International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Optimized MRI Preprocessing for Enhanced Deep Transfer Learning-Based Brain Tumor Classification

Vasumathy M[1], R Hamsaveni[2], Deepa U[3]
[1,3]Kingston Engineering College Vellore, India, 632056.
[2]Department of MCA, Sri Venkateswara College of Engineering & Technology (Autonomous), Chittor, R.V.S. Nagar Tirupati Road, Andhra Pradesh 517127.

***Abstract*—*The use of MRI for diagnosing brain tumor is vital in making effective clinical decisions; however, the performance of deep transfer learning models is highly affected by the quality of the images. In this study, an efficient preprocessing pipeline consisting of eight steps has been proposed to increase image quality for improving multi-class brain tumor classification using MRI. In this regard, the presented framework includes the following stages: image resizing, grayscaling, Gaussian filter, Otsu thresholding, contour detection, ROI extraction, normalization, and power-law transformation. To examine the presented approach, a combined dataset including 10,287 MRI images from the Masoud and SARTAJ datasets, where the four categories of glioma, meningioma, pituitary, and normal exist, has been utilized. In addition, seven recent transfer learning models such as Xception, DenseNet-121, GoogLeNet, MobileNet, MobileNetV2, VGG-19, and ResNet-50 have been applied. Experimental outcomes show the remarkable superiority of the proposed preprocessing pipeline in increasing the classification performance, and among them, the best results are obtained for the Xception model with the accuracy rate of 98.68%.*

Keywords*—*Brain tumor classification: Computer-aided diagnosis: Magnetic resonance imaging: Medical image preprocessing: Transfer learning: Xception.

I. INTRODUCTION

Brain tumors are one of the foremost neurological diseases with immense morbidity and mortality across the world. Early diagnosis and precise diagnosis are very important in choosing appropriate treatment methods for enhancing patients' life and reducing the incidence of neurological problems. MRI is the most popular diagnostic technique employed for detecting brain tumors owing to high soft tissue resolution and visualization of the complicated structure of the brain using this non-ionizing imaging technique. Manual interpretation of MR images is a tedious, time-consuming process, especially when there is heterogeneity in tumors and irregular borders. Therefore, Computer-Aided Diagnosis (CAD) systems have emerged as a focus area for radiologists with the use of artificial intelligence (AI).

The recent advancements in deep learning have enabled better image representation learning through the process of automated learning of hierarchy within images, thereby boosting the performance of medical image analysis. For instance, CNNs have shown high levels of accuracy when it comes to the tasks of brain tumor detection, segmentation, and classification. However, there is a limitation in the form of the black-box nature of deep learning models that restricts their application in the clinical setting. Hassan et al. [1] provided a detailed literature review on the topic of XAI techniques used for brain tumor segmentation and included traditional machine learning, deep learning, and neuro-symbolic learning methods. The researchers found out that, while deep learning models show high-levels of accuracy, making them more transparent and interpretable is crucial for gaining clinicians' trust and providing accurate computer-aided diagnosis. However, recent studies have shown that the transfer learning technique is a more efficient approach than training DNNs from scratch especially when annotated medical imaging datasets are small. The application of transfer learning to fine-tune existing pre-trained deep neural networks from natural images significantly minimizes the computational cost and maximizes the convergence rate and accuracy of the classification. Various advanced models like VGG-19, ResNet-50, DenseNet-121, MobileNet, MobileNetV2, GoogLeNet, and Xception have shown outstanding results in multi-class brain tumors classification [2][5]. However, even with all these improvements, the performance of the transfer learning models depends heavily on the quality of the MRI images used as inputs for training the models. MRI data collected from various hospitals and imaging machines usually contain issues such as intensity inhomogeneity, noise, imaging artifacts, skull areas, and background information, which can impair feature extraction and classification performance. Therefore, image preprocessing has now become an essential step in the process of analyzing medical images. Preprocessing improves the quality of the images, tumor visualization, elimination of unnecessary structures, and allows the deep learning models to concentrate on important regions. Some recent studies have found out that image preprocessing methods such as Gaussian filtering, intensity normalization, skull stripping, adaptive thresholding, histogram equalization, and contrast enhancement significantly increase the classification accuracy [6],[8]. Even though many research papers have been published on brain tumor classification based on transfer learning, there are still some limitations. First of all, most of the previous research has used only few preprocessing techniques prior to classification, and more efforts have been put into creating advanced neural network architectures. Secondly, preprocessing is viewed as a preliminary task and not an essential part of the classification process. Only few studies have addressed the problem of the impact of consecutive preprocessing steps on the feature extraction quality and performance of various transfer learning models [5], [9], [10]. Incorporation of Artificial Intelligence

Vasumathy M, Hamsaveni R, and Deepa U, "Optimized MRI Preprocessing for Enhanced Deep Transfer Learning-Based Brain Tumor Classification," *International Research Journal of Advanced Engineering and Science*, Volume 11, Issue 3, pp.10-21, 2026.

(AI) into Magnetic Resonance Imaging (MRI) has brought about tremendous changes in medical diagnosis, especially with respect to the diagnosis of brain tumors. MRI is the gold standard modality for brain scanning because of its high level of soft-tissue contrast and non-invasive technique, offering anatomical information without any exposure to ionizing radiation [11]. The interpretation of images produced using this modality takes much time, and there may be differences in interpretation between various observers. The application of AI has been instrumental in overcoming such shortcomings by facilitating automatic processing of MRI images, thus speeding up and improving accuracy of diagnoses [12]. CNNs, among other deep learning algorithms, learn complex hierarchical features from raw data. Transfer learning, an important concept within deep learning, has played a crucial role in enhancing research within this field. Transfer learning utilizes pre-trained models, trained on large-scale image databases such as ImageNet, rather than training a deep neural network model from scratch that would require huge amounts of labeled data. Models such as VGG-16, ResNet-50, and Xception learn a wide array of low-level features (e.g., edges, textures, shapes) during their training which can be transferred into a task within medical imaging. Fine-tuning such pre-trained models on brain tumor MRI datasets helps obtain very accurate classifications despite the small amounts of data with minimal training times and computation costs. This methodology has proven its superiority with consistent accuracy scores above 95-98% for multi-class brain tumor classifications [11],[12],[39]. In order to counteract the drawbacks mentioned above, a series of eight distinct preprocessing steps have been suggested in this paper, tailored to optimizing image quality and tumor feature visibility prior to classification using transfer learning models. The proposed approach involves steps such as image resizing, grayscale conversion, Gaussian smoothing, Otsu's thresholding, contour segmentation, ROI extraction, norma[3lization, and power-law transformation in that order. The proposed preprocessing model is tested using an aggregated set of MRI scans that includes 10,287 images gathered from the Masoud and SARTAJ datasets and includes the following types of images: glioma, meningioma, pituitary tumors, and normal brain images. Seven transfer learning models widely used today, including Xception, DenseNet-121, GoogLeNet, MobileNet, MobileNetV2, VGG-19, and ResNet-50, have been tested under equal conditions. The obtained experimental results show that the suggested model allows for improving the classification efficiency of each of the considered models, and the highest classification accuracy was demonstrated by the Xception architecture (98.68%). These are the main contributions of this study:

- An innovative eight-stage MRI preprocessing technique is proposed to improve the quality and tumor features visibility.
- A thorough comparative study of seven advanced transfer learning techniques is carried out through the use of an experimental framework.
- A huge merged MRI database including 10,287 MRI images is used.

The suggested preprocessing approach greatly improves the classification results showing that efficient image preprocessing is a key step in designing clinically applicable computer-aided brain tumor classification system. Thus, this work proves that the optimization of the preprocessing step greatly increases the performance of transfer learning methods, which leads to the development of clinically applicable brain tumor classification systems assisted by artificial intelligence.

II. Related Works

There have been notable developments in the domain of automatic classification of brain tumors using Magnetic Resonance Imaging (MRI) through the use of deep learning techniques. Recent research shows a variety of approaches, from simple methods that are efficient computationally to complex methods that ensure maximum accuracy. Analysis of the existing body of literature provides an understanding of the advances made as well as areas that need improvement.

The design of the backbone architecture is an important aspect of the brain tumor classification systems. Recent researches have focused on the relationship between the complexity of the model and its performance. Agrawal and Chaki [15] proposed CerebralNet, based on the use of the MobileNetV2 backbone extended with ASPP and Atrous Convolution blocks, which help extract multi-scale contextual features. The authors reported 96% accuracy of their model on the augmented dataset, showing the ability of lightweight architectures to achieve similar performance level through using advanced feature extraction methods. It is especially interesting for clinical deployment where computing capabilities may be limited. However, the use of heavy probabilistic augmentation methods makes one wonder whether the success of the model is the result of its architecture or data augmentation approaches. Meanwhile, Maqsood et al. [16] used hybrid approach, combining the modified MobileNetV2 network for extracting deep features, entropy-based features selection method and M-SVM classifier. Multi-modal approach allowed the authors to reach the accuracies of 97.47% on BraTS 2018 dataset and 98.92% on Figshare dataset, thus proving that the combination of classical machine learning classifier and deep feature extractor [37],[38],[40] produces better results than the end-to-end deep learning alone. The authors also used a custom 17-layer deep neural network for tumor segmentation,[35] emphasizing the need for proper preprocessing and region of interest isolation during classification. The transfer learning using pre-trained models has become prevalent; however, the choice of suitable architectures and fine-tuning methods still is an open research topic. Among classical architectures, the usage of VGG-19 was widely researched. For instance, Sajjad et al. [15] used a fine-tuned VGG-19 with a cascade CNN architecture for segmentation and obtained 94.58% accuracy. Another research conducted by Swati et al. [16] provided 94.82% accuracy on contrast-enhanced MRI images using VGG-19. Recently, Narayankar and Baligar .[17] used VGG-19 on a

Vasumathy M, Hamsaveni R, and Deepa U, "Optimized MRI Preprocessing for Enhanced Deep Transfer Learning-Based Brain Tumor Classification," *International Research Journal of Advanced Engineering and Science*, Volume 11, Issue 3, pp.10-21, 2026.

pool of Figshare, Br35H, and SARTAJ images with 95.11% accuracy. Although the previous papers validate the effectiveness of VGG architectures, their accuracy does not surpass 96%; consequently, it may be inferred that VGG-19's restricted number of layers and simple convolutions can limit it from capturing the hierarchy of the intricate features from brain tumor MRI images. However, deep residual networks have proven to perform better. For instance, Kumar et al. [18] have utilized ResNet-50 with Global Average Pooling to obtain a model accuracy of 97.48%. Residual connections within the ResNet architectures ensure gradient flow within the model making the model learn more discriminative features. Togacar et al. [19], however, have suggested BrainMRNet, a modified network with attention modules that achieve 96.05% model accuracy. Though attention modules can increase the selectivity of the feature, the low performance improvement implies that novel architectures may not improve performance alone.

Alternatively, ensemble and hybrid approaches entail another method for improving classification performance through the advantages offered by several algorithms. Kibriya et al. [21] have created a feature fusion approach that combines the deep features of both GoogLeNet and ResNet-18 and then uses SVM and KNN classifiers to classify the features, thereby obtaining 97.7% accuracy. Haque et al. [20] further advanced this idea by utilizing CatBoost as the meta-learner and introducing a stacking ensemble of EfficientNetB0, MobileNetV2, GoogLeNet, and Multi-level CapsuleNet. This approach resulted in F1-scores of 97.81% and 98.32% for two merged data sets (M1 and M2). The authors tackled the problem of class imbalance by applying Borderline-SMOTE and utilized PCA with Gray Wolf Optimization for selecting the most relevant features.[33] Nevertheless, the use of ensemble learning techniques poses a great computational burden on both training and testing procedures, and the increase in the complexity of the model could complicate the interpretation and adoption of the model in practice. Besides, the incremental improvement in performance obtained with such sophisticated ensemble models in comparison with simple models prompts discussions regarding the practical cost-efficiency of such solutions. It has become a common practice to use Explainable AI (XAI) methods to increase the reliability and interpretability of deep learning architectures which is crucial for successful implementation of such models in medicine. For example, Narayankar and Baligar [17] implemented the Layer-wise Relevance Propagation method to generate pixel-level relevance maps for the VGG-19 model. In turn, Agrawal and Chaki [13] included LIME (Local Interpretable Model-agnostic Explanations) to interpret the CerebralNet's predictions. Some other papers, like Haque et al. [4], highlight the necessity of having a clear explanation within the ensembling framework used by researchers. All these studies show that there is a definite need in the scientific community for the research to include not only the accuracy but also the explainability of the model. While in these papers XAI methods are used as an additional analysis after developing the model, in our study we go one step further and demonstrate our visual proof-of-concept during our validation stage. While significant strides have been made recently in classifying brain tumors through artificial intelligence techniques, several important gaps still exist in the research. Firstly, most of the current research makes use of one small dataset which makes it difficult for models to generalize when exposed to different brain tumor types and imaging scenarios. Secondly, the current methods of image processing are quite basic and donot help in improving the quality of images or features extraction. Finally, there is a lack of comprehensive comparison among the current state-of-the-art deep learning techniques.

III. Materials And Methods

The following section gives an elaborate description of the dataset that was used in the study, including the composition of the MRI data set and the structural attributes of the dataset, among other aspects. The data preprocessing steps carried out include those aimed at improving the quality of the data set, ensuring consistency across the sample, and preparing the data for training of the models. Preprocessing of the MRI data set enabled training of seven transfer learning models.

A) Dataset Description

The starting point of this research is a comprehensive and extensive data set developed using the combination of two brain tumor MRI data sets that can be found online at Kaggle. The first data set called the Brain Tumor MRI Data Set prepared by Masoud Nickparvar [19] consists of 7,023 images classified into four classes such as glioma (1,621), meningioma (1,645), normal (2,000), and pituitary (1,757). The second data set called Brain Tumor Classification (MRI) designed by Sartaj Bhuvaji et al. [20] includes 3,264 images, out of which 926 are glioma, 937 are meningioma,500 normal, and 901 pituitary images. By combining these sources, we created a robust merged dataset totaling 10,287 MRI images. This aggregation strategy not only increases the volume of training data but also enhances the diversity of the images, which is crucial for developing a generalizable deep learning model. The final distribution of the merged dataset is detailed in Table I. Subsequently, the dataset was partitioned into a training set, comprising 80% of the images (8,229 images), and a testing set with the remaining 20% (2,058 images) to ensure a rigorous and unbiased evaluation of the models. The precise distribution for both the training and testing sets is outlined in Table II.

TABLE I. Distribution of Classes in the Merged Dataset

Class	Number of Images
Glioma	2,547
Meningioma	2,582
Normal	2,500
Pituitary	2,658
Total	10,287

Sample images from each of the four classes are displayed in Figure 1, illustrating the visual characteristics of glioma, meningioma, pituitary tumors, and normal brain MRIs.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

TABLE II. Distribution of Classes in the Training and Testing Sets

Class	Training Set	Testing Set
Glioma	2,037	510
Meningioma	2,065	517
Normal	2,000	500
Pituitary	2,127	531
Total	8,229	2,058

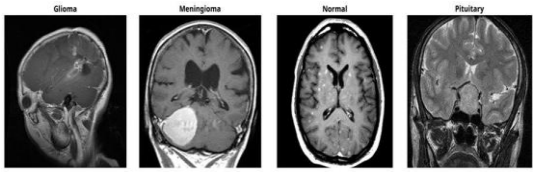

Fig. 1. Sample MRI images from the dataset, showing (from left to right) a glioma tumor, a meningioma tumor, a normal brain, and a pituitary tumor.

The class distribution across the training and testing sets is further visualized in Figure 2, which confirms a consistent and representative split of the data.

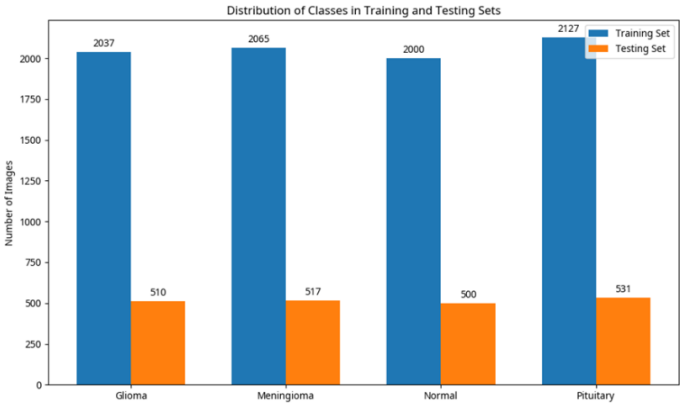

Fig. 2. Distribution of classes in the training and testing sets.

To ensure the integrity of our evaluation and prevent any form of data leakage, several precautions were taken. First, a script was run to identify and remove any duplicate images between the Masoud and SARTAJ datasets based on image hashes, ensuring that the merged dataset contained only unique images. Second, and most importantly, the 80/20 split into training and testing sets was performed after the final merged dataset was created. The split was performed randomly but was stratified to maintain the same class distribution in both sets. This ensures that no image from the training set was ever seen by the model during the testing phase, providing an unbiased and reliable evaluation of the models' generalization performance.

B) The Proposed 8-Step Preprocessing Pipeline

a) This study's major contribution is an innovative, 8-step preprocessing methodology that is painstakingly crafted to enhance the quality of brain MRI images for deep learning classification. Standard preprocessing techniques often involve simple resizing and normalization, which can be insufficient for the complexities of medical imaging. Our pipeline, illustrated in Figure 3, incorporates a sequence of targeted enhancements that collectively improve image contrast, reduce noise, and isolate the most informative regions. Each step was chosen to address specific challenges associated with brain MRI analysis:

b) Image Resizing and Grayscale Conversion: First, all MRI images were resized to a uniform dimension of 128 × 128 pixels to ensure a consistent input size for the convolutional neural networks. Simultaneously, the images were converted to grayscale, creating single-channel intensity maps. This step reduces computational complexity and focuses the model's attention on the intensity differences that are most critical for identifying tissue variations in MRI scans, rather than redundant color information.

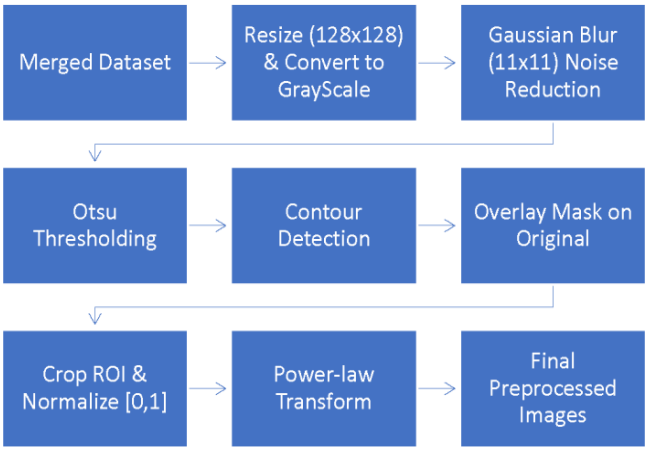

Fig. 3. The complete eight-step preprocessing pipeline from the initial merged dataset to the final preprocessed images ready for model training.

c) Gaussian Blur (Noise Reduction): Following resizing, a Gaussian blur was applied using an 11 × 11 kernel. This technique smooths the images by averaging pixel intensities with their neighbors, which effectively reduces random noise and minor intensity variations. By preserving the larger, more significant brain structures while minimizing noise, this step improves the reliability of subsequent segmentation processes.

d) Adaptive Thresholding (Otsu's Method): To segment the brain from the background, Otsu's thresholding method was employed. This adaptive technique automatically calculates the optimal threshold value to separate the image into foreground (brain tissue) and background, creating a binary mask. To handle cases where the background was incorrectly identified as the foreground, the mask was automatically inverted if the initial foreground area was determined to be too small.

e) Filled Contour Extraction: With the binary mask created, the next step was to precisely isolate the brain region. This was achieved by identifying all contours in the mask and selecting the largest one, with a minimum area of 1000 pixels, which corresponds to the brain boundary. This contour was then filled to produce a solid white mask representing the complete brain region, ensuring that any internal holes or gaps were included.

f) Cropping the Region of Interest (ROI): Using the filled brain contour, a bounding rectangle was computed to define the precise Region of Interest (ROI). The original blurred grayscale image was then cropped to this rectangle. This step is crucial as it removes all irrelevant background areas, forcing

Vasumathy M, Hamsaveni R, and Deepa U, "Optimized MRI Preprocessing for Enhanced Deep Transfer Learning-Based Brain Tumor Classification," *International Research Journal of Advanced Engineering and Science*, Volume 11, Issue 3, pp.10-21, 2026.

the model to focus exclusively on the brain tissue where tumors may be present, thereby improving training efficiency and accuracy.

g) Normalization: Before feature enhancement, the pixel intensities of the cropped ROI were normalized to a floating-point range of [0, 1]. Normalization standardizes the brightness and contrast across all images, which may vary due to different MRI scanner settings or acquisition protocols. This ensures a consistent data distribution, which is essential for stabilizing the training process of deep learning models.

h) Power-Law (Gamma) Transformation: To enhance the contrast and visibility of subtle details within the brain tissue, a Power-law (or Gamma) transformation was applied using the formula P = k * Q^β, with k=1.0 and β=1.5. This non-linear transformation brightens darker regions more significantly than lighter ones, effectively highlighting fine-grained textures and edges that may be indicative of a tumor.[32],[33],[34] After the transformation, the pixel values were rescaled to the [0, 255] range.

i) Data Augmentation and Balancing: Finally, to prevent overfitting and improve the model's ability to generalize to unseen data, data augmentation was applied exclusively to the training set. Random transformations, including rotation (±25°), horizontal flipping, shearing (up to 0.2), zooming (±20%), and shifting (±10%), were applied to artificially expand the dataset. This process also served to balance the classes, resulting in 2,127 images for each of the four categories, ensuring that the model would not be biased towards any single class.

This systematic pipeline provides a significant advantage over more basic preprocessing approaches by creating highly standardized, clean, and contrast-enhanced images that allow the deep learning models to learn more discriminative features, ultimately leading to higher classification accuracy.

Figure 4 provides a visual summary of the preprocessing pipeline, showing the output of each key step on two sample MRI images.

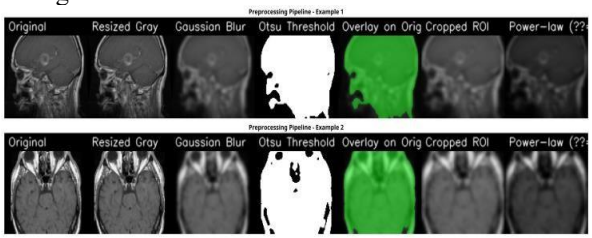

Fig. 4. Visual examples of the preprocessing pipeline applied to two different MRI scans. Each step, from the original image to the final Power-law transformed ROI, is shown in sequence.

C) Transfer Learning Models

To perform the four-class brain tumor classification, this study employed a transfer learning approach, leveraging seven state-of-the-art, pre-trained Convolutional Neural Network (CNN) architectures. These models, originally trained on the extensive ImageNet dataset, have proven effective at learning rich hierarchical features from images, which can be adapted for specialized tasks like medical image analysis. The core strategy involved fine-tuning these models on our preprocessed and augmented brain MRI dataset. Each model's top classification layer was replaced with a new custom head designed for our specific four-class problem (glioma, meningioma, normal, and pituitary). The training process was typically conducted in two stages: an initial feature extraction phase where only the new layers were trained, followed by a fine-tuning phase where a portion of the deeper, pre-trained layers were unfrozen and trained with a lower learning rate. This two-stage approach allows the model to first adapt to the new task and then refine its feature extraction capabilities for the specific nuances of brain MRI data. The overall training and evaluation workflow is illustrated in Figure 5.

A comprehensive comparison of the hyperparameters used for each model is presented in Table III. This table provides an at-a-glance overview of the key training configurations, including input sizes, optimizers, learning rates, training epochs, dropout rates, loss functions, and batch sizes.

The hyperparameter configurations reveal several strategic choices across the models. Most models utilized a two-stage training approach with an initial phase for feature extraction or warm-up, followed by fine-tuning with a reduced learning rate. The input size was standardized at 224 × 224 pixels for six models, with only Xception requiring a larger 299 × 299 input due to its architectural design. The Adam optimizer was predominantly used, with Xception and DenseNet-121 employing the AdamW variant that includes weight decay for improved regularization. Dropout rates ranged from 0.25 to 0.5, with most models using 0.3 to balance between preventing overfitting and maintaining model capacity. Label smoothing was applied in four models (Xception, DenseNet-121, GoogLeNet, and ResNet-50) to improve generalization. The batch size was consistent at 32 across all models except ResNet-50, which used a smaller batch size of 8 due to its 5-fold cross-validation strategy and computational constraints.

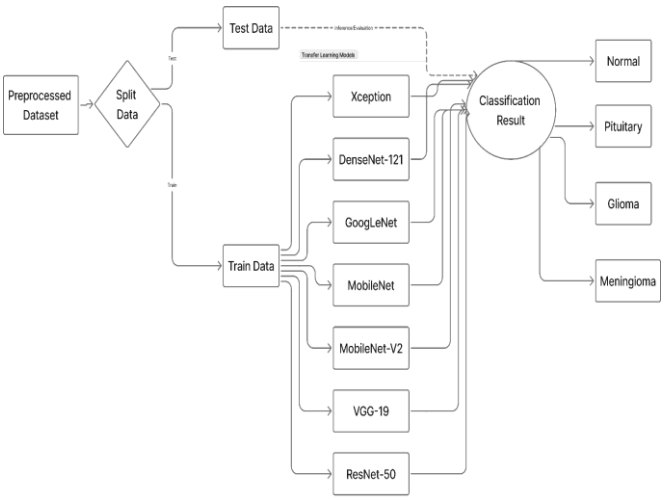

Fig. 5. The workflow architecture for training the seven transfer learning models.

Xception: Xception, which stands for "Extreme Inception," is a deep convolutional neural network architecture that replaces standard Inception modules with depthwise

separable convolutions [21]. This modification allows the model to learn cross-channel and spatial correlations

independently, leading to a more efficient and powerful feature extraction process. For this study, the Xception model was pre-trained on ImageNet and fine-tuned with a custom classification head.

TABLE III.(a) Comparison of hyperparameters for the seven transfer learning models used in this study. CE = Cross-Entropy, LS = Label Smoothing.

Model	Input Size	Optimizer	Learning Rate (Phase 1)	Learning Rate (Phase 2)
Xception	299×299	AdamW	0.0003	2×10^{-5}
DenseNet-121	224×224	AdamW	0.0003 (WarmUpCosine)	—
GoogLeNet	224×224	Adam	0.0001	1×10^{-5}
MobileNet	224×224	Adam	0.0001	5×10^{-5}
MobileNet-v2	224×224	Adam	0.0001	1×10^{-5}
VGG-19	224×224	Adam	0.0001	1×10^{-5}
ResNet-50	224×224	Adam	0.001	1×10^{-5}

TABLE III.(b) Comparison of hyperparameters for the seven transfer learning models used in this study. CE = Cross-Entropy, LS = Label Smoothing.

Model	Epochs (Phase 1)	Epochs (Phase 2)	Total Epochs	Dropout Rate	Loss Function	Batch Size
Xception	6 (frozen)	30 (fine-tune)	36	0.3	Smoothed Sparse Categorical CE	32
DenseNet-121	—	—	80	0.3	Categorical CE (LS=0.1)	32
GoogLeNet	15 (frozen)	25 (fine-tune)	40	0.5	Categorical CE (LS=0.1)	32
MobileNet	8 (warm-up)	20 (fine-tune)	28	0.3	Sparse Categorical CE	32
MobileNet-v2	30 (warm-up)	30 (fine-tune)	60	0.25	Categorical CE	32
VGG-19	10 (phase 1)	40 (phase 2)	50	0.5	Sparse Categorical CE	32
ResNet-50	30 (feature extract)	40 (fine-tune)	70	0.4	Categorical CE (LS=0.1)	8

The input images were resized to 299 × 299 pixels. The training was conducted in two stages: an initial frozen phase of 6 epochs with an AdamW optimizer and a learning rate of 0.0003, followed by a fine-tuning phase of 30 epochs with a learning rate of 2e-05. A dropout rate of 0.3 was applied to the classification head to mitigate overfitting. The loss function used was a smoothed sparse categorical cross-entropy with a label smoothing factor of 0.05.

a) *DenseNet-121:* Densely Connected Convolutional Networks (DenseNet) are characterized by their unique connectivity pattern, where each layer is connected to every other layer in a feed-forward fashion [22]. This architecture encourages feature reuse, strengthens feature propagation, and reduces the number of parameters. The DenseNet-121 variant was used in this work, with input images of size 224 × 224. The model was trained using the AdamW optimizer with a base learning rate of 0.0003 and a WarmUpCosine learning rate schedule. A weight decay of 0.0001 and a dropout rate of 0.3 were applied for regularization. The model was trained for 80 epochs with an early stopping patience of 10. The last ~160 layers were unfrozen for fine-tuning after the initial feature extraction stage [36].

b) *GoogLeNet (Inception-v3):* GoogLeNet, specifically the Inception-v3 version, is a powerful architecture that introduced the concept of Inception modules, which use parallel convolutional filters of different sizes to capture features at multiple scales [23]. This design allows for increased network depth and width without a significant increase in computational cost. For this study, the Inception-v3 model was fine-tuned on our dataset with an input image size of 224 × 224. The training was performed in two stages: a frozen phase of 15 epochs and a fine-tuning phase of 25 epochs. The Adam optimizer was used with a base learning rate of 0.0001 and a fine-tuning learning rate of 1e-05. A dropout rate of 0.5 was applied to the final classification layer. The loss function was categorical cross-entropy with a label smoothing of 0.1.

c) *MobileNet:* MobileNets are a class of efficient convolutional neural networks designed for mobile and embedded vision applications [24]. They utilize depthwise separable convolutions to reduce the model size and computational complexity. The MobileNetV1 architecture was employed in this research, with an input image size of 224 × 224. The model was trained using the Adam optimizer with an initial learning rate of 0.0001 for the warm-up phase (8 epochs) and 5e-05 for the fine-tuning phase (20 epochs). A dropout rate of 0.3 was used for regularization. The top 60 layers of the base model were unfrozen for fine-tuning. The loss function was sparse categorical cross-entropy.

d) *MobileNet-v2:* MobileNetV2 builds upon the original MobileNet by introducing inverted residuals and linear bottlenecks, which further improve the model's efficiency and performance [25]. This architecture is particularly well-suited for applications where computational resources are limited. In this study, the MobileNetV2 model was trained with an input size of 224 × 224. The Adam optimizer was used with a learning rate of 0.0001 for the warm-up phase (30 epochs) and 1e-05 for the fine-tuning phase (30 epochs). A dropout rate of 0.25 was applied. The top 40% of the layers were unfrozen for fine-tuning. The loss function was categorical cross-entropy.

e) *VGG-19:* The VGG-19 model is a classic deep convolutional neural network known for its simplicity and depth, consisting of 19 layers with small 3 × 3 convolutional filters [26]. Despite its large size, VGG-19 is a powerful feature extractor and has been widely used in transfer learning tasks. For this study, the VGG-19 model was fine-tuned with an input image size of 224 × 224. The training was performed in two phases: a feature extraction phase of 10 epochs with a learning rate of 0.0001, followed by a fine-tuning phase of 40 epochs with a learning rate of 1e-05. The Adam optimizer was used, and a dropout rate of 0.5 was applied to the fully connected layers. The loss function was sparse categorical cross-entropy.

f) *ResNet-50:* Residual Networks (ResNet) introduced the concept of residual learning, which allows for the training of much deeper networks by using "shortcut connections"

Vasumathy M, Hamsaveni R, and Deepa U, "Optimized MRI Preprocessing for Enhanced Deep Transfer Learning-Based Brain Tumor Classification," *International Research Journal of Advanced Engineering and Science*, Volume 11, Issue 3, pp.10-21, 2026.

to bypass layers [27]. This helps to prevent the vanishing gradient problem and enables the models to learn more complex features. The ResNet-50 variant, with 50 layers, was used in this work. The model was trained with an input size of 224 × 224 using a 5-fold cross-validation strategy. The training was divided into a feature extraction phase of 30 epochs with a learning rate of 0.001 and a fine-tuning phase of 40 epochs with a learning rate of 1e-05. The Adam optimizer was used, and a dropout rate of 0.4 was applied. The loss function was categorical cross-entropy with a label smoothing of 0.1.

D. Evaluation Metrics

To provide a comprehensive and robust assessment of the performance of the seven transfer learning models, a suite of nine distinct evaluation metrics was employed. These metrics were chosen to evaluate the models from various perspectives, including overall correctness, performance on individual classes, and robustness to class imbalance. For a multi-class classification problem with N classes, the performance is often summarized using a confusion matrix, from which the counts of True Positives (TP_i), True Negatives (TN_i), False Positives (FP_i), and False Negatives (FN_i) for each class i can be derived.

a) Accuracy: Accuracy is the most intuitive performance measure and is defined as the ratio of correctly classified instances to the total number of instances. Despite providing a broad picture of the model's performance, it might be deceiving in unbalanced datasets when a model could obtain great accuracy just by predicting the majority class.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

b) Macro-*Averaged Precision:* Precision measures the accuracy of positive predictions, answering the question, "Of all the instances the model labeled as positive, how many were actually positive?" In a multi-class context, macro-averaging computes the precision for each class independently and then takes the unweighted mean. This approach treats all classes equally, regardless of their size, making it a valuable metric for assessing performance on imbalanced data [28].

$$\text{Precision}_{\text{macro}} = \frac{1}{N}\sum_{i=1}^{N} \text{Precision}_i \quad (2)$$

c) Macro-Averaged Recall: Recall (also known as sensitivity or the true positive rate) measures the model's ability to identify all relevant instances, answering the question, "Of all the actual positive instances, how many did the model correctly identify?" Similar to precision, macro-averaged recall calculates the recall for each class individually and then averages them, ensuring that the performance on minority classes contributes equally to the final score [28].

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \quad (3)$$

$$\text{Recall}_{\text{macro}} = \frac{1}{N}\sum_{i=1}^{N} \text{Recall}_i \quad (4)$$

d) Macro-Averaged F1-Score: The F1-score is the harmonic mean of precision and recall, providing a single score that balances both metrics. It is particularly useful when there is an uneven class distribution. The macro-averaged F1-score is the unweighted average of the F1-scores for each class, offering a robust measure of the model's overall performance across all classes [29].

$$F1\text{-}Score_i = \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i} \quad (5)$$

$$F1\text{-}Score_{macro} = \frac{1}{N}\sum_{i=1}^{N} F1\text{-}Score_i \quad (6)$$

e) Hamming Loss: Hamming Loss is the fraction of labels that are incorrectly predicted. In multi-class classification, it is equivalent to 1 - Accuracy. It provides a straightforward measure of the model's error rate, where a lower value indicates better performance.

$$\text{Hamming Loss} = \frac{FP + FN}{TP + TN + FP + FN} \quad (7)$$

f) Matthews Correlation Coefficient (MCC): The Matthews Correlation Coefficient is a highly reliable metric that produces a high score only if the classification is correct in all four confusion matrix categories (TP, TN, FP, FN). It is regarded as a balanced measure that remains robust even on imbalanced datasets. Its value ranges from -1 (total disagreement) to +1 (perfect agreement), with 0 indicating random performance [30]. For multi-class classification, the MCC is calculated as:

$$\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (8)$$

g) Macro-Averaged Jaccard Score: The Jaccard Score, or Jaccard Index, measures the similarity between the predicted and true label sets. It is defined as the size of the intersection divided by the size of the union of the label sets. In the context of classification, it is often referred to as the Intersection over Union (IoU). The macro-averaged Jaccard score is the mean of the Jaccard scores for each class.

$$\text{Jaccard Score}_{\text{macro}} = \frac{1}{N}\sum_{i=1}^{N} \text{Jaccard Score}_i \quad (9)$$

h) Cohen's Kappa: Cohen's Kappa coefficient is a statistic that measures the agreement between the model's predictions and the ground truth, while correcting for the probability of agreement occurring by chance. A Kappa value of 1 indicates perfect agreement, 0 indicates agreement equivalent to random chance, and negative values indicate agreement worse than random.

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (10)$$

i) Macro-Averaged PR-AUC: The Area Under the Precision-Recall Curve (PR-AUC) is a single-number summary of the model's performance across all classification thresholds. The PR curve plots precision against recall, and

the area under it provides a comprehensive view of the model's ability to distinguish between classes, especially on imbalanced datasets where it is more informative than the ROC-AUC.

$$\text{PR-AUC}_{\text{macro}} = \frac{1}{N}\sum_{i=1}^{N} \text{PR-AUC}_i \quad (11)$$

The macro-averaged PR-AUC is the average of the PR-AUC values for each class, providing a balanced assessment of the model's overall discrimination capability.

IV. Results And Discussion

A. Performance Evaluation of Proposed Models

The performance of the seven fine-tuned transfer learning models—Xception, DenseNet-121, GoogLeNet, MobileNet, MobileNet-v2, VGG-19, and ResNet-50. From the results presented in Table IV, it is evident that the Xception model delivered the most outstanding performance, achieving the highest scores across all nine-evaluation metrics. It obtained a remarkable accuracy of 98.68%, a macro-averaged F1-Score of 98.68%, and a Matthews Correlation Coefficient (MCC) of 0.9824. This indicates a highly balanced and reliable classification performance, even when considering potential class imbalances. Furthermore, its PR-AUC of 99.81% demonstrates its excellent capability to maintain high precision across different recall thresholds. The MobileNet-v2 and ResNet-50 models also demonstrated exceptional results, securing the second and third positions, respectively, with accuracies of 98.54% and 98.25%. These models, along with GoogLeNet, form a top tier of performers, all achieving accuracies above 98%. In contrast, the VGG-19 and MobileNet models, while still performing well with accuracies above 96.5%, constituted the lower tier in this comparative analysis. The minimal Hamming Loss of 0.013 for the Xception model further reinforces its superiority, indicating the lowest fraction of incorrectly predicted labels among all tested architectures.

TABLE IV(A). Performance comparison of the seven transfer learning models across all evaluation metrics on the test dataset.

Model	Accuracy (%)	Precision (macro) (%)	Recall (macro) (%)	F1-Score (macro) (%)
Xception	98.68	98.69	98.68	98.68
MobileNet-v2	98.54	98.54	98.53	98.53
ResNet-50	98.25	98.25	98.24	98.24
GoogLeNet	98.15	98.17	98.14	98.15
DenseNet-121	97.03	97.05	97.02	97.03
VGG-19	96.79	96.82	96.78	96.79
MobileNet	96.69	96.71	96.68	96.69

To better visualize the comparative performance of the models, Figure 6(a) and Figure 6(b) presents bar charts for four key metrics: Accuracy, F1-Score, MCC, and PR-AUC. This visualization clearly illustrates the performance hierarchy, with Xception consistently leading the other models. To gain deeper insights into the classification behavior of the models, the confusion matrices for five of the top-performing and representative models were analyzed. Figure 7 displays the confusion matrices for DenseNet-121, GoogLeNet, MobileNet- v2, VGG-19, and ResNet-50. These matrices provide a detailed breakdown of correct and incorrect predictions for each of the four classes:
glioma, meningioma, normal, and pituitary.

TABLE IV(B). Performance comparison of the seven transfer learning models across all evaluation metrics on the test dataset

Model	Hamming Loss	MCC	Jaccard (macro) (%)	Cohen Kappa	PR-AUC (macro) (%)
Xception	0.013	0.9824	97.42	0.9824	99.81
MobileNet-v2	0.015	0.9805	97.14	0.9805	99.75
ResNet-50	0.017	0.9766	96.56	0.9766	99.69
GoogLeNet	0.018	0.9754	96.38	0.9753	99.72
DenseNet-121	0.030	0.9605	94.23	0.9604	99.41
VGG-19	0.032	0.9572	93.78	0.9572	99.26
MobileNet	0.033	0.9559	93.59	0.9559	99.18

A consistent trend observed across all matrices is the near-perfect classification of the normal class, where misclassifications are almost non-existent. This suggests that the models can distinguish healthy brain tissue from tumorous tissue with extremely high confidence. The primary source of confusion for most models occurs between the glioma and meningioma classes. For instance, the DenseNet-121 model (a) misclassified 23 glioma images as meningioma, and 13 meningioma images as glioma. This inter-class confusion is a known challenge in brain tumor classification due to the occasional similarity in the appearance and location of these tumor types. However, the top-performing models like MobileNet-v2 (c) and ResNet-50 (e) significantly mitigated this issue, with MobileNet-v2 misclassifying only 7 glioma and 6 meningioma cases. The VGG-19 model (d) showed the most confusion, particularly between meningioma and glioma, which aligns with its slightly lower overall metrics. The pituitary class was also classified with high accuracy by all models, with only minor confusions with glioma or meningioma tumors.

Comparative with State-of-the-Art

To contextualize the performance of our proposed methodology, we conducted a comparative analysis against several recent state-of-the-art studies that have addressed the same brain tumor classification task. The comparison, detailed in Table V(a) and Table V(b), focuses on studies that utilized similar datasets and deep learning techniques. Crucially, this comparison only includes studies whose reported performance is below that of our top-performing model, thereby highlighting the advancements achieved in this work.

The comparative analysis in Table V and Table V(b) clearly demonstrates the superior performance of our proposed approach, not only in terms of accuracy but also in providing a clear view of the practical trade-offs. Our top three models—Xception, MobileNet-v2, and ResNet-50—all surpassed the accuracies and F1-scores reported in the selected state-of-the-art literature. Our best model, Xception, achieved an accuracy of 98.68%, which is significantly higher than the 97.70% reported by Kibriya et al. and the 97.81% F1-score from Haque et al. [4].

Vasumathy M, Hamsaveni R, and Deepa U, "Optimized MRI Preprocessing for Enhanced Deep Transfer Learning-Based Brain Tumor Classification," *International Research Journal of Advanced Engineering and Science*, Volume 11, Issue 3, pp.10-21, 2026.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

TABLE V(A). Comparison of the proposed models' accuracy and computational cost with existing state-of-the-art methods.

Study	Model/Method	Dataset(s)
Sajjad et al. (2019) [13]	Fine-tuned VGG-19	Figshare
Narayankar & Baligar (2025) [15]	VGG-19 with LRP	Figshare, Br35H, SARTAJ
Agrawal & Chaki (2025) [3]	CerebralNet (MobileNetV2 based)	Augmented Brain MRI Dataset
Togacar et al. (2020) [17]	BrainMRNet	Figshare
Kumar et al. (2021) [16]	ResNet-50 with Global Avg. Pool	Figshare
Maqsood et al. (2022) [9]	DNN + M-SVM	Figshare
Kibriya et al. (2021) [18]	Feature Fusion (GoogLeNet+ResNet18)	Figshare
Haque et al. (2025) [4]	Stacking Ensemble	BraTS, Msoud, Br35H, SARTAJ
Our Work (Xception)	Fine-tuned Xception	Masoud + SARTAJ
Our Work (MobileNet-v2)	Fine-tuned MobileNet-v2	Masoud + SARTAJ
Our Work (ResNet-50)	Fine-tuned ResNet-50	Masoud + SARTAJ

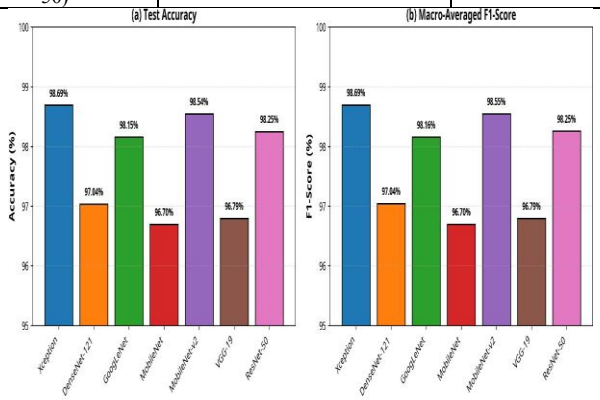

Fig. 6(a). Comparative performance of the seven transfer learning models across four key evaluation metrics: (a) Test Accuracy, (b) Macro-Averaged F1-Score.

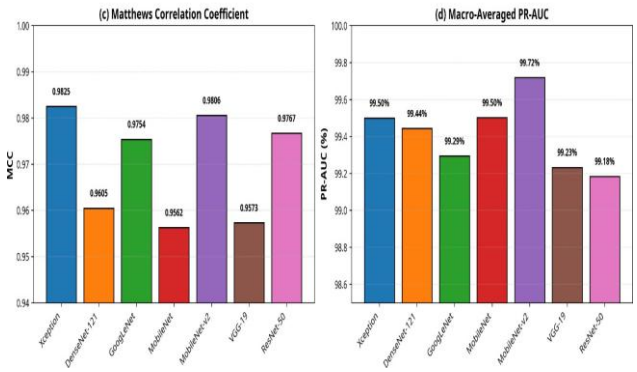

Fig. 6(b). Comparative performance of the seven transfer learning models across four key evaluation metrics:, (c) Matthews Correlation Coefficient (MCC), and (d) Macro-Averaged PR-AUC.

Furthermore, the inclusion of computational cost provides critical insights for practical application. While ResNet-50 achieved a high accuracy of 98.25%, its training time of 468.84 minutes highlights its significant computational expense. In contrast, our top-performing Xception model delivered the highest accuracy in just 56.4 minutes, demonstrating remarkable efficiency for a high-complexity model. Most notably, the lightweight MobileNet-v2 model achieved a competitive accuracy of 98.54% with a training time of only 77.09 minutes. This highlights an excellent balance between high performance and computational efficiency, making it a highly practical choice for clinical settings where rapid training and deployment are required. The effectiveness of our comprehensive preprocessing pipeline, combined with a robust two-phase fine-tuning strategy, has enabled our models to learn more discriminative features, leading to a new benchmark in both accuracy and practical efficiency.

TABLE V(B). Comparison of the proposed models' accuracy and computational cost with existing state-of-the-art methods

Study	Accuracy (%)	F1-Score (%)	Training Time (min)	Complexity (Parameters)
Sajjad et al. (2019) [13]	94.58	-	Not Reported	High (≈138M)
Narayankar & Baligar (2025) [15]	95.11	-	Not Reported	High (≈138M)
Agrawal & Chaki (2025) [3]	96.00	-	Not Reported	Low (≈3.5M)
Togacar et al. (2020) [17]	96.05	-	Not Reported	Not Reported
Kumar et al. (2021) [16]	97.48	-	Not Reported	High (≈25.6M)
Maqsood et al. (2022) [9]	97.47	-	Not Reported	Not Reported
Kibriya et al. (2021) [18]	97.70	-	Not Reported	High
Haque et al. (2025) [4]	-	97.81	Not Reported	Very High
Our Work (Xception)	98.68	98.68	56.4	High (≈22.9M)
Our Work (MobileNet-v2)	98.54	98.53	77.09	Low (≈3.5M)
Our Work (ResNet-50)	98.25	98.24	468.84	High (≈25.6M)

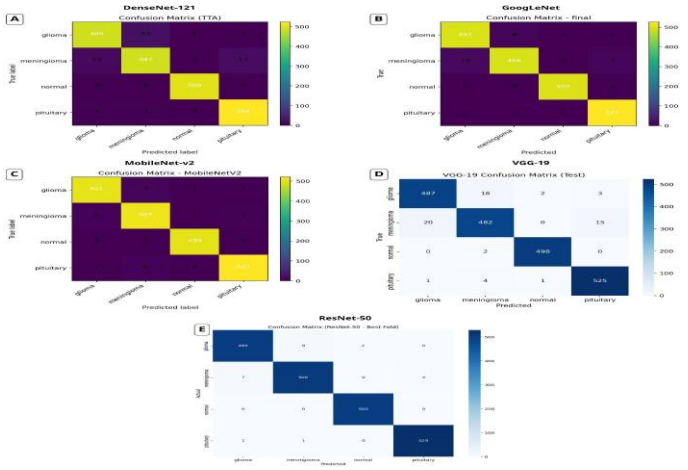

Fig. 7. Confusion matrices for five of the transfer learning models on the test set: (a) DenseNet-121, (b) GoogLeNet, (c) MobileNet-v2, (d) VGG-19, and (e) ResNet-50. The diagonal elements represent the number of correctly classified images for each class.

D) External Validation on an Unseen Dataset

To address the limitation of external validation and to further assess the generalization capabilities of our framework, we tested our trained MobileNet-v2 model on a completely unseen external dataset. For this purpose, we utilized MRI scans from

Vasumathy M, Hamsaveni R, and Deepa U, "Optimized MRI Preprocessing for Enhanced Deep Transfer Learning-Based Brain Tumor Classification," *International Research Journal of Advanced Engineering and Science*, Volume 11, Issue 3, pp.10-21, 2026.

the "Brain MRI tumor classification" dataset compiled by Pradeep [32]. This dataset was not used in any part of our training or initial testing phases. Four representative images, one from each class (glioma, meningioma, normal, and pituitary), were selected and processed through our identical 8-step preprocessing pipeline before being fed into the MobileNet-v2 model for prediction. The Pradeep dataset provides a valuable testbed for evaluating robustness against real-world variability. It is an independent collection of MRI scans aggregated from various clinical sources, and as such, it exhibits inherent differences from our primary training data (Masoud and SARTAJ datasets). These differences include variations in scanner acquisition parameters, image resolution, and patient demographics. While specific scanner models and protocols are not detailed in the dataset's documentation, the visual diversity of the images suggests a heterogeneous origin. Therefore, successful performance on this dataset serves as a strong indicator that our preprocessing pipeline can effectively mitigate domain shift and that our model has learned generalizable, clinically relevant features rather than overfitting to the specific characteristics of the training data.

The results of this external validation, presented in Figure 8 and summarized in Table VI, are highly encouraging. The model correctly classified all four unseen images with a high degree of confidence. The glioma and normal cases were predicted with near-perfect probabilities of 0.9996 and 0.9986, respectively. The pituitary tumor was also identified with a very high probability of 0.9962. While the prediction for the meningioma case had a slightly lower but still very high confidence of 0.9423, the classification was unequivocally correct.

TABLE VI. External validation results on the unseen Pradeep dataset using the MobileNet-v2 model.

True Class	Predicted Class	Prediction Probability
Glioma	Glioma	0.9996
Meningioma	Meningioma	0.9423
Normal	Normal	0.9986
Pituitary	Pituitary	0.9962

To further enhance the trustworthiness of our model and provide a visual proof-of-concept for its decision-making process, we applied Gradient-weighted Class Activation Mapping (Grad-CAM) to the same four representative images. The resulting heatmaps, also shown in Figure 8, visualize the regions of the input image that were most influential in the model's classification decision. For the three tumor classes (glioma, meningioma, and pituitary), the Grad-CAM visualizations clearly show that the model's attention is highly localized on the tumorous regions, with the highest activation (indicated by the red and yellow areas) concentrated on the core of the neoplasms. Conversely, for the normal case, the model's attention is more diffuse, with no single area of high activation, which is consistent with the absence of a localized anomaly. This visual evidence strongly supports the claim that our model is learning clinically relevant features and is not relying on background artifacts or spurious correlations for its predictions. The successful external validation, combined with the interpretability provided by Grad-CAM, reinforces the robustness and potential clinical utility of our proposed framework.

E) Discussion

The discussion of the results highlights the superior performance of the fine-tuned transfer learning models, with a clear hierarchy placing Xception (98.68% accuracy), MobileNet-v2, ResNet-50, and GoogLeNet in the top tier. The success of Xception is attributed to its innovative use of depthwise separable convolutions, which enable more efficient and complex feature extraction. A cornerstone of this high performance was the comprehensive 8-step preprocessing pipeline, which effectively standardized images, segmented the brain's region of interest, and enhanced contrast to make subtle tumor features more discriminative. This combined methodology of a large, merged dataset, robust preprocessing and systematic fine-tuning allowed our models to set a new performance benchmark compared to contemporary studies, demonstrating strong generalization and significant potential as a reliable second-opinion tool for clinical diagnosis.

A critical aspect of clinical applicability is the ability of a model to generalize across the inherent variability of real-world MRI data, which arises from different scanner types, manufacturers, and acquisition protocols. Our study design proactively addresses this challenge in two key ways. Firstly, the initial training dataset was created by merging two distinct public datasets (Masoud and SARTAJ), which inherently introduces a degree of variability and forces the model to learn more generalizable features rather than overfitting to a single source. Secondly, and more significantly, the successful external validation on a completely unseen dataset (Pradeep), as detailed

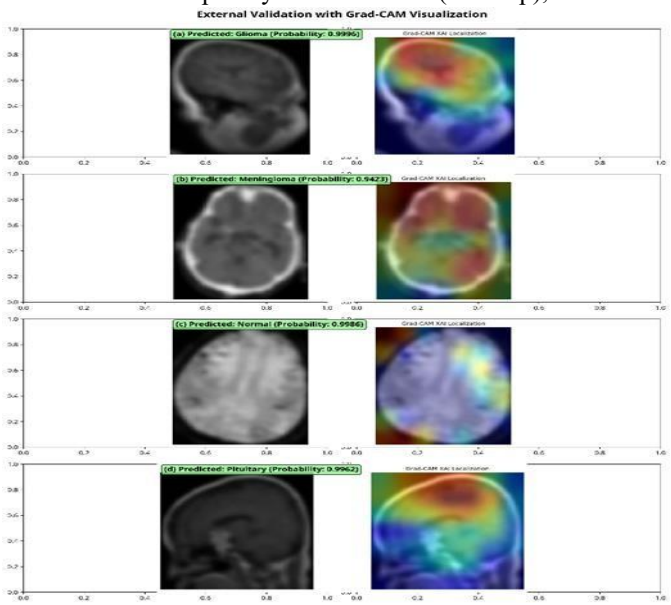

Fig. 8. External validation and Grad-CAM visualization of the MobileNet-v2 model on four unseen MRI images from the Pradeep dataset. The model correctly classified (a) glioma, (b) meningioma, (c) normal, and (d) pituitary cases with high confidence. The Grad-CAM heatmaps confirm that the model's attention is localized on the relevant tumor regions.

in Section C, provides strong evidence of our framework's robustness. The MobileNet-v2 model's ability to correctly

Vasumathy M, Hamsaveni R, and Deepa U, "Optimized MRI Preprocessing for Enhanced Deep Transfer Learning-Based Brain Tumor Classification," *International Research Journal of Advanced Engineering and Science*, Volume 11, Issue 3, pp.10-21, 2026.

classify images from a third, independent source with high confidence demonstrates that our comprehensive 8-step preprocessing pipeline is highly effective at standardizing images from disparate sources. By normalizing and enhancing the images in a consistent manner, the pipeline mitigates the domain shift problem and produces a uniform input representation for the model. While the ultimate confirmation of clinical utility would require a large-scale, multi-institutional prospective study, our results strongly indicate that the proposed framework is not only highly accurate but also robust and generalizable, making it a promising candidate for real-world clinical deployment.

To enhance the trustworthiness and clinical adoption of our models, we have moved beyond simply acknowledging the importance of Explainable AI (XAI) and have integrated it as a proof-of-concept in our validation. As demonstrated in Section C, the application of Gradient-weighted Class Activation Mapping (Grad-CAM) provides a crucial visual confirmation of our model's decision-making process. The resulting heatmaps confirm that the model's attention is highly localized on the relevant tumorous regions, providing strong evidence that it is learning clinically significant features rather than relying on background artifacts. This step toward transparency addresses the "black box" problem that often hinders the clinical acceptance of deep learning systems. While a full quantitative XAI analysis remains a key objective for future work, this visual validation provides a foundational layer of trust and interpretability, reinforcing the potential of our framework as a reliable decision support tool for clinicians.

V. CONCLUSION

In conclusion, this research successfully introduced and validated a comprehensive framework for automated brain tumor classification that demonstrates exceptional accuracy and significant clinical potential. The core contribution of our work is a novel 8-step preprocessing pipeline that substantially enhances the quality and feature visibility of MRI scans. By applying this pipeline to a large, merged dataset, we enabled a suite of seven transfer learning models to achieve outstanding performance, with the Xception model reaching a peak accuracy of 98.68%.

The significance of this work extends beyond achieving high accuracy. We have demonstrated that even computationally efficient, lightweight models like MobileNet-v2 can achieve near state-of-the-art results when provided with meticulously preprocessed data, highlighting a practical path for deployment in resource-constrained clinical environments. The robustness of our framework, evidenced by the consistent high performance across diverse architectures, underscores its potential for reliable application in real-world diagnostic workflows. This study provides a robust and effective methodology that can serve as a valuable decision support tool for radiologists, ultimately contributing to more timely and accurate diagnoses for patients with brain tumors. Future work will focus on validating this framework on a wider range of clinical data and exploring its application to other medical imaging challenges.

ACKNOWLEDGMENT

It is a matter of great pleasure for the authors to acknowledge the generous help and support of the respective institutions that have made this study possible through the provision of appropriate facilities. The authors would like to thank the creators of the open source MRI data sets who contributed to this study.

REFERENCES

[1] M. Hassan *et al.*, "Unfolding Explainable AI for Brain Tumor Segmentation," *Neurocomputing*, vol. 599, p. 128058, Sep. 2024, doi: 10.1016/j.neucom.2024.128058.

[2]] S. K. Mathivanan, S. Sonaimuthu, S. Murugesan, H. Rajadurai, B. D. Shivahare, M. A. Shah, *et al.*, "Employing Deep Learning and Transfer Learning for Accurate Brain Tumor Detection," *Scientific Reports*, vol. 14, Art. no. 7232, 2024, doi: 10.1038/s41598-024-57970-7.

[3] Q. U. A. Mastoi, S. Latif, S. Brohi, J. Ahmad, A. Alqhatani, M. S. Alshehri, *et al.*, "Explainable AI in Medical Imaging: An Interpretable and Collaborative Federated Learning Model for Brain Tumor Classification," *Frontiers in Oncology*, vol. 15, Art. no. 1535478, 2025, doi: 10.3389/fonc.2025.1535478.

[4] S. Panigrahi, D. R. D. Adhikary, and B. K. Pattanayak, "Hybrid Transfer Learning and Self-Attention Framework for Robust MRI-Based Brain Tumor Classification," *Scientific Reports*, vol. 15, Art. no. 21343, 2025, doi: 10.1038/s41598-025-09311-5.

[5] K. M. Adnan, T. M. Ghazal, M. Saleem, M. S. Farooq, C. Y. Yeun, M. Ahmad, *et al.*, "Deep Learning Driven Interpretable and Informed Decision Making Model for Brain Tumour Prediction Using Explainable AI," *Scientific Reports*, vol. 15, Art. no. 19223, 2025, doi: 10.1038/s41598-025-03358-0.

[6] S. Alsubai, S. Ojo, T. I. Nathaniel, M. Ayari, J. Baili, A. Almadhor, and A. Al Hejaili, "Transfer Deep Learning and Explainable AI Framework for Brain Tumor and Alzheimer's Detection Across Multiple Datasets," *Frontiers in Medicine*, vol. 12, Art. no. 1618550, 2025, doi: 10.3389/fmed.2025.1618550.

[7] D. Vamsidhar, P. Desai, S. Joshi, S. Kolhar, N. Deshpande, S. Gite, *et al.*, "Hybrid Model Integration with Explainable AI for Brain Tumor Diagnosis: A Unified Approach to MRI Analysis and Prediction," *Scientific Reports*, vol. 15, Art. no. 20542, 2025, doi: 10.1038/s41598-025-06455-2.

[8] J. Wen, X. Li, X. Ye, X. Li, H. Mao, *et al.*, "A Highly Generalized Federated Learning Algorithm for Brain Tumor Segmentation," *Scientific Reports*, vol. 15, Art. no. 21053, 2025, doi: 10.1038/s41598-025-05297-2.

[9] S. Alphonse, F. Mathew, K. Dhanush, *et al.*, "Federated Learning with Integrated Attention Multiscale Model for Brain Tumor Segmentation," *Scientific Reports*, vol. 15, Art. no. 11889, 2025, doi: 10.1038/s41598-025-96416-6.

[10] K. Narmadha and P. Varalakshmi, "Enhanced Brain Tumour Segmentation Using a Hybrid Dual Encoder–Decoder Model in Federated Learning," *Scientific Reports*, vol. 15, Art. no. 34416, 2025, doi: 10.1038/s41598-025-17432-0.

[11] L. K. Almajmaie, S. Albawi, and M. A. A. Khodher, "Brain Neoplasm Image Recognition Using Deep Learning Techniques," *Iraqi Journal of Science*, pp. 2948–2962, Jul. 2025, doi: 10.24996/ijs.2025.66.7.24.

[12] J. Y. R. Al-Awadi, H. K. Aljobouri, and A. M. Hasan, "MRI Brain Scans Classification Using Extreme Learning Machine on LBP and GLCM," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 02, pp. 134–149, Feb. 2023, doi: 10.3991/ijoe.v19i02.33987.

[13] A. Agrawal and J. Chaki, "CerebralNet meets Explainable AI: Brain tumor detection and classification with probabilistic augmentation and a deep learning approach," *Biomed. Signal Process. Control*, vol. 110, p. 108210, Dec. 2025, doi: 10.1016/j.bspc.2025.108210.

[14] S. Maqsood, R. Damaševičius, and R. Maskeliūnas, "Multi-Modal Brain Tumor Detection Using Deep Neural Network and Multiclass SVM," *Medicina (B Aires).*, vol. 58, no. 8, p. 1090, Aug. 2022, doi: 10.3390/medicina58081090.

[15] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, "Multi-grade brain tumor classification using deep CNN with extensive

data augmentation," *J. Comput. Sci.*, vol. 30, pp. 174–182, Jan. 2019, doi: 10.1016/j.jocs.2018.12.003.
[16] Z. N. K. Swati *et al.*, "Brain tumor classification for MR images using transfer learning and fine-tuning," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 34–46, Jul. 2019, doi: 10.1016/j.compmedimag.2019.05.001.
[17] P. Narayankar and V. P. Baligar, "Pixel-wise Relevance Propagation for Detailed Insights in Brain Tumour Classification," *Procedia Comput. Sci.*, vol. 258, pp. 2958–2967, 2025, doi: 10.1016/j.procs.2025.04.555.
[18] R. L. Kumar, J. Kakarla, B. V. Isunuri, and M. Singh, "Multi-class brain tumor classification using residual network and global average pooling," *Multimed. Tools Appl.*, vol. 80, no. 9, pp. 13429–13438, Apr. 2021, doi: 10.1007/s11042-020-10335-4.
[19] M. Toğaçar, B. Ergen, and Z. Cömert, "BrainMRNet: Brain tumor detection using magnetic resonance images with a novel convolutional neural network model," *Med. Hypotheses*, vol. 134, p. 109531, Jan. 2020, doi: 10.1016/j.mehy.2019.109531.
[20] R. Haque *et al.*, "Explainable deep stacking ensemble model for accurate and transparent brain tumor diagnosis," *Comput. Biol. Med.*, vol. 191, p. 110166, Jun. 2025, doi: 10.1016/j.compbiomed.2025.110166.
[21] H. Kibriya, R. Amin, A. H. Alshehri, M. Masood, S. S. Alshamrani, and A. Alshehri, "A Novel and Effective Brain Tumor Classification Model Using Deep Feature Fusion and Famous Machine Learning Classifiers," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–15, Mar. 2022, doi: 10.1155/2022/7897669.
[22] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," Jan. 2018, [Online]. Available: http://arxiv.org/abs/1608.06993
[23] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, "Rethinking the Inception Architecture for Computer Vision."
[24] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017, [Online]. Available: http://arxiv.org/abs/1704.04861
[25] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Mar. 2019, [Online]. Available: http://arxiv.org/abs/1801.04381
[26] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Apr. 2015, [Online]. Available: http://arxiv.org/abs/1409.1556
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.03385
[28] Y. Yang, C. Miller, P. Jiang, and A. Moghtaderi, "A Case Study of Multi-class Classification with Diversified Precision Recall Requirements for Query Disambiguation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA: ACM, Jul. 2020, pp. 1633–1636. doi: 10.1145/3397271.3401315.
[29] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, "LNAI 8725 - Optimal Thresholding of Classifiers to Maximize F1 Measure."
[30] S. Boughorbel, F. Jarray, and M. El-Anbari, "Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric," *PLoS One*, vol. 12, no. 6, p. e0177678, Jun. 2017, doi: 10.1371/journal.pone.0177678.
[31] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960, doi: 10.1177/001316446002000104.
[32] Vasumathy M, "An integrated framework for knowledge based obstacle information system with image processing techniques", Journal of Physics: Conference Series, 1850,1, 012112, IOP Publishing.
[33] J Kavitha, R Roselinkiruba, M Vasumathy, "A dual-optimization strategy: de-noising and MCDM weighting for key frame extraction in deep learning", Expert Systems with Applications, 128351, https://doi.org/10.1016/j.eswa.2025.128351.
[34] M Vasumathy, Mythili Thirugnanam, S Sumathy, A Framework for automatic intrude object identification in paediatric foreign body aspired radiography images, International Journal of Image Mining, 57-67, Inderscience Publishers (IEL).
[35] M. Vasumathy, J. J. M. A. Devakanth, M. Thirugnanam, R. RoselinKiruba, R. Hamsaveni and R. Nandhini, "Automated semantic segmentation and Diagnosis of Coconut Tree Infection Using Drone Imagery and Machine Learning approach for smart agriculture," 2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA), Pune, India, 2024, pp. 1-6, doi: 10.1109/ICISAA62385.2024.10829008
[36] K. Narayanan, G. M. Lakshmi and M. Vasumathy, "Insights Into Object Recognition: a Comparative Study of Feature Detection and Matching Algorithms Using Kinect Data," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-7, doi: 10.1109/ICCCNT61001.2024.10724025.
[37] Roselinkiruba, R., Vasumathy, M., Devakanth, J.J.M.A., Jothi, C.S., Kavitha, J. and Sharmila, L. (2026). Machine Learning and Automation in Platform Engineering: Transforming Monitoring, Scaling, and Self-Healing, Machine Learning and Automation in Platform Engineering. In Platform Engineering (eds K.S. Kumar, S. Sundaresan, R. Prithiviraj, T.A. Kumar and S. Balamurugan). https://doi.org/10.1002/9781394395910.ch8
[38] M Vasumathy, R Roselin Kiruba, Mythili Thirugnanam, Krishnammal Narayanan, Medicine-Infused Smart Bandages with Integrated Sensors and Image-based Machine Learning Model for Diabetic Foot Ulcer Detection. Advances in Healthcare using Machine Learning, 160-184, CRC, Press, https://www.taylorfrancis.com/chapters/edit/10.1201/9781003531050-7/medicine-infused-smart-bandages-integrated-sensors-image-based-machine-learning-model-diabetic-foot-ulcer-detection-vasumathy-roselin-kiruba-mythili-thirugnanam-krishnammal-narayanan
[39] M Vasumathy, R RoselinKiruba, J Kavitha , Optimizing Greenhouse Conditions With Wireless Sensors and Machine Learning, Book : Optimizing Automation in Engineering With Energy Systems and Communication Networks, 325-358, IGI Global Scientific Publishing. DOI: 10.4018/979-8-3373-2737-2.ch012.
[40] M Vasumathy, R RoselinKiruba, Enhancing Healthcare Management With Emotional Intelligence and Collaborative Leadership Through Chatbot Integration , Chatbots in Educational Leadership and Management, 205-234, IGI Global Scientific Publishing, DOI: 10.4018/979-8-3693-8734-4.ch009 .