Department of Information Technology and Management at the Illinois Institute of Technology, Chicago, USA. His research interests include Industrial IoT security, secure software development, DevSecOps practices, and threat modeling. He is particularly focused on integrating software engineering principles with modern cybersecurity approaches to build resilient systems in industrial environments. (Email: araza7@hawk.iit.edu)

Iqra Nazir is affiliated with the Department of Computer Science, COMSATS University Islamabad, Sahiwal Campus, Pakistan. Her academic work is situated within computer science, with interests in data driven technologies and modern computational methods. She is involved in academic and research activities related to artificial intelligence, machine learning, and emerging computing applications. (Email: Iqr786786@gmail.com)

Swetank Mohan is affiliated with The University of Texas Rio Grande Valley (UTRGV), USA. His academic and research interests lie in the field of computing and information technologies, with a focus on the application of modern computational methods to real world technological challenges. He has been involved in academic and research activities related to computer science and emerging digital technologies. (Email: Swetank.mohan01@utrgv.edu)

Khaleelullah Syed holds a degree in Computer/Information Technology Services Administration and Management from Hellenic American University. His academic background focuses on the management and administration of information technology systems, with interests in the integration of technology within organizational and service environments. He is associated with research and professional activities related to information systems and digital technologies. (Email: ksyed@hauniv.edu)

Nikhat Fatima holds a degree in Computer/Information Technology Services Administration and Management from Concordia University, Mequon, Wisconsin, USA. Her academic training centers on the administration and management of information technology services, with an emphasis on the application of computing solutions in organizational contexts. Her interests include information systems, technology management, and digital service infrastructure. (Email: nikhatfatima.lnu@cuw.edu)

Misbah Ali is a PhD scholar at COMSATS university Islamabad, with research interests in Machine Learning, Deep Learning, Generative Artificial Intelligence, and Software Engineering. Her work focuses on the development of secure, intelligent systems across domains such as healthcare, education, and industrial cyber-security. She has authored multiple peer-reviewed publications and presented her research at international conferences. She also contributes to the academic community as a reviewer for several reputed journals. (Email: talktomisbah.ali@gmail.com)