Comparative analysis of Xception and SVM for brain tumor classification on MRI images

Nurul Huda [a,b,1,*], Ku Ruhana Ku-Mahamud [c,2]

[a] Universiti Muhammadiyah Malaysia, 02100 Padang Besar, Perlis, Malaysia
[b] Institut Teknologi Statistika dan Bisnis Muhammadiyah Semarang, Semarang, Indonesia
[c] Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia
[1] p5240024@student.umam.edu.my; [2] ruhana@uum.edu.my
* corresponding author

ARTICLE INFO

Article history
Received March 4, 2025
Revised October 21, 2025
Accepted December 14, 2025
Available online February 28, 2026

Keywords
Brain Tumor Classification
Convolutional Neural Network
Deep Learning
Magnetic Resonance Imaging
Xception Architecture

ABSTRACT

Brain tumor classification from magnetic resonance imaging (MRI) plays a critical role in supporting radiologists during diagnosis and treatment planning. However, many existing automated approaches employ limited preprocessing, single-stage transfer learning, or evaluation on a single dataset, which restricts robustness and clinical applicability. This study proposes an enhanced transfer-learning framework based on the Xception architecture for multiclass brain tumor classification and compares its performance with baseline models under identical experimental conditions. The framework integrates a comprehensive preprocessing pipeline consisting of normalization, adaptive noise filtering, contrast enhancement, and targeted data augmentation, together with a structured two-phase fine-tuning strategy. A total of 6,537 MRI images were used, employing five-fold cross-validation, independent testing, and validation on an additional benchmark dataset. The proposed model achieved a mean cross-validation accuracy of 0.8994 ± 0.089 and 99.06% accuracy, precision, and recall on the independent test set, demonstrating strong stability and generalization ability. Evaluation on the CE-MRI Figshare dataset further confirmed robustness, yielding 98.45% accuracy, 98% precision, and 98% recall. In contrast, when re-evaluated in the same experimental setting, baseline models performed considerably worse: the SVM classifier achieved 21.41% accuracy, and ResNet50 reached 75.27%, both substantially lower than Xception. Although higher accuracies for these models have been reported in prior studies under different conditions, the present findings highlight their limited generalization under unified evaluation. Overall, the proposed Xception-based framework provides a reliable and generalizable solution for automated brain tumor classification, with strong potential to support clinical workflows such as triage prioritization and second-opinion assistance.

1. Introduction

Brain tumors represent one of the most critical neurological disorders, associated with high morbidity and mortality rates [1]. These abnormal growths, whether benign or malignant, can significantly impair neurological function, with malignant tumors posing greater risks owing to their aggressive and infiltrative nature [2]. Early and accurate classification of brain tumors is crucial for effective diagnosis, prognosis, and treatment planning [3]. Advanced imaging techniques, particularly MRI, have become the gold standard for brain tumor detection and evaluation due to their enhanced soft-tissue contrast and their ability to delineate tumor morphology and anatomical structures in detail [4]. Despite these advantages, the manual classification of MRI images remains labor-intensive, time-consuming, and

subject to inter-observer variability, highlighting the necessity for automated and reliable classification systems [5], [6].

Deep learning has emerged as a transformative approach in medical image analysis, particularly through CNN, which excel at extracting complex hierarchical features from images. Numerous studies have demonstrated the effectiveness of CNN in handling diverse tasks, such as object recognition, segmentation, and tumor classification [7]–[9]. CNN has been widely applied to brain tumor classification, offering a robust alternative to traditional machine learning methods. However, persistent challenges, such as tumor heterogeneity [10] , textural similarities between normal and pathological tissues [11], and differences in imaging modalities [12]–[14], continue to limit model performance and generalizability.

Several recent studies have sought to address these challenges using diverse CNN-based frameworks. Sarkar et al. proposed a hybrid CNN approach using AlexNet as a feature extractor combined with classical machine learning classifiers, achieving accuracies of up to 100% with AlexNet-RF [15]. Despite these promising results, reliance on small datasets raises concerns regarding overfitting and limited real-world applicability [16], [17]. Similarly, Malakouti et al. introduced an ensemble feature fusion strategy that leverages multiple pre-trained CNNs with Support Vector Machine (SVM) and Random Forest (RF) classifiers, demonstrating strong performance on large MRI datasets. Nonetheless, issues of class imbalance and susceptibility to overfitting persist, undermining clinical generalization. Azzahra et al. reported 93% accuracy using a single CNN, although the misclassification of meningioma tumors underscored the difficulty of distinguishing morphologically similar tumor types [18], [19]. In another study, Alanazi et al. developed isolated CNN models from scratch and later extended them via transfer learning for multiclass classification. While achieving accuracies exceeding 95%, their models were heavily dependent on machine-specific MRI characteristics, which limited their robustness across diverse imaging conditions [20], [21]. Traditional hybrid approaches have also been explored, for example, Ayadi et al. integrated intensity normalization with handcrafted feature extraction (DSURF, HoG) and SVM classification, reporting an accuracy of 90.27 % [22]. While statistically reliable, this method still faces limitations in handling tumor heterogeneity and dataset diversity.

Other researchers have proposed alternative deep learning and hybrid solutions. Afshar et al. [23] employed Capsule Networks (CapsNet) for brain tumor classification, addressing several inherent limitations of conventional CNNs, but the model achieved only 86.56% accuracy on 3,064 MRI images. Lakshmi and Rao [24] applied the Inception-v3 architecture for brain tumor detection, achieving 89% accuracy on the same dataset but with limited validation accuracy and restricted classification of only three tumor categories. Garg and Garg [25] introduced A hybrid approach integrating RF, K-Nearest Neighbor (KNN), and Decision Tree (DT) was proposed, achieving 97.3% accuracy on 2,556 MRI images. However, it was limited to binary classification of benign versus malignant tumors. Deepak and Ameer [26] developed an automated brain tumor classification framework using deep transfer learning combined with SVM and KNN classifiers, reaching 97.1% accuracy on 3,064 MRI images, but their approach was prone to misclassification, particularly for meningioma, and exhibited overfitting with smaller training sets.

The consistent limitations of these studies emphasized the need for improved frameworks that addressed tumor heterogeneity, mitigated data imbalance, and minimized overfitting while ensuring generalization across diverse datasets. Transfer learning offered a promising solution, enabling the use of large-scale pretrained CNN architectures for domain-specific tasks. Recent advances in CNN, such as Xception, which adopts depthwise separable convolutions to optimize feature extraction performance, demonstrated strong potential for medical imaging applications [27]. Kesav et al. highlighted the versatility of CNN in visual recognition and medical imaging [5], while other researchers [4], [28], [29] explored architectural enhancements and integrations with RNN. Transfer learning approaches further accelerated convergence and improved performance in the context of data scarcity [30], [31]. Building on these insights, this study adopted the Xception model within a transfer learning framework to enhance the accuracy and generalizability of brain tumor classification.

Table 1 summarizes recent studies on brain tumor classification using deep learning and hybrid machine learning approaches. Overall, CNN-based models such as Inception-v3 [24] CapsNet [23] demonstrated moderate performance, with accuracies ranging from 86% to 89%, indicating the challenges of relying solely on single deep learning architectures. Hybrid approaches that combine CNNs with classical machine learning classifiers, including SVM, KNN, and Random Forest, have achieved higher accuracies, often exceeding 97% [25], [26]. However, these methods were frequently limited by small or imbalanced datasets, leading to overfitting and poor generalization, as reported by Sarkar and Malakouti [15], [32]. More recent ensemble and transfer learning frameworks [20], [32] achieved robust multiclass classification, with high accuracies, though performance remained sensitive to dataset heterogeneity and machine-specific imaging characteristics. Collectively, these studies highlight both the potential and limitations of existing methods, underscoring the need for models that achieve high accuracy while generalizing across diverse tumor types and imaging conditions.

Table 1. Overview of Existing Techniques of Brain Tumor Classification

Techniques	Brain Tumor Classifications			
	∑MRI Datasets	*Accuracy*	*Contributions*	*Limitations*
AlexNet + ML classifiers (RF, BayesNet, SMO, Naïve Bayes) [15]	3.600	Up to 100% (AlexNet-RF)	Demonstrated hybrid DL-ML effectiveness	Overfitting due to a small dataset, poor generalizability
Ensemble of multiple pretrained CNNs + SVM/RF [32]	3.762	High (not specified)	Strong performance with feature fusion	Class imbalance, overfitting
Single CNN architecture [18]	3.265	93%	Showed CNN's effectiveness	Misclassification of meningioma
Isolated CNN models (later transfer learning) [20]	3.000	>95%	Robust multiclass classification	Dependent on machine-specific MRI features
Hybrid (Intensity normalization + DSURF + HoG + SVM) [22]	3.064	90.27%	Handcrafted + ML hybrid	Limited in handling heterogeneity & dataset diversity
CNN+CapsNet [23]	3.064	86.56%	Applied Capsule Networks to overcome CNN limitations	Relatively low accuracy
Inception-V3 [24]	3.064	89%	Applied Inception-v3, improving over handcrafted features.	Low validation accuracy, limited to 3 tumor types.
Random Forest, K-Nearest Neighbour, Decision Tree [25]	2.556	97.30%	Combined ML classifiers for enhanced accuracy.	Restricted to binary classification (benign vs malignant).
Deep Transfer Learning, Support Vector Machine, K-Nearest Neighbor [26]	3.064	97.1%	Automated system integrating DL and ML classifiers.	Misclassification in meningioma, overfitting on small data.

Despite significant advances in deep learning for medical imaging, many current methods for brain tumor classification still struggle to achieve both high accuracy and reliable generalization across different tumor types. To overcome these limitations, this study proposes an Xception-based transfer-learning framework for automated brain tumor classification from MRI scans. The contributions of this work are threefold:

- the implementation of a comprehensive preprocessing and augmentation pipeline to enhance model generalization,
- the adoption of a two-phase training strategy that balances efficiency with improved feature representation, and
- a comparative evaluation against conventional classifiers and transfer learning models to validate the proposed approach.

By addressing both accuracy and generalization, this study aims to advance the development of reliable computer-aided diagnostic systems for clinical applications in brain tumor detection. Automated multiclass brain tumor classification can improve radiological workflows by increasing diagnostic efficiency, helping in case triage, and providing quick second-opinion guidance, especially in areas with limited access to experts. By clearly distinguishing between gliomas, meningiomas, pituitary tumors, and no-tumor categories, these systems can reduce reporting times and differences in interpretation among observers. However, even with accuracy rates close to 99%, there is still a risk of misclassification due to unusual cases, artifacts, or variations in imaging quality. Therefore, it is important to see the model as a helpful tool instead of a replacement for diagnosis. Future efforts that incorporate methods to better understand, estimate uncertainty, and validate across multiple centers will be crucial for building trust, transparency, and preparation for clinical use.

2. Method

This section presents the methodological framework adopted to develop and evaluate the proposed brain-tumor classification model. The framework comprises four main stages: dataset preparation, preprocessing, model development, and performance evaluation. Each stage was designed to ensure reproducibility, robustness, and clinical relevance.

2.1. Research Framework

Fig. 1 presents an overview of the research workflow. The workflow starts with acquiring a publicly available brain MRI dataset, which is subsequently split into training, validation, and test sets. Image resizing, normalization, and augmentation were applied during preprocessing to improve generalization. The model development stage, Xception architecture was employed within a transfer learning paradigm using a two-phase training strategy: (i) freezing the convolutional base to optimize the fully connected classification head and (ii) fine-tuning the deeper layers with a reduced learning rate. Finally, the performance was evaluated using standard metrics (accuracy, precision, recall, and F1-score) with 5-fold cross-validation, and the results were compared with those of ResNet50 and SVM classifiers.

Algorithm : Research Framework for Brain Tumor Classification using Xception Transfer Learning

Input: MRI dataset D with four classes {glioma, meningioma, pituitary, no tumor}
Output: Trained Xception model with optimal classification performance

1. **Load dataset MRI**
2. **Split MRI into training (70%), validation (20%), and testing (10%) subsets**
3. **Preprocess images:**
    a. *Resize each image to 128 × 128 × 3*
    b. *Normalize pixel values to [0, 1]*
    c. *Apply data augmentation (rotation, shearing, zooming, flipping)*
4. **Define cross-validation scheme:**
    a. Perform 5-fold cross-validation on training + validation sets
    b. For each fold:
        i. *Assign one subset as validation, remaining as training*
        ii. *Initialize Xception with pretrained ImageNet weights*
        iii. *Freeze convolutional base and train classifier head*
        iv. *Fine-tune last 30% of convolutional layers with reduced learning rate*
        v. *Evaluate fold performance (accuracy, precision, recall, F1-score)*
5. **Aggregate performance across all folds**
6. **Select final trained model**
7. **Evaluate final model on independent test set**
8. **Compare results with baseline models (ResNet50, SVM)**
9. **Report classification metrics and analyze outcomes**

Fig. 1. Pseudocode of Research Framework

2.2. Dataset

The experiments in this study utilized a publicly available brain tumor MRI dataset obtained from Kaggle, which contains a total of 6,537 images categorized into four classes: glioma, meningioma,

pituitary tumor, and no tumor [2], [7]. Each image was originally provided in grayscale at varying resolutions while maintaining consistent quality suitable for deep learning-based classification tasks. This dataset has been widely employed in prior research, ensuring comparability and reproducibility across different studies [33], [34].

The dataset was divided into three subsets following a stratified split to preserve the class balance. Specifically, 70% of the images (4,576 samples) were allocated for training, 20% (1,307 samples) for validation, and the remaining 10% (654 samples) for an independent test set. To ensure robust model evaluation and minimize bias, the training and validation subsets (5,883 images in total) were subjected to a 5-fold cross-validation scheme. In each fold, approximately 4,706 images were used for training and 1,177 images for validation, while the test set remained untouched throughout training to provide an unbiased final performance evaluation.

Fig. 2 shows representative MRI images from the four classes in the dataset. Glioma and meningioma images generally present with irregular tumor boundaries, whereas pituitary tumors appear in the Sellar region with relatively distinct contours. No-tumor class, on the other hand, represents healthy brain scans with no abnormal growth. Such variations in appearance across the four categories underscore the importance of robust feature extraction methods for accurate classification.

This partitioning strategy, combined with diverse examples across all classes, not only enabled a comprehensive assessment of the model's generalizability but also mitigated potential overfitting by ensuring that the classifier was trained and validated on multiple data splits before being evaluated on the unseen test set.

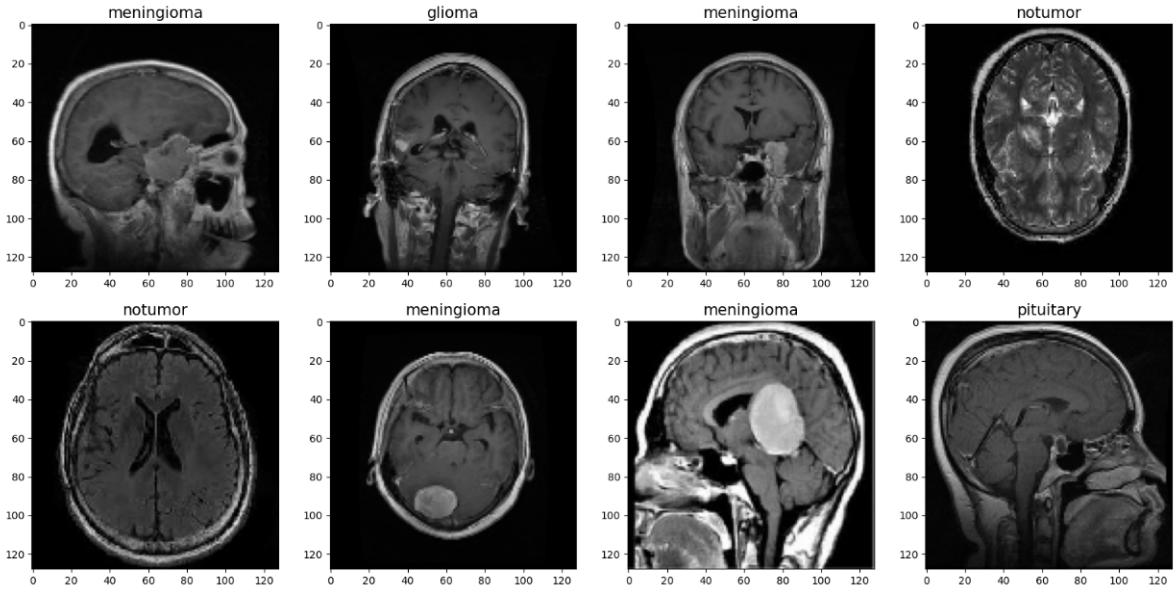


Fig. 2. Example of dataset (meningioma, glioma, notumor, pituitary)

2.3. Preprocessing

Image preprocessing is a crucial step in preparing the MRI dataset for training CNN models. This ensures that the images are standardized, denoised, and enhanced for better feature extraction, ultimately leading to improved classification accuracy. The preprocessing pipeline in this study consisted of four main stages: normalization, noise filtering, image enhancement, and data augmentation. These steps collectively optimize model training, improve convergence, and reduce the risk of overfitting.

a. Normalization

To maintain consistency across the dataset, Pixel intensity values of all MRI images were scaled to the range [0,1] as part of the normalization process [3], [17], [30]. This transformation is achieved by dividing each pixel value by the maximum possible intensity (usually 255 for 8-bit grayscale images).

Normalization reduces the effects of varying brightness and contrast levels across scans and accelerates convergence during training by stabilizing weight updates and preventing large gradient fluctuations.

Computation of normalization is as in (1) [30].

$$I_{normalized} = \frac{I_{original}}{255} \quad (1)$$

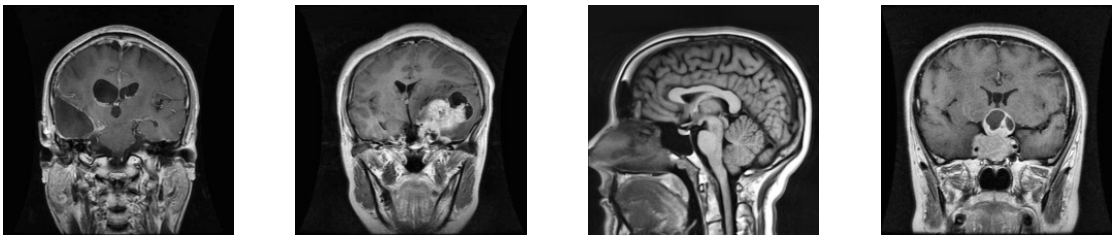
where$I_{original}$ represents the raw pixel intensity, and $I_{normalized}$ is the scaled *intensity* value between 0 and 1.

b. Noise Filtering

MRI images are often affected by acquisition noise, such as Gaussian or Rician noise, which may obscure subtle tumor features. To address this, a noise filtering step was applied to reduce unwanted artifacts while preserving the important structural details of the brain. In this study, median filtering and Gaussian smoothing were employed to suppress high-frequency noise without distorting the tumor boundaries [35], [36]. Noise reduction improves the signal-to-noise ratio (SNR), allowing the CNN models to extract more reliable features and enhance the robustness of the classification.

c. Image Enhancement

Following noise reduction, image enhancement techniques are applied to improve the visibility of the tumor regions. Histogram equalization and contrast-limited adaptive histogram equalization (CLAHE) are used to enhance the local contrast, ensuring that small or low-intensity tumor regions are more distinguishable from the surrounding healthy tissues [3], [37], [38]. This step ensured that subtle variations in the grayscale intensity were emphasized, thereby facilitating more accurate feature extraction by the CNN. Fig. 3 illustrates the dataset after noise filtering and image enhancement.



(a) Glioma (b) Meningioma (c) Notumor (d) Pituitary

Fig. 3. Images after noise filtering and image enhancement

d. Data Augmentation

Data augmentation techniques were applied to the training dataset to enhance model generalization and mitigate overfitting. Augmentation artificially expands the dataset by generating modified versions of the original images, introducing variability that helps the model to recognize tumors under different imaging conditions. In this study, several augmentation techniques were utilized, including [39], [40]: rotation (up to 30°), width and height shifts (up to 20%, shearing and zooming, horizontal flipping, and brightness adjustment was performed within a range of 0.8–1.2.

Brightness adjustment modifies the intensity levels of MRI images by randomly increasing or decreasing the brightness brightness [39], [41], ensuring that the model remains robust against scanner variability and acquisition conditions. Transformations, such as rotation, shifting, and flipping, introduce spatial diversity, allowing the model to detect tumors from different orientations and positions.

By implementing normalization, noise filtering, image enhancement, and data augmentation, the preprocessing pipeline ensured that the CNN models were trained on standardized, clean, and diverse inputs. This not only enhanced the robustness of the classification framework but also improved its accuracy and reliability across diverse MRI scans. The originality of this study comes from creating a complete Xception-based framework optimized for multiclass brain tumor classification using MRI scans. Unlike earlier work that used single preprocessing steps or limited model tuning, this research

introduces a step-by-step preprocessing pipeline. It combines normalization, adaptive noise filtering, contrast improvement, and targeted augmentation to enhance image consistency and maintain important tumor features across various data sources. A two-phase fine-tuning strategy further boosts performance by stabilizing classifier learning before unfreezing deeper layers. This method reduces overfitting and improves feature extraction, especially for less common tumor classes. The model was also tested against several deep learning architectures and an SVM classifier. Its reliability was confirmed using an additional benchmark dataset (CE-MRI Figshare). To prevent performance inflation and ensure methodological rigor, data augmentation was applied strictly to the training subset within each fold during five-fold cross-validation, ensuring that neither augmented images nor their original counterparts appeared in validation or independent test sets. Together, this integrated design, controlled evaluation strategy, and cross-dataset validation provide a more clinically relevant and reliable contribution compared to earlier Xception-based approaches. Fig. 4 illustrates examples of data augmentation and enhancement applied to the MRI data.

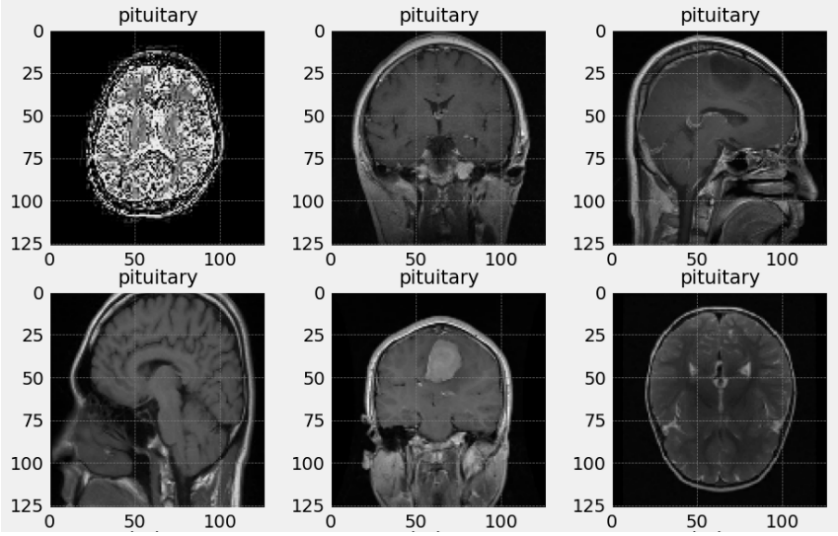


Fig. 4. Dataset after the augmentation process

2.4. Model Architecture

The Xception architecture is a convolutional neural network designed to improve feature extraction efficiency by factorizing standard convolutional operations into depthwise and pointwise components, allowing effective learning with reduced computational complexity. This design significantly reduces the computational complexity while preserving the representational capacity of the network, making Xception highly suitable for large-scale image classification tasks.

The Xception model consists of three primary components: an **entry flow** that captures low-level spatial features; a **middle flow** comprising repeated depthwise separable convolutional blocks to learn increasingly abstract representations; and an **exit flow** responsible for aggregating high-level features before classification. This hierarchical structure enables the network to effectively model both the local and global characteristics of brain tumor MRI images.

Xception, which stands for *"Extreme Inception,"* builds upon the Inception architecture by replacing its standard modules with depthwise separable convolutions, thereby allowing for more efficient learning of spatial and cross-channel correlations [42]. Its architectural design allows for faster convergence during training, reduces parameter overhead compared to traditional CNN, and enhances the capacity to extract discriminative features from complex medical images. In this study, Xception demonstrated the best results, achieving superior accuracy, precision, and recall compared with the baseline methods. Its ability to extract fine-grained features contributed to its robustness in distinguishing glioma, meningioma, pituitary tumors, and non-tumor cases.

To benchmark performance, two baseline models were employed: **ResNet50**, a residual learning-based CNN widely recognized for its robustness in deep architectures, and **SVM**, a classical machine learning classifier frequently applied in medical image analysis but lacking the representational depth of CNN. These comparisons highlight the superiority of Xception in extracting meaningful patterns and achieving reliable classification across the four tumor types. An overview of the Xception architecture employed in this study, highlighting the entry, middle, and exit flows that form the backbone of the model, is illustrated in Fig. 5, while Fig. 6 presents the proposed framework, detailing the sequential stages from input data preprocessing to classification and evaluation along with the architectural components of Xception.

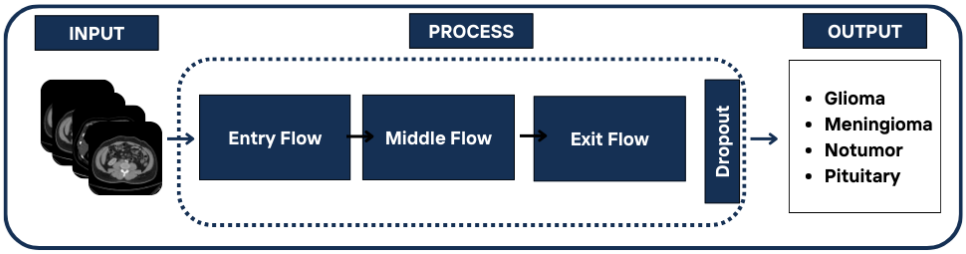


Fig. 5. Xception architecture

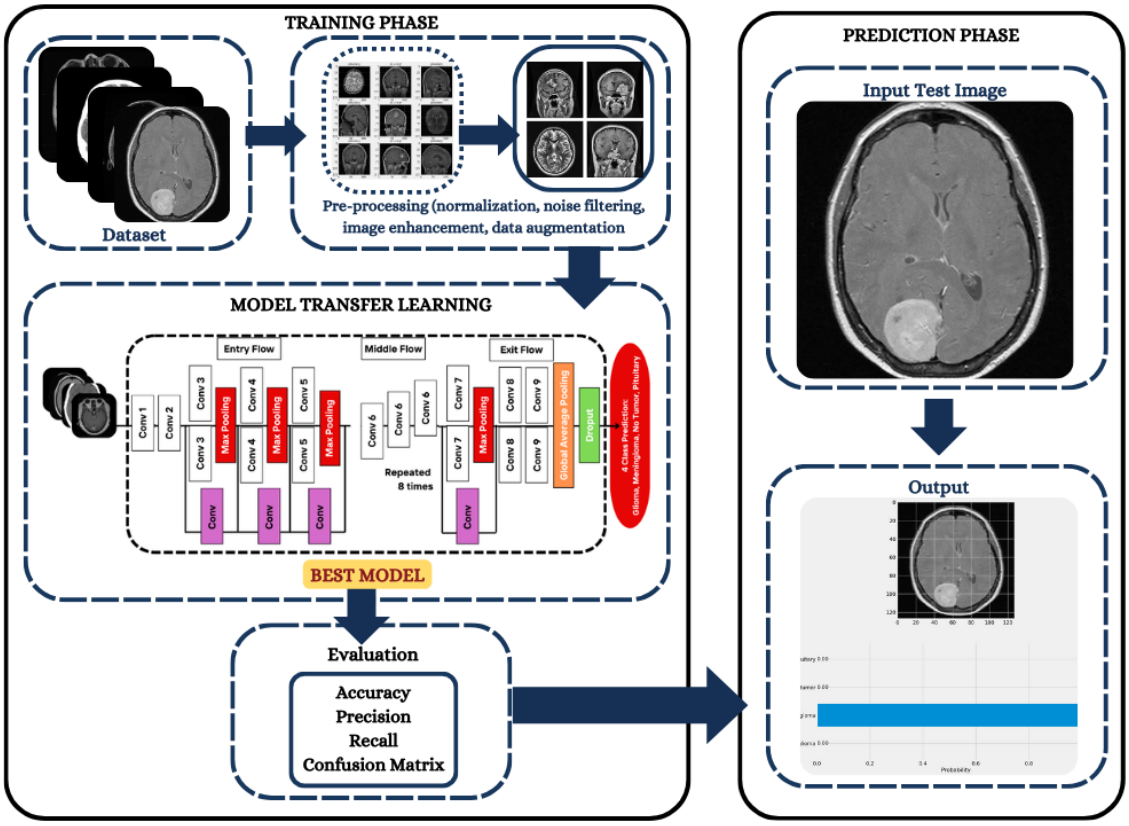


Fig. 6. Proposed Framework

Building on the strengths of the Xception architecture and addressing the limitations observed in prior studies, this work proposes a comprehensive brain tumor classification framework designed for robust and clinically relevant performance. The framework integrates systematic preprocessing of MRI images, data augmentation for improved generalization, and a carefully designed two-phase training strategy applied to the Xception model. Furthermore, to ensure reliability and reduce the risk of overfitting, the study incorporates a 5-fold cross-validation scheme on the combined training and validation sets while keeping the independent test set untouched for unbiased evaluation. By benchmarking Xception against ResNet50 and SVM classifiers, the framework not only demonstrates

the advantages of transfer learning with depthwise separable convolutions but also establishes a reliable pipeline for classifying four distinct tumor categories: glioma, meningioma, pituitary tumor, and no tumor. The computation for Xception Architecture are as follows:

- Standard Convolution Operation [27]. In a typical convolutional layer, the feature map, Y, is computed as in eq. (2):

$$Y = \sigma(W * X + b) \tag{2}$$

where X represents the input feature map containing the data or features to be processed. W denotes the convolutional kernel or filter, which is applied to the input feature map to extract important spatial features. The symbol * indicates the convolution operation, where the kernel slides over the input feature map to compute feature responses. The parameter b represents the bias term, which is added to the convolution result to adjust the activation output. Finally, σ denotes the activation function, such as ReLU (Rectified Linear Unit), which introduces non-linearity into the model and helps the network learn complex patterns from the data.

- Depthwise Separable Convolution [27] using (3) and Xception replaces standard convolution with depthwise separable convolution, which consists of: (i) Depthwise Convolution (Y_{dw}) that applies a single filter per input channel (3) and Pointwise Convolution Y_{pw} that uses a 1x1 convolution to combine the output of depthwise convolution using (4).

$$Y_{dw} = \sigma(W_d * X + b_d) \tag{3}$$

where W_d is the depthwise filter applied to each channel independently.

$$Y_{pw} = \sigma(W_p * Y_{dw} + b_p) \tag{4}$$

where W_p is the pointwise filter. The overall effect is a reduction in computational cost while maintaining expressiveness.

- Residual Connection in Xception [27]. The architecture integrates skip connections similar to ResNet and computed using (5).

$$Y = F(X) + X \tag{5}$$

where $F(X)$ is the output of several convolutional layers, and X is the original input. This helps prevent gradient vanishing and improves training stability.

- Global Average Pooling layer [27]. Instead of fully connected layers, Xception uses a Global Average Pooling, GAP(X), operation which is computed using (6).

$$GAP(X) = \frac{1}{H\,x\,W} \sum_{i=1}^{H} \sum_{j=1}^{W} Xi,j \tag{6}$$

where H and W are the height and width of the feature map, i and j are the row and column indexes of the feature map. This reduces overfitting and enhances model generalization.

2.5. Transfer Learning Strategy

Transfer learning was employed to leverage the representational power of a pretrained Xception model on the ImageNet dataset. A two-phase training strategy was designed to balance the efficiency, generalization, and accuracy of brain tumor MRI image classification.

Phase 1: Frozen Convolutional Base

In the initial phase, the convolutional backbone of Xception was frozen, and only the fully connected classification head was trained for the model. The fully connected classification head consisted of a Global

Average Pooling layer, followed by fully connected dense layers and a softmax output layer representing the four tumour classes. Through this approach, pretrained features were adapted to the medical imaging domain without altering the generalized feature extraction layers, thereby reducing the risk of overfitting.

Phase 2 – Fine-Tuning:

After the fully connected classification head stabilized, the final 30% of the convolutional layers of Xception were unfrozen and fine-tuned using a smaller learning rate. Fine-tuning enabled the model to adjust higher-level feature representations more specifically to the MRI dataset while retaining the general visual features learned from the ImageNet dataset. This hierarchical adaptation enabled more accurate discrimination of nuanced differences across the four tumor classes. The model was trained using the Adamax optimizer with an initial learning rate of 0.001 during Phase 1 and a reduced learning rate of 0.0001 during Phase 2 to ensure a stable convergence. A batch size of 16 was selected to balance the computational efficiency and gradient stability. The model was trained for 50 epochs, with early stopping employed to prevent overfitting based on the validation loss. The categorical cross-entropy loss was employed as it is appropriate for multi-class classification tasks.

This two-phase strategy enabled the efficient use of pretrained weights while adapting the model effectively to the specific characteristics of the MRI brain tumor dataset.

2.6. Evaluation Metrics

Multiple established evaluation metrics were applied to ensure a thorough performance evaluation of the proposed framework. These metrics provide complementary perspectives on classification performance, ensuring both accuracy and clinical reliability.

a. Accuracy

Accuracy is the most straightforward performance metric, representing the ratio of correctly predicted samples to the total number of samples. It measures the overall correctness of the model's predictions across all classes. Accuracy is calculated using (7).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

where *TP, TN, FP,* and *FN* are True Positive, True Negative, False Positive, and False Negative, respectively, that represent the classification outcomes. Although accuracy is a useful general indicator, it may not fully reflect model performance on imbalanced datasets, as it can be skewed by the majority class.

b. Precision

Precision evaluates the model's ability to correctly identify positive cases, focusing on the proportion of true positive predictions relative to all positive predictions [38]. In clinical decision-making, minimizing false positives is essential to prevent unwarranted interventions. Precision is calculated as shown in (8).

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

A high precision score indicates that the model has a low false positive rate, meaning it rarely misclassifies healthy cases as tumors or confuses one tumor type with another.

c. Recall (Sensitivity)

Recall, also known as sensitivity or the true positive rate, measures the model's ability to identify all relevant cases within a dataset. In the context of brain tumor classification, recall assesses how well the model detects actual tumor cases, minimizing false negatives (i.e., failing to detect a tumor) [43]. Recall is calculated as in (9).

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

3. Results and Discussion

Two experimental settings were conducted to evaluate the proposed framework: one using a small dataset of 1,639 MRI images and another with a larger dataset of 6,537 images. This dual setup demonstrates how dataset size influences classification accuracy, convergence, and generalization. The comparison provides evidence of the framework's effectiveness while underscoring the importance of sufficient data volume in medical image analysis.

3.1. Training and Validation Performance

To study how dataset scale affects transfer-learning performance, we conducted two experiments with the same Xception pipeline but different data volumes: (i) a balanced small dataset of 1,639 MRI images and (ii) a large dataset of 6,537 images (split 70/20/10 into train/validation/test). Holding preprocessing, augmentation, optimizer, learning-rate schedule, and fine-tuning policy constant allows a clean assessment of how sample size impacts convergence, stability, and generalization. The results confirm that small datasets materially constrain performance and training dynamics, whereas scaling to several thousand images markedly improves both accuracy and robustness.

3.1.1. Experiment 1 – 1,639 images (small data).

Training progressed gradually from ~59% to 92.1% accuracy by epoch 50, while validation rose from ~65% to a peak of 90.2%. Reduce-on-plateau events (e.g., around epochs 16, 22, 25, 29) helped curb overfitting and unlocked further gains after mid-training plateaus. The generalization gap (train – val) remained modest (~2% at the end), suggesting that augmentation and cautious fine-tuning mitigated severe overfitting despite limited data. Nevertheless, the test set (held out) reached 88.4% accuracy, 88.7% precision, and 86.3% recall, indicating that with scarce data the model still underfits harder decision boundaries and is more sensitive to class confusability (historically worst for meningioma). Overall, convergence was slower and more variance-prone, requiring ~40–50 epochs to approach its ceiling.

3.1.2. Experiment 2 – 6,537 images (large data).

With substantially more samples, Xception exhibited rapid convergence and high stability: training accuracy exceeded 99.5% after just a few epochs. It often reached 100%, while validation accuracy ranged from 97.2% to 99.45% across epochs, peaking at 99.45% with consistently low loss. Despite occasional small oscillations late in training, precision and recall co-tracked validation accuracy (≈98–99% most epochs), reflecting broadly improved class separability. On the independent test set of 654 images, the model achieved 99.06% accuracy, 99.06% precision, and 99.06% recall, demonstrating excellent generalization beyond the validation folds. These dynamics, fast ascent, early saturation, tight train/val coupling, and high external test accuracy, are characteristic of transfer learning on adequately large, diverse datasets.

3.1.3. Critical comparison and interpretation.

Across identical architectures and hyperparameters, the larger dataset delivered (a) faster, smoother convergence (reliable >98% validation within ~2–5 epochs vs. ~45–50 epochs to reach ~90% on small data), (b) higher ceilings (≈99–99.5% vs. ≈90%), and (c) lower variance and tighter generalization. Mechanistically, more data improves the coverage of tumor appearance (shape, texture, intensity, scanner variability), reducing the estimation variance of batch statistics and enabling Xception's depthwise-separable filters to learn finer, class-specific features without over-specializing. By contrast, with 1,639 images the model is capacity-limited by sample scarcity: augmentation helps but cannot fully emulate real distributional diversity; learning-rate reductions are needed to avoid premature convergence; and the classifier remains more vulnerable to borderline cases and inter-class similarities (especially meningioma vs. others). In short, data scale is the primary driver of both performance and training stability in this pipeline; architectural strength and careful regularization only realize their full potential

at the larger scale. In summary, dataset size strongly impacts transfer learning with Xception. Larger datasets enable faster convergence, higher accuracy, and more stable generalization, while smaller datasets limit performance despite augmentation and fine-tuning. The detailed results of both experiments are presented in Table 2.

Table 2. Summary of training/validation/test performance on small vs. large datasets (Xception, identical pipeline)

Experiment	Xception Model							
	Dataset Size	*Convergence*	*Final Train Acc*	*Peak Val Acc*	*Val Acc range*	*Test Acc*	*Test Prec*	*Test Rec*
Experiment 1 (small)	1.639	Slow, multiple LR reductions; steady late gains	92.1% (epoch50)	90.1% (epoch 47-50)	65%-90%	88.4%	88.7%	86.3%
Experiment 2 (large)	6.537	Very fast; stable high plateau by epoch 2-5	99.9%	99.45% (epoch 22)	97.2-99.45%	99.06%	99.06%	99.06%

3.2. Test Set Evaluation

The independent test set comprising 654 MRI images was used to provide an unbiased evaluation of model generalization. The proposed Xception-based framework consistently outperformed the baseline models. Specifically, Xception achieved a test accuracy of 99.06%, with precision and recall also reaching 99.06%, demonstrating robust and reliable classification performance. This performance aligns with the high training accuracy (99.87%) and validation accuracy (99.19%) obtained during the cross-validation phase, confirming that the model generalized effectively without overfitting.

In comparison, the baseline models delivered markedly inferior results. The SVM classifier, despite its popularity in medical image analysis, struggled with the high-dimensional MRI features and achieved only 21.41% accuracy on the test set. ResNet50, implemented within a transfer learning setting, performed better but still lagged behind Xception, reaching 75.27% test accuracy. These findings highlight the advantage of the Xception architecture, where depthwise separable convolutions enable efficient feature representation and lead to more accurate and reliable tumor classification. The training and validation performance of the proposed Xception framework is further illustrated in Fig. 7, which shows the training accuracy and loss trends on the large dataset.

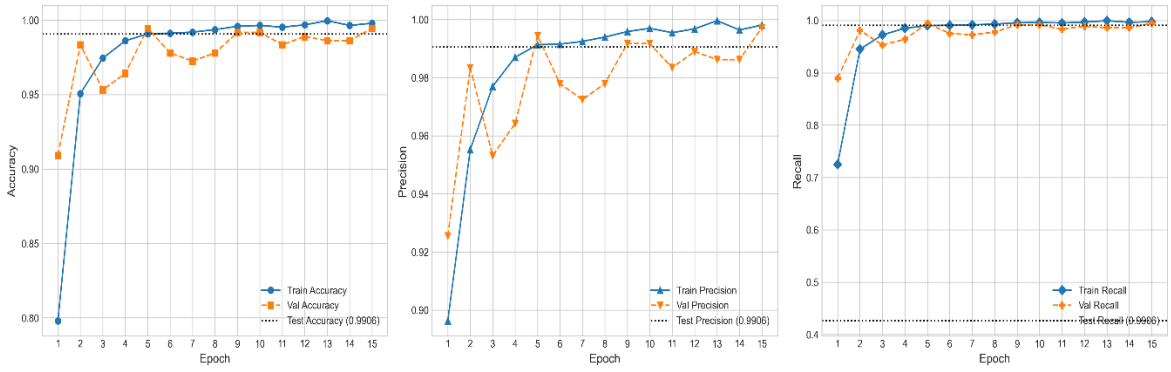


Fig. 7. Experiment result on a large dataset used Xception

To emphasize the impact of dataset size, Fig. 8 depicts the performance trends when training on the smaller dataset (1,639 images), where higher fluctuations in accuracy and loss indicate less stability and weaker generalization. For comparison, Fig. 9 presents a combination of bar and line plots depicting the accuracy, precision, and recall of the baseline SVM and ResNet50 models. The visualization demonstrates

that both models exhibit limited discriminative capacity compared to Xception, particularly in capturing complex, fine-grained tumor characteristics.

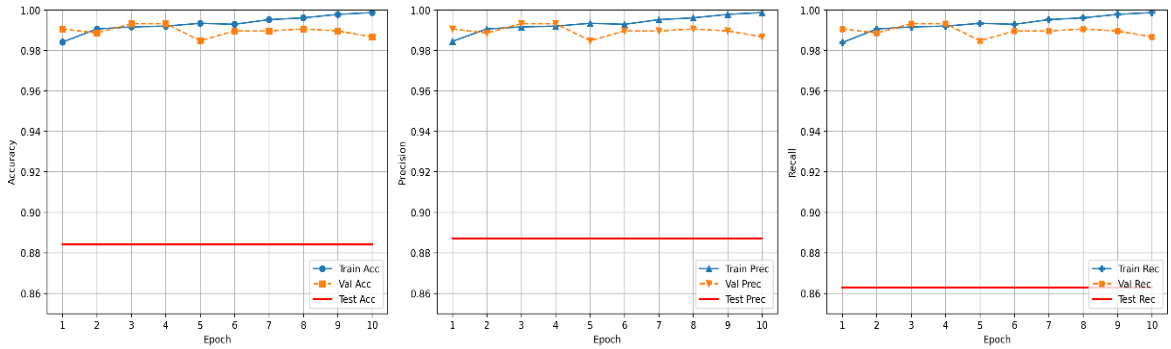


Fig. 8. Experiment results on small dataset used Xception

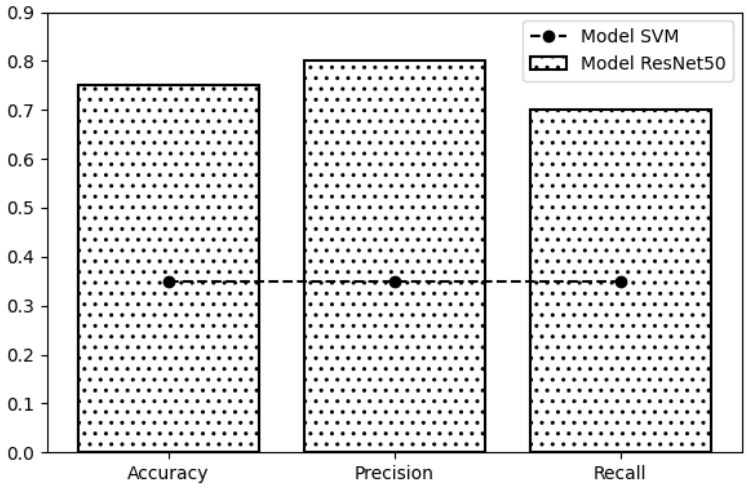


Fig. 9. Results of SVM and ResNet50

To ensure a rigorous and unbiased performance assessment, two complementary evaluation strategies were utilized. First, the combined training and validation dataset was evaluated using five-fold cross-validation (5,883 images), and the fold-wise outcomes are summarized in Table 3. Although performance varied across folds, this variation reflects natural differences in MRI acquisition characteristics and tumor morphology rather than methodological issues or data leakage. The aggregated cross-validation result achieved an accuracy of 0.8994 ± 0.089, indicating stable and consistent learning behavior across folds. The second strategy evaluated the final optimized model on an independent test set comprising 654 previously unseen images. As shown in Table 4, the model achieved 99.06% accuracy on this unbiased evaluation set, confirming strong generalization capability on unseen data.

Table 3. 5-Fold cross-validation results on large dataset

Fold	Fold-wise performance of the Xception model with 5-fold cross-validation			
	Accuracy	*Precision*	*Recall*	*F1-Score*
1	0.7527	0.8047	0.7037	0.7500
2	0.8506	0.8677	0.8201	0.8430
3	0.9280	0.9322	0.9242	0.9280
4	0.9750	0.9770	0.9730	0.9750
5	0.9906	0.9906	0.9906	0.9906
Mean ± Std	**0.8994 ± 0.089**	**0.9144 ± 0.0068**	**0.8823 ±0.108**	**0.8973 ± 0.089**

Table 4. Comparison results between the independent test and the 5-fold

Fold	Fold-wise performance of the Xception model with 5-fold cross-validation			
	Accuracy (%)	*Precision (%)*	*Recall (%)*	*Notes*
Independent Test Set	99.06%	99.06%	99.06%	Final performance on unseen data
5-Fold Cross Validation	0.8994 ± 0.089	0.9144± 0.0068	0.8823 ±0.108	Mean ± SD across folds (5.883 images)

The results of this study demonstrate the superior performance of the Xception architecture in brain tumor classification, achieving 99.06% test accuracy with equally high precision and recall. This performance substantially exceeded that of the baseline models, with ResNet50 attaining 75.27% accuracy and SVM achieving only 21.41%. These findings highlight the effectiveness of depthwise separable convolutions in capturing both fine-grained spatial details and cross-channel dependencies, leading to more reliable feature extraction for MRI classification. By leveraging transfer learning, extensive preprocessing, and a large dataset of 6,537 images combined with fivefold cross-validation, our framework minimized overfitting and achieved robustness that is often absent in models trained on smaller datasets.

When positioned against prior studies summarized in the Introduction, the advantages of our approach become more evident. Sarkar et al. [15] achieved 100% accuracy with AlexNet-RF, but their reliance on small datasets raised concerns about overfitting and limited generalizability. Malakouti et al. [32] achieved strong performance with ensemble feature fusion but were constrained by class imbalance. Other approaches such as Capsule Networks [23] and Inception-v3 [24] reported lower accuracies (86.56% and 89%, respectively) or limited classification capability. Similarly, hybrid machine learning methods [25], [26] achieved accuracies near 97% but were restricted to binary classification or showed weaknesses in distinguishing meningiomas. By contrast, our Xception-based framework delivered a generalizable multiclass solution with higher accuracy and robustness, surpassing both CNN-based and hybrid approaches. Another key outcome of this study is the critical role of dataset size. Experiments with the smaller dataset (1,639 images) revealed instability and reduced generalization, echoing prior concerns about the limitations of deep learning models on insufficient data. Training with the larger dataset (6,537 images) ensured stable convergence, improved validation accuracy, and higher fold-wise performance, underscoring the necessity of large, diverse datasets for clinically reliable deep learning in medical imaging.

The comparative results are summarized across multiple experimental analyses. Fig. 10 illustrates the accuracy, precision, and recall of Xception, ResNet50, and SVM on the independent test set, where Xception consistently outperforms both baseline models across all metrics.

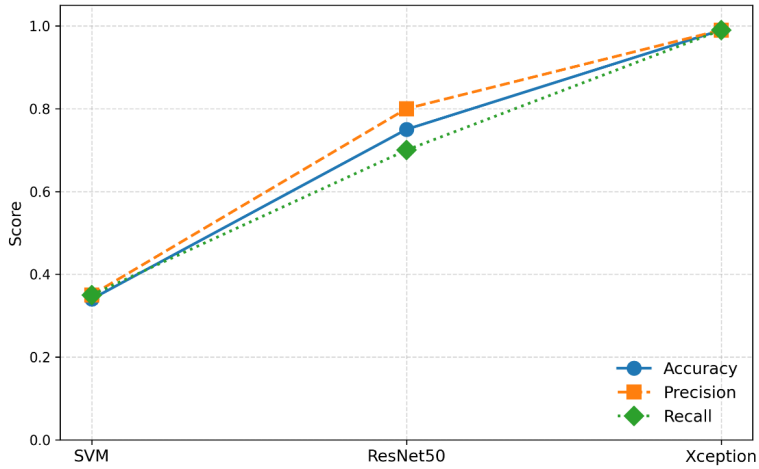


Fig. 10. Comparison results for large dataset

Fig. 11 further presents qualitative prediction examples, visually demonstrating the capability of Xception to correctly classify complex tumor cases. Table 4 contrasts five-fold cross-validation performance with independent test evaluation, providing insight into the model's generalization stability. Finally, Table 5 presents a broader comparison between the proposed framework and representative methods reported in prior studies, including CNN-based, ResNet-based, and SVM approaches, while also distinguishing baseline re-implementations conducted under identical conditions in this work. The proposed Xception-based framework achieves the highest accuracy on both the main dataset and the CE-MRI Figshare benchmark, highlighting its robustness and improved generalization relative to existing methods.

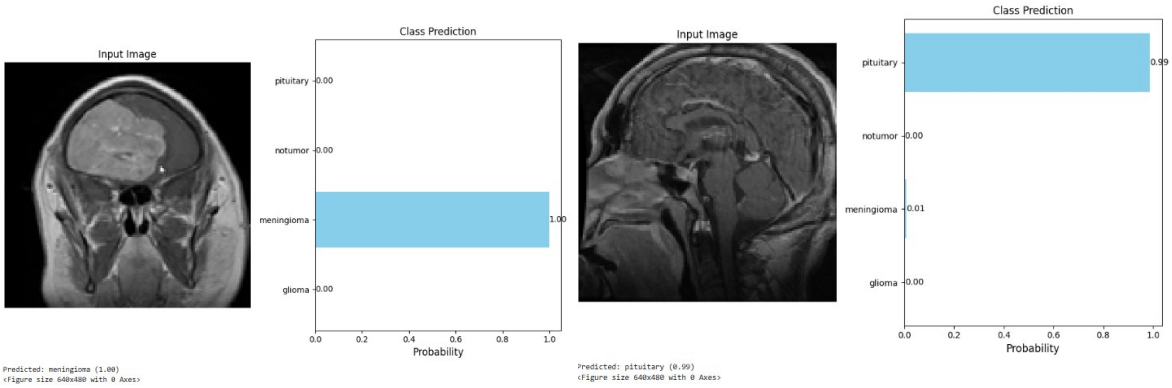


Fig. 1. Prediction sample test on proposed framework

This study establishes Xception as an effective and reliable architecture for multiclass brain tumor classification, consistently outperforming traditional machine learning techniques and contemporary CNN-based models. Despite these promising results, challenges remain, including occasional misclassification of meningioma tumors and reliance on predominantly single-center data. Future work will focus on multi-center validation, multimodal MRI integration, and exploration of attention-based or transformer architectures to further enhance generalization and facilitate clinical translation.

Table 5. Performance comparison of the proposed framework and existing approaches for brain tumor classification

Methods	Dataset	Dataset	Accuracy
Inception ResnetV2 [18]	MRI Kaggle	3.265 images	93.4%
CNN [44]	MRI Kaggle	4.195 images	93.81%
ResNet-50 [45]	MRI Kaggle	7.023 images	97.86%
CNN [20]	MRI Kaggle	3.064 images	95.75%
SVM [22]	MRI Kaggle	3.064 images	93.52%
ResNet-50 (Baseline this study)	MRI Kaggle	6.537 images	75.27%
SVM (Baseline this study)	MRI Kaggle	6.537 images	21.41%
Proposed framework (Xception)	CE-MRI Figshare	3.064 images	98.45%
Proposed framework (Xception)	MRI Kaggle	6.537 images	**99.06%**

4. Conclusion

The effectiveness of the proposed enhanced Xception-based deep learning framework was demonstrated for multiclass classification of brain tumors from MRI. By integrating a comprehensive preprocessing pipeline with a structured two-phase fine-tuning strategy, the framework consistently outperformed multiple deep learning architectures and an SVM classifier. The model achieved an almost perfect test accuracy of 99.06% and demonstrated strong generalization through evaluation on an additional benchmark dataset (CE-MRI Figshare), again with around 98% each for accuracy, precision, and recall. These results highlight the advantages of combining modern convolutional neural network

architectures with sequential preprocessing and cross-dataset evaluation to improve feature representation, reduce overfitting, and enhance reliability, particularly for complex or less common tumor types. Despite its strong performance, the model has some limitations, including occasional misclassifications in complex cases and a lack of integrated explainability features. Future work will focus on adding interpretability methods such as Grad-CAM, LIME, and SHAP, examining multimodal MRI inputs, and expanding validation to multi-center datasets to enhance transparency, reliability, and clinical readiness. Overall, these findings suggest that the proposed framework represents a promising, clinically relevant advancement in AI-assisted brain tumor diagnosis. It has the potential to support radiologists by aiding in triage, providing second opinions, and improving workflow efficiency.

Acknowledgment

The authors would like to express their sincere gratitude to Universiti Muhammadiyah Malaysia, Institut Teknologi Statistika dan Bisnis Muhammadiyah Semarang, and Universiti Utara Malaysia for their invaluable support throughout this research.

Declarations

Author contribution. Nurul Huda contributed to the conceptualization, methodology, data analysis, and manuscript writing. Ku Ruhana Ku-Mahamud served as the reviewer, providing critical revisions, expert guidance, and valuable feedback to enhance the quality of this research.
Funding statement. This research received no external funding.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

Data and Software Availability Statements

The dataset used in this study is publicly available on **Kaggle** and can be accessed at: **Brain Tumor Classification - CNN Dataset**. The software used for data analysis and model implementation is based on **Python (TensorFlow/Keras)** and is available upon reasonable request.

References

[1] Q. U. A. Ishfaq *et al.*, "Automatic smart brain tumor classification and prediction system using deep learning," *Sci. Rep.*, vol. 15, no. 1, pp. 1–31, 2025, doi: 10.1038/s41598-025-95803-3.

[2] R. Ali *et al.*, "Learning Architecture for Brain Tumor Classification Based on Deep Convolutional Neural Network: Classic and ResNet50," *Diagnostics*, vol. 15, no. 5, 2025, doi: 10.3390/diagnostics15050624.

[3] J. Shreeharsha, "Brain Tumor Segmentation and Classification Using Binomial Thresholding-Based Bidirectional-Long-Short Term Memory," *Int. J. Intell. Eng. Syst.*, vol. 17, no. 3, pp. 149–158, 2024, doi: 10.22266/ijies2024.0630.13.

[4] Y. Sankararao and S. Khasim, "An Effective analysis of brain tumor detection using deep learning," *EAI Endorsed Trans. Pervasive Heal. Technol.*, vol. 10, pp. 1–10, 2024, doi: 10.4108/eetpht.10.5627.

[5] O. H. Kesav and G. K. Rajini, "Enhancing Brain Tumor Detection and Classification with Reduced Complexity Spatial Fusion Convolutional Neural Networks," *Int. J. Intell. Eng. Syst.*, vol. 17, no. 1, pp. 263–277, 2024, doi: 10.22266/ijies2024.0229.25.

[6] A. Kirimtat and O. Krejcar, "GPU-Based Parallel Processing Techniques for Enhanced Brain Magnetic Resonance Imaging Analysis: A Review of Recent Advances †," *Sensors*, vol. 24, no. 5, 2024, doi: 10.3390/s24051591.

[7] P. K. Jena, B. Khuntia, and T. Panigrahi, "A Novel Saliency Region Enhanced Technique for Biomedical Image Indexing Using Deep Learning," *Rev. d'Intelligence Artif.*, vol. 37, no. 6, pp. 1397–1405, 2023, doi: 10.18280/ria.370603.

[8] R. Rajkumar, D. Shanthi, and K. Manivannan, "Efficient Guided Grad-CAM Tuned Patch Neural Network for Accurate Anomaly Detection in Full Images," *Inf. Technol. Control*, vol. 53, no. 2, pp. 355–371, 2024, doi: 10.5755/j01.itc.53.2.34525.

[9] Z. N. K. Swati *et al.*, "Brain tumor classification for MR images using transfer learning and fine-tuning," *Comput. Med. Imaging Graph.*, vol. 75, pp. 34–46, 2019, doi: 10.1016/j.compmedimag.2019.05.001.

[10] N. Elazab, W. A. Gab-Allah, and M. Elmogy, "A multi-class brain tumor grading system based on histopathological images using a hybrid YOLO and RESNET networks," *Sci. Rep.*, vol. 14, no. 1, pp. 1–20, 2024, doi: 10.1038/s41598-024-54864-6.

[11] G. Yue, M. Li, and H. Zheng, "Image Style Transfer Based On Generative Adversarial Network And Feature Transformation For Modern Home Design," *J. Appl. Sci. Eng.*, vol. 28, no. 2, pp. 257–263, 2025, doi: 10.6180/jase.202502_28(2).0005.

[12] R. Jalloul, H. K. Chethan, and R. Alkhatib, "A Review of Machine Learning Techniques for the Classification and Detection of Breast Cancer from Medical Images," *Diagnostics*, vol. 13, no. 14, 2023, doi: 10.3390/diagnostics13142460.

[13] Y. M. A. Mohammed, S. El Garouani, and I. Jellouli, "A survey of methods for brain tumor segmentation-based MRI images," *J. Comput. Des. Eng.*, vol. 10, no. 1, pp. 266–293, 2023, doi: 10.1093/jcde/qwac141.

[14] S. Dayarathna, K. T. Islam, S. Uribe, G. Yang, M. Hayat, and Z. Chen, "Deep learning based synthesis of MRI, CT and PET: Review and analysis," *Med. Image Anal.*, vol. 92, no. June 2023, p. 103046, 2024, doi: 10.1016/j.media.2023.103046.

[15] A. Sarkar, M. Maniruzzaman, M. A. Alahe, and M. Ahmad, "An Effective and Novel Approach for Brain Tumor Classification Using AlexNet CNN Feature Extractor and Multiple Eminent Machine Learning Classifiers in MRIs," *J. Sensors*, vol. 2023, 2023, doi: 10.1155/2023/1224619.

[16] G. Bompem and D. Pandluri, "Batch Normalization Based Convolutional Neural Network for Segmentation and Classification of Brain Tumor MRI Images," *Int. J. Intell. Eng. Syst.*, vol. 17, no. 2, pp. 39–49, 2024, doi: 10.22266/ijies2024.0430.04.

[17] T. Gayathri and K. S. Kumar, "Brain Tumor Segmentation and Classification Using Cnn Pre-Trained Vgg-16 Model in Mri Images," *IIUM Eng. J.*, vol. 25, no. 2, pp. 196–211, 2024, doi: 10.31436/iiumej.v25i2.2963.

[18] T. S. Azzahra, J. J. Cerelia, F. Azhar, L. Nugraha, and A. A. Pravitasari, "MRI-Based Brain Tumor Classification Using Inception Resnet V2," *Enthusiastic Int. J. Appl. Stat. Data Sci.*, vol. 3, no. 2, pp. 163–175, 2023, doi: 10.20885/enthusiastic.vol3.iss2.art4.

[19] T. Feng, S. Deng, Q. Duan, and Y. Mao, "Application of Local Search Particle Swarm Optimization Based on the Beetle Antennae Search Algorithm in Parameter Optimization," *Actuators*, vol. 13, no. 7, p. 270, 2024, doi: 10.3390/act13070270.

[20] M. F. Alanazi *et al.*, "Brain Tumor/Mass Classification Framework Using Magnetic-Resonance-Imaging-Based Isolated and Developed Transfer Deep-Learning Model," *Sensors*, vol. 22, no. 1, 2022, doi: 10.3390/s22010372.

[21] A. H. Saleh, Ü. Atila, and O. Menemencioğlu, "Multimodal Fusion for Enhanced Semantic Segmentation in Brain Tumor Imaging: Integrating Deep Learning and Guided Filtering Via Advanced 3D Semantic Segmentation Architectures," *Int. J. Imaging Syst. Technol.*, vol. 34, no. 5, 2024, doi: 10.1002/ima.23152.

[22] W. Ayadi, I. Charfi, W. Elhamzi, and M. Atri, "Brain tumor classification based on hybrid approach," *Vis. Comput.*, vol. 38, no. 1, pp. 107–117, 2022, doi: 10.1007/s00371-020-02005-1.

[23] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "Brain Tumor Type Classification Via Capsule Networks," *2018 25th IEEE Int. Conf. Image Process.*, pp. 3129–3133, 2018, doi: 10.1109/ICIP.2018.8451379.

[24] M. J. Lakshmi and S. Nagaraja Rao, "Brain tumor magnetic resonance image classification: a deep learning approach," *Soft Comput.*, vol. 26, no. 13, pp. 6245–6253, 2022, doi: 10.1007/s00500-022-07163-z.

[25] G. Garg and R. Garg, "Brain Tumor Detection and Classification based on Hybrid Ensemble Classifier," no. 3, pp. 1–18, 2021, [Online]. Available: http://arxiv.org/abs/2101.00216.

[26] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Comput. Biol. Med.*, vol. 111, no. March, p. 103345, 2019, doi: 10.1016/j.compbiomed.2019.103345.

[27] Ş. Kaba, H. Haci, A. Isin, A. Ilhan, and C. Conkbayir, "The Application of Deep Learning for the Segmentation and Classification of Coronary Arteries," *Diagnostics*, vol. 13, no. 13, 2023, doi: 10.3390/diagnostics13132274.

[28] I. Galić, M. Habijan, H. Leventić, and K. Romić, "Machine Learning Empowering Personalized Medicine: A Comprehensive Review of Medical Image Analysis Methods," *Electron.*, vol. 12, no. 21, 2023, doi: 10.3390/electronics12214411.

[29] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," *Front. Public Heal.*, vol. 11, no. November, pp. 1–28, 2023, doi: 10.3389/fpubh.2023.1273253.

[30] A. Rohini *et al.*, "Multimodal hybrid convolutional neural network based brain tumor grade classification," *BMC Bioinformatics*, vol. 24, no. 1, pp. 1–20, 2023, doi: 10.1186/s12859-023-05518-3.

[31] S. Raza *et al.*, "Brain Tumor Detection and Classification Using Deep Feature Fusion and Stacking Concepts," *J. Popul. Ther. Clin. Pharmacol.*, vol. 31, no. 01, pp. 1339–1356, 2024, doi: 10.53555/jptcp.v31i1.4179.

[32] S. M. Malakouti, M. Bagher Menhaj, and A. Abolfazl Suratgar, "Machine learning and transfer learning techniques for accurate brain tumor classification," *Clin. eHealth*, vol. 7, pp. 106–119, 2024, doi: 10.1016/j.ceh.2024.08.001.

[33] A. M. Al-Zoghby, E. M. K. Al-Awadly, A. Moawad, N. Yehia, and A. I. Ebada, "Dual Deep CNN for Tumor Brain Classification," *Diagnostics*, vol. 13, no. 12, 2023, doi: 10.3390/diagnostics13122050.

[34] L. Gaur, M. Bhandari, T. Razdan, S. Mallik, and Z. Zhao, "Explanation-Driven Deep Learning Model for Prediction of Brain Tumour Status Using MRI Image Data," *Front. Genet.*, vol. 13, no. March, pp. 1–9, 2022, doi: 10.3389/fgene.2022.822666.

[35] M. Petwan and K. R. Ku-Mahamud, "Feature selection for sky image classification based on self adaptive ant colony system algorithm," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 6, pp. 7036–7047, 2023, doi: 10.11591/ijece.v13i6.pp7037-7047.

[36] S. Bhatia *et al.*, "A Robust NIfTI Image Authentication Framework Based on DST and Multi-Scale Otsu Thresholding," *IEEE Access*, vol. 10, pp. 132608–132620, 2022, doi: 10.1109/ACCESS.2022.3225908.

[37] M. S. Ullah, M. A. Khan, A. Masood, O. Mzoughi, O. Saidani, and N. Alturki, "Brain tumor classification from MRI scans: a framework of hybrid deep learning model with Bayesian optimization and quantum theory-based marine predator algorithm," *Front. Oncol.*, vol. 14, no. February, pp. 1–21, 2024, doi: 10.3389/fonc.2024.1335740.

[38] J. Paul, J. Jerusalin Carol, and T. S. Sivarani, "Enhancing Brain Tumor Classification with VGG-19 in Deep Learning Paradigms," *SSRG Int. J. Electron. Commun. Eng.*, vol. 11, no. 4, pp. 41–50, 2024, doi: 10.14445/23488549/IJECE-V11I4P105.

[39] K. Sowjanya, K. R. Reddy, and M. Raveena, "A New Distinctive Methodology for the Classification of Brain MR Images Using Histogram Based Local Feature Descriptors," *Int. J. Comput. Digit. Syst.*, vol. 13, no. 1, pp. 1301–1315, 2023, doi: 10.12785/ijcds/1301106.

[40] R. G. Dumitru, D. Peteleaza, and C. Craciun, "Using DUCK-Net for polyp image segmentation," *Sci. Rep.*, vol. 13, no. 1, pp. 1–12, 2023, doi: 10.1038/s41598-023-36940-5.

[41] M. Abdel-Basset, R. Mohamed, M. Abouhawwash, S. S. Askar, and A. A. Tantawy, "An Efficient Multilevel Threshold Segmentation Method for Breast Cancer Imaging Based on Metaheuristics Algorithms: Analysis and Validations," *Int. J. Comput. Intell. Syst.*, vol. 16, no. 1, 2023, doi: 10.1007/s44196-023-00282-x.

[42] L. Gamage, U. Isuranga, D. Meedeniya, S. De Silva, and P. Yogarajah, "Melanoma Skin Cancer Identification with Explainability Utilizing Mask Guided Technique," *Electron.*, vol. 13, no. 4, pp. 1–30, 2024, doi: 10.3390/electronics13040680.

[43] M. Aggarwal, A. K. Tiwari, M. P. Sarathi, and A. Bijalwan, "An early detection and segmentation of Brain Tumor using Deep Neural Network," *BMC Med. Inform. Decis. Mak.*, vol. 23, no. 1, pp. 1–12, 2023, doi: 10.1186/s12911-023-02174-8.

[44] S. Srinivasan, D. Francis, S. K. Mathivanan, H. Rajadurai, B. D. Shivahare, and M. A. Shah, "A hybrid deep CNN model for brain tumor image multi-classification," *BMC Med. Imaging*, vol. 24, no. 1, pp. 1–21, 2024, doi: 10.1186/s12880-024-01195-7.

[45] J. Jain, M. Kubadia, M. Mangla, and P. Tawde, "Comparison of Transfer Learning Techniques to Classify Brain Tumours Using MRI Images †," *Eng. Proc.*, vol. 59, no. 1, 2023, doi: 10.3390/engproc2023059144.