International Journal of Technology

http://ijtech.eng.ui.ac.id

Research Article

Optimizing the Classification of Brain Tumors Using Ensemble Learning and Gradient-weighted Class Activation Mapping on Multi-section Magnetic Resonance Imaging Images

I Gede Susrama Mas Diyasa[1*], Denisa Septalian Alhamda[2], Hanifudin Sukri[3], Muhammad Salsabeela Rusdi[4], Deshinta Arrova Dewi[5], Hana Titania Sastrian[2], Sayyidah Humairah[6], Kraugusteeliana Kraugusteeliana[7]

[1]Department of Master Information Technology, Faculty of Computer Science, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya 60294, Indonesia

[2]Department of Data Science, Faculty of Computer Science, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya 60294, Indonesia

[3]Department of Electrical Engineering, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan 69162, Indonesia

[4]Department of Medical Research, Faculty of Medicine, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya 60294, Indonesia

[5]Center for Data Science and Sustainable Technologies, INTI International University, Putra Nilai 71800, Malaysia

[6]Department of Electrical & Computer Engineering, University of Patras, Patras 26504, Greece

[7]Department of Information Systems, Faculty of Computer Science, University of Pembangunan Nasional Veteran Jakarta, Jakarta 12450, Indonesia

*Corresponding author: igsusrama.if@upnjatim.ac.id; Tel.: +62-81999471017

Abstract:: A brain tumor is a condition in which cells in or around the brain grow abnormally and uncontrollably. Magnetic resonance imaging (MRI) scans are essential for diagnosing brain tumors, where rapid and accurate categorization is key to effective treatment. Automated systems utilizing CNNs provide an effective image classification solution due to their flexibility and architecture. This study proposes an ensemble learning methodology to classify images by combining three CNN architectures, namely Xception, EfficientNetV2S, and ResNet50, to increase classification performance while decreasing overfitting probability. EfficientNetV2S optimizes network components using NAS and progressive learning. Xception is based on the concept of depthwise separable convolution, which uses fewer computational resources while retaining the same discriminative power. ResNet50 relies on residual learning to alleviate the vanishing gradient problem and enhance the representation of deeper features. To provide better explainability, Grad-CAM focuses on critical regions of the MRI scans to assist the experts in confirming the classification outcomes. The ensemble technique generated significant performance metrics, including an accuracy of 99.18%, precision of 99.19%, recall of 99.18%, F1-score of 99.18%, and AUC-ROC of 0.9993, proving minimal false classifications. Combining ensemble learning and Grad-CAM offers the most explainable and accurate model for the classification of brain tumors into four categories: glioma, meningioma, non-tumor, and pituitary tumors. The research outcomes provide further evidence to support the use of AI in smart diagnostic devices, providing improved generalization and reliability for use in the medical field.

Keywords: Brain tumor classification; Cancer; EfficientNetV2S; Ensemble learning; MRI multi-section

https://doi.org/10.14716/ijtech.v17i2.8360
Received January 2026; Revised February 2026; Accepted February 2026

1. Introduction

Brain tumors are a serious health concern that pose a significant threat to human life. A brain tumor is a condition in which cells in or around the brain grow abnormally and uncontrollably. This type of tumor is hazardous, accounting for approximately 1.35% of all malignant tumor cases and contributing to 29.5% of cancer-related deaths worldwide (Lapointe et al., 2018). Tumor symptoms can vary and depend on tumor size and location, how quickly it grows and, over time, how much intracranial pressure increases, as noted in the most recent research from 2020 (Alther et al., 2020). Tumors are categorized in two ways. The first type of tumor is primary brain tumors, which develop from the tissue of the brain, while the second type is secondary brain tumors, which are metastatic growths from other cancers in the body (Ghazvini et al., 2024; Tandel et al., 2019).

Today, the incidence of brain tumors is increasing compared with years ago. According to the Global Cancer Observatory, in 2020, there were 308,102 documented brain tumor cases in the world, 54.2% of which were located in Asia. Additionally, the mortality rate of brain tumors remains unacceptably high, at 71% for patients aged 60 years (Sung et al., 2021). Even with these staggering numbers, there is still a notable lack of research on brain tumors and their early detection. Current methods still rely on biopsies and manual inspection and can take a long time as different specialists are needed to arrive at the same diagnosis. This increases the focus on the need for better diagnostic methods (T. Singh et al., 2024).

The capability of AI is rapidly evolving and expanding into various fields, including medicine and, more specifically, medical image classification. MRI imaging is essential for diagnosing brain tumors due to its respective advantages. It can accurately identify soft and hard brain tissues. Using deep learning and machine learning, MRI scans can be analyzed and interpreted with a level of accuracy that conventional approaches often do not achieve (Guo et al., 2024). Convolutional neural network (CNN) is one of the most frequently used algorithms for medical image classification, which is particularly effective at recognizing fine patterns in medical imaging datasets (Mahdi et al., 2023).

Among the numerous available CNN architectures, the EfficientNet model has gained popularity due to its high level of computational efficiency and accuracy through optimized depth, width, and resolution scaling (Zhang et al., 2021). In 2021, Google launched EfficientNetV2 with further enhancements to the training rate, speed, and inference accuracy. The most recent studies on the classification of brain tumors in MRIs indicate that EfficientNetV2, augmented with global attention, has attained some of the highest available accuracy (Ghaffar et al., 2024). Studies have shown that EfficientNetV2 achieved accuracy improvements of as much as 99% on complex medical datasets when compared with other models of the same parametric value (Priyadarshini et al., 2024). Along with EfficientNetV2, Xception and ResNet50 are examples of deep learning architectures widely used for image classification. Xception has been successful in multiple image recognition assignments, such as brain tumor classification in MRI images (Gawali and Dhongade, 2024; Amarnath et al., 2024; Dishar and Muhammed, 2023). Similarly, ResNet50 has various medical applications and efficiently predicts brain tumor classification (Musa et al., 2024; Amarnath et al., 2024; Mathivanan et al., 2024).

Recent studies have incorporated attention mechanisms with CNN architectures to optimize MRI-based brain tumor classification. For instance, studies have augmented EfficientNetV2 with the Global Attention Mechanism (GAM) and Efficient Channel Attention (ECA) to improve the model's attention to salient features, attaining a remarkable 99.76% accuracy on public datasets (Pacal et al., 2024). Furthermore, Grad-CAM has been used for model visualization to provide clear and comprehensible explanations of the rationale behind the model's decisions, which is essential for clinical reasoning (Vo et al., 2024; Nazir et al., 2024). Comparative frameworks have clarified the choice of visualization strategies. For instance, it was recently demonstrated that while SHAP provides granular pixel-level attribution, Grad-CAM offers superior spatial localization for defining anatomical tumor boundaries, achieving a higher clinical relevance rating than SHAP (Gharaibeh, 2025). These studies contribute to the body of work on the application

of deep learning to the analysis of medical images, especially for the classification of brain tumors, while underscoring the importance of attention mechanisms and interpretability in improving the accuracy of diagnostic systems.

Each CNN architecture has unique advantages and disadvantages for a given image classification assignment. The performances of individual classification techniques can be improved using ensemble learning (Basha et al., 2025; Younis et al., 2022; Alsubai et al., 2022). This method combines outcomes from multiple architectures or classification methods to improve accuracy and robustness. To improve the prediction accuracy and reduce the prediction error of single-model systems, ensemble learning uses diverse architectures to capture the unique and complementary features of each.

The current state-of-the-art landscape increasingly emphasizes this integration. For example, a hybrid ensemble concatenating the Xception model with parallel deep CNNs (PDCNN) was recently proposed (Yoon, 2025). This approach achieved a classification accuracy of 99.09% by leveraging Xception's depthwise separable convolutions for fine-grained feature extraction and PDCNN's varied kernel sizes for broader spatial details. Similarly, studies focusing on advanced optimization strategies have demonstrated the value of dynamic weighting in ensemble frameworks. A novel method using PSO was introduced to determine the optimal contribution weights for five distinct CNN architectures was introduced (Çetin-Kaya and Kaya, 2024). This PSO-based weighted ensemble successfully mitigated overfitting and achieved superior performance, reaching 99.92% accuracy on large-scale magnetic resonance imaging datasets.

Other recent works have also examined ensemble applications in MRI classification of brain tumors. For example, an ensemble attention mechanism combining the MobileNetV3 and EfficientNetB7 architectures obtained 98.94% and 98.48% accuracy on the Figshare and BRATS 2019 datasets, respectively (Pacal et al., 2024). Likewise, a deep ensemble learning model based on transfer learning with DenseNet121 and InceptionV3 has improved interpretability without sacrificing classification accuracy (Sterniczuk and Charytanowicz, 2024). However, while these methods demonstrate the utility of hybrid architectures and heuristic optimization, they typically rely on fixed, globally assigned ensemble weights that are not re-evaluated across different data partitions. Therefore, frameworks that select ensemble fusion weights adaptively per data partition and are validated across multiple independent experimental runs to demonstrate reproducible robustness are needed.

Consequently, several research gaps remain in the current brain tumor classification approaches. First, most existing ensemble learning studies concentrate primarily on optimizing model performance during the training phase and do not adequately investigate the models' generalizability to diverse MRI brain tumor datasets (Sharif et al., 2024; Ismail et al., 2024). Second, these approaches frequently lack validation on heterogeneous MRI data that varies in tumor size, imaging parameters, and scanner type, limiting their applicability in real-world clinical settings where such variations are common (Kesuma et al., 2023). Third, although individual models and some ensemble approaches have achieved high accuracy, there is a need for methods that simultaneously address overfitting, validate statistical reliability, and provide clinical interpretability through explainable AI. These gaps highlight the need for a robust ensemble approach that can maintain high performance across diverse imaging conditions while offering transparent decision-making processes suitable for clinical adoption.

To address these limitations, this study introduces an ensemble learning approach that integrates EfficientNetV2S, Xception, and ResNet50 through a per-fold grid search weight optimization strategy on the four-class brain tumor MRI dataset (Nickparvar, 2023), encompassing glioma, meningioma, no tumor, and pituitary categories. Different combinations of model weights are systematically tested for each cross-validation fold, and the set that gives the highest validation accuracy is selected for that fold. This allows the fusion weights to be adjusted to each data split instead of using one fixed set of weights for all folds. The ensemble evaluation is repeated using three different random seeds to ensure reliable results, producing 15 cross-validation runs. The final performance is reported as the mean standard deviation along with

95% confidence intervals.

Grad-CAM is applied as a systematic post-prediction interpretability layer over the trained ensemble to support clinical applicability, generating spatial activation heatmaps that highlight the tumor-relevant regions driving each classification decision. Unlike prior studies that present Grad-CAM visualizations qualitatively, this work quantitatively validates spatial alignment using IoU scores computed against expert-annotated segmentation masks, providing an objective measure of whether the model attends to clinically meaningful regions. To ensure reliable and comparable assessment, performance is evaluated using standard metrics, including accuracy, precision, recall, and F1 score.

The remainder of this paper is structured as follows: Section 2 discusses the proposed methodology, which consists of the dataset, preprocessing, model architecture, ensemble learning, Grad-CAM, and metrics. Section 3 presents the experimental results and discussion. Finally, Section 4 provides the conclusions and outlines opportunities for future research.

2. Methods

The research workflow is illustrated in Figure 1, starting from obtaining an MRI image dataset, which was classified into four categories: glioma, meningioma, No tumor and pituitary The dataset was obtained from Kaggle, an open-access data repository. The subsequent steps are data preprocessing and image enhancement to make the dataset suitable for model training. A denoising autoencoder is used to process noisy images to yield a set of cleaner images, and data augmentation is applied through rotation and flipping to improve the variety of the training dataset. After preprocessing, the dataset is split into three parts: training, validation, and testing. Moreover, the model employs the Grad-CAM (Gradient Class Activation Mapping) technique to highlight the important areas of MRI images related to the classification of tumors. The explainable AI (XAI) precursor applied to the model, helps users understand the model by explaining what areas of the image the model focused on for classification and why, thereby providing functional transparency. Grad-CAM is used to focus on certain parts of the images based on the model's reasoning to construct a post hoc explainable AI (XAI) approach. The model's performance is assessed using the confusion metric.

This study aims to create a brain tumor classification system that is both accurate and robust to overfitting. With this classification system, we hope to assist medical personnel in the quicker, more efficient, and more dependable diagnosis of brain tumors and contribute to the advancement of AI-driven diagnostic technologies.

2.1 Data Acquisition

At this stage, we acquired the MRI Image Data from Kaggle, which consists of various types of brain tumors, including glioma, meningioma, and pituitary tumors, as well as non-tumor cases. The dataset for this study was sourced from Kaggle's "Brain Tumor MRI Dataset" created by Nickparvar in 2023 (Nickparvar, 2023). The MRI dataset contains a total of 7,023 images in the ".jpg" format. The breakdown is as follows: 1,621 images of glioma tumors, 1,645 images of meningioma tumors, 2,000 no-tumor images, and 1,757 images of pituitary tumors. Each image is 512x512 pixels in size and is in the grayscale channel, which represents a suitable foundation for training the brain tumor classification model.

Meningioma tumors originate in the lining of the brain and spinal cord inside the skull. In contrast, pituitary tumors abnormally develop in the area around the pituitary gland, which is located inside the skull between the brain and nasal cavity. Both tumor types are considered benign tumors, meaning they do not spread to other cells, tissues, or body parts. In contrast, gliomas are malignant tumors that can spread to other organs (Gokila et al., 2024). The dataset becomes more representative of this variety of tumor types, allowing the model to detect more profound differences between benign and malignant tumors and providing a solid basis for more accurate classification. Figure 4 (A) shows the types of brain tumors and their respective

brain locations. This diagram shows the various locations in which tumors, such as glioma, meningioma, and pituitary, can occur, helping to understand the classification of different types of tumors.

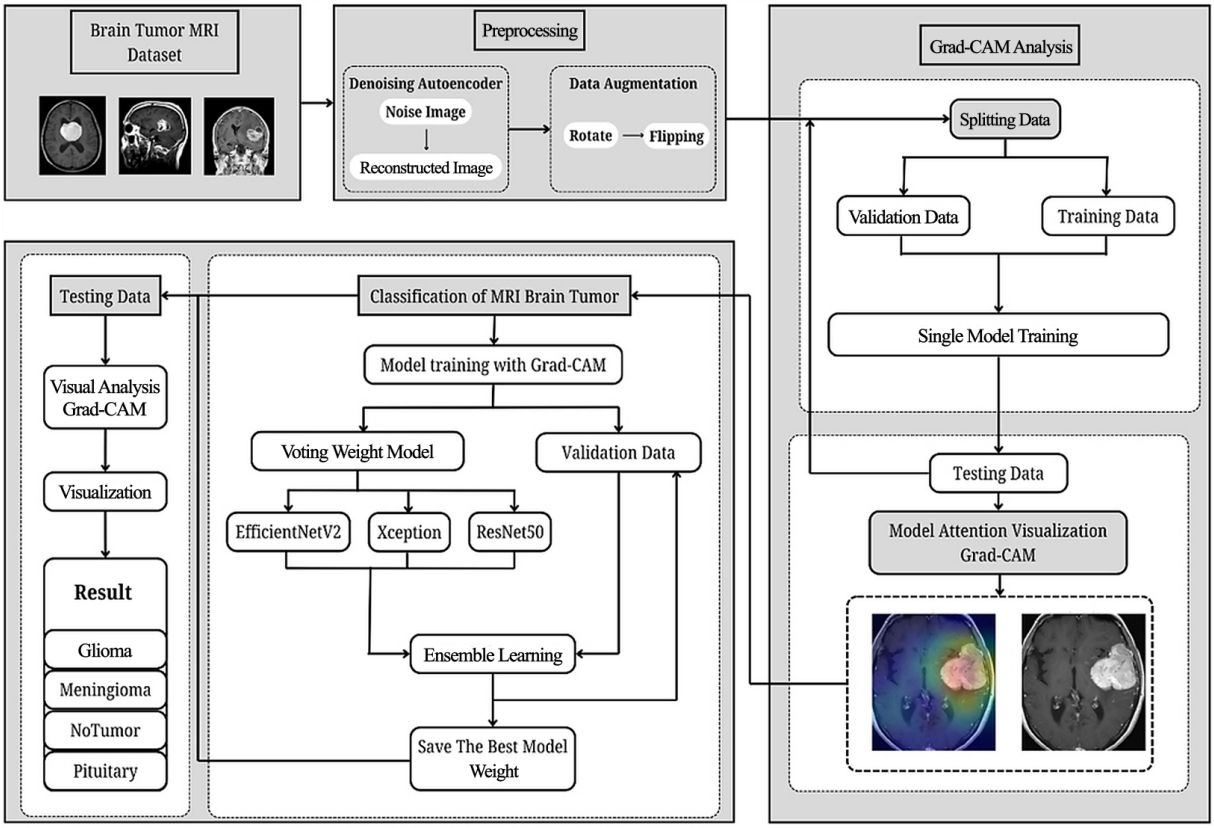

Figure 1 Research workflow in the MRI brain tumor image classification process

2.2 Preprocessing Denoising Image

The denoising autoencoder works on the MRI data by removing the details that create the noise in the images that are being enhanced (Subburaj and Bhavana, 2024; Pfaff et al., 2023). The model works by taking MRI images of organisms and has been constructed to output and reconstruct enhanced images. This process involves encoding and then decoding the images. The aim is to keep most of the information in the MRI images and discard the lesser detailed visual noise that can negatively affect the classifier performance (Radke et al., 2023; Murugan, 2023).

2.3 Data Augmentation

The purpose of data augmentation is to provide more variety in the image dataset to improve the performance of the model in the classification (Mikołajczyk and Grochowski, 2018). The augmentation techniques used in this study included random horizontal flipping, random rotation (±15°), and color jittering with brightness and contrast adjustments (Tatar et al., 2024; Mumuni and Mumuni, 2022).

These augmentation measures also keep the dataset diverse so that the model can learn from a broader variety of data to improve its performance (Shorten and Khoshgoftaar, 2019). Figure 4 (B) illustrates an example of the output from the MRI brain tumor image data augmentation.

2.4 Data Partitioning

All experiments use 5-fold stratified cross-validation for individual models to ensure statistically robust performance evaluation. In each fold, approximately 80% of the data is used for training and 20% for validation, with stratified splitting preserving class distribution across all folds (Gharaibeh, 2025). This design eliminates dependence on any single train-test split and provides stable performance estimates across data partitions. For the ensemble, cross-validation was additionally repeated across 3 independent random seeds, yielding 15 independent training-evaluation runs in total. All performance metrics are reported as mean standard deviation with 95% confidence intervals computed across fold scores, providing a statistically reliable summary of model performance.

2.5 Heat map area data using the Grad-CAM

The Grad-CAM visualization technique highlights critical regions in MRI images, such as tumor locations, that significantly contribute to the classification process. Grad-CAM allows the model to focus on critical features using spatial attention based on Grad-CAM activation maps (Shinde et al., 2021).

The process starts by performing a forward pass on the input image x to obtain a class prediction y_c from the model:

$$y_c = f(x) \tag{1}$$

A backward pass is then performed to calculate the gradient of the target y_c class against the feature map $\partial A_k(x)$ in the last convolutional layer:

$$\frac{\partial y_c}{\partial A_k(x)} \tag{2}$$

The channel weight α_k results from averaging the gradients:

$$\alpha_k = \frac{1}{HW} \sum_i \sum_j \frac{\partial y_c}{\partial A_k(x)} \tag{3}$$

The activation map of LGrad-CAM is obtained by summing the weights and feature map. ReLU function is then applied to ensure that only positive contributions are calculated:

$$L_{\text{Grad-Cam}} = \text{ReLU}\left(\sum_k \partial_k A_k(x)\right) \tag{4}$$

This activation map is then mapped back to the input image dimension x to produce an HGrad-CAM heatmap, which shows the important areas that affect the classification:

$$H_{\text{Grad-Cam}} = \text{Upsample}(L_{\text{Grad-Cam}}, x) \tag{5}$$

Grad-CAM provides an effective way to interpret CNN models, allowing visualization of image areas that most influence model decisions. It captures gradient-weighted activation maps for each predicted class without modifying the model weights. The resulting heatmaps are overlaid on input MRI images to highlight the most influential spatial regions in the classification decision, enabling post hoc interpretability of both individual model and ensemble predictions.

2.6 EfficienetV2S Model

The EfficientNetV2S model uses a combined approach with Fused-MBConv and MBConv. Fused-MBConv combines depthwise and standard convolutions, thereby improving the training efficiency in the early layers. MBConv is predominant in the late layers to expand the depth of the model (Priyadarshini et al., 2024).

EfficientNetV2S applies progressive learning, a strategy in which image size and regulation (such as dropout rate) gradually increase during training. This technique allows the model to reach convergence faster without sacrificing accuracy, as adaptive regulation maintains a balance between overfitting and underfitting at each training stage (R. Singh et al., 2025). The EfficientNetV2S family of models uses compound scaling expressed as follows:

$$F(\alpha, \beta, \gamma) = \alpha^d \times \beta^w \times \gamma^r \tag{6}$$

α = scale factor of depth
γ = scaling factor for the resolution
β = scale factor for the width

These factors are determined through an architectural search to optimize the balance between accuracy and computational efficiency. This composite scaling approach allows the model to be dynamically adjusted by simultaneously changing depth, width, and resolution to achieve the best performance. The EfficientNetV2S architecture and scaling principles significantly improve the performance of datasets such as ImageNet, CIFAR-10, and CIFAR-100, surpassing the previous ConvNet and Transformer models in terms of speed and accuracy (Sivaz and Aykut, 2024; Tan and Le, 2019).

2.7 Xception Model

The Xception model is named after its innovation of combining extreme inception with depthwise separable convolutions. The model ensures a better trade-off between the number of parameters and execution speed compared to regular convolutions (Lu et al., 2022). In this approach, the two principal steps of convolution (mapping of input to output values) are executed separately. The first is a depth convolution where a filter is applied individually to each channel, and the second is a pointwise convolution that combines the results of the first step and utilizes 1x1 filtering. This method is a great advantage to the model as it enables the model to focus more on efficient feature extraction (Muhammad et al., 2021). A Depthwise Convolution of an input X with dimensions H×W×C_in (height, width, and input channel) is conducted using a filter of size k×k with respect to each input channel.

$$Y_{dw}(i, j, c) = \sum_{m=-k/2}^{k/2} \sum_{n=-k/2}^{k/2} X(i+m, j+n, c) \times W_{dw}(m, n, c) \tag{7}$$

Here, i and j are the output pixel coordinates, and c is the input channel. Following depthwise convolution, pointwise convolution W_{pw} is executed with a $1 \times 1 \times C_{in} \times C_{out}$ sized filter to enhance the output.

$$Y_{pw}(i, j, c_{out}) = \sum_{c=1}^{C_{in}} Y_{dw}(i, j, c) \times W_{pw}(c, c_{out}) \tag{8}$$

In this situation, Xception employs a mix of convolution types, enabling both a decrease in parameters and an enhancement in the computational efficiency of the model.

2.8 ResNet50 Model

ResNet50 is a member of the Residual Networks (ResNet) family proposed by He et al., 2016. It is perhaps most well known for its implementation of residual or skip connections, which help alleviate the vanishing gradient phenomenon in severe deep learning models. The fundamental idea of ResNet is to use direct residual connections between the model's higher- and lower-order layers. This facilitates more effective training of the model by reducing the gradient flow and loss of valuable information. The residual connection can be formally defined as follows. Consider a residual block parameterized by an input x and output F(x). The residual block output is defined as the addition of the input and the output:

$$y = F(x) + x \tag{9}$$

F(x) typically takes a number of convolutional layers as input. ResNet50 has 50 layers, each with several residual blocks with 2 or 3 convolutional layers. ResNet50 increases network depth while maintaining training stability and enhancing performance on difficult image recognition tasks thanks to these residual connections. The model can streamline the optimization of its parameters by residual connections and avert the complications associated with deeper networks. The depth of the network architecture can be explained using a composite formula, which is also the case with ResNet50.

$$d_{\text{total}} = d_{\text{residual}} + d_{\text{skip connection}} \tag{10}$$

In this case, ResNet50 employs a significant depth of the network architecture, but overfitting and vanishing gradients are averted because of the residual connections. This model has shown outstanding performance on numerous datasets such as ImageNet and has become a standard for computer vision tasks such as object detection and image segmentation that require a high level of accuracy.

2.9 Ensemble Learning

The ensemble learning technique combines the outputs of all three trained backbone models: EfficientNetV2S, Xception, and ResNet50 to produce a final classification decision that outperforms any individual model. Each trained model independently processes an input MRI image and produces a class probability vector through its softmax output layer, representing the model's estimated likelihood for each of the four tumor categories: glioma, meningioma, no tumor, and pituitary.

The three softmax outputs are then merged through a weighted voting approach, where each model's probability vector is multiplied by an assigned weight that controls its relative contribution to the final prediction, and the three weighted vectors are summed to produce a single combined probability estimate for each class (Sterniczuk and Charytanowicz, 2024; Kesuma et al., 2023). The class with the highest combined probability across all three models is considered the final predicted label.

The contribution weights are determined through a per-fold grid search: for each cross-validation fold, all possible weight combinations at discrete steps are evaluated on the validation set, and the combination that maximizes validation accuracy for that specific data partition is selected (Sankar et al., 2024; Hosny et al., 2024). This allows the fusion weights to adapt across folds rather than applying a single fixed value to all data. As a reference baseline, weights can also be assigned in proportion to the mean cross-validation accuracy of each model, which in the final evaluation yields near-equal contributions from all three models, reflecting their comparable individual performance. The ensemble technique is expected to improve prediction accuracy and generalization to previously unseen data by leveraging the three architectures' complementary feature representations.

2.10 Evaluation

Different methods can be employed to evaluate the predictive performance of a classification system. One of these methods is the confusion matrix. A confusion matrix is used to plot prediction results and compare them with the classifications. A confusion matrix summarizes the performance of a model and offers key metrics such as accuracy, precision, recall, and the F1-score. All these metrics are based on the observed actual values (Diyasa et al., 2024; Lovell et al., 2023). The measurement matrix is composed of the following four values: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These values represent events with positive (P) or negative (N) probabilities. Based on these values, the evaluation metrics are calculated using equations (11), (12), (13), and (14).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

$$F1\text{-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{14}$$

Grad-Cam received an evaluation concerning the prediction and ground truth areas using the intersection over union (IoU) score, which is calculated by measuring the overlapping and union areas (Table 1). The formula is (15)

$$IoU = \frac{\text{Intersection}}{\text{Union}} = \frac{A \cap B}{A + B - A \cap B} \tag{15}$$

A = area of the ground truth (expert-annotated tumor region). B = is the area of the predicted region (Grad-CAM map).

Table 1 IoU score ranges and their interpretation

IoU Range	Description
IoU = 1	Perfect prediction.
IoU > 0.7	Good.
IoU 0.5–0.7	Sufficient.
IoU < 0.5	Model improvement

A = area of the ground truth (expert-annotated tumor region). B = is the area of the predicted region (Grad-CAM map).

3. Experiment and Results

The experimental setup was standardized across all three backbone architectures. Each model was trained with a maximum of 20 epochs, a batch size of 32, and the Adam optimizer (learning rate = 1e-4, weight decay = 5e-4), using a CosineAnnealingLR schedule (T_max = 20). To reduce overconfident predictions, label smoothing was applied to the cross-entropy loss. Early stopping (patience = 5; validation accuracy was monitored) was used to prevent overfitting. Mixed-precision training (16-bit) was used in all experiments. Transfer learning used ImageNet-pretrained weights with the last 2 feature blocks and classifier head unfrozen. The input images were resized per the model specifications: EfficientNetV2S at 384×384, Xception at 299×299, and ResNet50 at 224×224. The individual models were evaluated via 5-fold stratified cross-validation, and the ensemble was evaluated across 3 independent seeds × 5 folds = 15 independent training runs. Model training was performed using a single NVIDIA Tesla T4 GPU to accelerate deep learning computations through parallel processing. The system was equipped with approximately 30 GB of system memory (RAM), enabling efficient dataset and model training procedures. The training data were analyzed for accuracy and loss during evaluation, using cross-entropy for loss calculation. This technique aids in adopting a uniform evaluation paradigm, which fosters an equitable assessment of the models in relation to their performances and the identification of the model with the optimum performance in terms of loss and accuracy.

3.1 Classification structure for individual model experiments

Figure 2 illustrates the performance of EfficientNetV2s, Xception, and ResNet50 in the training and validation phases for brain tumor MRI classification. The training accuracy is represented by the blue line, and the validation accuracy is represented by the orange line. The first experiment featured individual models to understand the performance of each model on brain tumor MRIs. Model training on individual brain tumor MRI datasets is each model using the transfer learning technique for each respective architecture.

Figure 2 (A) shows that the EfficientNetV2s architecture exhibits a notable training accuracy that advanced in every epoch. The training accuracy started at 88% and advanced to 99% by the final epoch. The validation data accuracy also advanced to 99% from 88% at the start, demonstrating that the model can learn without overfitting. The training and validation loss graphs for this architecture are stable and low past, indicating the efficiency of this architecture in learning the data patterns.

Figure 2 (B) shows the volatility exhibited in the accuracy for the Xception architecture. The model achieved a training accuracy of 88% in the 1st epoch and 98% in the 10th epoch. On the other hand, the validation accuracy was much higher in the beginning, at 43% in the 3rd epoch and 9% in the 10th epoch, in addition to exhibiting high volatility in the 1st and 2nd epochs. Overall, despite the impressive final validation accuracy of 98%, the training of the Xception architecture requires more smoothing and stability in the training process.

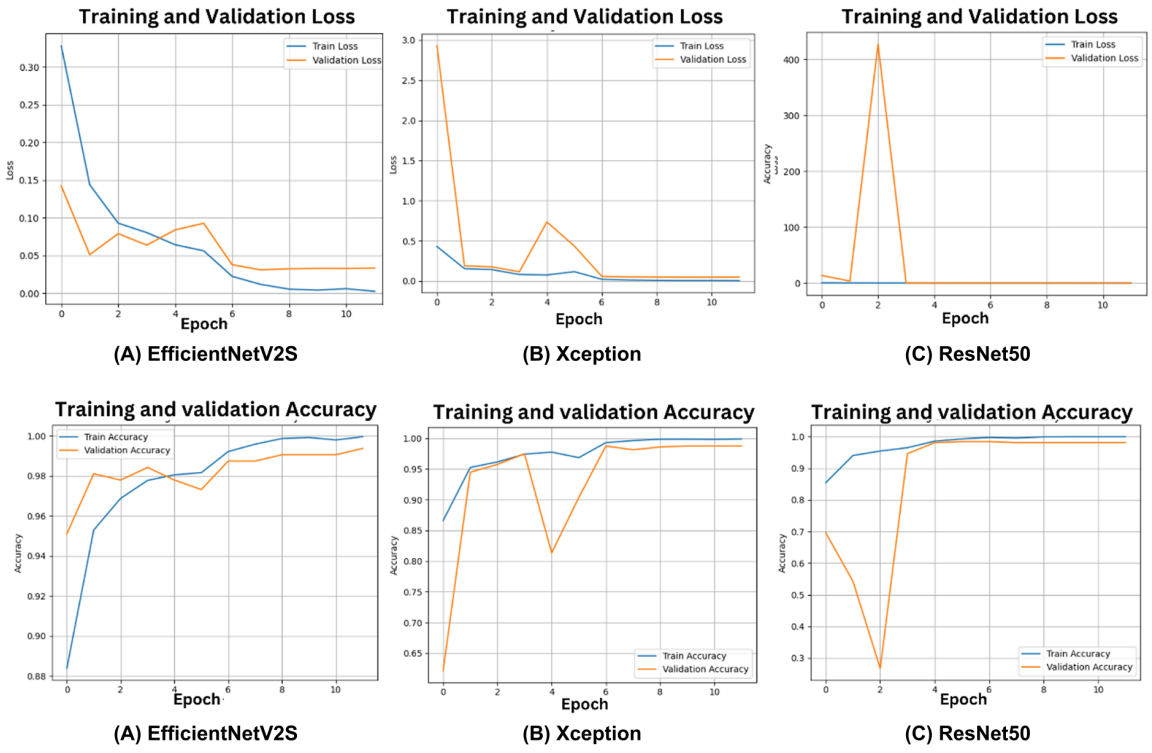

Figure 2 Training accuracy of different architectures

Figure 2 (C) shows the accuracy for the ResNet50 architecture. The accuracy of the training data for the architecture shows a consistent growth, going from 88% to 98% in the 10th epoch. However, the accuracy of the validation data shows it has certainly lagged behind. The validation accuracy also increases from 88% to 98%, but the training-to-validation loss ratio shows that the architecture could be under the overfitting regime. The architecture was and still is overfitting to the training data as it consistently held a higher training loss compared to the validation loss.

The final analysis indicates that EfficientNetV2S recorded the best results, exhibiting the most consistent graphs of accuracy and loss, followed by Xception, who, while attaining high

accuracy, left much to be desired regarding early training fluctuation. Xception also exhibited less than stellar final accuracy. Despite showing good final accuracy, the training of the ResNet50 demonstrated signs of overfitting and generalization loss on the validation data.

3.2 Ensemble learning classification

Figure 3 (A) shows a graph of the increase in train and validation accuracy, while Figure 3 (B) shows a graph on the left illustrating the decrease in train and validation losses as the number of epochs increases.

The accuracy graph shows a significant improvement in both the training and validation data. The accuracy starts from approximately 80% at the first epoch and increases rapidly to reach 99% at the fifth epoch. After that, the accuracy stabilizes until the end of training, indicating that the ensemble model has a consistent performance on the never-before-seen data (validation).

In the loss graph, the loss values for the training and validation data decrease dramatically in the early epochs, from around 3.0 to less than 0.1 at the fifth epoch, and remain stable until the end of training. This trend shows that the ensemble model can efficiently learn patterns from the data with a good level of generalization without signs of overfitting. The results of the confusion matrix test are shown in Figure 3 (C).

The confusion matrix confirms that the model accurately classifies most samples across all four classes. Glioma was correctly predicted in 167 samples, with only three errors for meningioma. No tumor had a perfect prediction of 203 samples. For meningioma, 172 samples were correctly predicted with two errors to the pituitary, whereas 154 samples were correctly predicted with two errors to no tumor and meningioma. This low error rate supports the model's high accuracy and superior performance. Overall, the ensemble approach of the three architectures provides superior performance over each individual model, resulting in a robust and reliable model for the brain tumor MRI image classification task.

This ROC plot illustrates the effective performance of the ensemble model (EfficientNetV2, Xception, ResNet50) in classifying brain tumors using MRI images. With a low FPR and a high TPR, the curve moves toward the upper left corner, which indicates the discrimination ability of the model. A weighted AUC-ROC of 0.9993 indicates that the model performs almost perfectly and does not overfit (Figure 3 (D)). The stability of the curve and the distance from the random diagonal line show that the ensemble learning approach enhances accuracy and generalizes well to the test data.

3.3 Brain tumor ensemble classification

Figure 4 (D) shows the comparison result of each model, and Table 2 shows a comparison of the accuracy values of each model. Based on the classification report in Figure 3 (D), the ensemble model achieved an overall accuracy of 99%, with equally high precision, recall, and F1 values for all four classes. The Grad-CAM visualization supports these results by highlighting the regions that contribute most to the model's decisions in classifying brain tumor MRI images. Red and yellow indicate areas with the highest contribution that correspond to the primary location of the tumor, respectively, while green and blue indicate moderate to low contribution. Areas in purple or without activation indicate an insignificant part in the model's decision, such as in the "No Tumor" image, which has no areas in red or yellow. The model seems to focus on the correct areas, such as in meningioma and glioma, where red and yellow activation is clearly visible around the tumor, whereas low activation indicates the absence of abnormal patterns in "No Tumor." Grad-CAM aids in the interpretation of deep learning models in medical diagnosis by ensuring that the model makes decisions based on relevant features.

Figure 4 (C) shows the results of brain tumor classification using a manually segmented mask from an expert or radiology specialist as ground truth. The first row shows the original MRI image, the second row shows the expert-validated tumor area mask, and the third row uses

Grad-CAM to visualize the model's focal area during prediction.

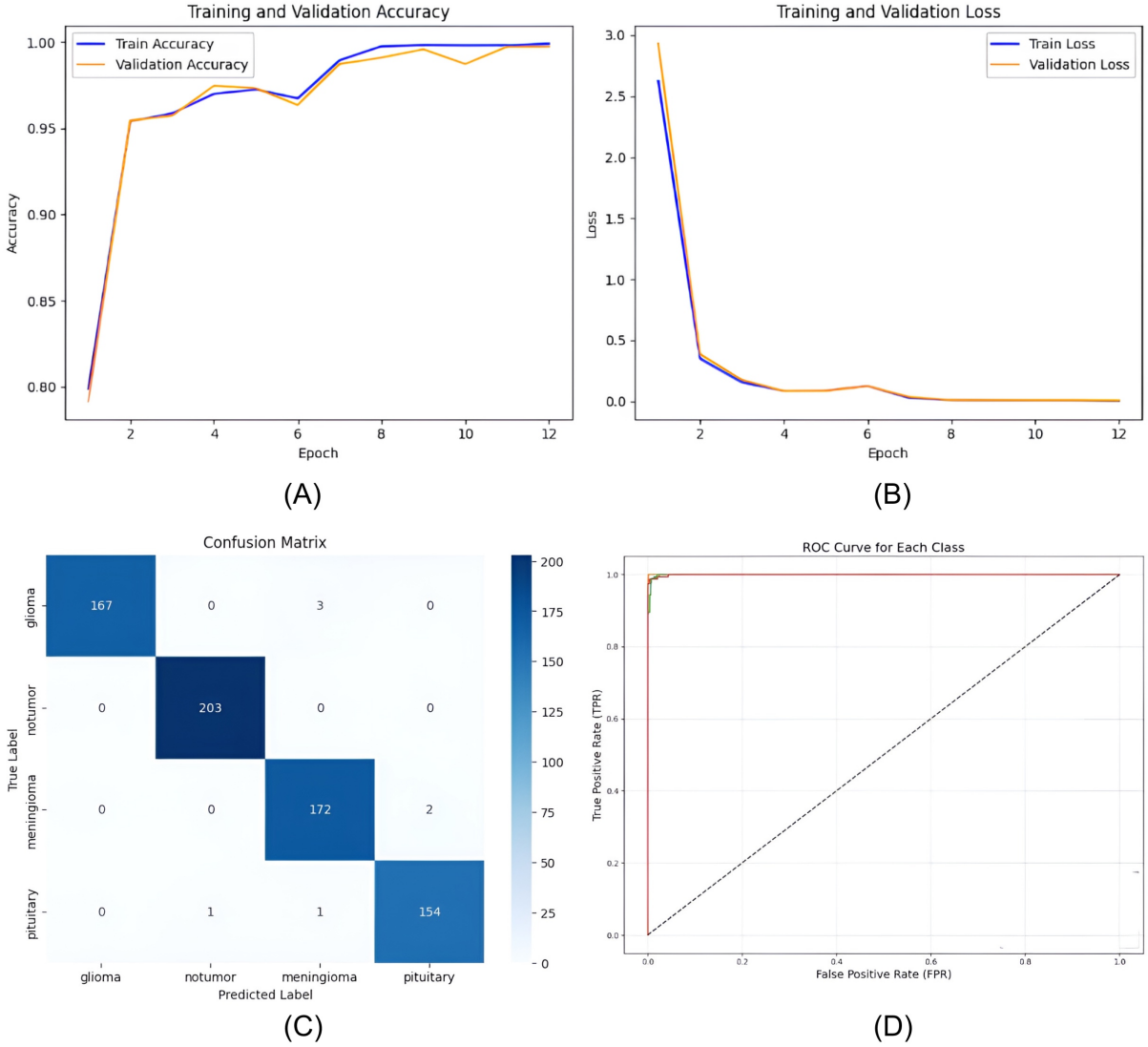

Figure 3 (A) Graph of the accuracy classification training process using the ensemble learning, (B) Graph of the validation classification training process using the ensemble learning, (C) Confusion matrix ensemble learning model, (D) AUC-ROC Result Graph of ensemble learning

Table 2 Training and validation loss (top row) and accuracy (bottom row) curves for (A) EfficientNetV2S, (B) Xception, and (C) ResNet50 across training epochs on the brain tumor MRI dataset

Model	Accuracy (%)	95% CI (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetV2S	98.84±0.52	[98.19, 99.48]	98.85	98.84	98.84
Xception	98.23±0.49	[97.63, 98.83]	98.25	98.23	98.23
ResNet50	98.82±0.46	[98.25, 99.39]	98.83	98.82	98.82
Ensemble Learning	99.23±0.40	[98.73, 99.73]	99.24	99.23	99.23
Ensemble Learning (3 seeds)	99.18±0.34	[98.99, 99.37]	99.19	99.18	99.18

The correct classification is shown for each category. The model performance was further evaluated using the IoU metric, as presented in Table 1. The IoU results presented in Table 3 show strong performance for most categories. The IoU for meningioma was 0.95, indicating almost perfect alignment between the Grad-CAM area and the expert-marked mask. Although the pituitary class received an alignment that was slightly lower than the rest, an IoU of 0.75

shows decent performance. 0.85 for glioma, which was also considered good. For No Tumor, the IoU was N/A because there was no region for comparison. Thus, the results show that the pituitary class is the only class that shows room for improvement, while the model does well to localize most the tumor classes.

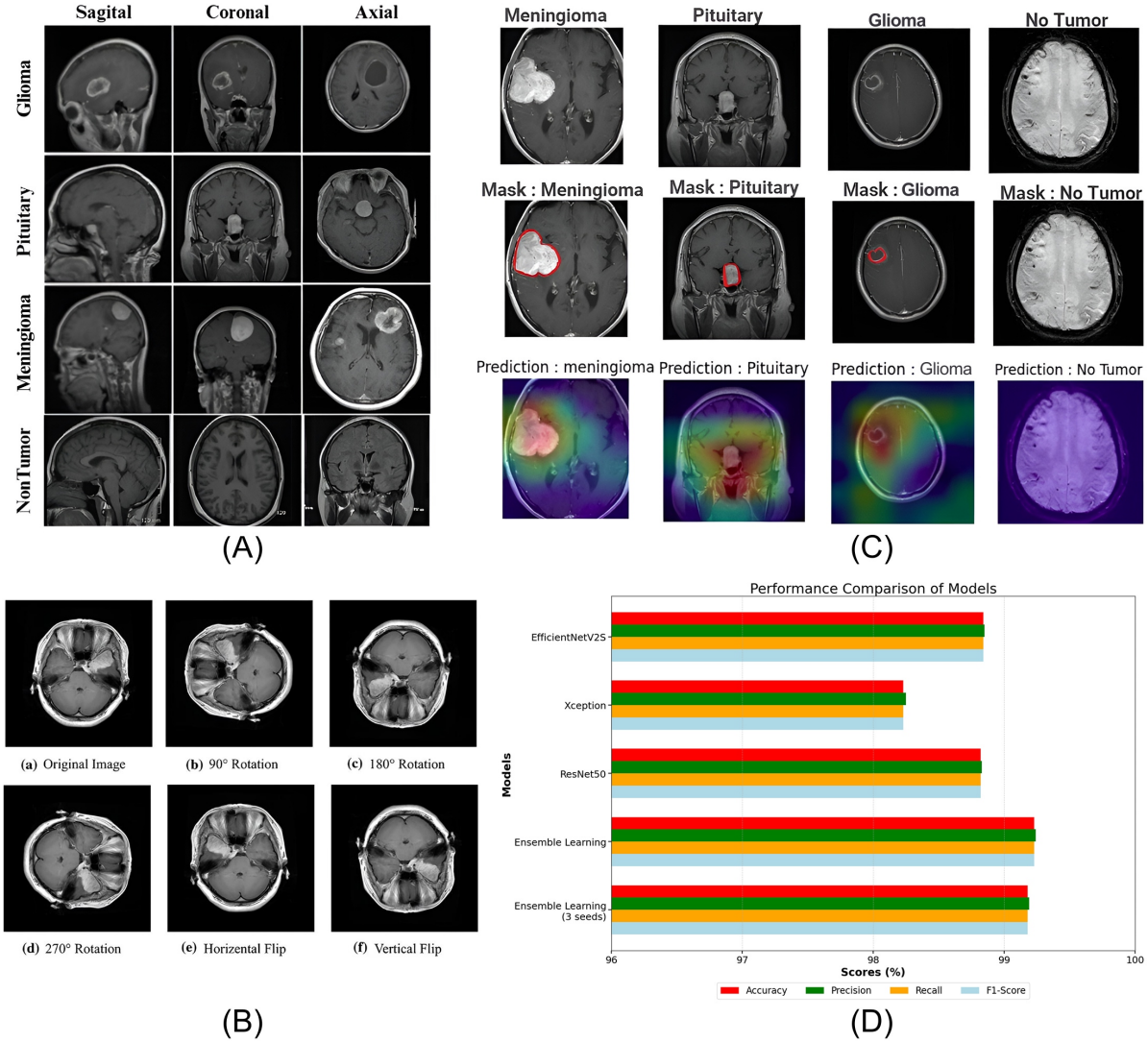

Figure 4 (A) Dataset sample images for each class, (B) data augmentation output, (C) Grad-CAM visualizations, and (D) model performance comparison

Table 3 IoU Score Output for Grad-CAM

Class	Expert Marked Area (A)	Grad-CAM Area (B)	Intersection (A ∩ B)	Union (A ∪ B)	IoU
Meningioma	200	210	200	210	0.95
Pituitary	150	200	150	200	0.75
Glioma	140	165	140	165	0.85
No Tumor	0	0	0	0	N/A

3.4 Evaluation of the model

This study assesses the ensemble learning + Grad-CAM method while juxtaposing it with earlier works. The results of brain tumor classification and MRI are presented in Table 4.

Table 4 Comparison of image classification results with those of previous studies

Method	Year	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Ensemble deep learning (Aziz et al., 2021)	2021	99.02	98.75	98.98	98.86
The ensemble attention mechanism (Pacal et al., 2024)	2024	98.94	98.91	98.93	98.91
Comparative analysis of Ensemble Learning (Sharif et al., 2024)	2024	99.33	100	98.32	99.15
Ensemble transfer learning F1-WMVE (Sterniczuk and Charytanowicz, 2024)	2024	94.14	94.11	94.02	94.09
Ensemble Learning + Grad-CAM	**2025**	**99.18**	**99.19**	**99.18**	**99.18**

The proposed ensemble learning + Grad-CAM method achieved a mean cross-validated accuracy of 99.18% ± 0.34%, with macro precision of 99.19%, recall of 99.18%, and F1-score of 99.18%. As shown in Table 4, this result is competitive with the current state of the art. It exceeds the ensemble attention mechanism (Pacal et al., 2024) at 98.94% and the ensemble transfer learning F1-WMVE method (Sterniczuk and Charytanowicz, 2024) at 94.14%, and is comparable to the comparative ensemble analysis (Sharif et al., 2024) at 99.33% and the ensemble deep learning approach (Aziz et al., 2021) at 99.02%.

A key distinction of the proposed method is that its 99.18% figure is derived from 15 independent cross-validation runs rather than a single train-test split, providing a statistically conservative and reproducible estimate. The compared methods report single-split results, which may reflect more favorable data partitions. The integration of Grad-CAM with quantitative IoU validation adds a clinical interpretability dimension not present in the compared methods, making the proposed framework more transparent for potential diagnostic support applications. The primary remaining error cases involve meningioma–glioma boundary misclassifications, consistent with the overlapping MRI morphology of these two classes.

4. Conclusions

In this study, we integrated Grad-CAM and EL to optimize multi-section MRI-based brain tumor classification under a repeated cross-validation evaluation framework. When EfficientNetV2S, Xception, and ResNet50 were combined through per-fold grid-search weighted voting, classification accuracy and reliability were consistently improved compared with those of individual models. The proposed method achieved a mean accuracy of 99.18% ± 0.34%, a precision of 99.19%, a recall of 99.18%, and an F1 score of 99.18%, with a weighted AUC-ROC of 0.9993, demonstrating a strong capability to discriminate between tumor classes with minimal classification errors. The use of multiple complementary architectures reduced overfitting and improved generalization, as reflected by stable performance across 15 independent cross-validation runs. The model provided a high AUC-ROC score along with a stable ROC curve, indicating that the data were learned rather than memorized. Grad-CAM also improves model interpretability, which is especially critical for models used in the medical field, as it illustrates which parts of the MRI the model is analyzing in which regions. The current research strengthens the advances in AI-aided diagnostics, providing the potential for faster and more accurate medical analysis. In the future, larger data sets should be incorporated and ensembles applied to other diseases to increase the scope of clinical applicability.

Acknowledgements

The Institute for Research and Community Service (LPPM) at Universitas Pembangunan Nasional Veteran Jawa Timur supported this work under Grant Number SPP/207/UN.63.8/LT/IX/2025.

Conflict of Interest

The authors declare no conflicts of interest.

Declaration of AI

Artificial intelligence tools were used for grammar checking and language editing. All scientific content and analysis were performed and validated by the authors.

References

Alsubai, S., Khan, H. U., Alqahtani, A., Sha, M., Abbas, S., & Mohammad, U. G. (2022). Ensemble deep learning for brain tumor detection. *Frontiers in Computational Neuroscience*, *16*, 1–14. https://doi.org/10.3389/fncom.2022.1005617

Alther, B., Mylius, V., Weller, M., & Gantenbein, A. (2020). From first symptoms to diagnosis: Initial clinical presentation of primary brain tumors. *Clinical and Translational Neuroscience*, *4*(2), 2514183X20968368. https://doi.org/10.1177/2514183x20968368

Amarnath, A., Bataineh, A. A., & Hansen, J. A. (2024). Transfer-learning approach for enhanced brain tumor classification in mri imaging. *BioMedInformatics*, *4*(3), 1745–1756. https://doi.org/10.3390/biomedinformatics4030095

Aziz, A., Attique, M., Tariq, U., Nam, Y., Nazir, M., Jeong, C. W., Mostafa, R. R., & Sakr, R. H. (2021). An ensemble of optimal deep learning features for brain tumor classification. *Computers, Materials & Continua*, *69*(2), 2653–2670. https://doi.org/10.32604/cmc.2021.018606

Basha, S. A. K., Vincent, P. M. D. R., Mohammad, S. I., Vasudevan, A., Soon, E. E. H., Shambour, Q., & Alshurideh, M. T. (2025). Exploring deep learning methods for audio speech emotion detection: An ensemble mfccs, cnns and lstm. *Applied Mathematics & Information Sciences*, *19*(1), 75–85. https://doi.org/10.18576/amis/190107

Çetin-Kaya, Y., & Kaya, M. (2024). A novel ensemble framework for multi-classification of brain tumors using magnetic resonance imaging. *Diagnostics*, *14*(4), 383. https://doi.org/10.3390/diagnostics14040383

Dishar, H. K., & Muhammed, L. A. (2023). Detection brain tumor disease using a combination of xception and nasnetmobile. *International Journal of Advanced Soft Computing and its Applications*, *15*(2), 325–336. https://doi.org/10.15849/IJASCA.230720.22

Diyasa, I. G. S. M., Saputra, W. S. J., Gunawan, A. A. N., Herawati, D., Munir, S., & Humairah, S. (2024). Abnormality determination of spermatozoa motility using gaussian mixture model and matching-based algorithm. *Journal of Robotics and Control*, *5*(1), 103–116. https://doi.org/10.18196/jrc.v5i1.20686

Gawali, B. S., & Dhongade, V. S. (2024). Enhancing brain tumor classification: A cnn-based approach with inceptionv3 and xception. *International Journal of Advanced Research in Science, Communication and Technology*, *4*(2), 492–504. https://doi.org/10.48175/ijarsct-18174

Ghaffar, A., Javid, M. A., Arshad, S., & Azeem, W. (2024). Enhanced efficientnet model for multiclass brain tumor prognostication using advanced mr image analysis techniques. *Research Square*, 1–12. https://doi.org/10.21203/rs.3.rs-4809509/v2

Gharaibeh, N. (2025). Enhancing interpretability in brain tumor detection: Leveraging grad-cam and shap for explainable ai in mri-based cancer diagnosis. *Applied Computer Science*, *21*(3), 182–197. https://doi.org/10.35784/acs_7375

Ghazvini, M., Vahab, D., Arash, P., & Meysam, S. (2024). Diagnosis and classification of brain tumors from mri images using the svm algorithm. *Journal of Clinical Research in Paramedicine Sciences*, *13*(1), 1–8. https://doi.org/10.5812/jcrps-148703

Gokila, P., Nagarasan, M., Anandhi, R., Manikandan, B., Sharan, V. J., & Daniel, S. S. (2024). Efficientnet transfer learning approach for multi-class brain tumor classification. *International Journal of Scientific Research in Engineering and Management*, *8*(5), 1–5. https://doi.org/10.55041/ijsrem32835

Guo, X., Liu, T., & Chi, Q. (2024). Brain tumor diagnosis in mri scans images using residual/shuffle network optimized by augmented falcon finch optimization. *Scientific Reports*, *14*(1), 1–21. https://doi.org/10.1038/s41598-024-77523-2

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hosny, K. M., Mohammed, M. A., Salama, R. A., & Elshewey, A. M. (2024). Explainable ensemble deep learning-based model for brain tumor detection and classification. *Neural Computing and Applications*, *37*(3), 1289–1306. https://doi.org/10.1007/s00521-024-10401-0

Ismail, K. A. A., Dutta, A. K., & Sait, A. R. W. (2024). Ensemble learning-based multiple sclerosis detection technique using magnetic resonance imaging. *Journal of Disability Research*, *3*(6), 1–8. https://doi.org/10.57197/jdr-2024-0078

Kesuma, L. I., Ermatita, & Erwin. (2023). Elrei: Ensemble learning of resnet, efficientnet, and inception-v3 for lung disease classification based on chest x-ray image. *International Journal of Intelligent Engineering and Systems*, *16*(5), 149–161. https://doi.org/10.22266/ijies2023.1031.14

Lapointe, S., Perry, A., & Butowski, N. A. (2018). Primary brain tumours in adults. *The Lancet*, *392*(10145), 432–446. https://doi.org/10.1016/S0140-6736(18)30990-5

Lovell, D., Miller, D., Capra, J., & Bradley, A. (2023). Never mind the metrics – what about the uncertainty? visualising binary confusion matrix metric distributions to put performance in perspective. *Proceedings of the 40th International Conference on Machine Learning (ICML)*, *202*, 22702–22757. https://proceedings.mlr.press/v202/lovell23a.html

Lu, G., Zhang, W., & Wang, Z. (2022). Optimizing depthwise separable convolution operations on gpus. *IEEE Transactions on Parallel and Distributed Systems*, *33*(1), 70–87. https://doi.org/10.1109/TPDS.2021.3084813

Mahdi, H. A., Shujaa, M. I., & Zghair, E. M. (2023). Diagnosis of medical images using fuzzy convolutional neural networks. *Mathematical Modelling of Engineering Problems*, *10*(4), 1345–1351. https://doi.org/10.18280/mmep.100428

Mathivanan, S. K., Sonaimuthu, S., Murugesan, S., Rajadurai, H., Shivahare, B. D., & Shah, M. A. (2024). Employing deep learning and transfer learning for accurate brain tumor detection. *Scientific Reports*, *14*(1), 1–15. https://doi.org/10.1038/s41598-024-57970-7

Mikołajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, 117–122. https://doi.org/10.1109/IIPHDW.2018.8388338

Muhammad, W., Aramvith, S., & Onoye, T. (2021). Multi-scale xception based depthwise separable convolution for single image superresolution. *PLoS One*, *16*(8), 1–20. https://doi.org/10.1371/journal.pone.0249278

Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, *16*, 100258. https://doi.org/10.1016/j.array.2022.100258

Murugan, D. V. (2023). Image enhancement of magnetic resonance imaging under clustering environment. In *Future trends in computer technology and data science* (pp. 118–126, Vol. 2). https://doi.org/10.58532/v2bs18p2ch4

Musa, M. N., Sanusi, M. B., Odion, P., & Shitu, S. S. (2024). Mri-based brain tumor classification using resnet-50 and optimized softmax regression. *Jurnal Infotel*, *16*(3), 598–614. https://doi.org/10.20895/INFOTEL.V16I3.1175

Nazir, M. I., Akter, A., Wadud, M. A. H., & Uddin, M. A. (2024). Utilizing customized cnn for brain tumor prediction with explainable ai. *Heliyon*, *10*(20), e38997. https://doi.org/10.1016/j.heliyon.2024.e38997

Nickparvar, M. (2023). Brain tumor mri dataset [Kaggle dataset]. https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

Pacal, I., Celik, O., Bayram, B., & Cunha, A. (2024). Enhancing efficientnetv2 with global and efficient channel attention mechanisms for accurate mri-based brain tumor classification. *Cluster Computing*, *27*(8), 11187–11212. https://doi.org/10.1007/s10586-024-04532-1

Pfaff, L., Hossbach, J., Preuhs, E., Wagner, F., Camejo, S. A., Kannengiesser, S., Nickel, D., Wuerfl, T., & Maier, A. (2023). Self-supervised mri denoising: Leveraging stein's unbiased risk estimator and spatially resolved noise maps. *Scientific Reports*, *13*(1), 1–13. https://doi.org/10.1038/s41598-023-49023-2

Priyadarshini, P., Kanungo, P., & Kar, T. (2024). Multigrade brain tumor classification in mri images using fine tuned efficientnet. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, *8*, 100498. https://doi.org/10.1016/j.prime.2024.100498

Radke, K. L., Kamp, B., Adriaenssens, V., Stabinska, J., Gallinnis, P., Wittsack, H. J., Antoch, G., & Müller-Lutz, A. (2023). Deep learning-based denoising of cest mr data: A feasibility study on applying synthetic phantoms in medical imaging. *Diagnostics*, *13*(21), 3326. https://doi.org/10.3390/diagnostics13213326

Sankar, M., Baiju, B., Preethi, D., Kumar, A. S., Mathivanan, S. K., & Shah, M. A. (2024). Efficient brain tumor grade classification using ensemble deep learning models. *BMC Medical Imaging*, *24*(1), 297. https://doi.org/10.1186/s12880-024-01476-1

Sharif, R., Azam, M., Ali, A., Hashmi, M. U., & Uzair, M. (2024). Comparative analysis of ensemble learning techniques for brain tumor classification. *Informatica*, *48*(20), 41–50. https://doi.org/10.31449/inf.v48i20.6714

Shinde, S., Tupe-Waghmare, P., Chougule, T., Saini, J., & Ingalhalikar, M. (2021). Predictive and discriminative localization of pathology using high resolution class activation maps with cnns. *PeerJ Computer Science*, *7*, 1–14. https://doi.org/10.7717/peerj-cs.622

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0197-0

Singh, R., Prabha, C., Malik, M., & Goyal, A. (2025). A robust deep learning model for brain tumor detection and classification using efficient net: A brief meta-analysis. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *49*(2), 26–51. https://doi.org/10.37934/araset.49.2.2651

Singh, T., Nair, R. R., Babu, T., Wagh, A., Bhosalea, A., & Duraisamy, P. (2024). Brainnet: A deep learning approach for brain tumor classification. *Procedia Computer Science*, *235*, 3283–3292. https://doi.org/10.1016/j.procs.2024.04.310

Sivaz, O., & Aykut, M. (2024). Combining efficientnet with ml-decoder classification head for multi-label retinal disease classification. *Neural Computing and Applications*, *36*(23), 14251–14261. https://doi.org/10.1007/s00521-024-09820-w

Sterniczuk, B., & Charytanowicz, M. (2024). An ensemble transfer learning model for brain tumors classification using convolutional neural networks. *Advances in Science and Technology Research Journal*, *18*(8), 204–216. https://doi.org/10.12913/22998624/193627

Subburaj, T., & Bhavana, S. (2024). Image noise reduction with auto-encoders using tensorflow. *International Journal of Advanced Research in Science, Communication and Technology*, 86–91. https://doi.org/10.48175/ijarsct-19016

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality

worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Tan, M., & Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning*, *97*, 10691–10700. https://proceedings.mlr.press/v97/tan19a.html

Tandel, G. S., Biswas, M., Kakde, O. G., Tiwari, A., Suri, H. S., Turk, M., Laird, J. R., Asare, C. K., Ankrah, A. A., Khanna, N. N., Madhusudhan, B. K., Saba, L., & Suri, J. S. (2019). A review on a deep learning perspective in brain cancer classification. *Cancers*, *11*(1), 1–32. https://doi.org/10.3390/cancers11010111

Tatar, A., Haghighi, M., & Zeinijahromi, A. (2024). Experiments on image data augmentation techniques for geological rock type classification with convolutional neural networks. *Journal of Rock Mechanics and Geotechnical Engineering*. https://doi.org/10.1016/j.jrmge.2024.02.015

Vo, H. T., Thien, N. N., Mui, K. C., & Tien, P. P. (2024). Enhancing confidence in brain tumor classification models with grad-cam and grad-cam++. *Indonesian Journal of Electrical Engineering and Informatics*, *12*(3), 926–939. https://doi.org/10.52549/ijeei.v12i3.5977

Yoon, S. (2025). Brain tumor classification using a hybrid ensemble of xception and parallel deep cnn models. *Informatics in Medicine Unlocked*, *54*, 101629. https://doi.org/10.1016/j.imu.2025.101629

Younis, A., Qiang, L., Nyatega, C. O., Adamu, M. J., & Kawuwa, H. B. (2022). Brain tumor analysis using deep learning and vgg-16 ensembling learning approaches. *Applied Sciences*, *12*(14), 7282. https://doi.org/10.3390/app12147282

Zhang, Z., Li, G., Xu, Y., & Tang, X. (2021). Application of artificial intelligence in the mri classification task of human brain neurological and psychiatric diseases: A scoping review. *Diagnostics*, *11*(8), 1402. https://doi.org/10.3390/diagnostics11081402