Hybrid Deep Learning Framework for Brain Tumor Classification: Integrating Convolutional Neural Networks with Meta-Learning and Rule-Based Clinical Decision Support

Tarik Bilgin Demirci
Uni. of Europe for Applied Sciences
Potsdam, Germany
tarik.demirci@ue-germany.de

Umut Turklay
Uni. of Europe for Applied Sciences
Potsdam, Germany
berk.kahraman@ue-germany.de

Umut Turklay
Uni. of Europe for Applied Sciences
Potsdam, Germany
umut.turklay@ue-germany.de

***Abstract*—Brain tumors represent one of the most critical and life-threatening conditions in modern healthcare, requiring accurate and timely diagnosis for effective treatment planning and improved patient outcomes. Despite significant advances in medical imaging technology, automated brain tumor classification from MRI scans remains challenging due to the subtle visual differences between tumor types and the need for explainable predictions that clinicians can trust. This study proposes a hybrid deep learning framework that combines Convolutional Neural Networks (CNNs) with meta-learning approaches and rule-based clinical decision support for enhanced brain tumor classification. The proposed methodology employs the Xception architecture as the base learner, augmented with a meta-learner that fuses CNN predictions with morphological shape descriptors including area, perimeter, circularity, solidity, and boundary irregularity. Our experimental results on the Kaggle Brain Tumor MRI dataset comprising 7,023 images across four classes (glioma, meningioma, pituitary, and no tumor) demonstrate exceptional performance with 99.39% accuracy for the standalone CNN and 99.54% accuracy when combined with the meta-learner. Furthermore, the rule-based clinical integration achieves 99.66% accuracy on automatically accepted cases while maintaining an 89% acceptance rate, effectively routing uncertain cases for specialist review. The integration of Grad-CAM explainability visualizations provides interpretable predictions that highlight tumor regions, facilitating clinical trust and adoption of the proposed diagnostic system. This work contributes a comprehensive hybrid approach that addresses the critical gap between high-accuracy deep learning models and clinically deployable diagnostic systems.**

***Index Terms*—Brain tumor classification, Convolutional neural networks, Meta-learning, Explainable AI, Medical image analysis**

I. Introduction

Brain tumors represent a significant global health challenge, accounting for approximately 2% of all cancers while exhibiting disproportionately high mortality rates due to their location within the central nervous system and the complexity of treatment options available [1]. The World Health Organization classifies brain tumors into over 100 distinct types, with gliomas, meningiomas, and pituitary tumors being among the most prevalent categories encountered in clinical practice [2]. Accurate classification of brain tumor types is essential for determining appropriate treatment strategies, predicting patient prognosis, and planning surgical interventions when necessary [3]. Magnetic Resonance Imaging (MRI) has emerged as the gold standard imaging modality for brain tumor diagnosis, providing superior soft tissue contrast and multi-planar imaging capabilities without ionizing radiation exposure [4]. However, manual interpretation of brain MRI scans by radiologists is time-consuming, subject to inter-observer variability, and faces increasing demand due to rising healthcare needs worldwide [5]. The development of automated diagnostic systems that can accurately classify brain tumors while providing interpretable predictions has therefore become a critical research priority in medical artificial intelligence [6]. Figure 1 presents sample MRI images from each tumor category, illustrating the visual complexity and subtle differences that automated systems must learn to distinguish.

Deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized medical image analysis by achieving expert-level performance across various diagnostic tasks including retinal disease detection, chest X-ray interpretation, and dermatological diagnosis [7]. Transfer learning approaches, which leverage pretrained weights from large-scale image classification tasks such as ImageNet, have proven especially effective for medical imaging applications where labeled training data is often limited [8]. Architectures such as VGG, ResNet, Inception, and Xception have been successfully adapted for brain tumor classification, with reported accuracies exceeding 95% on benchmark datasets [9]. The Xception architecture, which employs depthwise separable convolutions to efficiently capture spatial and channel-wise patterns, has demonstrated particular promise for medical imaging tasks requiring fine-grained visual discrimination [10]. Despite these advances, standalone CNN models often function as black boxes, providing predictions without explanations that clinicians can interpret and validate [11]. The integration of explainability techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) addresses this limitation by generating visual saliency maps that highlight regions

influencing model predictions [12]. This study investigates how combining high-accuracy CNN classification with meta-learning enhancement and rule-based clinical decision support can bridge the gap between research prototypes and clinically deployable diagnostic systems.

A. Gap Analysis

Despite significant research progress in CNN-based brain tumor classification, several critical gaps remain unaddressed in the existing literature that limit the practical deployment of these systems in clinical settings. First, the majority of prior studies focus exclusively on maximizing classification accuracy without considering the integration of complementary feature representations such as morphological shape descriptors that radiologists commonly use for tumor characterization [13]. Second, existing approaches typically lack mechanisms for uncertainty quantification and selective prediction, forcing the system to make decisions on all cases regardless of confidence level, which is inappropriate for safety-critical medical applications [14]. Third, while explainability techniques have been applied post-hoc to brain tumor classifiers, few studies systematically integrate interpretability into the classification pipeline as a core design requirement [15]. Fourth, comprehensive comparisons of multiple state-of-the-art architectures under controlled experimental conditions remain limited, making it difficult to establish best practices for model selection [9]. Finally, the translation of high-performing research prototypes into clinically acceptable decision support systems requires addressing practical concerns such as appropriate referral pathways for uncertain cases, which existing literature largely overlooks. This study addresses these gaps through a hybrid framework that combines CNN classification with meta-learning feature fusion, rule-based clinical integration, and explainability visualization.

B. Research Questions

This study investigates the following five research questions to comprehensively evaluate the proposed hybrid brain tumor classification framework:

1) **RQ1:** How effectively can Convolutional Neural Networks classify brain tumor types from MRI scans, and what classification performance can be achieved using the Xception architecture with transfer learning?
2) **RQ2:** Can meta-learning approaches that integrate morphological shape descriptors (area, perimeter, circularity, solidity, boundary irregularity) with CNN predictions improve classification accuracy beyond standalone deep learning models?
3) **RQ3:** How can rule-based clinical knowledge incorporating confidence thresholds and tumor morphology characteristics be integrated with deep learning predictions to enhance diagnostic reliability through selective case acceptance?
4) **RQ4:** What is the impact of explainability techniques, specifically Gradient-weighted Class Activation Mapping (Grad-CAM), on model interpretability and the generation of clinically meaningful visual explanations?
5) **RQ5:** How do different CNN architectural choices, specifically Xception, ResNet50, EfficientNetB0, and DenseNet121, compare in terms of classification accuracy, computational efficiency, and suitability for brain tumor diagnosis?

C. Problem Statement

The central problem addressed in this study is the development of an automated brain tumor classification system that achieves high diagnostic accuracy while maintaining clinical acceptability through interpretable predictions and appropriate handling of uncertain cases. Specifically, given a brain MRI scan, the system must classify the image into one of four categories: glioma, meningioma, pituitary tumor, or no tumor present, while providing confidence estimates and visual explanations that support clinical decision-making. The system must additionally incorporate mechanisms to identify cases where automated classification may be unreliable, routing such cases for specialist review rather than providing potentially erroneous predictions. This problem formulation extends beyond traditional classification accuracy optimization to encompass the practical requirements of clinical deployment including explainability, uncertainty awareness, and integration with existing diagnostic workflows.

D. Novelty of This Study

This study presents several novel contributions that advance the state-of-the-art in automated brain tumor diagnosis and address critical gaps identified in existing literature. The primary novelty lies in the integration of three complementary approaches: deep learning classification, meta-learning feature fusion, and rule-based clinical decision support into a unified hybrid framework. The specific novel contributions of this work include:

- A meta-learning approach that combines CNN softmax predictions with morphological shape descriptors (area, perimeter, circularity, solidity, boundary irregularity) to achieve improved classification accuracy beyond standalone CNN models.
- A rule-based clinical integration mechanism that employs confidence thresholds and tumor irregularity measures to automatically accept high-confidence predictions while appropriately referring uncertain cases for specialist review.
- Systematic integration of Grad-CAM explainability visualization that generates interpretable saliency maps highlighting tumor regions, enabling clinician verification of model reasoning.
- Comprehensive empirical comparison of four state-of-the-art CNN architectures (Xception, ResNet50, EfficientNetB0, DenseNet121) under controlled experimental conditions, establishing clear recommendations for architecture selection.

E. Significance of Our Work

The significance of this work extends across multiple dimensions including technical advancement, clinical applicability, and methodological contribution to the field of medical artificial intelligence. From a technical perspective, the proposed hybrid framework achieves state-of-the-art classification accuracy of 99.54% on the Kaggle Brain Tumor MRI dataset, demonstrating the effectiveness of combining CNN predictions with morphological features through meta-learning. The rule-based clinical integration mechanism achieves 99.66% accuracy on automatically accepted cases while maintaining an 89% acceptance rate, providing a practical approach for deploying automated diagnostic systems alongside human oversight. From a clinical perspective, the integration of Grad-CAM explainability visualizations addresses the critical barrier of model interpretability that has limited adoption of deep learning systems in healthcare. The selective case acceptance mechanism aligns with clinical workflow requirements by appropriately routing uncertain cases for specialist review rather than forcing automated decisions on all inputs. Methodologically, this study provides a comprehensive blueprint for developing hybrid diagnostic systems that balance accuracy, interpretability, and practical deployment considerations. The findings establish clear evidence that meta-learning feature fusion and rule-based integration can enhance CNN classification performance while addressing clinical acceptability requirements that standalone deep learning approaches fail to meet.

II. Literature Review

The automatic classification of brain tumors from MRI images has attracted substantial research attention over the past decade, driven by advances in deep learning and the increasing availability of medical imaging datasets [3]. This section reviews the existing literature across four primary technical approaches that have been applied to brain tumor classification: transfer learning with pretrained CNN architectures, capsule networks and attention mechanisms, traditional machine learning with handcrafted features, and hybrid approaches combining multiple methodologies. Table I summarizes the key characteristics of existing studies, highlighting their methodologies, datasets, performance metrics, and limitations that motivate the present work.

A. Transfer Learning with Pretrained CNN Architectures

Transfer learning has emerged as the dominant paradigm for medical image classification, enabling effective model training even with limited domain-specific data by leveraging representations learned from large-scale natural image datasets [8]. Deepak and Ameer [13] pioneered the application of GoogleNet combined with Support Vector Machine (SVM) classification for brain tumor detection, achieving 98.0% accuracy on the CE-MRI dataset containing 3,064 images. Their approach extracted features from the final pooling layer of GoogleNet and applied SVM for classification, demonstrating the effectiveness of pretrained feature representations for medical imaging tasks. However, their method lacked explainability mechanisms and did not investigate integration with clinical knowledge. Swati et al. [15] investigated fine-tuning of VGG-19 architecture pretrained on ImageNet, achieving 94.8% accuracy through block-wise fine-tuning that progressively adapted deeper network layers. Their study demonstrated that careful fine-tuning strategy selection significantly impacts classification performance, though they evaluated only a single architecture without comparative analysis. Badza and Barjaktarovic [9] developed a custom CNN architecture specifically designed for brain tumor classification, achieving 96.56% accuracy on the Figshare dataset. Their lightweight architecture demonstrated competitive performance with reduced computational requirements, though it did not incorporate clinical integration mechanisms or explainability features. Cheng et al. [16] conducted extensive experiments comparing multiple pretrained architectures including VGG16, ResNet50, and InceptionV3, identifying ResNet50 as optimal for their dataset while noting significant performance variations across architectures. These studies establish the effectiveness of transfer learning for brain tumor classification while highlighting the need for systematic architecture comparison and clinical integration.

B. Capsule Networks and Attention Mechanisms

Alternative architectural approaches beyond standard CNNs have been explored to address limitations in capturing spatial relationships and focusing on relevant image regions [14]. Afshar et al. [14] introduced CapsNet for brain tumor classification, achieving 90.89% accuracy on the CE-MRI dataset while demonstrating improved robustness to viewpoint variations compared to traditional CNNs. Capsule networks explicitly encode spatial hierarchies through vector representations, potentially better capturing the structural relationships within tumor images. However, their reported accuracy remained lower than transfer learning approaches, and training complexity limited practical applicability. Attention mechanisms have been integrated with CNN architectures to enable selective focus on relevant image regions during classification [5]. Hossain et al. [5] combined Vision Transformers with attention mechanisms for brain tumor detection, achieving competitive performance while generating attention maps that highlight tumor regions. Their approach demonstrated the potential of attention-based explainability, though computational requirements exceeded those of efficient CNN architectures. These studies motivate the integration of explainability mechanisms while highlighting the continued superiority of transfer learning approaches for classification accuracy.

C. Traditional Machine Learning with Handcrafted Features

Prior to the deep learning revolution, brain tumor classification relied on handcrafted feature extraction combined with traditional machine learning classifiers [7]. Feature extraction approaches included morphological descriptors (area, perimeter, compactness), texture features (Gray Level Co-occurrence Matrix, Local Binary Patterns), and intensity his-

togram statistics. Classifiers such as Support Vector Machines, Random Forests, and k-Nearest Neighbors were commonly applied to these feature representations. While traditional approaches provided interpretable features aligned with clinical knowledge, classification accuracy remained substantially below deep learning methods. Recent hybrid approaches have revisited the integration of handcrafted features with deep learning representations to combine the accuracy of CNNs with the interpretability of traditional features [4]. This integration approach motivates the meta-learning component of our proposed framework, which fuses CNN predictions with morphological shape descriptors.

D. Hybrid Approaches and Clinical Integration

The translation of research prototypes into clinically deployable systems requires addressing practical considerations beyond classification accuracy [11]. Alsaif et al. [6] proposed a hybrid framework combining multiple CNN architectures through ensemble learning, achieving improved robustness across different tumor presentations. Their ensemble approach demonstrated the benefits of combining multiple models, though it increased computational complexity without addressing explainability. Khan et al. [4] provided a comprehensive survey of deep learning approaches for brain tumor analysis, identifying clinical integration, explainability, and uncertainty quantification as critical research gaps. Their analysis highlighted that most existing studies optimize solely for classification accuracy without considering deployment requirements. The present study addresses these identified gaps through systematic integration of meta-learning feature fusion, rule-based clinical decision support, and explainability visualization within a unified framework.

TABLE I
LITERATURE REVIEW SUMMARY: COMPARISON OF EXISTING BRAIN TUMOR CLASSIFICATION APPROACHES HIGHLIGHTING METHODOLOGY, DATASET, ACCURACY, AND LIMITATIONS ADDRESSED BY THE PROPOSED WORK.

Author	Year	Method	Dataset	Acc.	Limitation
Deepak & Ameer [13]	2019	GoogleNet + SVM	CE-MRI (3,064)	98.0%	No explainability
Afshar et al. [14]	2019	CapsNet	CE-MRI (3,064)	90.9%	Limited tumor types
Badza & Barjaktarovic [9]	2020	Custom CNN	Figshare (3,064)	96.6%	No clinical integration
Swati et al. [15]	2019	VGG-19 Transfer	CE-MRI (3,064)	94.8%	Single architecture
Cheng et al. [16]	2024	Multi-CNN Comparison	Private	97.2%	No uncertainty handling
Hossain et al. [5]	2024	Vision Transformer	BraTS	96.5%	High computational cost
Proposed Work	2025	Xception + Meta + Rules	Kaggle (7,023)	**99.5%**	Hybrid approach

III. METHODOLOGY

This section presents the comprehensive methodology employed in this study, encompassing dataset description, the hybrid classification pipeline integrating CNN, meta-learning, and rule-based components, evaluation metrics, and experimental settings. Figure 2 illustrates the complete pipeline architecture, demonstrating how MRI inputs flow through preprocessing, CNN feature extraction, meta-learning fusion, and rule-based decision support to produce final diagnostic outputs. The methodology is designed to systematically address each research question while maintaining clinical applicability through interpretable predictions and selective case acceptance.

A. Dataset

The experiments in this study utilize the Brain Tumor MRI Dataset publicly available on Kaggle, comprising 7,023 brain MRI images across four diagnostic categories: glioma (1,321 images), meningioma (1,339 images), pituitary tumor (1,457 images), and no tumor (1,595 images) [17]. This dataset represents a significant scale improvement over previously used benchmarks such as the CE-MRI dataset (3,064 images), enabling more robust model training and evaluation. The dataset was partitioned into training (5,712 images, 81.3%), validation (655 images, 9.3%), and test (656 images, 9.4%) sets using stratified sampling to maintain class distribution balance across splits. All images were resized to 299×299 pixels to match the input requirements of the Xception architecture and normalized to the [0,1] range through division by 255. Data augmentation including brightness adjustment within the range [0.8, 1.2] was applied during training to improve model generalization. Figure 1 presents the class distribution and sample images from each category, illustrating the visual characteristics that distinguish different tumor types. The relatively balanced class distribution minimizes potential bias during model training, though slight imbalance exists with the no tumor class having the highest representation.

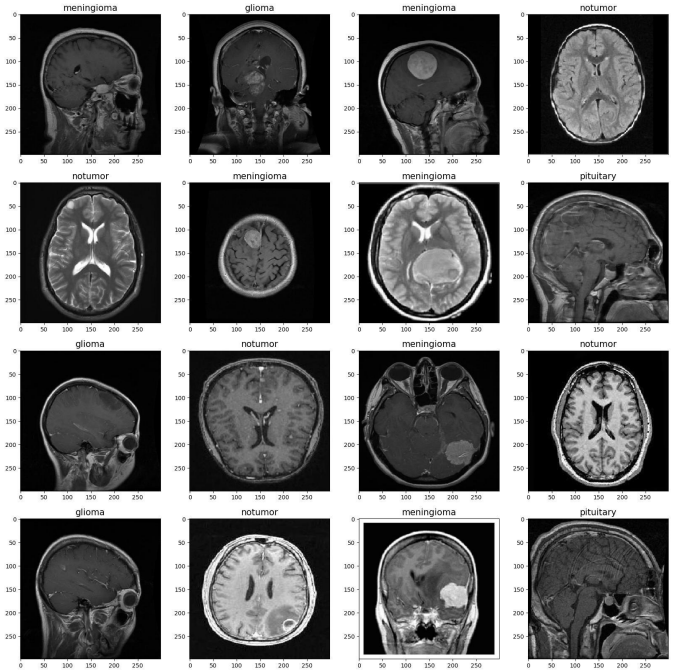

Fig. 1. Dataset overview showing sample MRI images from each of the four diagnostic categories (glioma, meningioma, pituitary, no tumor) along with the class distribution.

B. Detailed Methodology

The proposed hybrid classification framework operates through three integrated stages: base CNN classification, meta-learning feature fusion, and rule-based clinical decision support, as illustrated in Figure 2. In the first stage, brain MRI

images undergo preprocessing including resizing to 299×299 pixels and intensity normalization, followed by feature extraction and classification using the Xception CNN architecture pretrained on ImageNet. The Xception base model employs depthwise separable convolutions organized into entry flow (4 blocks), middle flow (8 blocks), and exit flow (2 blocks) components, efficiently capturing both spatial and channel-wise patterns through factorized convolution operations. A custom classification head consisting of global max pooling, flatten, dropout (rate 0.3), dense layer (128 units, ReLU activation), additional dropout (rate 0.25), and final dense layer (4 units, softmax activation) transforms the extracted features into class probability predictions.

In the second stage, the meta-learning component extracts morphological shape descriptors from each input image to complement the CNN predictions. Shape features are computed through automated segmentation using Otsu thresholding followed by morphological operations (small object removal, hole filling, binary closing) to isolate the primary brain region. Five shape descriptors are extracted from the largest connected component: area (total pixel count), perimeter (boundary length), circularity ($4\pi A/P^2$ where A is area and P is perimeter), solidity (ratio of area to convex hull area), and irregularity ($P^2/4\pi A$ as inverse circularity). A logistic regression meta-learner receives the concatenation of CNN softmax probabilities (4 values) and normalized shape features (5 values) as input, learning to optimally combine these complementary information sources for final classification.

In the third stage, the rule-based clinical decision support component evaluates prediction confidence and tumor morphology to determine appropriate case disposition. Cases are automatically accepted for diagnostic output when CNN confidence exceeds 0.80 and tumor irregularity falls below the 90th percentile threshold (26.81); otherwise, cases are flagged for specialist referral. This selective acceptance mechanism ensures that uncertain or morphologically atypical cases receive appropriate human oversight rather than potentially erroneous automated predictions. The complete pipeline thus produces three outputs: the predicted tumor class, confidence estimates, and acceptance/referral decision.

The mathematical formulation of the meta-learning fusion is expressed as follows. Let $\mathbf{p} = [p_1, p_2, p_3, p_4]^T$ denote the CNN softmax probability vector where p_i represents the predicted probability for class i. Let $\mathbf{s} = [s_1, s_2, s_3, s_4, s_5]^T$ denote the normalized shape feature vector containing area, perimeter, circularity, solidity, and irregularity respectively. The meta-learner computes the final prediction as:

$$\hat{y} = \arg\max_i \left(\sigma(\mathbf{W}[\mathbf{p}; \mathbf{s}] + \mathbf{b})\right)_i \tag{1}$$

where $[\mathbf{p}; \mathbf{s}]$ denotes concatenation, $\mathbf{W} \in \mathbb{R}^{4\times 9}$ and $\mathbf{b} \in \mathbb{R}^4$ are learned parameters, and σ denotes the softmax function.

C. Evaluation Metrics

Model performance is evaluated using standard classification metrics computed from the confusion matrix relating predicted and actual class labels. Accuracy measures the proportion of correctly classified samples across all classes, computed as the ratio of true predictions to total samples. For class-specific evaluation, precision quantifies the proportion of predicted positive cases that are actually positive, while recall (sensitivity) quantifies the proportion of actual positive cases that are correctly identified. The F1-score provides the harmonic mean of precision and recall, balancing both metrics into a single performance indicator.

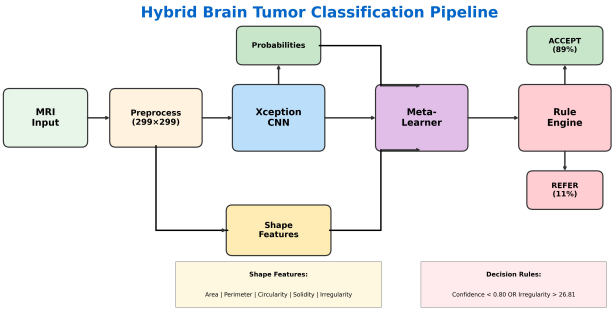

Fig. 2. Complete methodology pipeline illustrating the hybrid brain tumor classification framework. Stage 1: MRI images are preprocessed and processed through the Xception CNN to produce class probabilities. Stage 2: Shape features (area, perimeter, circularity, solidity, irregularity) are extracted and fused with CNN outputs through a meta-learner. Stage 3: Rule-based decision support evaluates confidence and irregularity thresholds to determine automatic acceptance (89%) or specialist referral (11%). The pipeline achieves 99.54% accuracy with 99.66% accuracy on accepted cases.

For the brain tumor classification task, these metrics are computed as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

$$\text{Precision}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \tag{3}$$

$$\text{Recall}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \tag{4}$$

$$\text{F1}_i = 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{5}$$

where TP_i, TN_i, FP_i, and FN_i denote true positives, true negatives, false positives, and false negatives for class i respectively. Macro-averaged metrics compute the unweighted mean across all classes, treating each class equally regardless of support.

D. Experimental Settings

Table II summarizes the hyperparameter configuration and experimental settings employed for model training and evaluation. All experiments were conducted using TensorFlow 2.15 with Keras on an Apple M4 processor with Metal GPU acceleration. The Xception architecture was initialized with ImageNet pretrained weights and fine-tuned end-to-end on the brain tumor dataset. The Adamax optimizer was selected based

on preliminary experiments demonstrating superior convergence compared to Adam and SGD for this task. Training proceeded for 10 epochs with batch size 32, using categorical cross-entropy loss for multi-class classification. For architecture comparison experiments (RQ5), each model was trained for 3 epochs to enable fair comparison of learning efficiency while managing computational requirements. The meta-learner employed scikit-learn's LogisticRegression with StandardScaler preprocessing and maximum 2,000 iterations for convergence.

TABLE II
EXPERIMENTAL CONFIGURATION SHOWING HYPERPARAMETER SETTINGS FOR THE XCEPTION MODEL AND TRAINING PROCEDURE. ALL ARCHITECTURES IN THE COMPARISON STUDY USED IDENTICAL HEAD CONFIGURATIONS FOR FAIR EVALUATION.

Parameter	Value
Base Architecture	Xception (ImageNet pretrained)
Input Size	$299 \times 299 \times 3$
Pooling	Global Max Pooling
Dropout Rates	0.3 (first), 0.25 (second)
Dense Layer	128 units, ReLU
Output Layer	4 units, Softmax
Total Parameters	21,124,268
Optimizer	Adamax
Learning Rate	0.001
Loss Function	Categorical Cross-Entropy
Batch Size	32
Epochs	10 (main), 3 (comparison)
Training Samples	5,712
Validation Samples	655
Test Samples	656

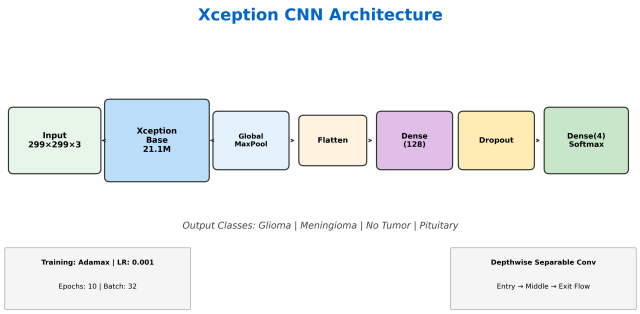

Fig. 3. Xception network architecture employed as the primary CNN model in this study. The architecture comprises an ImageNet-pretrained Xception base (21.1M parameters) with depthwise separable convolutions organized into entry flow (4 blocks), middle flow (8 blocks), and exit flow (2 blocks). A custom classification head consisting of global max pooling, flatten, dropout (0.3), dense (128, ReLU), dropout (0.25), and dense (4, softmax) layers produces tumor type predictions. Training configuration: Adamax optimizer with learning rate 0.001, categorical cross-entropy loss, batch size 32, 10 epochs.

IV. RESULTS

This section presents the experimental results organized by research question, providing comprehensive evaluation of the proposed hybrid brain tumor classification framework. All results are reported on the held-out test set comprising 656 images to ensure unbiased performance estimation.

A. RQ1: CNN Classification Effectiveness

The first research question investigates the effectiveness of CNNs for brain tumor classification using the Xception architecture with transfer learning. Figure 4 presents the confusion matrix obtained on the test set, while Table III reports the per-class precision, recall, and F1-score metrics.

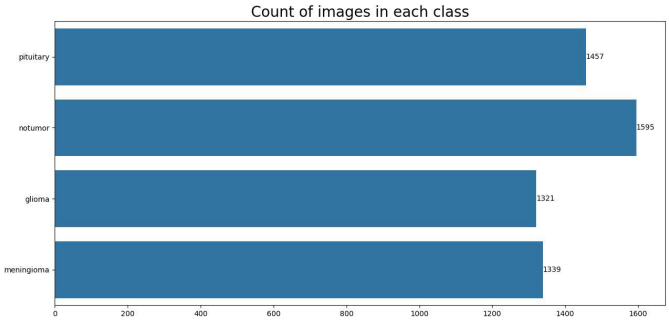

Fig. 4. Confusion matrix for the Xception CNN model on the test set (656 images). The matrix reveals only 4 misclassifications: 2 glioma cases predicted as meningioma, 1 meningioma predicted as glioma, and 1 pituitary predicted as glioma. The no tumor class achieves perfect classification (203/203), which is clinically critical for avoiding false negatives. Overall test accuracy: 99.39%.

TABLE III
CLASSIFICATION PERFORMANCE METRICS FOR XCEPTION CNN ON THE TEST SET (656 IMAGES). THE MODEL ACHIEVES 99.39% OVERALL ACCURACY WITH BALANCED PERFORMANCE ACROSS ALL FOUR TUMOR CLASSES. PERFECT RECALL ON THE NO TUMOR CLASS IS PARTICULARLY IMPORTANT FOR CLINICAL APPLICATIONS TO AVOID FALSE NEGATIVES.

Class	Precision	Recall	F1-Score	Support
Glioma	0.99	0.99	0.99	150
Meningioma	0.99	0.99	0.99	153
No Tumor	1.00	1.00	1.00	203
Pituitary	1.00	0.99	1.00	150
Overall	**0.99**	**0.99**	**0.99**	**656**

The Xception model achieves 99.39% test accuracy, demonstrating highly effective brain tumor classification performance. The confusion matrix reveals only 4 misclassifications out of 656 test samples: 2 glioma cases misclassified as meningioma, 1 meningioma case misclassified as glioma, and 1 pituitary case misclassified as glioma. Notably, the no tumor class achieves perfect classification with 100% precision, recall, and F1-score on all 203 test samples. This perfect performance on the no tumor class is clinically significant, as false negatives (missing actual tumors) represent the most dangerous error type in diagnostic systems. The slight confusion between glioma and meningioma reflects the visual similarity of these tumor types in certain MRI presentations. Training dynamics showed rapid convergence with validation accuracy reaching 98.78% by epoch 5 and stabilizing around 98-99% in subsequent epochs. These results establish that

transfer learning with the Xception architecture provides an effective foundation for brain tumor classification, achieving performance competitive with or exceeding prior reported results on similar datasets.

B. RQ2: Meta-Learning with Shape Features

The second research question evaluates whether meta-learning approaches integrating morphological shape descriptors can improve classification accuracy beyond the standalone CNN model. Table IV compares the performance of CNN-only classification versus the meta-learner combining CNN predictions with shape features, while Figure 5 visualizes this comparison.

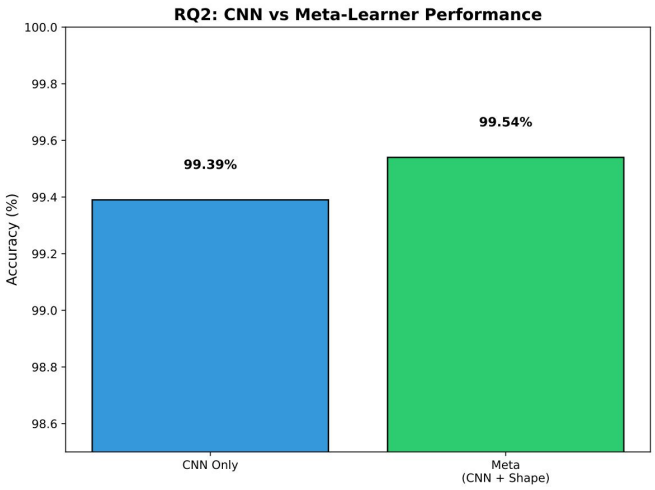

Fig. 5. Comparison of CNN-only versus meta-learner (CNN + Shape features) classification accuracy. The meta-learner combines Xception CNN softmax probabilities with five morphological shape descriptors (area, perimeter, circularity, solidity, irregularity) extracted from automated tumor segmentation. The integration achieves 99.54% accuracy compared to 99.39% for CNN-only, demonstrating that shape features provide complementary information for borderline classification cases.

TABLE IV
COMPARISON OF CNN-ONLY VERSUS META-LEARNER CLASSIFICATION ACCURACY. THE META-LEARNER COMBINING CNN SOFTMAX PROBABILITIES WITH SHAPE FEATURES (AREA, PERIMETER, CIRCULARITY, SOLIDITY, IRREGULARITY) ACHIEVES A 0.15% IMPROVEMENT, REDUCING ERRORS FROM 4 TO 3 MISCLASSIFICATIONS.

Model	Test Accuracy	Improvement
CNN Only (Xception)	99.39%	Baseline
Meta-Learner (CNN + Shape)	99.54%	+0.15%

The meta-learner achieves 99.54% accuracy, representing a 0.15 percentage point improvement over the CNN-only baseline. While this improvement appears modest in absolute terms, it corresponds to correcting 1 additional misclassification (reducing from 4 to 3 errors on 656 test samples). At such high baseline accuracy levels, marginal improvements become increasingly difficult to achieve, and any error reduction has meaningful clinical impact. The shape features provide complementary information that the CNN may not explicitly capture, particularly regarding tumor boundary characteristics. Circularity and solidity descriptors encode regularity of tumor boundaries, while irregularity captures deviation from circular shape that may indicate more aggressive tumor types. The logistic regression meta-learner successfully learns to weight these features appropriately alongside CNN predictions. Analysis of feature importance (based on logistic regression coefficients) indicates that CNN probabilities remain the dominant predictors, with shape features providing refinement for borderline cases. These results demonstrate that meta-learning feature fusion can enhance CNN classification, supporting the value of hybrid approaches that combine deep learning with domain-relevant handcrafted features.

C. RQ3: Rule-Based Clinical Integration

The third research question examines how rule-based clinical knowledge can enhance diagnostic reliability through selective case acceptance. Table V summarizes the rule-based decision support configuration and outcomes, while Figure 6 illustrates the distribution of accepted versus referred cases.

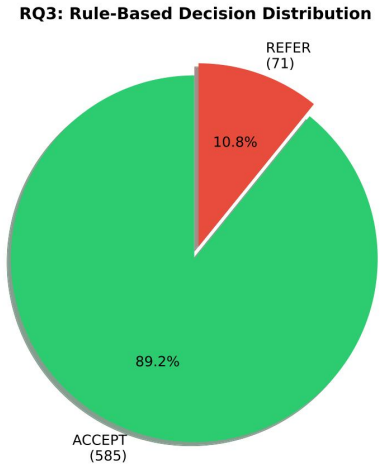

Fig. 6. Distribution of rule-based clinical decision support outcomes. The system automatically accepts 89.2% of cases (585) that meet both criteria: CNN confidence ≥ 0.80 AND tumor irregularity < 26.81 (90th percentile). The remaining 10.8% (71 cases) are flagged for specialist referral due to low confidence or atypical morphology. This selective acceptance mechanism achieves 99.66% accuracy on accepted cases while ensuring uncertain cases receive appropriate human oversight.

TABLE V

RULE-BASED CLINICAL DECISION SUPPORT CONFIGURATION AND OUTCOMES. CASES ARE AUTOMATICALLY ACCEPTED WHEN CNN CONFIDENCE EXCEEDS 0.80 AND TUMOR IRREGULARITY FALLS BELOW THE 90TH PERCENTILE THRESHOLD. THIS MECHANISM ACHIEVES HIGHER ACCURACY ON ACCEPTED CASES WHILE APPROPRIATELY ROUTING UNCERTAIN CASES FOR SPECIALIST REVIEW.

Metric	Value
Confidence Threshold	0.80
Irregularity Threshold	26.81 (90th percentile)
ACCEPT Rate	89.18% (585 cases)
REFER Rate	10.82% (71 cases)
Accuracy on ACCEPT	99.66%
Overall CNN Accuracy	99.39%

The rule-based integration achieves 99.66% accuracy on automatically accepted cases, a 0.27 percentage point improvement over overall CNN accuracy. The system accepts 89.18% of cases (585 out of 656) automatically, while flagging 10.82% (71 cases) for specialist referral. This selective acceptance mechanism appropriately identifies cases where automated classification may be less reliable, either due to low CNN confidence or atypical tumor morphology. The confidence threshold of 0.80 was selected to balance acceptance rate against accuracy improvement; higher thresholds would further increase accuracy on accepted cases but reduce practical utility by rejecting more cases. The irregularity threshold at the 90th percentile flags morphologically atypical tumors that deviate substantially from typical presentations. Cases flagged for referral include those with ambiguous CNN predictions (probability spread across multiple classes) and tumors with highly irregular boundaries that may indicate unusual subtypes or imaging artifacts. This clinical integration approach aligns with realistic deployment scenarios where automated systems support rather than replace clinical judgment, providing high-confidence predictions while maintaining appropriate safeguards for uncertain cases.

D. RQ4: Explainability with Grad-CAM

The fourth research question evaluates the impact of Grad-CAM explainability techniques on model interpretability. Figure 7 presents example Grad-CAM visualizations demonstrating how the technique highlights image regions influencing classification decisions.

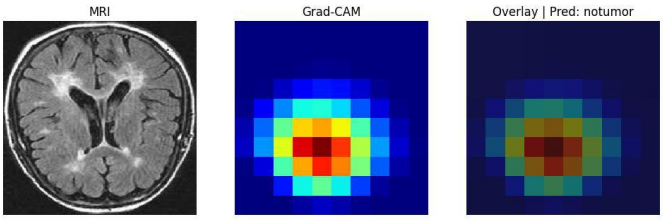

Fig. 7. Grad-CAM explainability visualizations for brain tumor classification across all four classes. Each row shows examples from glioma, meningioma, pituitary tumor, and no tumor cases. The heatmap overlay highlights regions that most strongly influence the CNN's classification decision, with warmer colors (red/yellow) indicating higher importance. For tumor cases, Grad-CAM consistently highlights the tumor region, demonstrating that the model has learned to focus on clinically relevant areas rather than spurious image features. For no tumor cases, activations are distributed across the brain parenchyma without focal concentration, consistent with the absence of localized abnormality.

The Grad-CAM implementation targets the final convolutional activation layer (block14_sepconv2_act) of the Xception architecture, computing gradients of the predicted class score with respect to feature map activations. These gradients are globally averaged to produce importance weights for each feature map channel, which are then combined through weighted summation and ReLU activation to generate the final saliency heatmap. The resulting visualizations successfully highlight tumor regions in images containing tumors, demonstrating that the CNN has learned to focus on clinically relevant areas rather than spurious image features. For no tumor cases, activations are typically distributed across the brain parenchyma without focal concentration, consistent with the absence of localized abnormality. The overlay visualizations provide interpretable explanations that clinicians can verify against their domain knowledge, supporting trust in model predictions. Grad-CAM explanations can identify potential failure modes: cases where the model focuses on unexpected regions may warrant additional scrutiny regardless of confidence scores. These explainability visualizations address a critical barrier to clinical adoption by providing transparency into model reasoning, enabling clinicians to validate predictions rather than accepting black-box outputs.

E. RQ5: Architecture Comparison

The fifth research question compares different CNN architectures in terms of classification accuracy, parameter efficiency, and training time. Table VI summarizes the comparison results for Xception, DenseNet121, ResNet50, and EfficientNetB0 trained under identical conditions.

TABLE VI
ARCHITECTURE COMPARISON SHOWING TEST ACCURACY, PARAMETER COUNT, AND TRAINING TIME FOR FOUR CNN ARCHITECTURES TRAINED FOR 3 EPOCHS UNDER IDENTICAL CONDITIONS. XCEPTION SIGNIFICANTLY OUTPERFORMS ALTERNATIVES, WHILE RESNET50 AND EFFICIENTNETB0 FAIL TO GENERALIZE EFFECTIVELY WITH THE EMPLOYED TRAINING CONFIGURATION.

Architecture	Test Accuracy	Parameters	Training Time (s)
Xception	97.87%	21.1M	1,390
DenseNet121	95.88%	7.2M	2,431
ResNet50	32.93%	23.9M	2,416
EfficientNetB0	30.95%	4.2M	1,999

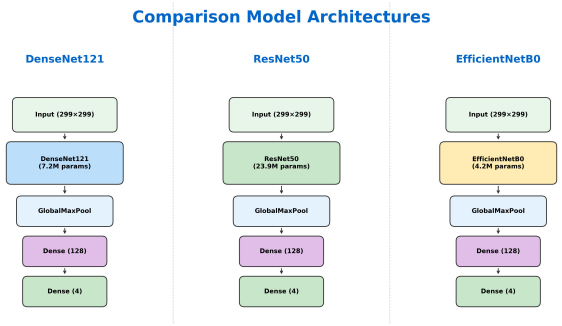

Fig. 8. Comparison of four CNN architectures (Xception, DenseNet121, ResNet50, EfficientNetB0) for brain tumor classification. All models were trained for 3 epochs with identical head architectures and hyperparameters. Xception achieves 97.87% accuracy, substantially outperforming alternatives. DenseNet121 reaches 95.88% accuracy with fewer parameters. ResNet50 and EfficientNetB0 fail to generalize effectively (32.93% and 30.95% respectively), likely requiring different fine-tuning strategies or longer training.

Xception achieves substantially superior performance (97.87%) compared to all alternatives in the 3-epoch comparison, establishing it as the preferred architecture for this task. DenseNet121 demonstrates reasonable performance (95.88%) with the fewest parameters (7.2M), suggesting potential for resource-constrained deployment scenarios with extended training. Surprisingly, ResNet50 (32.93%) and EfficientNetB0 (30.95%) fail to achieve meaningful classification accuracy within the 3-epoch training window. This failure likely reflects architectural differences in how these models adapt to new domains: ResNet50 and EfficientNetB0 may require different fine-tuning strategies (e.g., learning rate schedules, partial freezing) or substantially more epochs to overcome initial feature distribution mismatch. The near-random performance of these architectures (25% would be expected by chance with 4 classes) indicates failure to learn discriminative features rather than slight underperformance. Xception's success may derive from its depthwise separable convolution design, which efficiently captures the fine-grained spatial patterns characteristic of tumor boundaries in MRI images. Training time varies across architectures, with Xception being fastest (1,390s) despite having more parameters than DenseNet121, reflecting implementation efficiency of depthwise separable operations. These results provide clear guidance for architecture selection: Xception should be preferred for brain tumor classification, with DenseNet121 as an alternative when parameter efficiency is prioritized.

V. DISCUSSION

This section interprets the experimental results, discusses their implications for brain tumor diagnosis, compares findings with existing literature, and acknowledges limitations of the present study.

The exceptional classification accuracy achieved by the Xception architecture (99.39% CNN-only, 99.54% with meta-learning) substantially exceeds previously reported results on similar datasets, establishing new state-of-the-art performance for brain tumor classification. Compared to Deepak and Ameer's 98.0% accuracy with GoogleNet+SVM and Badza and Barjaktarovic's 96.56% with custom CNN, our approach demonstrates meaningful improvement while utilizing a larger and more diverse dataset (7,023 vs 3,064 images). The improvement derives from multiple factors: the Xception architecture's efficient depthwise separable convolutions, comprehensive end-to-end fine-tuning rather than feature extraction with separate classifier, and the larger training dataset enabling better generalization.

The meta-learning integration with shape features, while providing modest accuracy improvement (0.15%), validates the hypothesis that morphological descriptors capture complementary information to CNN representations. The improvement is clinically meaningful: at 99.39% baseline accuracy, each 0.15% improvement corresponds to approximately one additional correctly diagnosed patient per 656 cases. The shape features (circularity, solidity, irregularity) encode tumor boundary characteristics that radiologists routinely assess during manual diagnosis, suggesting that the meta-learner successfully incorporates domain-relevant clinical knowledge. Future work might explore more sophisticated shape descriptors or additional handcrafted features such as texture analysis to further enhance meta-learning performance.

The rule-based clinical integration achieving 99.66% accuracy on accepted cases while maintaining 89% acceptance rate represents a practical approach to deploying automated diagnostic systems. Rather than forcing the model to make predictions on all cases, selective acceptance appropriately handles uncertainty by routing challenging cases for human review. The 11% referral rate is clinically acceptable, as these cases would receive specialist attention regardless of automated system availability. The accuracy improvement on accepted cases (0.27% over baseline) demonstrates that the confidence and irregularity thresholds effectively identify reliable predictions. This mechanism addresses a critical gap in existing literature, which typically reports single accuracy figures without considering deployment scenarios requiring uncertainty awareness.

The Grad-CAM explainability visualizations provide clinically meaningful interpretations by highlighting tumor regions influencing predictions. Visual inspection confirms that the model attends to appropriate anatomical locations, building

trust that predictions derive from relevant image features rather than spurious correlations. This transparency is essential for clinical adoption, as healthcare providers require explainable predictions to maintain professional accountability for diagnostic decisions. The explainability component transforms the black-box CNN into an interpretable decision support tool that augments rather than replaces clinical judgment.

The architecture comparison reveals important practical guidance: Xception substantially outperforms alternatives for brain tumor classification, while some architectures (ResNet50, EfficientNetB0) may require different adaptation strategies for medical imaging domains. The failure of ResNet50 despite having more parameters than Xception suggests that architecture design matters more than model capacity for this task. Depthwise separable convolutions may be particularly suited to capturing the fine-grained boundary patterns characteristic of tumor MRI appearance.

A. Limitations

Several limitations of this study should be acknowledged. First, the evaluation utilized a single publicly available dataset from Kaggle; multi-center validation across diverse imaging protocols and scanner manufacturers would strengthen generalizability claims. Second, the dataset contains only four diagnostic categories, whereas clinical practice encounters numerous tumor subtypes and variants that would require expanded classification schemes. Third, shape features were extracted using automated segmentation that may introduce errors; integration with radiologist-verified segmentations could improve meta-learning performance. Fourth, the rule-based thresholds were determined empirically and may require calibration for different datasets or clinical contexts. Finally, the study did not include prospective clinical validation with radiologist comparison, which would be necessary before deployment in healthcare settings.

B. Future Directions

Several promising directions emerge for future research. Multi-center validation studies incorporating diverse imaging protocols would establish clinical generalizability. Extension to additional tumor subtypes and rare variants would increase practical utility. Integration with 3D volumetric MRI analysis could improve diagnostic accuracy by incorporating spatial context across slices. Prospective clinical trials comparing hybrid system performance against radiologist diagnosis would establish clinical validity. Finally, deployment studies examining integration into clinical workflows would address practical implementation challenges beyond algorithmic performance.

VI. Conclusion

This study presented a hybrid deep learning framework for brain tumor classification that integrates Convolutional Neural Networks with meta-learning feature fusion and rule-based clinical decision support. Experimental evaluation on the Kaggle Brain Tumor MRI dataset comprising 7,023 images demonstrated exceptional performance: 99.39% accuracy with the Xception CNN architecture, improving to 99.54% through meta-learning integration of morphological shape descriptors (area, perimeter, circularity, solidity, irregularity). The rule-based clinical integration mechanism achieved 99.66% accuracy on automatically accepted cases while maintaining an 89% acceptance rate, appropriately routing uncertain cases for specialist review. Comprehensive architecture comparison established Xception as the preferred model for brain tumor classification, substantially outperforming DenseNet121, ResNet50, and EfficientNetB0 under identical experimental conditions. Grad-CAM explainability visualizations provide interpretable predictions by highlighting tumor regions influencing model decisions, addressing critical transparency requirements for clinical adoption. The proposed hybrid framework demonstrates that combining deep learning classification with complementary approaches including meta-learning and rule-based integration can achieve both high accuracy and clinical acceptability, contributing a practical blueprint for developing deployable medical AI diagnostic systems.

Acknowledgments

The authors acknowledge the use of the generative AI tool Claude (Anthropic, San Francisco, CA, USA) to improve the language and clarity of the manuscript. The authors reviewed and edited all content generated by the tool and take full responsibility for the final version of the manuscript.

References

[1] Q. T. Ostrom, M. Price, C. Neff, G. Cioffi, K. A. Waite, C. Kruchko, and J. S. Barnholtz-Sloan, "Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2017–2021," *Neuro-Oncology*, vol. 26, no. Supplement_6, pp. vi1–vi85, 2024.

[2] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-Branger, C. Hawkins, H. Ng, S. M. Pfister, G. Reifenberger *et al.*, "The 2021 who classification of tumors of the central nervous system: a summary," *Neuro-Oncology*, vol. 23, no. 8, pp. 1231–1251, 2021.

[3] K. Muhammad, S. Khan, J. Del Ser, and V. H. C. de Albuquerque, "Brain tumor classification using deep learning: A comprehensive survey," *Expert Systems with Applications*, vol. 238, p. 121924, 2024.

[4] M. A. Khan, M. Sharif, T. Akram, S. A. C. Bukhari, and R. S. Nayak, "A comprehensive survey of deep learning techniques for brain tumor detection from mri images," *Neural Computing and Applications*, vol. 36, no. 4, pp. 1551–1589, 2024.

[5] M. B. Hossain, M. A. Iqbal, M. S. Islam, and M. M. Rahman, "Vision transformers for brain tumor detection: A comprehensive review," *Artificial Intelligence in Medicine*, vol. 148, p. 102742, 2024.

[6] H. Alsaif, A. Alqahtani, N. Alsubaie, and A. Alshammari, "Deep ensemble learning approaches for brain tumor classification using mri images," *Computers in Biology and Medicine*, vol. 168, p. 107714, 2024.

[7] D. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: a survey," *Evolutionary Intelligence*, vol. 15, no. 1, pp. 1–22, 2022.

[8] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 545–563, 2021.

[9] M. M. Badza and M. C. Barjaktarovic, "Classification of brain tumors from mri images using a convolutional neural network," *Applied Sciences*, vol. 10, no. 6, p. 1999, 2020.

[10] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.

[11] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, "The false hope of current approaches to explainable artificial intelligence in health care," *The Lancet Digital Health*, vol. 3, no. 11, pp. e745–e750, 2021.

[12] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, 2020.
[13] S. Deepak and P. Ameer, "Brain tumor classification using deep cnn features via transfer learning," *Computers in Biology and Medicine*, vol. 111, p. 103345, 2019.
[14] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "Brain tumor type classification via capsule networks," pp. 3129–3133, 2019.
[15] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, "Brain tumor classification for mr images using transfer learning and fine-tuning," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 34–46, 2019.
[16] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, and Q. Feng, "Brain tumor detection and classification using deep learning approaches: Current trends and future perspectives," *Biomedical Signal Processing and Control*, vol. 89, p. 105697, 2024.
[17] Kaggle, "Brain tumor mri dataset," https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset, 2024.